IBM Storage Virtualize for SAN Volume Controller and FlashSystem Family

Concept Guide



Note

Before using this information and the product it supports, read the information in Chapter 1, “Notices ,” on page 1.

This edition applies to version 8, release 7, modification x, and to all subsequent modifications until otherwise indicated in new editions.

© Copyright International Business Machines Corporation 2024.
US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Figures ix

Tables xiii

Chapter 1. Notices 1
Trademarks 2

Chapter 2. System monitoring 5
Dashboard 5
Monitoring capacity 7
Performance statistics 11
Statistics collection 13
IBM Storage Insights 24
Preventive monitoring 25

Chapter 3. Pools 27
Arrays 31
Array configurations 31
Expanding an MDisk in a distributed array 43
RAID write response time 44
MDisks 44
Troubleshooting medium errors and bad blocks 46
External storage 47
Planning an external storage system 48
Attaching an external storage system 50
Discovering an external storage system 51
Importing an external storage system 53
Managing an external storage system 53
Removing an external storage system 55
External storage system zones 57
Pool configuration guidelines 57
Managing pools 58
Managing throttles for pools 59
Creating data reduction pools 60
Linked pools 62
Easy Tier 62
Easy Tier modes of operation 62
Easy Tier automatic data placement requirements, recommendations, and limitations 65
Easy Tier reports 66
Easy Tier settings for storage pools and volumes 72

Chapter 4. Volumes 75
Basic volumes 78
Mirrored volumes 78
Custom volumes 80
Standard-provisioned volumes 81
Thin-provisioned volumes 81
Compressed volumes 83
Deduplicated volumes 84

Volume protection 84
Expand or shrink volumes.... 85
Dependent volumes....86
Managing throttles for volumes....87
Volume groups.... 88
Transparent cloud tiering....89
Supported cloud service providers.... 89
Security considerations for cloud accounts.... 90
Configuring transparent cloud tiering.... 91
Cloud volumes and snapshots.... 96
Transparent cloud tiering restrictions.... 102
Snapshots.... 102
Cloning snapshots....102
Refresh from a snapshot....103
Restoring from snapshots....103
Safeguarded snapshots.... 104
Snapshot properties....104
Managing snapshots.... 105

Chapter 5. Hosts 107
Host mapping....107
Determining the volume name from the device identifier on the host....109
Volume reservations.... 110
Host portsets.... 110
Host connectivity.... 111
Host clusters....112
Managing host clusters 113
Managing hosts....114
Managing WWPN changes of replaced host HBAs....114
Managing throttles for hosts....115

Chapter 6. Host attachment 117
Using VMware.... 117
VMware Virtual Volumes (vVols)....118
Using OpenStack....124
Using Kubernetes and OpenShift.... 125
Copy offload by using ODX on Microsoft Windows.... 125
SCSI Fibre Channel host attachments.... 126
IBM Power Systems AIX.... 126
IBM Power Systems with Virtual I/O Server....128
IBM z Systems that run the Linux operating system....129
IBM z Systems running the IBM z/VSE operating system....131
Hosts that run the Linux operating system.... 132
Hosts that run the Microsoft Windows Server operating system.... 135
Hosts that run the VMware ESXi operating system....139
Hosts that run the Microsoft Hyper-V operating system.... 142
Citrix XenServer hosts....143
Zoning and portsets.... 144
Fibre Channel port-name identification.... 146
Queue depth in Fibre Channel hosts.... 148
iSCSI Ethernet host attachment....148
iSCSI overview....148
Example iSCSI configurations....150
iSCSI configuration details for host connections.... 152
Configuring the system for iSCSI hosts.... 154
iSCSI support (considerations)....156
Setting up the host server as an iSCSI initiator....157

Linux iSCSI host attachment....159
Microsoft Windows iSCSI host attachment....165
IBM AIX iSCSI host attachment....170
VMware iSCSI host attachment....173
Queue depth in iSCSI hosts....178
Multisession....180
NVMe over Fibre Channel host attachments....182
NVMe hosts that run the Linux operating system....182
VMware ESXi installation and configuration for NVMe over Fibre Channel hosts....187
NVMe hosts that run the Microsoft Windows operating system....188
IBM Power Systems AIX....189
NVMe over RDMA and NVMe over TCP host attachments....191
Linux....191
VMware ESXi....192

Chapter 7. Policies....195
Provisioning policies....195
Creating provisioning policy and assigning to pools....197
Managing provisioning policies....198
Snapshot policies....199
Replication policy....200

Chapter 8. Replication....201
High availability....201
Planning high availability....202
Implementing high availability....203
Supported host operating systems for high availability....205
Disaster recovery....206
Planning replication for disaster recovery....206
Implementing replication for disaster recovery....207
Managing replication for disaster recovery....208
Migrating replication for disaster recovery to use partitions....209
High availability and disaster recovery (3-site replication)....210
Partnerships....213
Pool links....214
Replication policies....215
Replication statistics....216

Chapter 9. Access....217
Ownership groups....217
Configuring ownership groups....220
Managing ownership groups....222
Users....227
Locking user accounts....228
Managing users and user groups....230
Locking and unlocking user account automatically....234
Locking and unlocking user accounts manually....235
Locking and unlocking superuser accounts....236
Working with local and remote users....239
Audit log....239
Login message....240

Chapter 10. FlashSystem grid....241

Chapter 11. Storage partitions....243

Chapter 12. Data migration..... 245
Migrating storage partitions between systems..... 245
Moving volumes to a data reduction pool..... 247
Migrating extents using the CLI..... 248
Migrating volumes to an encrypted pool..... 249
Migration methods for image mode volumes..... 252
Migrating volumes between pools using the CLI..... 252

Chapter 13. Stretched systems..... 255
Stretched system configuration details..... 255
Planning for a stretched topology system..... 256
Requirements for stretched systems with Fibre Channel connections..... 257
Stretched system configuration by using interswitch links..... 259
Stretched system configuration without using interswitch links..... 261
Configuring an enhanced stretched system 262
Configuring an external iSCSI storage system..... 264
Manual quorum override in a stretched system..... 264

Chapter 14. Legacy copy services..... 267
FlashCopy function..... 267
FlashCopy mapping guidelines for storage systems..... 268
FlashCopy mapping events..... 268
Grains and the FlashCopy bitmap..... 270
Thin-provisioned FlashCopy..... 270
Legacy Safeguarded Copy based on FlashCopy..... 271
IBM Copy Services Manager..... 272
IBM Storage Copy Data Management..... 272
Configuring Safeguarded Copy function..... 276
Managing the Safeguarded Copy function..... 283
Changing the system topology..... 287
FlashCopy statistics collection 287
Converting a volume to a stretched volume 289

Chapter 15. Notifications..... 291
Syslog notifications..... 291
SNMP notifications..... 293
SNMP over TLS..... 294
Examples of configuration using net-snmp..... 295

Chapter 16. Network..... 297
Ethernet connectivity..... 297
Routing considerations..... 297
Configuring VLAN..... 297
System IP addresses..... 299
Priority flow control..... 300
Modifying VLAN for priority flow control..... 302
Planning for a direct-attached iSCSI configuration..... 304
Fibre channel connectivity..... 305
Fabric Device Management Interface..... 305
Fibre Channel port numbers and worldwide port names..... 305
N-Port ID Virtualization..... 305
Planning for connectivity..... 309
Zoning examples..... 315
Fibre Channel port masking..... 320
Fibre Channel SAN configuration details..... 322
System configuration by using SAN fabrics with long-distance fiber connections 327

Determining and changing the Fibre Channel network speed....328
HTTP proxy server....328
Portsets....329
Ethernet Portsets....329
Creating a portset....331
Resolving a problem with PFC settings....333
Partnerships....334
Partnerships using Fibre Channel connectivity....336
Partnerships using IP Connectivity....337
Intersystem link for host traffic....351
Long-distance requirements for partnerships....351

Chapter 17. Cyber resilience....353

Chapter 18. Security....355
Security overview....355
Managing security....358
Security protocol levels....359
SSH rules....389
System Certificates....389
Configuring system certificates....391
Updating an internally signed certificate....391
Requesting and installing an externally signed certificate....394
Export certificates....398
Creating and managing certificate authority store by using CLI....398
Removing certificate authority store by using CLI....399
Multifactor authentication....399
Configuring multifactor authentication....400
Changing MFA credentials of system....419
Changing the failmode parameter for multifactor authentication....420
User authentication....421
Remote Authentication....422
Single Sign-on....425
Configuring single sign-on....427
Managing proxy server for SSO login....446
Changing SSO credentials of system....446
Encryption....447
Encryption overview....447
Planning for encryption....450
Configuring encryption....453
Licensing encryption....454
Encryption with key servers....455
Encryption with USB flash drives....470
Encryption recovery key....474
Decommissioning encryption....478
Secure data deletion....479
Password policy....481
Two person integrity....481

Chapter 19. System....485
I/O group....485
Nodes....486
Quorum....491
IP quorum application....493
System topology....494
System layers....495
Licensed functions....496

Bitmap space configuration....498
Throttles....501

Chapter 20. Support....503
Call Home....503
Support assistance....504
Support packages....506

Chapter 21. Software update....509
Obtaining packages....509
System update....511
Updating the system manually....512
Patch installation....512
Drive update....513

Chapter 22. Management GUI....515

Chapter 23. Ansible collection....517

Chapter 24. Storage Virtualize RESTful API....519
Storage Virtualize Performance Statistics....523
Performance data files....523
Getting performance statistics....526

Chapter 25. Conventions used in this IBM Documentation....541
Notifications....541
Emphasis....541
Numbering conventions....541

Figures

1. Storage pool..... 27

2. Two member drive storage array with no rebuild area..... 35

3. Two member drive storage array with a failed drive..... 36

4. Rebuild-in-place process on the two member drive.....37

5. Distributed RAID array (RAID 6 level) 38

6. Distributed RAID array (RAID 6 level) with a failed drive..... 39

7. Storage systems and MDisks..... 45

8. Storage pools and volumes.....76

9. Hosts, WWPNs, IQNs or EUIs, and volumes..... 108

10. Hosts, WWPNs, IQNs or EUIs, volumes, and SCSI mappings..... 108

11. Host cluster that contains private and shared volumes 113

12. Transmitting SCSI over TCP/IP.....149

13. iSCSI configuration with a single subnet.....151

14. iSCSI configuration that uses multiple subnets and provides alternative configuration interfaces.... 151

15. iSCSI configuration with host multipathing..... 152

16. Four two-session iSCSI multisession configurations for a single subnet..... 180

17. Two four-session iSCSI multisession configurations for a single subnet.....181

18. Four two-session iSCSI multisession configurations for multiple subnets and dual configuration.....181

19. Two four-session iSCSI multisession configurations for a multiple subnets and dual configuration.. 181

20. Ownership group inheritance..... 218

21. Example: Child pools are configured on the system.....223

22. Example: Child pools are not configured on the system..... 225

23. A stretched system with a quorum disk at a third site..... 258

24. Stretched system nodes with ISLs, DS3000, DS4000, and DS5000 connected to both fabrics..... 260

25. Recovering Safeguarded backups..... 275

26. A configuration that is not valid for use with NPIV..... 306

27. A configuration that is valid for use with NPIV and (WWPN-based) zoning.....307

28. A single switch configuration that is not valid for use with NPIV and hard (port-based) zoning..... 308

29. A single switch configuration that is valid for use with NPIV and hard (port-based) zoning 309

30. Fabric port configuration..... 313

31. An example of a host zone.....317

32. An example of a storage system zone.....318

33. An example of a system zone..... 318

34. Fabric with ISL between nodes in a system.....323

35. Fabric with ISL in a redundant configuration..... 323

36. Fibre Channel forwarder linked to existing Fibre Channel SAN.....324

37. Fibre Channel forwarder linked to hosts and storage systems without an existing Fibre Channel SAN.....324

38. Fibre Channel host connects into the Fibre Channel ports on the Fibre Channel forwarder..... 325

39. Fibre Channel host connects into the Fibre Channel ports on the Fibre Channel forwarder without an existing Fibre Channel SAN.....325

40. Simple SAN configuration.....326

41. SAN configuration with a medium-sized fabric.....326

42. SAN configuration with a large fabric.....327

43. SAN configuration across two sites.....327

44. Redundant fabrics.....337

45. One intersite link, one I/O group per system.....343

46. Two intersite links, one I/O group per system..... 344

47. Example of a write operation in an I/O group..... 486

48. Example of a standard system topology 494

49. Example of a stretched system topology 495

50. Example configuration with systems in different layers This figure shows an overview of system layers 496

Tables

1. Low space warning percentages for compressed arrays 8

2. Statistics collection for MDisks for individual nodes 13

3. Statistic collection for volumes for individual nodes 15

4. Statistic collection for node ports 16

5. Statistic collection for nodes 17

6. Performance statistics for volume 19

7. Statistic collection for volume cache per individual nodes 19

8. XML statistics for an IP partnership port 20

9. ODX VDisk and node level statistics 21

10. Statistics collection for cloud per cloud account ID 22

11. Statistics collection for cloud per VDisk 23

12. Minimum overhead capacity requirements for data reduction pools 30

13. Pool states 30

14. MDisk status 45

15. Bad block errors 47

16. Storage-system port-selection algorithm 49

17. Data move statistic descriptions 67

18. Workload categorization descriptions 69

19. Easy Tier settings for storage pools and volumes 72

20. Volume states 77

21. Volume cache modes 78

22. Configuration maximums for NMP on IBM Power Systems AIX hosts 127

23. Configuration maximums for LVM1 130

24. Registry key parameters for QLogic models..... 137

25. Configuration file parameters for the Emulex HBA.....137

26. Registry key parameters for QLogic models..... 142

27. Comparison of iSCSI and Fibre Channel components.....148

28. System registry..... 167

29. Common queue depth configurations.....179

30. NVMe over RDMA minimum supported Ethernet cards and firmware versions for VMware ESXi 7.0 UP2.....192

31. Example error scenarios.....211

32. Statistics collection for volume group.....216

33. Statistics collection for node statistics..... 216

34. Statistics collection for volume statistics.....216

35. Supported access control for management interfaces and functions with user groups..... 230

36. FlashCopy mapping events.....268

37. Statistics collection for volume aggregate..... 288

38. Statistics collection for volume count..... 288

39. Security levels and required credentials for SNMP version 3 servers using the user-based security model.....293

40. Examples of WWPN values for the system 305

41. Port allocation scheme..... 310

42. Two hosts and their ports..... 316

43. Two storage systems and their ports..... 316

44. Six hosts and their ports.....319

45. Three storage systems and their ports.....319

46. Port masking configuration.....321

47. Limits for Host Bus Adapters.....330

48. System limits.....346

49. Maximum supported round-trip latency between sites..... 352

50. Intersystem heartbeat traffic in Mbps..... 352

51. Supported SSL/TLS security levels..... 360

52. SSH algorithms supported at each security level..... 371

53. Protocols supported at level 7.....375

54. Java SSL ciphers supported at security level 7.....375

55. OpenSSL ciphers supported at level 7 (chsecurity -sslprotocol 7)..... 376

56. Protocols supported at level 6.....376

57. Java SSL ciphers supported at security level 6.....376

58. OpenSSL ciphers supported at level 6 (chsecurity -sslprotocol 6)..... 376

59. Protocols supported at level 5 376

60. Java SSL ciphers supported at security level 5.....377

61. OpenSSL ciphers supported at level 5 for TLS 1.3 (chsecurity -sslprotocol 5)..... 377

62. OpenSSL ciphers supported at level 5 for TLS 1.2 (chsecurity -sslprotocol 5)..... 378

63. Protocols supported at level 4 379

64. Java SSL ciphers supported at security level 4.....379

65. OpenSSL ciphers supported at level 4 (chsecurity -sslprotocol 4)..... 379

66. Protocols supported at level 3 379

67. Java SSL ciphers supported at security level 3.....379

68. OpenSSL ciphers supported at level 3 (chsecurity -sslprotocol 3)..... 381

69. Protocols supported at level 2.....382

70. Java SSL ciphers supported at level 2.....382

71. OpenSSL ciphers supported at level 2 (chsecurity -sslprotocol 2)..... 384

72. TCP and UDP ports that are supported..... 385

73. SSH algorithms supported at each security level..... 387

74. Recommended delimiters..... 394

75. Recommended delimiters..... 397

76. General tab.....402

77. Sign-on tab.....402

78. Add API Client action.....404

79. Authentication methods comparison 422

80. Authentication methods summary 426

81. Register an application page..... 428

82. General tab.....430

83. Sign-on tab.....430

84. Welcome.....433

85. Server application.....434

86. Configure Application Credentials.....435

87. Configure Web API.....435

88. Choose Access Control Policy..... 436

89. Configure Application Permissions..... 436

90. Register an application page..... 438

91. General settings on the app integration page.....440

92. Assignments settings on the app integration page.....440

93. Metadata section..... 442

94. Relying Party section..... 442

95. OIDC Response section.....443

96. Settings section.....444

97. Comparison of methods to securely delete data from drives..... 480

98. Node status.....488

99. System topology and volume summary.....495

100. Comparison of storage layer and replication layer systems.....495

101. Supported SCU licenses, drive classes, and ratios.....497

102. Example of SCU usage.....497

103. Product comparison of required I/O group memory for policy-based replication.....498

104. Bitmap space default configuration.....499

105. Examples of memory required 500

106. Distributed RAID requirements.....501

107. Supported network configurations for Call Home with cloud services.....503

108. Firewall rule configuration.....506

109. Firewall rule configuration for remote access.....506

110. Support packages.....507

111. Packages published to FixCentral.....509

112. Upgrade time for a four-node cluster.....511

113. 520

114. Calculations.....527

115. Calculations.....528

116. Calculations.....531

117. Calculations.....534

118. Calculations.....536

Chapter 1. Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan, Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US

© Copyright IBM Corp. 2024

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

All IBM prices shown are IBM's suggested retail prices, are current and are subject to change without notice. Dealer prices may vary.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM, the IBM logo, and ibm.com® are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at Copyright and trademark information at www.ibm.com/legal/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linux® and the Linux logo is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other product and service names might be trademarks of IBM or other companies.

Chapter 2. System monitoring

The following topics describe how to monitor the system.

Dashboard

The **Dashboard** on the management GUI provides a way to quickly assess the overall condition of the system and view notifications of any critical issues that require immediate action.

The **Dashboard** is the home page that displays after you log in to the system. It contains high-level information about the system and is not a replacement for the **Performance** page or the system view. The **Dashboard** contains sections for performance, capacity, and system health that provide an overall understanding of what is happening on the system.

Performance

The **Performance** section on the **Dashboard** displays a graph with up to 5 minutes of data from the performance of a specific system or a comparison of the performance of the nodes. Select either **System** or **Node Comparison** to switch the view between system-level performance and node-level performance.

The performance graph is updated with new data every 5 seconds. Select any of the available metrics to change what is displayed by the graph. To view a metric not available on the **Dashboard**, go to the **Performance** page in the navigation.

Capacity

The **Capacity** section on the **Dashboard** provides an overall view of system capacity. This section displays usable capacity, provisioned capacity, and capacity savings.

Usable Capacity

Usable capacity indicates the total capacity in all storage on the system. Usable capacity includes all the storage the system can be virtualized and assigned to pools. Usable capacity is displayed in a bar graph and is divided into three categories: **Stored Capacity**, **Available Capacity**, and **Total**. If the system supports self-compressing drives, certain system configurations make determining accurate usable capacity on the system difficult. For example, if the system contains self-compressed drives and data reduction pools without compression enabled, the system cannot determine the accurate amount of usable capacity that is used on the system. In this case, overprovisioning and losing access to write operations is possible. If this condition is detected by the system, the **Usable Capacity** section of the **Dashboard** page displays a message instead of capacity information. To recover from this condition, you need to ensure that all thin-provisioned volumes and thin-provisioned volumes that are deduplicated are migrated to volumes with compression enabled in the data reduction pools. Alternatively, you can migrate the volumes to standard-provisioned volumes and use the drive compression to save capacity.

Stored Capacity

Stored Capacity indicates the amount of capacity that is used on the system after capacity savings. The system calculates the stored capacity by subtracting the available capacity and any reclaimable capacity from the total capacity that is allocated to MDisks. To calculate the percentage, the stored capacity is divided by the total capacity that is allocated to MDisks. On the left side of the bar graph, the stored capacity is displayed in both the total capacity and as a percentage.

Available Capacity

The total **Available Capacity** displays on the right side of the bar graph. Available capacity is calculated by adding the available capacity and the total reclaimable capacity. To calculate the percentage of available capacity on the system, the available capacity is divided by the total amount of capacity that is allocated to MDisks.

© Copyright IBM Corp. 2024

Total Capacity

The **Total Capacity** capacity displays on the right under the bar graph and shows all the capacity available on the system. The bar graph is a visual representation of capacity usage and availability and can be used to determine whether more storage needs to be added to the system. Select **MDisks** to view more information about the usable capacity of the system on the **MDisks by Pools** page. You can also select **Compressed Volumes**, **Deduplicated Volumes**, or **Thin-Provisioned Volumes**.

If the system supports self-compressing drives, certain system configurations make determining accurate usable capacity on the system difficult.

Provisioned Capacity

Provisioned capacity is the total capacity of all virtualized storage on the system. Provisioned capacity is displayed in a bar graph and is divided into two categories: **Written Capacity** and **Available Capacity** capacity. **Written Capacity** displays on the left side of the bar graph and indicates the amount of capacity that has data that is written to all the configured volumes on the system. The system calculates the written capacity for volumes by adding the stored capacity to capacity savings. The percentage of written capacity for volumes is calculated by dividing the written capacity by the total provisioned capacity for volumes on the system. The **Available Capacity** displays on the right side of the bar graph and indicates the capacity on all configured volumes that is available to write new data. The available capacity is calculated by subtracting the written capacity for volumes from the total amount of capacity that is provisioned for volumes. The percentage of available capacity is calculated by dividing the available capacity for volumes by the total amount of capacity that is provisioned to volumes on the system. The **Total Provisioned** capacity displays under the **Available Capacity** and indicates the total amount of capacity that is allocated to volumes. The Provisioned Capacity also displays the percentage for over-provisioned volumes. The **Overprovisioned** value indicates the percentage of provisioned capacity that is increased because of capacity savings.

Capacity Savings

Capacity Savings indicates the amount of capacity that is saved on the system by using compression, deduplication, and thin-provisioning. The percentage value for each of these capacity savings methods compares the stored capacity before capacity savings is applied to the stored capacity after capacity savings is applied. Compression shows the total capacity savings gained from using compression on the system. Deduplication indicates the total capacity savings that the system is saved from all deduplicated volumes. Thin-Provisioning displays the total capacity savings for all thin-provisioned volumes on the system. You can view all the volumes that use each of these technologies. Different system models can have additional requirements to use compression or deduplication. These requirements can include additional hardware, licenses, or both to use these functions. Verify all system requirements before using these functions. The system supports several compression technologies, such as data reduction pool and self-compressing drives, where data is compressed in the hardware as it is written to a drive. A single capacity savings percentage for systems that use multiple compression technologies cannot be determined. Select **View Compression Details** to view savings details for each of the compression technologies that are currently used on the system.

System Health

The **System Health** section on the **Dashboard** provides a holistic view of the system through tiles of data that represent different components that make up the system. A tile contains one type of component, but it can contain multiple items of the same type.

The current software version number and the system ID are displayed above the **Connectivity components** category.

System Health tiles

The tiles of data in the **System Health** section of the Dashboard represent different components that make up the system, providing a holistic view of the system.

Each tile contains one type of component, but it can contain multiple items of the same type. For example, a Fibre Channel port is a component, so it is contained in a tile, but the Fibre Channel tile can contain multiple Fibre Channel ports. Tiles with errors and warnings are displayed first so that

components that require attention have higher visibility. Healthy tiles are sorted in order of importance in day-to-day use.

The following list defines the categories of tiles:

- **Hardware components** display the health of all components that are specific to the physical hardware.
- **Logical components** display the health of all logical and virtual components in the management GUI.
- **Connectivity components** display the health of all components that are related to the system's connectivity and the relationship between other components or systems.

Monitoring capacity

The system supports multiple ways of monitoring and adjusting capacity for your data. Use these methods to determine whether new storage resources are needed or if existing resources need to be reallocated.

Monitoring usable capacity

The system supports several ways to monitor usable capacity to ensure that storage is sufficient for host workloads.

Monitoring system-level capacity

The **Capacity** section on the **Dashboard** provides an overall view of system capacity. This section displays usable capacity, provisioned capacity, and capacity savings.

If you are using the command-line interface to determine usable capacity on your system, several parameter values are used from the **`lssystem`** command to calculate stored, available, and total capacities. Stored capacity is calculated with the values in the **`total_mdisk_capacity`**, **`total_free_space`**, **`total_reclaimable_capacity`** by using the following formula:

- Total stored capacity = **`total_mdisk_capacity`** - **`total_free_space`** - **`total_reclaimable_capacity`**

To calculate the available capacity, use the values in **`total_free_space`** and **`total_reclaimable_capacity`** in the following formula:

- Total available capacity = **`total_free_space`** + **`total_reclaimable_capacity`**

The value in the **`total_mdisk_capacity`** in the **`lssystem`** command indicates the total usable capacity is on the system.

Monitoring external storage systems, pool, and MDisk level capacity

The system supports virtualizing capacity that is on external storage systems that are attached to the system. For external storage systems, administrators must configure out-of-space alerts that notify the system when usable capacity on the external storage system reaches a defined threshold. Without these thresholds defined on the external storage systems, external storage on these systems can become overprovisioned and risk running out of usable capacity that is used for host operations. Overprovisioned external storage occurs when the sum of the provisioned capacity of all the volumes in a system or pool is greater than the total usable capacity of the system or pool that is allocated from the external storage system. The system detects potential overprovisioned external storage and presents a warning on the **Dashboard** page. If this warning is displayed, select **Overprovisioned External Storage** to display pools that contain storage that is nearing these overprovisioned scenarios. Depending on how external storage is configured in pools, different strategies and best practices can be used to ensure that usable capacity is sufficient and reported accurately to the system.

Configuration Scenario 1: One pool contains storage from one external storage system

In this configuration, storage administrators have the same concern of running out of capacity as when data is written directly to external storage systems. Storage administrators must set up out-of-space alerts on external storage and monitor usage to ensure that the I/O operations do not exceed the usable capacity. Consult documentation for your external storage system for specific guidelines on

these thresholds. If an out-of-space condition occurs, then usable capacity must be freed by deleting data or volumes.

Configuration Scenario 2: Multiple pools contain a single tier of storage from multiple external storage systems

In this configuration, multiple pools use the same type of storage, or tier, across several external storage systems. If capacity from different external storage systems is shared across multiple pools, then provisioning groups are created. Provisioning groups are objects that identify whether storage is shared across multiple pools. The **MDisks by Pools** page in the management GUI, displays all pools and their assigned MDisks. If you are not sure whether external storage systems are shared across several pools, right-click the pool and select **View Resources** to display the provisioning group that is associated with the pool. In this configuration, the system spreads extent allocations across all the external storage within the provisioning group to ensure that space is used evenly across all the external storage systems. However, usable capacity on external storage systems within the provisioning groups can still become overprovisioned, so storage administrators need to configure out-of-space alerts and monitor usable capacity to determine how capacity is being used on the system. If low-space warning occurs on virtualized MDisks on external storage and usable capacity is available on other external storage system in the pool, you can remove some of these MDisks from the pool until the usable capacity is within reasonable limits. This process migrates data to other MDisks in the pool. Ensure that enough usable capacity is available on the other external storage systems in the pool so that they do not run out of space during this operation.

Configuration Scenario 3: Multiple pools contain different tiers of storage from multiple external storage systems

In this configuration, different tiers of storage are present on the external storage systems within the pools. As with the previous configuration, provisioning groups are created to identify the external storage that is shared within the pools. However, in this case the system attempts to allocate the entire usable capacity in the top tier rather than spreading extents across provisioning groups. Storage administrators need to monitor the usable capacity in these top tiers of storage to ensure that space is sufficient for the workloads. If low-space warnings occur on virtualized MDisks on external storage and usable capacity is available on other external storage system in the pool, you can remove some of these MDisks from the pool until the usable capacity is within reasonable limits. This process moves data to other MDisks in the pool. Ensure that enough usable capacity is available on the other external storage systems in the pool so that they do not run out of space during this operation.

Configuration Scenario 4: Data reduction pools

Systems that support data reduction pools, which use data reduction technologies, like compression, can configure these types of pools to minimize overallocation on the external storage system. Data reduction pools also support the ability to reclaim unused capacity from host unmap operations and volume deletions. In the management GUI, select **Pools** > **MDisks by Pools** to view all the pools configured on the system. Data reduction pools are shown with **Data Reduction** set to **Yes**. When you use data reduction pools that virtualize external storage that can run out of space, always enable compression when you create volumes in these pools. The external storage system must be configured to present only usable capacity to the system, as the size of the written data cannot be reduced any further.

To assist with storage planning, Table 1 on page 8 lists the low space warnings that can be generated by the system for compressed arrays. Error 1246 is not raised on Easy Tier® pools or data reduction pools.

Table 1. Low space warning percentages for compressed arrays

Condition	Event ID	Error code	Percentage of usable capacity used
Compressed array is running low on usable capacity	020009	1246	90% used

Table 1. Low space warning percentages for compressed arrays (continued)			
Condition	**Event ID**	**Error code**	**Percentage of usable capacity used**
Critical level of compressed array is running low on usable capacity.	020010	1246	96%
1% usable capacity that is left for compressed array.	020011	1242	99%
Compressed array out of usable capacity.	020012	1241	100%

If your system configuration dedicates specific storage systems to individual pools, analyzing the usable capacity for the pool can determine whether more usable capacity is needed. On the **Pools** page, check each MDisk in the pool for capacity and use the value for **Storage System - LUN** to determine the external storage system that provides the MDisk with space.

Monitoring capacity savings

You can use the management GUI to determine the capacity savings that are created by thin-provisioning, compression, and deduplication.

The system supports several methods to save capacity. These methods can be used separately, but when used together they can provide greater capacity savings and efficiencies for the system. The system supports thin-provisioning, compression, and deduplication. With thin-provisioned volumes, the amount of capacity that can be provisioned to hosts is not limited by the usable capacity because the system eliminates the unused allocated capacity. A *thin-provisioned* volume presents a different capacity to mapped hosts than the capacity that the volume consumes in the storage pool. With thin provisioning, storage administrators can configure thin-provisioned volumes that can grow into the usable capacity they need based on the storage needs of the host. This flexibility saves on the costs to manage the storage capacity for host applications. Apart from standard-provisioned volumes, all volume types provide capacity savings and benefits from thin-provisioning. It can be used with compression and deduplication, or it can be used by itself. Thin-provisioning is supported in both standard pools and data reduction pools. However, data reduction pools support reclaiming capacity when it is no longer needed by hosts and then can redistribute it automatically for other uses.

Unlike thin-provisioning, which removes unused capacity, compression and deduplication increase the usable capacity by reducing the size of the data itself. The system supports several types of compression technologies that can be used on the system at the same time. Software-based compression is implemented when you create volumes, where compression can be enabled on the volume. With compressed volumes, data is compressed in the software then it is written to disk, saving more capacity. When data is read to hosts, the software decompresses the data. Compression is available through data reduction support as part of the system. If you want volumes to use compression as part of data reduction support, compressed volumes must belong to data reduction pools. Compressed volumes can also be created in standard pools, but data reduction pools support more capacity savings functions, such as automatic redistribution of reclaimed storage. The system also supports hardware-based compression technologies, such as self-compressing drives. This type of compression happens automatically within the hardware itself as data is written to the drive, and decompressed on the hardware as it is read to hosts. Depending on what technology handles the compression operations, a single capacity savings percentage for systems that use multiple compression technologies cannot be determined. Select **Compression Details** to view savings details for each of the compression technologies that are currently used on the system. If compressed volumes are configured in standard pools, you cannot create compressed volumes in data reduction pools within the same I/O group. Potential compression savings can be calculated for existing volumes in the management GUI through a built-in compression estimation

tool. This tool can be used to determine the capacity savings that are possible for existing data on the system by using compression.

The system also supports deduplication. Deduplication can be configured with volumes that use different capacity saving methods, such as thin-provisioning. When you create a volume, you can specify to include deduplication with other supported capacity savings methods. Deduplicated volumes must be created in data reduction pools. If you have existing volumes in standard pools, you can migrate them to data reduction pools to add deduplication to increase capacity savings for the volume.

In addition to capacity savings from these methods, data reduction pools add another benefit by reclaiming capacity and redistributing it to other uses. Reclaimable capacity is calculated as part of the usable capacity on the system. The amount of reclaimable capacity varies based on the number of write operations and how much reclaimable capacity currently exists on the system. The reclaimable capacity represents usable capacity that will eventually become available for overwriting with new data when it is "reclaimed" and made available to the system. This reclaimable capacity is calculated and displayed with the usable capacity of the system. Since reclaimable capacity is not used by production data and is eventually reclaimed for use, the amount of reclaimable capacity is subtracted from stored capacity and added to available capacity in the **Capacity** tile on the **Dashboard** in the management GUI. This calculation gives a more accurate view of capacity usage on the system. In systems that contain self-compressing drives, the actual usable capacity that reclaimable capacity consumes cannot be calculated unless it is already compressed by the data reduction pool. When the data is compressed on the drive, it is not possible to determine how much usable capacity that becomes available when the reclamation process completes. Each of these capacity savings methods requires capacity monitoring to ensure efficient capacity provisioning on the system. Capacity savings result from using these methods and establishing best practices for capacity management in your organization and actual savings is contingent on these practices. However, compression and deduplication might provide data reduction two times greater than without these methods and thin-provisioning might provide two times the data reduction. When these data reduction methods are used together, overall capacity savings are combined.

To determine capacity savings in the management GUI, complete these steps:

1. In the management GUI, select **Dashboard**.
2. Under **Capacity Savings**, you can view the total capacity savings that the system currently gains from using capacity savings methods. The total capacity savings is divided into thin-provisioning, compression, and deduplication savings. Click the link to view volumes that use each of these methods to save capacity. The management GUI calculates capacity savings for each of these methods. For thin-provisioning savings, all written capacity is subtracted from all the provisioned capacity on the system. Written capacity is the capacity that hosts write to storage before any data reduction methods are applied. Thin-provisioning savings indicate how much capacity is provisioned to hosts but is not currently used to store data from hosts. For compression savings, the system calculates savings from compressed volumes in data reduction pools. If your system supports compression in standard pools, the total compressed savings are calculated by combining the savings from both data reduction pool and from standard pools. The system supports several compression technologies, such as data reduction pool and self-compressing drives, where data is compressed in the hardware as it is written to a drive. A single capacity savings percentage for systems that use multiple compression technologies cannot be determined. Select **View Compression Details** to view savings details for each of the compression technologies that are currently used on the system. Deduplication savings are calculated for deduplicated volumes in data reduction pools. Deduplication is not supported in standard pools.

Recovering from configuration that prevents accurate usable capacity reporting

If the system supports self-compressing drives, certain system configurations make determining accurate usable capacity on the system difficult.

If the system contains self-compressed drives and data reduction pools without compressed volumes, the system cannot determine the accurate amount of usable capacity that is used on the system. In this case, overprovisioning and losing access to write operations is possible. If this condition is detected by the system, the **Usable Capacity** section of the **Dashboard** page displays a message instead of capacity

information. To recover from this condition, you need to ensure that all thin-provisioned volumes and thin-provisioned volumes that are deduplicated are migrated to volumes with compression enabled in the data reduction pools. Alternatively, you can migrate the volumes to standard-provisioned volumes and use the drive compression to save capacity.

You can use volume mirroring to enable compression on existing volumes in data reduction pools. Because data reduction pools also support deduplication, you can add that function to compressed volumes to gain extra capacity savings. After the original volumes are deleted from the data reduction pool, the capacity from those volumes is added to the reclaimable capacity for the system, and the error is removed from the **Dashboard**. Depending on the amount of reclaimable capacity on the system, the system can take some time to remove the reclaimable capacity from the system. During that process, the usable capacity tends to shrink in size. If you want to use self-compressing drives, convert these volumes to standard-provisioned volumes that can be used in a standard pool that uses self-compressing drives for storage. In the management GUI, you can use the **Modify Capacity Savings** action to quickly convert the capacity savings method for your volumes. This action creates a mirrored volume of the original, with the new capacity savings method enabled, then automatically deletes the original version from the pool.

If you have data reduction pools that contain volumes without compression on systems that contain self-compressing drives, you need to convert all these volumes to enable compression. Accurate reporting for usable capacity is unavailable for these configurations. To change the capacity savings on volumes in data reduction pools, complete these steps:

1. In the management GUI, select **Volumes** > **Volumes by Pools**. The left navigation displays all the configured pools on the system and expands the data reduction pool that contains the volumes that you want to change the capacity savings method for.
2. Right-click the volume that you want to convert and select **Modify Capacity Savings**. Since this process creates volume copies to manage the conversion, ensure that the pool has enough storage to accommodate these volume copies. For example, if you are converting a volume that has 2 GB allocated capacity, ensure that the pool has another 2 GB to accommodate both the original version and the copy during the conversion. After the conversion completes, the original copy is automatically deleted from the pool. You can also convert volumes in standard pools to data reduction pools to use the data reduction features that are available in these pools. For that method, you need to create a new data reduction pool and volume mirroring.
3. On the **Modify Capacity Savings** page, select **Compression** to change the currently configured capacity setting to enable compression. For additional capacity savings, you can also select **Deduplicated** to remove duplicate data from the pool.
4. Click **Modify**. After the original volumes are deleted from the data reduction pool, the capacity from those volumes is added to the reclaimable capacity for the system, and the error is removed from the **Dashboard**. Depending on the amount of reclaimable capacity on the system, the system can take some time to remove the reclaimable capacity from the system. During that process, the usable capacity tends to shrink in size.

Performance statistics

Real-time performance statistics provide short-term status information for the system.

To access these performance statistics, click **Monitoring** > **Performance** in the management GUI. In addition, the management GUI displays an overview of system performance, in the **Performance** section on the **Dashboard**.

You can use system statistics to monitor the bandwidth of all the volumes, protocols, and MDisks that are being used on your system. You can also monitor the overall System CPU (total) for the system. These statistics summarize the overall performance health of the system and can be used to monitor trends in bandwidth and System CPU (total). You can monitor changes to stable values or differences between related statistics, such as the latency between volumes and MDisks. These differences then can be further evaluated by performance diagnostic tools.

Additionally, with system-level statistics, you can quickly view bandwidth of volumes, protocols, and MDisks. Each of these charts displays the current bandwidth in megabytes per second (Mbps) and a view

of bandwidth over time. Each data point can be accessed to determine its individual bandwidth use and to evaluate whether a specific data point might represent performance impacts. For example, you can monitor the protocols to determine whether the host data-transfer rate is different from the expected rate.

You can also select node-level statistics, which can help you determine the performance impact of a specific node. As with system statistics, node statistics help you to evaluate whether the node is operating within normal performance expectations..

CPU Utilization

The System CPU (total) chart shows the current percentage of CPU usage and peaks in utilization.

A single spike often does not indicate a performance impact on the system; however, if a data point is consistently above 95% of utilization and I/O input is high, the system might be overloaded, which might indicate a need for more back-end storage. However, for compression utilization alone, it can be normal to see rates at 100%, especially if compression is used frequently on the system. If both compression and I/O input are high, the system might need more storage to accommodate the utilization.

Interface

MDisks and volumes

The MDisk charts display six metrics while the Volumes charts display another six metrics on the **Performance** section. These metrics are: Read IOPS, Write IOPS, Read bandwidth, Write bandwidth, Read latency, and Write latency. You can use these metrics to help determine the overall performance health of the volumes and MDisks on your system. Consistent unexpected results can indicate errors in configuration, system faults, or connectivity issues. Both the volumes and MDisk charts contain the same metrics to compare and use to evaluate performance; however, the data points for these metrics can be different due to the impact of system cache, RAID overhead, and Copy Services functions. You can view data points in megabytes per second (MBps) and I/O per second (IOPS).

If you select the read metric, data points on the chart indicate the average amount in MBps or IOPS for read operations that have been processed over the sample period. Read metrics represent how much data the system is processing (the bandwidth). The read latency metric for volumes measures the average amount of time in milliseconds that the system takes to respond to read requests over the sample period. The MDisks latency metrics measures the response time for back-end storage. Increased workload or error recovery can cause spikes in read latency. Values that are consistently higher than expected can indicate a fault situation or an overloaded system. Volume write latency tends to be lower than the MDisk write latency because of volume caching. If the volume write latency is equal to or greater than MDisk write latency, it suggests that the drives might be overloaded and more drives might be necessary to accommodate the increased workload. Volume latency may be lower than MDisk latency because of the volume cache. Alternatively volume latency may be higher than MDisk latency due to hosts using large I/O sizes. Neither of these conditions indicates a problem with the system.

Note: The management GUI supports latency metrics that are displayed in both microseconds and milliseconds. Measurements for metrics are rendered based on whether current values are less than 1.000 milliseconds. Values over 1.000 milliseconds are displayed as milliseconds and metrics under that value display as microseconds. Whereas in command-line interface (CLI) the values are always displayed in microseconds.

The difference between read and write IOPS shows the mixture of the workload that the system is executing. You can determine the average transfer sizes of data that the system is experiencing by dividing the reads and write operations in MBps by the read and write operations in IOPS. This information can be used for validating and predicting disk configuration for the system or input to a disk provisioning application. Write latency is the average time (in milliseconds) that the system writes data to volumes or MDisks but does not include the time for write operations that are used to keep volumes in Replication relationships synchronized. As with read latency, MDisk write latency tends to be higher than volumes

because of write caching and RAID overheads. For example, a write operation to a volume can result in additional read and write operations on the MDisk depending on the RAID type for the array.

Statistics collection

The system collects statistics over an interval and creates files that can be viewed.

Introduction

For each collection interval, the management GUI creates five statistics files: one for managed disks (MDisks), which is named **Nm_stats**; one for volumes and volume copies, which is named **Nv_stats**; one for volumes groups, which is named **Ng_stats**; one for nodes, which is named **Nn_stats**; and one for drives, which is named **Nd_stats**. The files are written to the `/dumps/iostats/` directory on the node. To retrieve the statistics files from the non-configuration nodes onto the configuration node, use the **svctask cpdumps** command.

A maximum of 16 files of each type can be created for the node. When the 17th file is created, the oldest file for the node is overwritten.

Fields

The following fields are available for user definition:

Interval
Specify the interval in minutes between the collection of statistics. You can specify 1 - 60 minutes in increments of 1 minute. Statistics must be set to 1 or 5 minute intervals when sending data to Storage Insights.

Tables

The following tables describe the information that is reported for individual nodes and volumes.

Table 2 on page 13 describes the statistics collection for MDisks, for individual nodes.

Table 2. Statistics collection for MDisks for individual nodes

Statistic name	Description
id	Indicates the name of the MDisk for which the statistics apply.
idx	Indicates the identifier of the MDisk for which the statistics apply.
pre	Indicates the peak of read external response time in milliseconds for each MDisk. The external response time for disk reads is calculated by starting a timer when a SCSI read command is issued and stopped when the command completes successfully.
pro	Indicates the peak of read queued response time in milliseconds for each MDisk. The value means the peak elapsed time that is taken for read commands to complete from the time they join the queue.
pwe	Indicates the peak of write external response time in milliseconds for each MDisk. The external response time for disk writes is calculated by starting a timer when a SCSI write command is issued and stopped when the command completes successfully.
pwo	Indicates the peak of write queued response time in milliseconds for each MDisk. The value means the peak elapsed time that is taken for write commands to complete from the time they join the queue.
rb	Indicates the cumulative number of blocks of data that was read (since the node started).

Table 2. Statistics collection for MDisks for individual nodes (continued)	
Statistic name	**Description**
re	Indicates the cumulative read external response time in milliseconds for each MDisk. The cumulative response time for disk reads is calculated by starting a timer when a **SCSI read** command is issued and stopped when the command completes successfully. The elapsed time is added to the cumulative counter.
ro	Indicates the cumulative number of MDisk read operations that were processed (since the node started).
rq	Indicates the cumulative read queued response time in milliseconds for each MDisk. This response is measured from above the queue of commands to be sent to an MDisk because the queue depth is already full. This calculation includes the elapsed time that is taken for **read** commands to complete from the time they join the queue.
ure	Indicates the cumulative read external response time in microseconds for each MDisk. The cumulative response time for disk reads is calculated by starting a timer when a SCSI read command is issued and stopped when the command completes successfully. The elapsed time is added to the cumulative counter.
urq	Indicates the cumulative read queued response time in microseconds for each MDisk. This response is measured from above the queue of commands to be sent to an MDisk because the queue depth is already full. This calculation includes the elapsed time that is taken for read commands to complete from the time they join the queue.
uwe	Indicates the cumulative write external response time in microseconds for each MDisk. The cumulative response time for disk writes is calculated by starting a timer when a SCSI write command is issued and stopped when the command completes successfully. The elapsed time is added to the cumulative counter.
uwq	Indicates the cumulative write queued response time in microseconds for each MDisk. This time is measured from above the queue of commands to be sent to an MDisk because the queue depth is already full. This calculation includes the elapsed time that is taken for write commands to complete from the time they join the queue.
wb	Indicates the cumulative number of blocks of data written (since the node started).
we	Indicates the cumulative write external response time in milliseconds for each MDisk. The cumulative response time for disk writes is calculated by starting a timer when a **SCSI write** command is issued and stopped when the command completes successfully. The elapsed time is added to the cumulative counter.
wo	Indicates the cumulative number of MDisk write operations that were processed (since the node started).
wq	Indicates the cumulative write queued response time in milliseconds for each MDisk. This time is measured from above the queue of commands to be sent to an MDisk because the queue depth is already full. This calculation includes the elapsed time that is taken for write commands to complete from the time they join the queue.

Note: MDisk statistics files for nodes are written to the `/dumps/iostats` directory on the individual node.

Table 3 on page 15 describes the volume information that is reported for individual nodes.

Note: The system only collects a subset of statistics for volumes that are used for snapshot function and policy-based replication.

Table 3. Statistic collection for volumes for individual nodes	
Statistic name	**Description**
id	Indicates the volume name for which the statistics apply.
idx	Indicates the volume for which the statistics apply.
rb	Indicates the cumulative number of blocks of data read (since the node started).
rl	Indicates the cumulative read response time in milliseconds for each volume. The cumulative response time for volume reads is calculated by starting a timer when a SCSI read command is received and stopped when the command completes successfully. The elapsed time is added to the cumulative counter.
rlw	Indicates the worst read response time in microseconds for each volume since the last time statistics were collected. This value is reset to zero after each statistics collection sample.
ro	Indicates the cumulative number of volume read operations that were processed (since the node started).
rxl	Indicates the cumulative read data transfer response times in milliseconds for each volume since the last time the node was reset. A high value for this statistic indicates that latency is likely caused by the fabric or the host, which is submitting more read commands than it can process.
ub	Indicates the cumulative number of blocks of data unmapped (since the node started).
ul	Indicates the cumulative unmap response time in milliseconds for each volume. The cumulative response time for volume unmaps is calculated by starting a timer when a SCSI unmap command is received and stopped when the command completes successfully. The elapsed time is added to the cumulative counter.
ulw	Indicates the worst unmap response time in microseconds for each volume. The worst response time for volume unmaps is calculated by starting a timer when a SCSI unmap command is received and stopped when the command completes successfully.
uo	Indicates the cumulative number of volume unmap operations that were processed (since the node started).
uou	Indicates the cumulative number of volume unmap operations that are not aligned on an 8 K boundary (according to the alignment/granularity setting in Block Limits VPD Page (0xb0).
wb	Indicates the cumulative number of blocks of data written (since the node started).
wl	Indicates the cumulative write response time in milliseconds for each volume. The cumulative response time for volume writes is calculated by starting a timer when a SCSI write command is received and stopped when the command completes successfully. The elapsed time is added to the cumulative counter.
wlw	Indicates the worst write response time in microseconds for each volume since the last time statistics were collected. This value is reset to zero after each statistics collection sample.
wo	Indicates the cumulative number of volumes write operations that were processed (since the node started).
wou	Indicates the cumulative number of volumes write operations that are not aligned on a 4 K boundary.

Table 3. Statistic collection for volumes for individual nodes (continued)	
Statistic name	**Description**
wxl	Indicates the cumulative write data transfer response times in milliseconds for each volume since the last time the node was reset. A high value for this statistic indicates that latency is likely caused by the fabric and/or the host responding slowly when the node requests the data to write to the volume.
xl	Indicates the cumulative read and write data transfer response time in milliseconds for each volume since the last time the node was reset. When this statistic is viewed for multiple volumes and with other statistics, it can indicate whether the latency is caused by the host or fabric.

Note: For unmap statistics, it is where an unmap operation is a **SCSI unmap** or **Write same with unmap** command.

Table 4 on page 16 describes the port information that is reported for individual nodes.

Table 4. Statistic collection for node ports	
Statistic name	**Description**
bbcz	Indicates the total time in microseconds for which the buffer credit counter was at zero. This statistic is only reported by 8 Gbps Fibre Channel ports. For other port types, this statistic is 0.
cbr	Indicates the bytes received from controllers.
cbt	Indicates the bytes transmitted to disk controllers.
cer	Indicates the commands that are received from disk controllers. **Note:** The cer metric is always 0.
cet	Indicates the commands that are initiated to disk controllers.
dtdc	Indicates the number of transfers that experienced excessive data transmission delay.
dtdm	Indicates the number of transfers that had their data transmission delay measured.
dtdt	Indicates the total time in microseconds for which data transmission was excessively delayed.
har	Indicates the count aborted host read operations aborted while a data transfer was in progress. This count includes I/O operations aborted by hosts, and those operations aborted internally by the system.
haw	Indicates the count aborted host write operations aborted while a data transfer was in progress. This count includes I/O operations aborted by hosts, and those operations aborted internally by the system.
hbr	Indicates the bytes received from hosts.
hbt	Indicates the bytes transmitted to hosts.
her	Indicates the commands that are received from hosts.
het	Indicates the commands that are initiated to hosts. **Note:** The het metric is always 0.
hsr	Indicates the count of data transfers that were considered slow for host read operations.

Table 4. Statistic collection for node ports (continued)	
Statistic name	**Description**
hsw	Indicates the count of data transfers that were considered slow for host write operations.
icrc	Indicates the number of CRC that is not valid.
id	Indicates the port identifier for the node.
itw	Indicates the number of transmission word counts that are not valid.
lf	Indicates a link failure count.
lnbr	Indicates the bytes received to other nodes in the same cluster.
lnbt	Indicates the bytes transmitted to other nodes in the same cluster.
lner	Indicates the commands that are received from other nodes in the same cluster.
lnet	Indicates the commands that are initiated to other nodes in the same cluster.
lsi	Indicates the lost-of-signal count.
lsy	Indicates the loss-of-synchronization count.
pspe	Indicates the primitive sequence-protocol error count.
tmp	Indicates the SFP temperature in degrees Celsius.
tmpht	Indicates the SFP temperature high alarm threshold in degrees Celsius.
txpwr	Indicates the TX power in microwatts (µW).
txpwrlt	Indicates the TX power low alarm threshold in microwatts (µW).
rmbr	Indicates the bytes received to other nodes in the other clusters.
rmbt	Indicates the bytes transmitted to other nodes in the other clusters.
rmer	Indicates the commands that are received from other nodes in the other clusters.
rmet	Indicates the commands that are initiated to other nodes in the other clusters.
rxpwr	Indicates the RX power in microwatts (µW).
rxpwrlt	Indicates the RX power low alarm threshold in microwatts (µW).
wwpn	Indicates the worldwide port name for the node.

Table 5 on page 17 describes the node information that is reported for each node.

Table 5. Statistic collection for nodes	
Statistic name	**Description**
cluster_id	Indicates the identifier of the cluster.
cluster	Indicates the name of the cluster.

Table 5. Statistic collection for nodes (continued)	
Statistic name	**Description**
cpu	busy - Indicates the total CPU average core busy milliseconds since the node was reset. This statistic reports the amount of the time that the processor spends polling, waiting for work versus doing work. This statistic accumulates from zero.
	system - Indicates the total CPU average core busy milliseconds since the node was reset. This statistic reports the amount of the time the processor spends polling, waiting for work versus doing work. This statistic accumulates from zero. This statistic is the same information as the information provided with the cpu busy statistic and eventually replaces the cpu busy statistic.
cpu_core	id - Indicates the CPU core ID.
	system - Indicates the per-core CPU average core busy milliseconds for system process cores since node was reset.
env	obj - Indicates whether enclosure (e) or node (n).
	id - Indicates the enclosure or node ID.
	p - Indicates the power consumption of the enclosure or node.
	t - Indicates the temperature (Celsius) of the enclosure or node.
dimm	id - Indicates the memory module ID.
	loc - Indicates the location of the memory module.
	manu - Indicates the manufacturer of the memory module.
	sn - Indicates the serial number of the memory module.
	ce - Indicates the number of corrected errors in the memory module.
id	Indicates the name of the node.
lrb	Indicates the number of logical bytes received from the other node.
lwb	Indicates the number of logical bytes sent to the other node.
node_id	Indicates the unique identifier for the node.
rb	Indicates the number of physical bytes received from the other node.
re	Indicates the accumulated receive latency, excluding inbound queue time. This statistic is the latency that is experienced by the node communication layer from the time that an I/O is queued to cache until the time that the cache gives completion for it.
ro	Indicates the number of messages or bulk data received.
rq	Indicates the accumulated receive latency, including inbound queue time. This statistic is the latency from the time that a command arrives at the node communication layer to the time that the cache completes the command.
wb	Indicates the number of physical bytes sent to the other node.
we	Indicates the accumulated send latency, excluding outbound queue time. This statistic is the time from when the node communication layer issues a message out onto the Fibre Channel until the node communication layer receives notification that the message arrived.
wo	Indicates the number of messages or bulk data sent.

Table 5. Statistic collection for nodes (continued)	
Statistic name	**Description**
wq	Indicates the accumulated send latency, including outbound queue time. This statistic includes the entire time that data is sent. This time includes the time from when the node communication layer receives a message and waits for resources, the time to send the message to the remote node, and the time that is taken for the remote node to respond.

There are three types of data reduction properties per data reduction pool.

- dca - these statistics are related to the data stored within the data reduction pool.
- rca - these statistics are related to I/O to manage the background garbage collection processes of the data reduction pool.
- jca - these statistics are related to journaling operations for the metadata that manages the data reduction pool.

Table 6 on page 19 provides details about the statistic collection for volume performance.

Table 6. Performance statistics for volume. This table lists the performance statistics that is reported for individual volume.	
Statistic name	**Description**
entav	Internal cyber resiliency statistic average
entcn	Internal cyber resiliency statistic count

Table 7. Statistic collection for volume cache per individual nodes. This table describes the volume cache information that is reported for individual nodes.	
Statistic name	**Description**
cm	Indicates the number of sectors of modified or dirty data that are held in the cache.
ctd	Indicates the total number of cache destages that were initiated writes, submitted to other components as a result of a volume cache flush or destage operation.
ctds	Indicates the total number of sectors that are written for cache-initiated track writes.
ctp	Indicates the number of track stages that are initiated by the cache that are prestage reads.
ctps	Indicates the total number of staged sectors that are initiated by the cache.
ctrh	Indicates the number of total track read-cache hits on prestage or non-prestage data. For example, a single read that spans two tracks where only one of the tracks obtained a total cache hit, is counted as one track read-cache hit.
ctrhp	Indicates the number of track reads received from other components, which are treated as cache hits on any prestaged data. For example, if a single read spans two tracks where only one of the tracks obtained a total cache hit on prestaged data, it is counted as one track that is read for the prestaged data. A cache hit that obtains a partial hit on prestage and non-prestage data still contributes to this value.
ctrhps	Indicates the total number of sectors that are read for reads received from other components that obtained cache hits on any prestaged data.
ctrhs	Indicates the total number of sectors that are read for reads received from other components that obtained total cache hits on prestage or non-prestage data.

Table 7. Statistic collection for volume cache per individual nodes. This table describes the volume cache information that is reported for individual nodes. *(continued)*	
Statistic name	**Description**
ctr	Indicates the total number of track reads received. For example, if a single read spans two tracks, it is counted as two total track reads.
ctrs	Indicates the total number of sectors that are read for reads received.
ctwft	Indicates the number of track writes received from other components and processed in flush through write mode.
ctwfts	Indicates the total number of sectors that are written for writes that are received from other components and processed in flush through write mode.
ctwfw	Indicates the number of track writes received from other components and processed in fast-write mode.
ctwfwsh	Indicates the track writes in fast-write mode that were written in write-through mode because of the lack of memory.
ctwfwshs	Indicates the track writes in fast-write mode that were written in write through due to the lack of memory.
ctwfws	Indicates the total number of sectors that are written for writes that are received from other components and processed in fast-write mode.
ctwh	Indicates the number of track writes received from other components where every sector in the track obtained a write hit on already dirty data in the cache. For a write to count as a total cache hit, the entire track write data must already be marked in the write cache as dirty.
ctwhs	Indicates the total number of sectors that are received from other components where every sector in the track obtained a write hit on already dirty data in the cache.
ctw	Indicates the total number of track writes received. For example, if a single write spans two tracks, it is counted as two total track writes.
ctws	Indicates the total number of sectors that are written for writes that are received from components.
ctwwt	Indicates the number of track writes received from other components and processed in write through write mode.
ctwwts	Indicates the total number of sectors that are written for writes that are received from other components and processed in write through write mode.
cv	Indicates the number of sectors of read and write cache data that is held in the cache.

Table 8 on page 20 describes the XML statistics specific to an IP Partnership port.

Table 8. XML statistics for an IP partnership port	
Statistic name	**Description**
ipbz	Indicates the average size (in bytes) of data that is being submitted to the IP partnership driver since the last statistics collection period.
iprc	Indicates the total bytes that are received before any decompression takes place.
ipre	Indicates the bytes retransmitted to other nodes in other clusters by the IP partnership driver.

Table 8. XML statistics for an IP partnership port (continued)	
Statistic name	**Description**
iprt	Indicates the average round-trip time in microseconds for the IP partnership link since the last statistics collection period.
iprx	Indicates the bytes received from other nodes in other clusters by the IP partnership driver.
ipsz	Indicates the average size (in bytes) of data that is being transmitted by the IP partnership driver since the last statistics collection period.
iptc	Indicates the total bytes that are transmitted after any compression (if active) takes place.
iptx	Indicates the bytes transmitted to other nodes in other clusters by the IP partnership driver.

Table 9 on page 21 describes the offload data transfer (ODX) Vdisk and node level I/O statistics.

Table 9. ODX VDisk and node level statistics		
Statistic name	**Acronym**	**Description**
Read cumulative ODX I/O latency	orl	Cumulative total read latency of ODX I/O per VDisk. The unit type is micro-seconds (US).
Write cumulative ODX I/O latency	owl	Cumulative total write latency of ODX I/O per VDisk. The unit type is micro-seconds (US).
Total transferred ODX I/O read blocks	oro	Cumulative total number of blocks that are read and successfully reported to the host, by ODX WUT command per VDisk. It is represented in blocks unit type.
Total transferred ODX I/O write blocks	owo	Cumulative total number of blocks that are written and successfully reported to the host, by ODX WUT command per VDisk. It is represented in blocks unit type.
Wasted ODX I/Os	oiowp	Cumulative total number of wasted blocks that are written by ODX WUT command per node. It is represented in blocks unit type.
WUT failure count	otrec	Cumulative total number of failed ODX WUT commands per node. It includes WUT failures due to a token revocation and expiration.

Table 10 on page 22 describes the statistics collection for cloud per cloud account ID.

Table 10. Statistics collection for cloud per cloud account ID		
Statistic name	**Acronym**	**Description**
id	id	Cloud account ID
Total Successful Puts	puts	Total number of successful PUT operations
Total Successful Gets	gets	Total number of successful GET operations
Bytes Up	bup	Total number of bytes successfully transferred to the cloud
Bytes Down	bdown	Total number of bytes successfully downloaded/read from the cloud
Up Latency	uplt	Total time that is taken to transfer the data to the cloud
Down Latency	dwlt	Total time that is taken to download the data from the cloud
Down Error Latency	dwerlt	Time that is taken for the GET errors
Part Error Latency	pterlt	Total time that is taken for part errors
Persisted Bytes Down	prbdw	Total number of bytes successfully downloaded from the cloud and persisted on the local storage that were part of successful GET operation
Persisted Bytes Up	prbup	Total number of bytes successfully transferred to the cloud and persisted on the cloud that were part of successful PUT operation. The difference is that you might have a 100 bytes file, of which you successfully had 80 bytes sent to the cloud through a PUT operation, but the last data transfer cycle carrying 20 bytes errored out, and the entire request failed. In that case, the statistics indicates: BYTES_UP = 80 and PERSISTED_BYTES_UP = 0
Persisted Down Latency	prdwlt	Total time that is taken to download the data from the cloud that were part of successful GET operation
Persisted Up Latency	pruplt	Total time that is taken to transfer the data to the cloud that were part of successful PUT operation
Failed Gets	flgt	Total number of failed GET operations
Failed Puts	flpt	Total number of failed PUT operations

Table 10. Statistics collection for cloud per cloud account ID (continued)		
Statistic name	**Acronym**	**Description**
Get Errors	gter	Total number of times a read from the cloud failed (including the last retry that failed the GET request)
Get Retries	gtrt	Total number of GET retries
Part Errors	pter	Total number of part errors. It is the count if multi part upload occurs. The part refers to the multi-part upload scenario.
Parts Put	ptpt	Total number of parts that are successfully transferred to the cloud
Persisted parts	prpt	Total number parts successfully persisted on the cloud that were part of successful put operation
Put retries	ptrt	Total number of PUT retries
Throttle upload latency	tuplt	Average delay introduced due to setting upload bandwidth limit
Throttle download latency	tdwlt	Average delay introduced due to setting download bandwidth limit
Throttle upload bandwidth utilization percentage	tupbwpc	Bandwidth utilization in percentage of configured upload bandwidth limit
Throttle download bandwidth utilization percentage	tdwbwpc	Bandwidth utilization in percentage of configured download bandwidth limit

Table 11 on page 23 describes the statistics collection for cloud per VDisk.

Table 11. Statistics collection for cloud per VDisk			
SNo	**Statistic name**	**Acronym**	**Description**
1	blocks up	bup	Number of blocks that are uploaded in the cloud.
2	blocks down	bdn	Number of blocks that are downloaded from cloud.

Note: A block is 512 bytes.

Retrieving statistics files

To view the statistics files, you need to retrieve them from the `/dumps/iostats/` directory on the configuration node.

Use secure copy (**scp**) to copy the statistics files from the configuration node to a local workstation. The system copies all the files from the `/dumps/iostats/` directory on the configuration node to the `/tmp/` directory on a local machine. To verify a particulate file in the individual partnership statistic details, check **`Nn_stats`**.

```
scp superuser@cluster_name:/dumps/iostats/* /tmp
```

IBM Storage Insights

IBM Storage Insights is a cloud-based service that allows you to analyze, protect, and optimize your storage infrastructure.

You can integrate IBM Storage Virtualize systems with IBM Storage Insights by using Call Home with cloud services or by installing lightweight applications that are called *data collectors* into your storage environment. The ability to register and connect to IBM Storage Insights is integrated with the management GUI or command-line interface on your IBM Storage Virtualize system so that you can configure it at any time.

To help you monitor a storage system, IBM Storage Insights collects status, configuration, capacity, and performance metadata about its managed disks (MDisks), volumes and volume copies, nodes, drives, and other resources. This metadata includes statistics and other information that's used for analysis, troubleshooting, and viewing in dashboards, charts, and reports.

A key part of monitoring your system includes the detection of potential ransomware attacks. To ensure that, you have the latest storage metadata for detecting those types of attacks, compression and cyber resiliency statistics for volumes are collected every 5 minutes. With these statistics, IBM Storage Insights builds a historical model of a storage system and uses its built-in intelligence and formulas to identify when and where ransomware attacks might be occurring. For more information on statistics, refer Chapter 17, “Cyber resilience,” on page 353.

There are two versions of the IBM Storage Insights service: IBM Storage Insights and IBM Storage Insights Pro. IBM Storage Insights is available at no cost if you own IBM Block Storage. With this version, you can monitor the basic health, status, and performance of various storage resources.

IBM Storage Insights Pro is a subscription-based product that provides a more comprehensive view of the performance, capacity, and health of your storage resources. In addition to the features offered by IBM Storage Insights, IBM Storage Insights Pro provides tools for intelligent capacity planning, storage reclamation, storage tiering, and performance troubleshooting services. Together, these features can help you reduce storage costs and optimize your data center.

To use Call Home with cloud services, ensure you have the following set up:

1. System memory requirement: Ensure that your storage system has at least 128 GB of system memory to support integrated IBM Storage Insights.
2. IBMid: Ensure that you have an IBMid. You can create an IBM account here Create an IBMid.
3. IBM Storage Insights registration: Register for the IBM Storage Insights service at https://ibm.biz/register.
4. Service name: Create a unique name for your IBM Storage Insights service. To ensure that the name is unique, you can use your company name and another identifying feature, such as a location or department. For example, *Bank Texas* or *Insurance Company Accounting*.

For more information about collecting metadata in IBM Storage Insights, see the following topics:

- Monitoring resources with Call Home, see Monitoring resources with Call Home
- Downloading and installing data collectors, see Downloading and installing data collectors

Related information

“Call Home” on page 503

Call Home connects the system to support personnel who can monitor and respond to system events to ensure that your system remains up and running.

Preventive monitoring

To prevent problems before they occur, you can monitor and verify the system hardware and software periodically.

Some of the following activities can be completed on a daily or weekly basis. Typically, you complete the following activities at least every six months.

System environment

Verify that the prerequisite conditions for the system environment are still met. In particular, verify that temperature and moisture limits are not exceeded, and that airflow and power requirements are still met.

Physically verify the system cabling.

- Inspect the system for any loose cables, open doors, and bent covers.
- Inspect the system for any open slots, which can affect airflow.
- Ensure that wrap plugs are installed on the unused optical ports.
- Ensure that optical cables do not have any bend that is less than 3 inches in radius.
- Check for any abnormal noise that is coming from the system.

Check the error and information LEDs on the system. A system attention light, an enclosure fault light, or a system information light indicates that a serviceable event occurred on the system.

Management GUI

In the management GUI, check the status indicators on the Dashboard. In particular, examine the **System Health** tiles section of the Dashboard. Each tile represents a type of component. Tiles with errors and warnings are displayed first.

In addition to the **System Health** tiles, you can monitor the performance statistics for processor (CPU) utilization, host and internal interfaces, volumes, and storage pools.

Check for updates to the system software, and update the software as needed.

Check that sufficient free physical space (`physical_free_capacity`) is available to meet demand until you plan to add more physical capacity.

Check that each storage pool does not share a resource provisioning group (`shared_resources`) with another storage pool. Check that each storage pool has enough free physical space (`physical_free_capacity`) to meet demand until you plan to add more physical capacity.

Check that each over-provisioned managed disk (`over_provisioned`) has enough free physical space (`physical_free_capacity`) to meet demand until you plan to add more physical capacity.

For more information on the event notifications, see Chapter 15, “Notifications,” on page 291.

Chapter 3. Pools

A pool or storage pool is an amount of capacity that is allocated to volumes that are created in that pool.

Use the Pools page in the management GUI to configure and manage storage pools, internal and external storage, MDisks, and to migrate existing storage to the system.

This figure shows a basic parent pool with associated child pools. In this graphic, the *usable capacity* for the parent group is divided between two child pools. Usable capacity is the amount of capacity that is available for storing data on a parent pool after formatting and RAID techniques are applied. Volumes can then be created by using either the capacity from the MDisks through the parent pool or from the child pool.

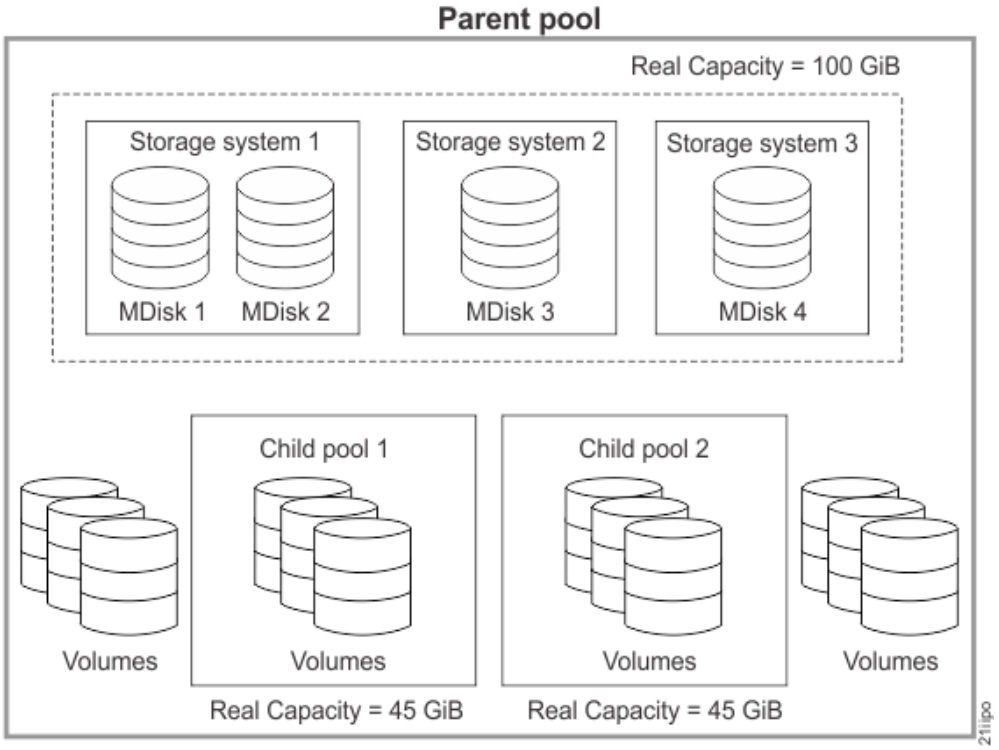


Figure 1. Storage pool

Parent Pools

A parent pool is a collection of managed disks (MDisks). When the pool is created, the managed disks are split into extents. Volumes are created from those extents with the data striped across managed disks.

When you create or manage a parent pool, consider the following general guidelines:

- You specify the extent size when you create a new parent pool. You cannot change the extent size later. The choice of extent size affects the total amount of storage that is managed by the system. Consider the following when determining the extent size for your pool:
  - A system can manage 2^22 extents. For example, with a 2048 MB extent size, the system can manage up to 2048 MB x 4,194,304 = 8 PB of storage.
  - When a volume is created, the storage capacity for the volume is rounded to a whole number of extents.

    Important: You can specify different extent sizes for different systems; however, you cannot migrate (volumes) between systems with different extent sizes. If possible, create all your systems with the same extent size.

© Copyright IBM Corp. 2024

- You can add MDisks to a pool at any time either to increase the number of extents that are available for new volume copies or to expand existing volume copies.

  Attention: If you add an MDisk to a system as an MDisk, any data on the MDisk is lost. If you want to keep the data on an MDisk (for example, because you want to import storage that was previously not managed), you must create image mode volumes instead.
- The **Pools** page in the management GUI displays the **Usable Capacity** and **Capacity Details**. *Usable capacity* indicates the amount of capacity that is available for volumes before any capacity savings methods are applied.
- You can specify a warning threshold for a pool. A warning event is generated when the amount of used capacity in the pool exceeds the warning threshold. The warning threshold is especially useful with thin-provisioned volumes that are configured to automatically use capacity from the pool.
- Volumes can have one or two volume copies. A volume copy is associated with just one pool, except when you migrate a volume copy between parent pools. A volume with two volume copies can have each volume copy in a different pool.
- You can only add MDisks that are in unmanaged mode. When MDisks are added to a parent pool, their mode changes from unmanaged to managed.
- If the volume is mirrored and the synchronized copies of the volume are all in one pool, the mirrored volume is destroyed when the storage pool is deleted. If the volume is mirrored and there is a synchronized copy in another pool, the volume remains after the pool is deleted.

Child Pools

Instead of being created directly from MDisks, child pools are created from existing usable capacity that is assigned to a parent pool. Volumes can be created that specifically use the usable capacity that is assigned to the child pool.

When a standard child pool is created, the usable capacity for a child pool is reserved from the usable capacity of the parent pool. The usable capacity for the child pool must be smaller than the usable capacity in the parent pool. After the child pool is created, the amount of usable capacity that is specified for the child pool is no longer reported as usable capacity of its parent pool. When a data reduction child pool is created, the usable capacity for the child pool is the entire usable capacity of the data reduction parent pool without limit. After a data reduction child pool is created, the usable capacity of the child pool and the usable capacity of the parent pool are reported as the same.

When you create or work with a child pool, consider the following general guidelines:

- You can create and manage child pools in the management GUI.
- As with parent pools, you can specify a warning threshold that alerts you when the used capacity of the child pool is reaching its upper limit. Use this threshold to ensure that access is not lost when the used capacity of the child pool is close to its usable capacity.
- You cannot shrink the usable capacity of a child pool below its used capacity. The system also resets the warning level when the child pool is shrunk and issues a warning if the level is reached when the usable capacity is shrunk.
- On systems with encryption enabled, you can create the standard child pools to migrate existing volumes in a non-encrypted pool to encrypted child pools only when you virtualize external storage. When you create a standard child pool after encryption is enabled, an encryption key is created for the child pool even when the parent pool is not encrypted. You can then use volume mirroring to migrate the volumes from the non-encrypted parent pool to the encrypted child pool. Encrypted data reduction child pools can be created only if the parent pool is encrypted. The data reduction child pool inherits an encryption key from the parent pool.
- The system supports migrating a copy of volumes between child pools within the same parent pool or migrating a copy of a volume between a child pool and its parent pool. Migrations between a source and target child pool with different parent pools are not supported. However, you can migrate a copy of the volume from the source child pool to its parent pool. The volume copy can then be migrated from the

parent pool to the parent pool of the target child pool. Finally, the volume copy can be migrated from the target parent pool to the target child pool.

Child pools can also be assigned to an ownership group. An ownership group defines a subset of users and objects within the system.

Ownership can be defined explicitly or it can be inherited from the user, user group, or from other parent resources, depending on the type of resource. Ownership of child pools must be assigned explicitly, and they do not inherit ownership from other parent resources. New or existing volumes that are defined in the child pool inherit the ownership group that is assigned for the child pool.

With the Safeguarded Copy function, child pools provide a Safeguarded backup location for a group of volumes that are associated with the parent pool. The Safeguarded backup location can contain many snapshots of volume data, each created at a specific interval and with a defined retention period to satisfy your recovery point objective. After the Safeguarded backup location is created, you need to create a volume group and assign a Safeguarded backup policy to the volume group. For more information, see "Creating Safeguarded backup locations" on page 281.

Data Reduction Pools

To use data reduction technologies on the system, you need to create a data reduction pool, create volumes with the data reduction pool, and map these volumes to hosts that support SCSI unmap commands.

Data reduction is a set of techniques that can be used to reduce the amount of usable capacity that is required to store data. An example of data reduction includes data deduplication. Data reduction can increase storage efficiency and performance and reduce storage costs, especially for flash storage. Data reduction reduces the amount of data that is stored on external storage systems and internal drives by reclaiming previously used capacity that are no longer needed by host systems. To estimate potential capacity savings that data reduction can provide on the system, use the Data Reduction Estimation Tool (DRET). This tool analyzes existing user workloads that are being migrated to a new system. The tool scans target workloads on all attached storage arrays, consolidates these results, and generates an estimate of potential data reduction savings for the entire system.

For more information about DRET, see https://www.ibm.com/support/pages/node/6217841. For more information about Comprestimator, see https://www.ibm.com/support/pages/node/6209688.

The system supports data reduction pools which can use different capacity savings methods simultaneously, increasing the capacity savings across the entire pool. Data reduction pools also support deduplication. When deduplication is specified for a volume, duplicate versions of data are eliminated and not written to storage, thus saving more usable capacity. Some models or software versions require specific hardware or software to use this function.

Support for the host SCSI unmap command is enabled by default.

Verify whether the storage system supports data reduction technologies, like data deduplication. If you use storage systems that support data reduction technologies, you can also configure data reduction on the storage systems. The storage system can reclaim that freed storage and reorganize the data on other volumes to more efficiently use the capacity. For standard-provisioned volumes, the system fully controls storage on these storage systems. When a volume is deleted, capacity is freed on the system and can be reallocated; the storage system is not aware of this freed space. However, if the storage system uses compression, thin-provisioning, or deduplication, the storage system controls the use of the usable capacity. In this configuration, when capacity is freed, the system notifies the storage system that capacity is no longer needed. The storage system can then reuse that capacity or free it as reclaimable capacity. The system also supports reclaimable capacity from certain internal drives, such as the 15 TB tier 1 flash drives, which can improve performance on these types of drives.

When you create a data reduction pool, ensure that the usable capacity of the pool includes *overhead capacity*. Overhead capacity is an amount of usable capacity that contains the metadata for tracking unmap and reclaim operations within the pool. A general guideline is to ensure that the provisioned capacity with the data reduction pool does not exceed 85% of the total usable capacity of the data

reduction pool. Table 12 on page 30 includes the minimum data reduction pool capacity that is required to be able to create a volume within the pool.

Table 12. Minimum overhead capacity requirements for data reduction pools

Extent size (in gigabytes)	**Overhead capacity requirements (in terabytes)[1]**
1 GB or smaller	1.1 TB
2 GB	2.1 TB
4 GB	4.2 TB
8 GB	8.5 TB

[1]Standard-provisioned volumes are not included into the minimum overhead capacity values. When you are planning usable capacity for data reduction pools, determine the usable capacity that is needed for any standard-provisioned volumes first, then ensure that the minimum usable capacity values for the data reduction pools are included.

Pool states

This table describes the operational states of a pool. Child pools adopt the state of the parent pool. States that indicate an error must be resolved on the parent pool.

Table 13. Pool states

State	**Description**
Online	The pool is online and available. All the MDisks in the pool are available.
Degraded paths	This state indicates that one or more nodes in the system cannot access all the MDisks in the pool. A degraded path state is most likely the result of incorrect configuration of either the storage system or the Fibre Channel fabric. However, hardware failures in the storage system, Fibre Channel fabric, or node might also be a contributing factor to this state. To recover from this state, follow these steps: 1. Verify that the fabric configuration rules for storage systems are correct. 2. Ensure that you configured the storage system properly. 3. Correct any errors in the event log.

Table 13. Pool states (continued)	
State	**Description**
Degraded ports	This state indicates that one or more 1220 errors were logged against the MDisks in the pool. The 1220 error indicates that the remote Fibre Channel port was excluded from the MDisk. This error might cause reduced performance on the storage system and usually indicates a hardware problem with the storage system. To fix this problem, you must resolve any hardware problems on the storage system and fix the 1220 errors in the event log. To resolve these errors in the log, click **Monitor** > **Events** in the management GUI. This action displays a list of unfixed errors that are currently in the event log. For these unfixed errors, select the error name to begin a guided maintenance procedure to resolve them. Errors are listed in descending order with the highest priority error listed first. Resolve highest priority errors first.
Offline	The pool is offline and unavailable. No nodes in the system can access the MDisks. The most likely cause is that one or more MDisks are offline or excluded.  **Attention:** If a single MDisk in a pool is offline and cannot be seen by any of the online nodes in the system, the pool of which this MDisk is a member goes offline. This causes all of the volume copies that are being presented by this pool to go offline. Take care when you create pools to ensure an optimal configuration.

Easy Tier

For more information, see “Easy Tier” on page 62.

A provisioning policy defines a set of rules for allocating capacity from a pool. For more information, see “Provisioning policies” on page 195.

Arrays

The system can manage a combination of internal and external storage and represents all the detected storage as drives.

You can select to have the system automatically assign all detected internal drives to pools, or select to customize the configuration. When the system automatically assigns drives, pools are created based on the available drive classes and recommended array configurations. In this case, the system might not use all available drives if they do not meet configuration recommendations. If you choose to customize, you can adjust the configuration based on the number of drives and drive criteria. These arrays are presented as MDisks, which can be added to pools.

Array configurations

An array is an ordered group of physical devices (drives) that is used to define logical volumes or devices. An array is a type of MDisk that is made up of drives; these drives are *members* of the array. A Redundant

Array of Independent Disks (RAID) is a method of configuring member drives to create high availability and high-performance systems.

The system supports distributed RAID array configurations. *Distributed RAID array* that contains more than two drives have a minimum of one rebuild area. All the drives in an array can process I/O data and provide faster rebuild times when a drive fails. The RAID level provides different degrees of redundancy and performance; it also determines the minimum and maximum number of member drives in the array.

In order to restore redundancy for degraded distributed RAID arrays, the *rebuild-in-place* process is used for reconstructing the data (or parity) directly back into the replaced member drive. Degraded distributed RAID arrays with only two member drives use the *rebuild-in-place* process to restore redundancy, copying the data directly back into the replaced member drive.

If encryption is enabled on your system, you can create RAID arrays that are encrypted. For more information, see "Configuring encryption" on page 453.

Distributed RAID arrays

Distributed RAID array configurations typically contain rebuild areas that are used to maintain redundancy after a drive failure. RAID array configurations that do not contain rebuild areas still maintain redundancy after a drive failure, by using *rebuild-in-place*. Data is rebuilt and redundancy is restored much more rapidly.

For more information and details about distributed RAID arrays, see "Distributed RAID array properties" on page 33.

You can expand distributed RAID arrays to increase the available capacity of the array or add usable capacity. As part of the expansion, the system automatically migrates data for optimal performance for the new expanded configuration. Expansion of distributed RAID arrays supports incremental growth of available capacity for arrays and is compatible with other functions, such as IBM Easy Tier and data migrations.

The type of RAID array configuration that the system uses is based on the type of drives that are detected through the serial-attached SCSI (SAS) or Non-Volatile Memory express (NVMe) protocols. If your system contains SAS-attached drives only, the system can configure distributed RAID arrays based on the number of drives available and the characteristics of those drives. Distributed RAID arrays are preferred because distributed RAID arrays provide better performance and rebuild times. For systems that support only NVMe drives, only distributed RAID arrays are created for array configuration since these drives have improved performance for high-demand drives. Distributed RAID arrays are also supported for compressing drives. NVMe drives are only supported within certain models of control enclosures. For systems that contain both NVMe and SAS drives, the system restricts mixing these different types of drives in the same RAID array configuration. Drive members within the array must be as similar as possible to avoid performance problems and issues replacing drives.

The system also supports NVMe drives that are self-compressing, where the drive itself compresses data as it is written to the drive. If the array includes self-compressing drives, the system uses the following rules when the array is created:

- Non-compressing drives and self-compressing drives cannot be in the same RAID array configuration.
- All drives in the array have the same usable capacity and use the same compression algorithm.
- If the array contains at least 7 drives, the system recommends a distributed RAID 6 array, which is preferred to ensure capacity for self-compressing drives during rebuild operations.
- Only one distributed RAID array of self-compressing drives is allowed per storage pool.

Distributed RAID array configuration guidelines

The management GUI automatically defaults to the recommended number of rebuild areas that is based on the current width and other settings for the distributed RAID array. The width of the array includes the number of physical drives and rebuild areas that are needed to ensure redundancy if drives fail. The system can recommend a maximum of four rebuild areas for distributed RAID arrays for SAS-attached drives and non-compressing NVMe distributed RAID arrays. Compressing NVMe distributed RAID arrays

support one rebuild area. Other settings, such as drive class, are also used to determine the best configuration for arrays. For example, if you are adding storage or expanding a distributed RAID array, the management GUI disables any drive classes that are incompatible with drives already in the pool. It then analyzes the current array configuration in the pool to recommend stripe widths and drive counts.

After system setup, you must configure storage by creating pools and assigning storage to specific pools. Ensure that a pool or pools are created before you assign storage. In the management GUI, select **Pools** > **Actions** > **Add Storage**. The **Add Storage** selection automatically configures existing drives into arrays based on array characteristics.

Use the **lsarrayrecommendation** command to display the system recommendations for configuring an array. To configure a distributed RAID array, you can use the **mkdistributedarray** command. For more information on arrays commands, refer to the **Array command**.

The management GUI does not provide array recommendation if the following conditions are present:

- The selected pool already contains a distributed RAID array with compressing drives.
- The selected drive class might not be compatible with other drives that are already in the pool.
- The number of selected drives is not enough to create a valid RAID array.
- The extent size is smaller than the minimum extent size.

If these conditions are present, you can take one of the following actions:

- Select a different drive class.
- Select a different storage pool.
- Select a different number of candidate drives.
- Expand an existing array in the pool instead of adding a new array.

For more information on supported array types and RAID levels, see Configuration limits.

Distributed RAID array properties

Distributed RAID arrays contain rebuild areas that are used to maintain redundancy after a drive fails.

This is true of all distributed RAID arrays, except for distributed RAID 1 arrays with only 2 drives. If not enough drives are available on the system, you cannot configure a distributed RAID array. Distributed RAID arrays solve rebuild bottlenecks in array configurations because rebuild areas are distributed across all the drives in the array.

The rebuild write workload is spread across all the drives rather than just the single spare drive, which results in faster rebuilds on an array. Distributed RAID arrays remove the need for separate drives that are idle until a failure occurs. Instead of allocating one or more drives as spares, the spare capacity is distributed over specific rebuild areas across all the member drives. Data can be copied faster to the rebuild area and redundancy is restored much more rapidly. Additionally, as the rebuild progresses, the performance of the pool is more uniform because all of the available drives are used for every volume extent. After the failed drive is replaced, data is copied back to the drive from the distributed rebuild areas. Unlike hot-spare drives, read/write requests are processed on other parts of the drive that are not being used as rebuild areas. The number of rebuild areas is based on the width of the distributed RAID array and on the technology of the member drives.

For example, a distributed RAID array that uses RAID 6 drives can handle two concurrent drive failures. After the two failed drives are rebuilt, the array can, once again, tolerate another two drive failures. If all of the rebuild areas are used, the array becomes degraded on the next drive failure.

The effective capacity of a distributed RAID 1 array that is made from the same number of underlying drives as a RAID 5 or RAID 6 array is in general less than the effective capacity of the latter. However, while capacity-wise, distributed RAID 5 and 6 are generally better, performance-wise, RAID 1 may be better for certain workloads.

Supported RAID levels

The system supports the following RAID levels for distributed RAID arrays:

Distributed RAID 1

Distributed RAID 1 arrays consist of two mirrored strips, which are distributed across all member drives. Distributed RAID 1 arrays can tolerate a single failed member drive when a rebuild area is in place. If a member drive fails in a distributed RAID 1 array contains two member drives, or the rebuild area is no longer available, the array becomes degraded. Unlike distributed RAID 5 and 6, distributed RAID 1 does not contain any parity strips. In order to restore redundancy for degraded distributed storage arrays, the *rebuild-in-place* process is used, reconstructing the data directly back into the replaced member drive. Degraded distributed RAID arrays with only two member drives use the *rebuild-in-place* process to restore redundancy, copying the data directly back into the replaced member drive.

Distributed RAID 5

Distributed RAID 5 arrays stripe data over the member drives with one parity strip on every stripe. These distributed RAID arrays can support 4 - 128 drives. RAID 5 distributed arrays can tolerate only one failed member drive. In order to restore redundancy for degraded distributed RAID arrays, the *rebuild-in-place* process is used for reconstructing the data (or parity) directly back into the replaced member drive.

Distributed RAID 6

Distributed RAID 6 arrays stripe data over the member drives with two parity strips on every stripe. Distributed RAID 6 arrays can tolerate any two concurrent member drive failures. In order to restore redundancy for degraded distributed RAID arrays, the *rebuild-in-place* process is used for reconstructing the data (or parity) directly back into the replaced member drive.

Example of a distributed RAID array with no rebuild areas

Note: While this example is showing a 2 member drive storage array (with distributed RAID 1), the same rebuild-in-place process occurs in any distributed storage arrays if the member drives have failed to an extent that no rebuild area is available. In these cases, the data is reconstructed, not copied, onto the replaced drive.

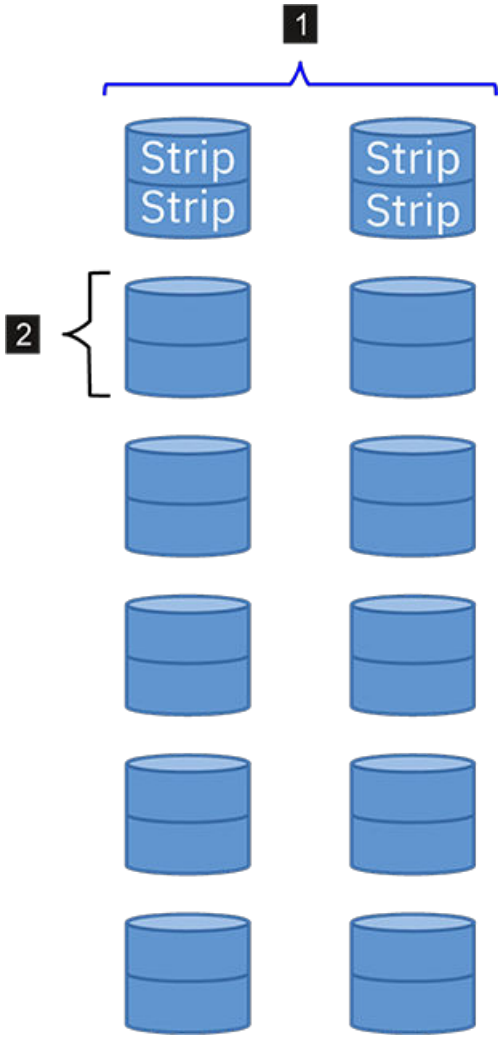


Figure 2. Two member drive storage array with no rebuild area

Figure 2 on page 35 shows an example of a distributed RAID array with two member drives and no rebuild area; both of the drives in the array are active.

1 (Minimum) two active drives and stripe width
2 Pack, with depth of two strips

Figure 3 on page 36 shows a distributed RAID array with two member drives that contains a failed drive. To recover data, the data copies the strip from the active drive to the new or previously failed drive.

1 Active drive
2 Failed drive

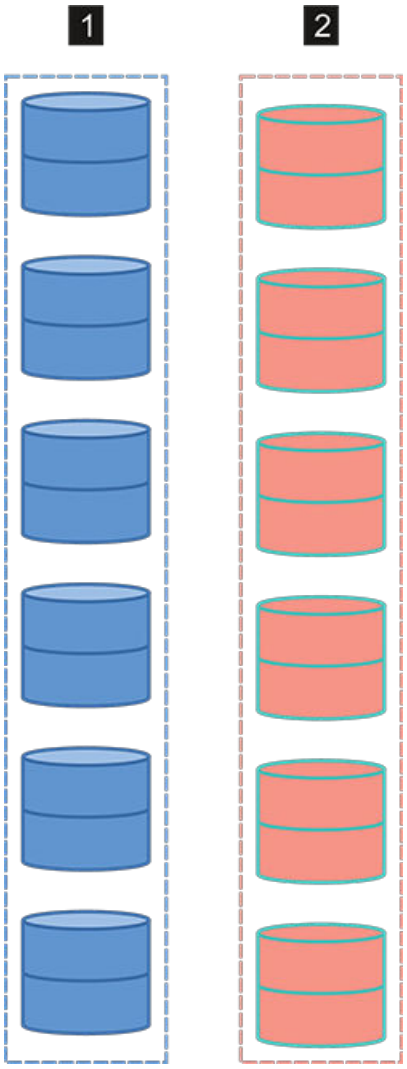


Figure 3. Two member drive storage array with a failed drive

Figure 4 on page 37 shows the rebuild-in-place process, with the new data being copied directly into the replaced drive.

1 Active drive
2 Data being copied directly into the replaced drive

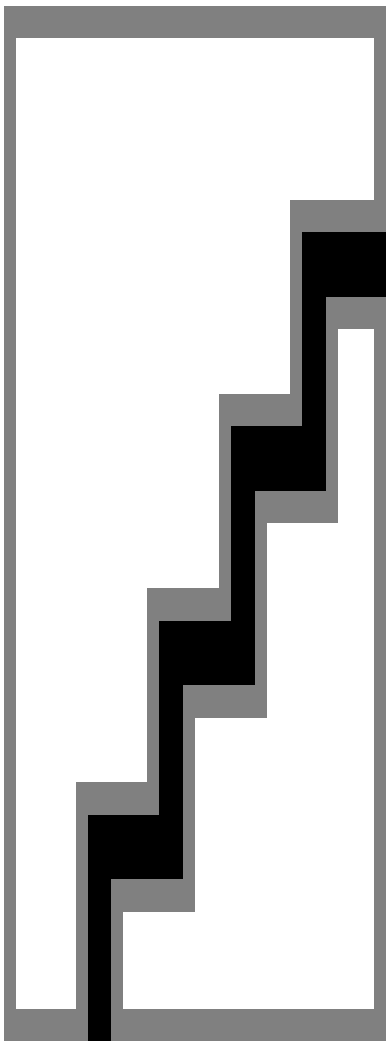


Figure 4. Rebuild-in-place process on the two member drive

Example of a distributed RAID array with rebuild areas

Figure 5 on page 38 shows an example of a distributed RAID array that is configured with RAID 6; all of the drives in the array are active. The rebuild areas are distributed across all of the drives and the drive count includes all of the drives.

1. An active drive
2. Rebuild areas, which are distributed across all drives.
3. Drive count, which includes all drives.
4. Stripes of data (two stripes are shown)
5. Stripe width
6. Pack, which equals the drive count that is multiplied by stripe width.
7. Additional packs in the array (not shown)

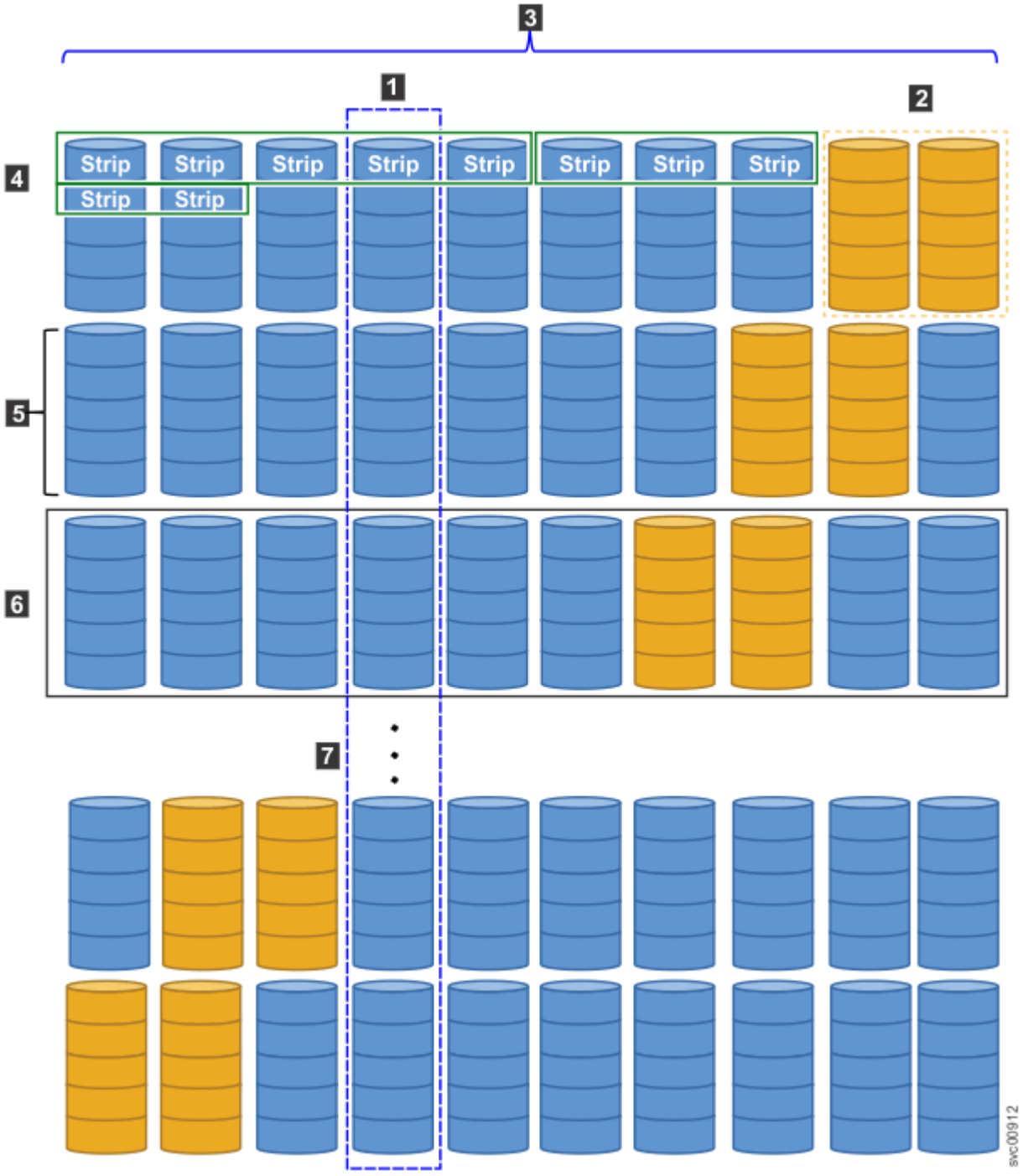


Figure 5. Distributed RAID array (RAID 6 level)

Figure 6 on page 39 shows a distributed RAID array that contains a failed drive. To recover data, the data is read from multiple drives. The recovered data is then written to the rebuild areas, which are distributed across all of the drives in the array. The remaining rebuild areas are distributed across all drives.

1 Failed drive
2 Rebuild areas, which are distributed across all drives.
3 Remaining rebuild areas rotate across each remaining drive.
4 Additional packs in the array (not shown)

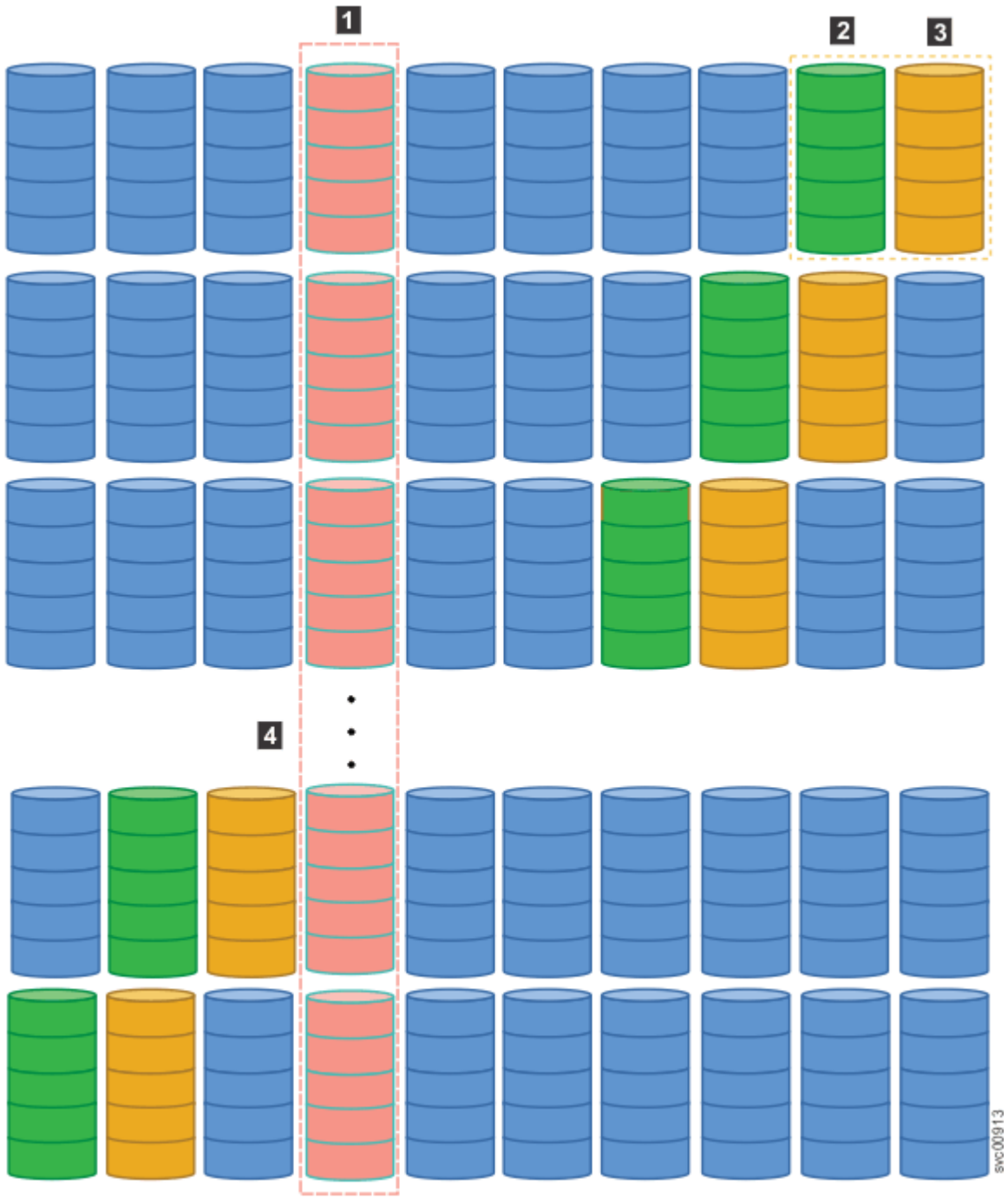


Figure 6. Distributed RAID array (RAID 6 level) with a failed drive

Array width

The *array width*, which is also referred to as the *drive count*, indicates the total number of drives in a distributed RAID array. This total includes the number of drives that are used for data capacity and parity, (if there are parity strips, like in distributed RAID 6), and the *rebuild area* (if existing) that is used to restore redundancy of the data.

Rebuild area

The *rebuild area* is the drive capacity that is reserved within a distributed RAID array to reconstruct data after a drive failure; it provides no usable capacity. The rebuild area is distributed across all of the

drives in the array. As data is rebuilt during the *copyback* process, the rebuild area contributes to the performance of the distributed RAID array because all of the member drives perform I/O requests.

Note: Distributed RAID 1 systems with only two member drives do not contain a rebuild area.

Rebuild-in-place

A distributed RAID array process that takes place when no rebuild area is available, which causes the array to become degraded. The *rebuild-in-place* process restores data redundancy by copying or reconstructing the data directly back into the replaced member drive by using the original data distribution.

Note: Not all distributed RAID array storage systems might support the *rebuild-in-place* process. For more information, see "Supported RAID levels" on page 34.

Stripe and stripe width

A *stripe*, which can also be referred to as a *redundancy unit*, is the smallest amount of data that can be addressed. Distributed RAID array strip size is 256 KiB.

The *stripe width* indicates the number of strips of data that can be written at one time when data is rebuilt after a drive fails. This value is also referred to as the *redundancy unit width*. In Figure 5 on page 38, the stripe width of the array is 5.

Drive class

To enhance performance of a distributed RAID array, all of the drives must come from the same, or superior, *drive class*. Each drive class is identified by its *drive_class_id*. The system uses the following information to determine the drive class of each drive:

Block size
Indicates the block size of the drive class. Valid block size is either 512 or 4096 bytes.

Capacity
Indicates the capacity of the drive class.

I/O group
Indicates the I/O group name that is associated with the drive class.

RPM speed
Indicates the speed of the drive class. Valid RPM speed can be 7.2 K, 10 K, or 15 K. For SSDs, this value is blank.

Technology
Indicates the technology for the drive class. The following technology types are supported:

Storage Class Memory
Storage Class Memory drives use persistent memory technologies that improve endurance and reduce latency of current flash storage device technologies. All SCM drives use NVMe architecture.

Tier 0 flash
Tier 0 flash drives are high-performance flash drives that process read and write operations and provide faster access to data than Enterprise or nearline drives. For most Tier 0 flash drives, the system monitors their wear level and issues warnings when the drive is nearing replacement. Some NVMe-attached drives, such as IBM FlashCore® Module drives, are considered Tier 0 flash drives.

Tier 1 flash
Tier 1 flash drives are lower-cost flash drives, typically with larger capacities, but slightly lower performance and write endurance characteristics. The system monitors their wear level and issues warnings when the drive is nearing replacement.

Enterprise disks
Enterprise disks are disk drives that are optimized for performance.

Nearline disks
Nearline disks are disk drives that are optimized for capacity.

To replace a failed member drive in the distributed RAID array, the system can use another drive that has the same drive class as the failed drive. The system can also select a drive from a superior drive class. For example, two drive classes can contain drives of the same technology type but different data capacities. In this case, the superior drive class is the drive class that contains the higher capacity drives.

To display information about all of the drive classes that are available on the system, use the **lsdriveclass** command. Example output from the lsdriveclass command shows two drive classes on the system. Drive class 0 contains drives with a capacity of 837.9 GB and drive class 1 contains drives with a capacity of 1.1 TB. As the drives have the same technology type and block size, drive class 1 is considered to be superior to drive class 0 as it has greater capacity.

Example output from the `lsdriveclass` command

```
id RPM   capacity IO_group_id IO_group_name tech_type        block_size  candidate_count
superior_count total_count  transport_protocol compressed  fips
0 10000  837.9GB  0           io_grp0       tier0_enterprise 512         11
22             11           sas              no          no
1 10000  1.1TB    0           io_grp0       tier0_enterprise 512         11
11             11           sas              no          no
```

Slow write priority settings

When a redundant array is doing read/write I/O operations, the performance of the array is bound by the performance of the slowest member drive. If the SAS network is unstable or if too much work is being driven to the array when drives do internal ERP processes, performance to member drives can be far worse than usual. In this situation, arrays that offer redundancy can accept a short interruption to redundancy to avoid writing to, or reading from, the slow component. Write operations that are mapped to a poorly performing drive are committed to the other copy or parity, and are then completed with good status (assuming no other failures). When the member drive recovers, the redundancy is restored by a background process of writing the strips that were marked out of sync while the member was slow.

This technique is governed by the setting of the **slow_write_priority** attribute of the distributed RAID array, which defaults to **latency** when the array is created. When set to **latency**, the array is allowed to become out of sync in an attempt to smooth poor member performance. You can use the command to change the **slow_write_priority** attribute to **redundancy**. When set to **redundancy**, the array is not allowed to become out of sync. However, the array can avoid suffering read performance loss by returning reads to the slow component from redundant paths.

When the array uses **latency** mode or attempts to avoid reading a component that is in **redundancy** mode, the system evaluates the drive regularly to assess when it becomes a reliable part of the system again. If the drive never offers good performance or causes too many performance failures in the array, the system fails the hardware to prevent ongoing exposure to the poor-performing drive. The system fails the hardware only if it cannot detect another explanation for the bad performance from the drive.

Distributed RAID array member drive replacement

If the fault LED light on a drive is lit, the drive is marked as failed and is no longer used in the distributed RAID array. When the system detects that a failed drive was replaced, it automatically removes the failed hardware from the array configuration. If the new drive is suitable (for example, in the same drive class), the system begins a copyback operation to make a rebuild area available again in the distributed RAID array.

Planning array configurations

When you plan your network, consideration must be given to the RAID configuration that you use. The system supports distributed RAID 1, 5, and 6 array configurations.

Distributed array

The concept of distributed RAID arrays is to distribute an array with width W across a set of X drives. For example, you might have a 9+P+Q+R RAID 6 array that is distributed across a set of 40 drives. The array type and width define the level of redundancy. In this example, a 25% capacity overhead for parity and rebuild areas exists.

If an array stride needs to be rebuilt, two component strips must be read to rebuild the data for the third component. The set size defines how many drives are used by the distributed array. Performance and usable capacity scale according to the number of drives in the set. The other key feature of a distributed array is that instead having a hot spare, the set includes distributed rebuild area strips that are also distributed across the set of drives. The data and distributed rebuild areas are distributed such that if one drive in the set fails redundancy can be restored by rebuilding data on to the spare distributed rebuild areas at a rate much greater than the rate of a single component.

Note: When you determine the array configuration for your system, plan to create arrays of 256 KiB strip size only.

Distributed arrays of NVMe drives support are used to create large-scale internal managed disks. As a result, rebuild times are dramatically reduced, which decreases the volumes' exposure to the extra load of recovering redundancy. Because the capacity of these managed disks is potentially so great, when they are configured in the system the overall limits change to allow them to be virtualized. For every distributed array, the space for 16 MDisk extent allocations is reserved. Therefore, 15 other MDisk identities are removed from the overall pool of 4096. Distributed arrays also aim to provide a uniform performance level. A distributed array can contain multiple drive classes if the drives are similar (for example, the drives have the same attributes, but the capacities are larger) to achieve this performance. All the drives in a distributed array must come from the same I/O group to maintain a simple configuration model.

A distributed array has the following key benefits:

- Quicker rebuild times with less impact to host I/O.
- More user flexibility in defining how many drives are used by an array (for example, a user can create 9+P arrays with 24 drives without having four drives left unused).
- Distributed rebuild areas means that no idle drives exist in the system, and thus performance improves slightly.
- Degraded distributed RAID arrays can use the *rebuild-in-place* process (if supported) to restore redundancy, copying or reconstructing the data directly back into the replaced member drive.

One disadvantage of a distributed array is that the array redundancy is covering a greater number of components. Therefore, mean time between failure (MTBF) is reduced. Quicker rebuild times improve MTBF; however, limits to how widely distributed an array can be before the MTBF becomes unacceptable remain.

Distributed array expansion

Distributed array expansion allows the conversion of a small, not-very-distributed array into a larger distributed array while it preserves volume configuration and restriping for optimal performance. Expansion offers the option of getting better rebuild performance with an existing configuration without the migration steps that might require excess capacity. Expanding a distributed array is preferable over creating a new small array.

Expansion can increase the capacity of an array, but it cannot change the basic parameter of stripe width. When you plan a distributed array configuration, it is necessary to plan for future array requirements. Additionally, a distributed array that might fit within the extent limit (16*128 K extents) at a particular

extent size might not fit if you expand it over time. Planning your extent size for the future is also important. The minimum (and recommended) storage pool extent size is 4096 MiB.

Expansion also benefits NVMe arrays for the same reasons. However, thin provisioned (compressing) NVMe drives add an extra layer of complexity when you calculate the available capacity in the array during an expansion. When you plan for the possible expansion of thin provisioned NVMe arrays, the drives must be the same physical and logical size. When you expand a thin provisioned NVMe array, the usable capacity is not immediately available, and the availability of new usable capacity does not track with logical expansion progress. The expansion process monitors usable capacity usage and analyzes the changes that are caused by the actions it takes during data restriping. This information is used to release the correct amount of usable capacity as it becomes available.

Expanding an MDisk in a distributed array

Distributed RAID expansion

You can expand distributed RAID arrays to increase the available capacity of the array or add usable capacity. As part of the expansion, the system automatically migrates data for optimal performance for the new expanded configuration. Expansion of distributed RAID arrays supports incremental growth of available capacity for arrays and is compatible with other functions, such as IBM Easy Tier and data migrations.

You can dynamically add one or more drives and increase the capacity of an existing distributed RAID array that is already in use. The expansion process is dynamic and nondisruptive, which means that you can continue to access the data on the distributed RAID array. New drives are integrated and data is re-striped to maintain the placement of strips across the existing and new components. *Stripe width* does not change during an expansion. This means that when an array is expanded, it would allocate parity in the same ratio. Therefore, not as much of the new drives would be additional capacity compared to an equivalent array with the same number of drives with a higher stripe width, as the ratio of data strips to parity strips would be lower. Expansion does not allow you to change the stripe width or geometry. For more information about *stripe width* and *strips*, see “Distributed RAID array properties” on page 33.

Note: When you expand a thin-provisioned NVMe array, the physical capacity is not immediately available, and the availability of new physical capacity does not track with logical expansion progress.

Using the management GUI

To expand a distributed RAID array, complete these steps:

1. In the management GUI, select **Pools** > **MDisks by Pools**.
2. Expand the pool and right-click an MDisk within a distributed RAID array and select **Expand**.
3. Select the drive class that is compatible with the drives in the original MDisk and enter the number of drives to expand the array.
4. Expand **Advanced Settings** to view the number of additional drives that is recommended for the new configuration.
5. Click **Expand**. After starting the expand operation, you can monitor the progress of the expansion by selecting **Monitoring** > **Background Tasks**.

   Note: The **Progress Remaining** value represents the expected time of completion for the data migrations that are required during the expansion of the array. The **Progress Remaining** value can be empty during processing when data migrations are not required or have not started. In these cases, use the percentage value in the **Progress** column or the value displayed in the **Additional Capacity Remaining** field to estimate completion of the expansion.

Using the command-line interface

RAID write response time

This function means that the RAID software layer, where redundancy exists to do so, can prevent drive bad behavior from having an unlimited impact on I/O performance. In addition, the system tries to avoid immediately committing to an array rebuild due to a brief offline event from a single drive, while there is full redundancy.

Enabling RAID latency mode

The system temporarily reduces the array redundancy when drive latency impacts a write task. There is a task that is associated with the repair operation to catch up the out of sync component. Other background tasks stop running when drives are slow to lower the workload and expedite recovery.

If the preference is that write I/O does not cause loss of sync, then the CLI supports this function that is disabled on an array-by-array basis through the **slowwritepriority** parameter to the **charray** and **mkdistributedarray** commands. This parameter does not affect the RAID layer efforts to return reads promptly when drives misbehave. The **slowwritepriority** parameter also does not prevent the array from attempting to avoid a full component rebuild by using an *incremental rebuild* when drives go offline (also displayed as a *resync* task).

If there is excessive use of the function due to I/O, then the availability of the data can be at risk. To warn of an excessive loss of sync, the array logs an event to help identify if a particular member is the cause. The system monitors drives and fails a drive that is consistently causing slow events.

The response time target is not active in the following situations:

- On a RAID 1 mirror where one mirror is rebuilding
- On a RAID 6 array with a rebuilding component, and it is that component that is slow
- During system events

MDisks

A *managed disk (MDisk)* is a logical unit of physical storage. MDisks are not visible to host systems.

An MDisk might consist of multiple physical disks that are presented as a single logical disk to a network. An MDisk always provides usable blocks of physical storage to the system even if it does not have a one-to-one correspondence with a physical disk.

Each MDisk is divided into a number of extents, which are numbered, from 0, sequentially from the start to the end of the MDisk. The extent size is a property of pools. When an MDisk is added to a pool, the size of the extents that the MDisk is divided into depends on the attribute of the pool to which it was added.

The access mode determines how the clustered system uses the MDisk.

Attention: If you observed intermittent breaks in links or if you replaced cables or connections in the SAN fabric or LAN configuration, you might have one or more MDisks in degraded status. If an I/O operation is attempted when a link is broken and the I/O operation fails several times, the system partially excludes the MDisk and changes the status of the MDisk to excluded. You must include the MDisk to resolve the problem.

Access modes

The following list describes the possible access modes for MDisks:

Unmanaged
The MDisk is not used by the system.

Managed
The MDisk is assigned to a storage pool and provides extents that volumes can use.

Image
The MDisk is assigned directly to a volume with a one-to-one mapping of extents between the MDisk and the volume.

Array

The MDisk represents a set of drives in a RAID from internal storage.

Attention: If you add an MDisk that contains existing data to a storage pool while the MDisk is in unmanaged or managed mode, you lose the data that it contains. The *image mode* is the only mode that preserves this data.

Figure 7 on page 45 shows examples of physical disks and MDisks.

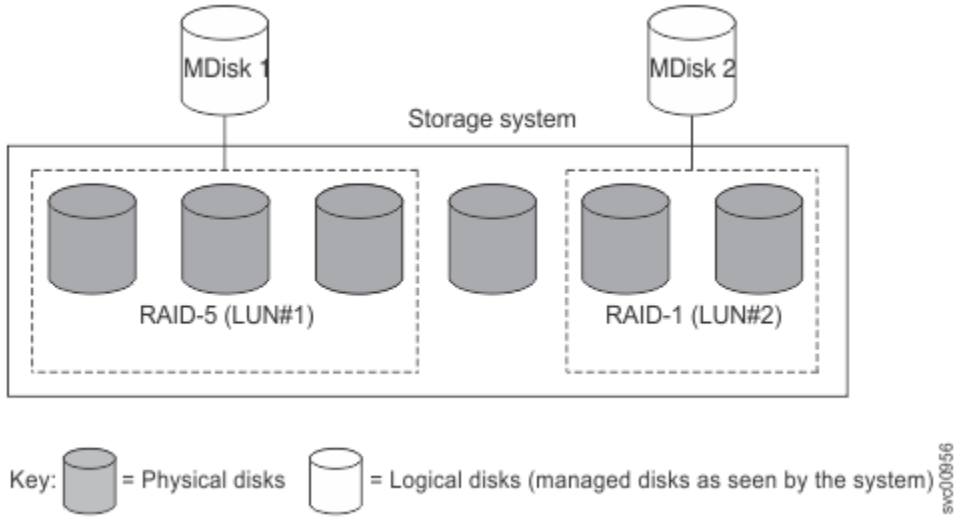


Figure 7. Storage systems and MDisks

Table 14 on page 45 describes the operational states of an MDisk.

Table 14. MDisk status

Status	Description
Online	The MDisk can be accessed by all online nodes. All the nodes that are currently working members of the system can access this MDisk. The MDisk is online when the following conditions are met: • All timeout error-recovery procedures complete and report the disk as online. • Logical unit number (LUN) inventory of the target ports correctly reported the MDisk. • Discovery of this LUN completed successfully. • All of the MDisk target ports report this LUN as available with no fault conditions. • For iSCSI connections, both nodes of at least one I/O group in the system can access the MDisk.

Table 14. MDisk status (continued)	
Status	**Description**
Degraded paths	The MDisk is not accessible to one or more nodes in the system. Degraded path status is most likely the result of incorrect configuration of either the storage system or the SAN fabric. However, hardware failures in the storage system, SAN fabric, or node might also be a contributing factor to this state. To recover from this state, follow these steps: 1. Verify that the fabric configuration rules for storage systems are correct. 2. Ensure that you configured the storage system properly. 3. Correct any errors in the event log.
Degraded ports	The MDisk has one or more 1220 errors in the event log. The 1220 error indicates that the remote Fibre Channel port was excluded from the MDisk. This error might cause reduced performance on the storage system and usually indicates a hardware problem with the storage system. To fix this problem, you must resolve any hardware problems on the storage system and fix the 1220 errors in the event log.
Excluded	The MDisk was excluded from use by the system after repeated access errors. Run the directed maintenance procedures to determine the problem.
Offline	The MDisk cannot be accessed by any of the online nodes. None of the nodes that are currently working members of the system can access this MDisk. This state can be caused by a failure in the SAN, storage system, or one or more physical disks that are connected to the storage system. The MDisk is reported as offline if all paths to the disk fail.

Extents

Each MDisk is divided into chunks of equal size called *extents*. Extents are a unit of mapping that provides the logical connection between MDisks and volume copies.

MDisk path

Each MDisk from external storage has an online path count, which is the number of nodes that have access to that MDisk. The path count represents a summary of the I/O path status between the system nodes and the storage device. The maximum path count is the maximum number of paths that were detected by the system at any point in the past. If the current path count is not equal to the maximum path count, the MDisk might be degraded. That is, one or more nodes might not see the MDisk on the fabric.

Troubleshooting medium errors and bad blocks

A storage system returns a medium error response to a host when it is unable to successfully read a block. The system response to a host read follows this behavior.

The system allocates volumes from the extents that are on the managed disks (MDisks). The MDisk can be a volume on an external storage controller or a RAID array that is created from internal drives. In either case, depending on the RAID level that is used, there is normally protection against a read error on a single drive. However, it is still possible to get a medium error on a read request if multiple drives have errors or if the drives are rebuilding or are offline due to other issues.

The system provides migration facilities to move a volume from one underlying set of physical storage to another. In all these cases, the migrated volume or the replicated volume returns a medium error to the host when the logical block address on the original volume is read. The system maintains tables of bad blocks to record where the logical block addresses that cannot be read are. These tables are associated with the MDisks that are providing storage for the volumes.

The **dumpmdiskbadblocks** command and the **dumpallmdiskbadblocks** command are available to query the location of bad blocks.

Important: The **dumpmdiskbadblocks** outputs the virtual medium errors that are created, and not a list of the actual medium errors on MDisks or drives.

It is possible that the tables that are used to record bad block locations can fill up. The table can fill either on an MDisk or on the system as a whole. If a table does fill up, the migration or replication that was creating the bad block fails because it was not possible to create an exact image of the source volume.

The system creates alerts in the event log for the following situations:

- When it detects medium errors and creates a bad block
- When the bad block tables fill up

This table lists the bad block error codes. The recommended actions for these alerts guide you in correcting the situation.

Table 15. Bad block errors

Error code	Description
1840	The managed disk has bad blocks. On an external controller, this error must be a copied medium error.
1226	The system fails to create a bad block because the MDisk already has the maximum number of allowed bad blocks.
1225	The system fails to create a bad block because the system already has the maximum number of allowed bad blocks.

Clear bad blocks by deallocating the volume disk extent, by deleting the volume or by issuing write I/O to the block. It is good practice to correct bad blocks as soon as they are detected. This action prevents the bad block from being propagated when the volume is replicated or migrated. However, it is possible for the bad block to be on part of the volume that is not used by the application. For example, it can be in part of a database that is not initialized. These bad blocks are corrected when the application writes data to these areas. Before the correction happens, the bad block records continue to use up the available bad block space.

External storage

An external storage system, or *storage controller*, is a device that coordinates and controls the operation of one or more disk drives.

A storage system synchronizes the operation of the drives with the operation of the system as a whole. IBM Storage Virtualize for Public Cloud software only supports cloud storage as external storage. Cloud storage is logical disks at the storage system level. Cloud-based storage is purchased and configured from a supported cloud-based provider.

The system can manage a combination of internal and external storage and represents all the detected storage as MDisks. The management GUI can be used to configure both internally and externally attached storage. Data that is stored on external storage can also be migrated to internal storage, simplifying data and storage management. The system supports Fibre Channel and iSCSI-attached storage systems. Fibre Channel attached storage is automatically detected through the SAN fabric, but iSCSI storage must be manually added to the system for virtualization.

External storage systems reside on the SAN fabric and are addressable by one or more worldwide port names (WWPNs). An external storage system might contain one or more logical units (LUs), each identified by a different logical unit number (LUN). External storage systems that are managed by the system typically contain multiple LUs.

Select **Pools** > **External Storage** in the management GUI to view all MDisks detected by the system. The MDisks are organized by the external storage system that presents them.

For the latest support information about external storage systems, see the IBM support website.

Planning an external storage system

To use Logical Units (LUs) of the external storage systems as Managed Disks (MDisks), several configuration requirements need to be followed to avoid performance and data integrity issues.

Use the following configuration guidelines that describe the system protocol for access to external storage systems LUs through multiple target ports.

External storage-system configuration-requirements

Follow the guidelines to configure the external storage systems arrays.

1. Avoid splitting arrays into multiple logical disks at the external storage systems level. Where possible, create a single logical disk LU from the entire capacity of the external storage systems array.
2. Depending on the redundancy requirements, create RAID-5 (RAID 5) arrays by using 5 - 8 data bits plus parity components that are, 5 + P, 6 + P, 7 + P, or 8 + P.
3. Complete the appropriate calculations to ensure that your external storage systems are configured correctly.
4. If any external storage system that is associated with an MDisk has the **`allowquorum`** parameter set to no, the **`chquorum`** command fails for that MDisk. Before you set the **`allowquorum`** parameter to yes on any external storage system, see IBM support website for external storage systems configuration-requirements.

Exporting external storage systems LUs to the system

Ensure that you are familiar and follow the following guidelines for exporting external storage systems LUs to the system.

- When you define the system as a host object to the external storage systems, you must include all ports on all system nodes and system candidate nodes.
- When you first create an external storage-system LU, you must wait until it is initialized before you export it to the system.
  - Do not present new LUs to the system until the array initialization format is complete. If you add a Logical Unit Number (LUN) to a storage pool before the array initialization format is complete, the storage pool goes offline. While the storage pool is offline, you cannot access the volumes that are in the storage pool.

    Attention: Failure to wait for the LUs to initialize can result in excessive system discovery time and an unstable view of the SAN.
- When you export an external storage-system LU to the system, the LU must be accessible through all ports on the external storage system that are visible to the system.
- The external storage-system LU must be mapped to all the system ports by using the same LUN for all external storage-system ports.

Attention:

The system uses the same LUN as the addressing method to identify an external storage-system LU through any access connection. Ensure that an external storage-system LU is mapped to all

the system ports by using the same LUN for all external storage-system ports. Failure to do this mapping can cause loss of access to the associated MDisk.

After exporting the LU to the system, a system discovery might need to be requested to discover the new external storage-system MDisks and add them to the MDisks view.

External storage systems LU access through multiple ports

If an MDisk LU is accessible through multiple external storage-system ports, the system ensures that all nodes that access the LU coordinate their activity. The system also ensures that nodes access the LU through the same external storage-system port or group of ports depending on the specific access capabilities of the external storage system.

When the system can access an LU through multiple external storage-system ports, it uses the following criteria to determine the accessibility of these ports:

- The system node is a member of a system.
- The system node has Fibre Channel or iSCSI connections to the external storage-system port.
- The system node successfully discovers the external storage-system LU.
- Slandering does not cause exclusion of system node access to the MDisk through the external storage-system port.

An MDisk path is presented for all system nodes that meet these criteria.

MDisk port selection for external storage systems

When an MDisk is created, the system selects a single port or group of ports of the external storage-system ports to access the MDisk.

Table 16 on page 49 describes the algorithm that the system uses to select the external storage-system port or group of ports.

Table 16. Storage-system port-selection algorithm

Criteria	**Description**
Accessibility	Creates an initial set of candidate storage-system ports that include the ports that are accessible by the highest number of nodes
Slandering	Reduces the set of candidate storage-system ports to the ports that have lowest number of slanders
Preference	Reduces the set of candidate storage-system ports to the preferred ports used by the external storage system
Load balance	Selects the single port or group of ports from the set of candidate storage-system ports that has the lowest MDisk access count

After the initial external storage-system port-selection is made for an MDisk, the following events can cause the port selection algorithm to rerun:

- A new node joins the system.
- Requests for discovery of MDisks.
- The external storage system changes its preferred port.
- New external storage-system ports are discovered.
- The currently selected external storage-system port becomes inaccessible.

- Slandering causes exclusion of system node access to the MDisk through the external storage-system port.

Related concepts

"Attaching an external storage system" on page 50
Attach your external storage systems configurations through Fibre Channel or iSCSI connections with system nodes.

"Discovering an external storage system" on page 51
System initialization includes discovery process that identifies external storage systems and the Logical Units (LUs) that they export.

Attaching an external storage system

Attach your external storage systems configurations through Fibre Channel or iSCSI connections with system nodes.

External storage systems attachment through Fibre Channel connections

All nodes in a system must be able to connect to the same set of ports on an external storage system. A system is considered degraded if it contains any two nodes that cannot connect to the same set of external storage-system ports. In such a situation, a system error is logged that requires a repair action.

Use one of the following techniques to control the access to the external storage systems:

- Switch zoning
- LUN-masking capability of the external storage systems

You can configure certain external storage systems to safely share resources between the system nodes and direct-attached hosts. This type of configuration is described as a split external storage system. In all cases, it is critical to configure the external storage system and SAN so that the system cannot access logical units (LUs) that a host or another system can also access. You can arrange the split external storage systems configuration by external storage-system logical unit number (LUN)-mapping and masking.

Configurations where an external storage system is split between two sets of system nodes are also supported. In all cases, it is critical to configure the external storage system and SAN so that the system nodes cannot access LUs that a host or another system can also access. You can use external storage systems LUN mapping and masking to arrange the configuration, failure to which can result in data corruption.

Attention: An external storage-system LU must not be shared between the system and a host or between more than one system. This configuration is not supported and is likely to cause undetected data loss or corruption. Ensure that only a single system can access each external storage-system LU

When an external storage system is detected on the Fibre Channel SAN, the system attempts to recognize it using its Inquiry data. If the device is not supported, the system configures the device as a generic device. A generic device might not function correctly when it is addressed by a node, especially under failure scenarios. However, the system does not regard accessing a generic device as an error condition and does not log an error. Managed disks (MDisks) that are presented by generic devices are not eligible to be used as quorum disks.

External storage systems attachment through iSCSI connections

Use the following guidelines to configure iSCSI-attached storage to the system:

- Unlike Fibre Channel connections, manually configure the connections between the source system and the target external storage systems.
- Discovery of the target IPs and session establishment is manually configured:
  - Either I/O group-wide - that is nodes from a specified I/O group discover and establish sessions, or

  - System-wide - that is all nodes discover and establish sessions.
- After successful discovery, manually establish the initiator sessions from the source system to the target external storage system through:
  - Nodes in a specific I/O group, or
  - All nodes across the system.
- A one-to-one mapping of source ports to target ports is required. The same numbered ports on each source system or node connect to only one port on each target system or node. Multiple initiator ports on the same system node cannot connect to the same target port (many-to-one connectivity). Similarly, one initiator port cannot be connected to multiple target ports on same target external storage system (one-to-many connectivity).
- Zoning is not required.
- Direct attachment between the system and external storage systems is not supported and requires Ethernet switches between the system and the external storage system.
- To avoid a single point of failure, the guidelines suggest use of a dual-switch configuration. For full redundancy, a minimum of two paths between each initiator node and target node must be configured with each path going through a separate switch.
- Extra paths can be configured to increase throughput if both initiator and target nodes support more ports.

External storage system direct attachment

If an external storage system with multiple World Wide Node Names (WWNN) is directly attached with the system, unexpected connectivity issues might occur for associated system controller objects in the system event log. Reason for the event logging is redundancy expectation of both initiator and target port connections between the system node and all the controller objects that are associated with the external storage system. In a direct attached configuration, the connection redundancy might not exist. To disable the unexpected event logs related to connection redundancy, use **chcontroller -directattachenhanced yes** command. For more information, see **chcontroller** command.

Discovering an external storage system

System initialization includes discovery process that identifies external storage systems and the Logical Units (LUs) that they export.

The discovery process systematically recognizes all visible ports on the SAN for devices that identify themselves as external storage systems. Each LU that the external storage system exports to the system is also identified. The LUs can contain new storage or a new path for previously discovered storage. The set of discovered LUs forms the managed disk (MDisk) view associated with external storage systems.

Requesting system discovery

The discovery process runs automatically when ports are added to or deleted from the SAN and when some error conditions occur.

To request the discovery process manually, use the **Discover Storage** function from the **External Storage** view in the management GUI. You can also use the **detectmdisk** command-line interface (CLI) command to request the discovery process manually. The request to perform discovery causes the clustered system to rescan the SAN. The rescan discovers any new MDisks that were added to the external storage systems and rebalances MDisk access across the available external storage-system ports.

The system discovery process completes asynchronously and might take a few minutes. When using the management GUI to perform system discovery, the GUI checks whether the discovery process is complete. When using the CLI to perform system discovery, use the **lsdiscoverystatus** command to check whether the discovery process is still in progress.

Discover new external storage system LUs

You can add newly exported external storage-system LUs to the system by using the system management GUI or the CLI.

When LUs are exported by creating LUN mappings on an external storage system, the system managed disk (MDisk) view is not always automatically updated. To update the view, you must request the system discovery process that is performed.

Note: If the data on the associated external storage-system LU needs to be preserved, then follow the Importing an external storage system procedure.

If newly added external storage systems do not appear in the MDisks display, ensure to cable the system correctly to the external storage systems.

Note: To achieve the best system performance, follow the guidelines for adding MDisks to pools as described in Pool configuration guidelines.

Discover external storage-system quorum-disks

During quorum disk discovery, the system assesses each external storage-system LU to determine its potential use as a quorum disk. From the set of eligible LUs, the system nominates three quorum candidate disks.

The MDisk associated with an external storage-system LU must meet the following criteria to be considered as a candidate for a quorum disk:

- It must be in managed mode.
- It must be visible to all nodes in the system.
- It must be presented by an external storage system that is approved for quorum disks.
- It must have sufficient free extents to hold the system state and the configuration metadata.

Note: Quorum disks cannot be placed on the iSCSI-attached external storage systems.

Quorum disk candidates are not found after the discovery because of one of the following situations:

- LU MDisks of the external storage system does not exist in managed space mode. An error is logged when this situation occurs.
- LU MDisks of the external storage system exists in managed space mode, but they do not meet the eligibility criteria. An error is logged when this situation occurs.

Related concepts

“Planning an external storage system” on page 48
To use Logical Units (LUs) of the external storage systems as Managed Disks (MDisks), several configuration requirements need to be followed to avoid performance and data integrity issues.

“Quorum” on page 491
A quorum device is used to break a tie when a SAN fault occurs, when exactly half of the nodes that were previously a member of the system are present. A quorum device is also used to store a backup copy of important system configuration data.

Related reference

“Importing an external storage system ” on page 53
You can use image mode volumes to import and migrate existing data that is managed by an external storage system into the system.

“Pool configuration guidelines” on page 57

A storage pool provides the pool of storage from which volumes are created. For optimal performance and reliability, you must ensure that the MDisks that make up each tier of the storage pool have the same characteristics.

Importing an external storage system

You can use image mode volumes to import and migrate existing data that is managed by an external storage system into the system.

You can either use the command-line interface (CLI) or the Storage Migration wizard in the management GUI to migrate data from external storage systems. The Storage Migration wizard simplifies the process by creating the image mode volume of the migrated data, host objects, host mappings, and moving that volume to a storage pool to be managed by the system.

The Storage Migration wizard guides through the appropriate changes to the system. After the migration is complete, remove the external storage system from the system by using the appropriate method for each supported connection type.

If the existing external storage systems do not follow the configuration guidelines, consider completing the data migration away from the image mode volume before you resume I/O operations on the host systems. If I/O operations are continued and the external storage system does not follow the guidelines, I/O operations might fail at the hosts, which might cause loss of access to the data.

The procedure for importing managed disks (MDisks) that contain existing data, depends on the amount of available capacity of the system. You must have the same amount of available capacity in the system as the size of the data that you want to migrate into the system. If you do not have the amount of available capacity, the migration causes the storage pool to have an uneven distribution of data because some MDisks are more heavily loaded than others. Further migration operations are required to ensure an even distribution of data and subsequent I/O loading.

Importing image mode volumes with an equivalent amount of available capacity

When you import an image mode volume that has a certain gigabytes capacity, and your system has that amount of available capacity in a single storage pool, then:

- Follow the Start New Migration wizard in the management GUI.
- Select **Pools** > **Import External Storage** to import the image mode volumes and to provide an even distribution of data.

Importing image mode volumes with a smaller amount of available capacity

When you import an image mode volume that has a certain gigabytes capacity, and your system does not have that amount of available capacity in a single storage pool, then:

- Follow the Start New Migration wizard in the management GUI.
- Select **Pools** > **Import External Storage** to import the image mode volumes.

Do not select the destination pool at the end of the wizard. If the destination pool is not selected, the system creates the image mode volumes but does not migrate the data away from the image mode volumes. Use volume mirroring or migration to move the data around as you want.

Managing an external storage system

You can manage the external storage-system resources that are exported to the system by using the system GUI or the system CLI. Managing the external storage system includes how to identify and change the system configuration of the external storage system resources.

Identifying an external storage system

To view the external storage-system serial-number in the management GUI, click **Pools** > **External Storage**. To identify the serial number, run the **`lscontroller`** command on CLI.

For Fibre Channel external storage systems, the serial numbers, the Fibre Channel worldwide node name (WWNN), or worldwide port name (WWPN) is displayed on the external storage system. The WWNN or WWPN can be used to identify the different external storage systems.

For iSCSI external storage systems, refer to the **iscsi_port_id** and ip fields in the output of **lscontroller**. Since the controller can have multiple Internet Protocol addresses (IP addresses), an iSCSI controller has multiple **iscsi_port_id** and ip listings. The ip field shows the IPv4 or IPv6 address for the iSCSI controller and is blank for other controllers. The **iscsi_port_id** refers to a row ID in the output of **lsiscsistorageport** command. The **lsiscsistorageport iscsi_port_id** lists the IQN of the target iSCSI controller in the **target_iscsiname** output field. Use the IP addresses and IQN gathered to determine the actual external storage system.

To verify the actual external storage system that has the IP addresses and IQN found:

- Start the local user interface, or
- Use the command-line tools that the external storage system provides.

Expanding the capacity of an external storage-system logical unit (LU)

On some external storage systems, you can expand the size of an LU by using vendor-specific disk-configuration software that is provided. The following steps are required for the system to use extra capacity that is provided in this way.

To ensure that the additional capacity is available to the system, follow these steps:

- Remove the managed disk (MDisk) from the storage pool.
- Wait for the migration to complete.
- Verify the specified MDisk is listed as unmanaged by selecting **Pools** > **MDisks by Pool** and expand Unassigned MDisk to display all unmanaged Mdisks.
- Use the vendor-specific, disk-configuration software to expand the size of the logical unit on the external storage system.
- Instigate a system discovery to rescan the Fibre Channel network by following the Discovering an External Storage System procedure.
- Wait for the discovery to complete.
- Check the specified MDisk displays the additional capacity that was expanded.
- Add the specified MDisk back to the pool.

The MDisk extra capacity becomes available for use by the system.

Changing the configuration of an existing external storage system

You can change the configuration of an existing external storage system when you want to delete and replace logical units (LUs).

To delete an existing external storage-system LU and replace it with a new LU:

- Remove the specific existing MDisk from the storage pool by following the procedure to remove an external storage-system LU outlined in the topic Removing an external storage system.
- Check the specified MDisk is no longer displayed in the **Pools** > **External Storage** view.
- Use the vendor-specific, disk-configuration software to create a new LU and add the new LU mapping on the external storage system so that the new LU is now visible to the system.
- Discover the newly exported external storage-system LU by following the procedure in Discovering an External Storage System.
- Check the newly created MDisk is displayed in the **Pools** > **External Storage** view.
- Add the newly created MDisk into a system pool.

The newly created MDisk capacity is now available for the system to use.

Important: When creating new external storage-system LU-mappings, always follow the system requirements as described in Planning an external storage system.

Modifying an external storage-system LU mapping

You can modify the LU mapping of an external storage-system LU. The LU mapping logical unit number (LUN) ID and LU unique identifier (UID) are linked and cannot be changed dynamically when the system is managing the MDisk.

To modify an LU-mapping LUN, complete the following steps:

- Before removing an MDisk from a storage pool, make sure to preserve the existing LU array data on the external storage system.
- Remove the specific existing MDisk from the storage pool by following the Removing an external storage system procedure to remove an external storage-system LU.
- Check the specified MDisk is no longer displayed in the **Pools** > **External Storage** view.
- Use the vendor-specific, disk-configuration software to add the new LU mapping on the external storage system so that the existing LU is now visible to the system through the new LU-mapping LUN.
- Discover the newly exported external storage-system LU by following the procedure in Discovering an external storage system.
- Check the newly re-created MDisk is displayed in the **Pools** > **External Storage** view correctly with the new LU-mapping LUN.
- Add the newly re-created MDisk into a system pool.

The newly re-created MDisk with the changed LUN is now available for use by the system.

Important:

When creating a new external storage-system LU-mapping, always follow the system requirements as described in Planning an external storage system.

Servicing external storage systems

External storage systems that are supported for attachment to the system are designed with redundant components and access paths to enable concurrent maintenance. Hosts have continuous access to their data during component failure and replacement.

Related concepts

“Planning an external storage system” on page 48
To use Logical Units (LUs) of the external storage systems as Managed Disks (MDisks), several configuration requirements need to be followed to avoid performance and data integrity issues.

“Discovering an external storage system” on page 51
System initialization includes discovery process that identifies external storage systems and the Logical Units (LUs) that they export.

Related tasks

“Removing an external storage system” on page 55
You can replace or decommission an external storage system or individual MDisks.

Removing an external storage system

You can replace or decommission an external storage system or individual MDisks.

About this task

During this procedure, you add a device, migrate off the data from the external storage system, and remove the old MDisks.

An alternative procedure is to migrate all volumes that are using storage in this pool to another pool. This procedure consolidates the volumes in a single or new pool. However, you can move one volume at a time. The following procedure describes migration of all the data through a single command.

You can also use this procedure to remove or replace a single MDisk in a pool. If an MDisk experiences a partial failure, such as a degraded array, and you can still read the data from the disk but cannot write to it, then replace that MDisk. Steps "1" on page 56 and "2" on page 56 detail how you can add or remove a single MDisk rather than a list of MDisks.

Complete the following steps to remove an external storage system:

Procedure

1. Add the new MDisks to the pool by completing the following steps:
   a) Click **Pools** > **MDisks by Pool**.
   b) Select the pool that you want to add the new MDisks to and select **Add Storage**.
   c) In the Assign Storage wizard, select **External**.
      Select the external storage system and the MDisks that you want to add. Ensure that you add the number of MDisks so that the capacity of the new MDisks is the same or exceeds the capacity of the old MDisks. Click **Assign**.
2. Force the deletion of the old MDisks from the pool to migrate all the data from the old MDisks to the new MDisks.
   a) Click **Pools** > **MDisks by Pool**.
   b) Expand the pool that you added the new MDisks to and right-click the MDisks to remove and select **Remove**.
      Click **Yes** to migrated data from the removed MDisks to the MDisks that you added to the pool.

      Note: If the MDisk is in image mode and you want to preserve the data on the external storage-system logical unit (LU) instead of converting the volume to a striped volume, then for step "2" on page 56:

      - Go to **Volumes** > **Volumes** view in the management GUI and delete the image mode volume, or
      - If using command-line interface (CLI), issue the **rmvolume** command to delete the volume.
3. When all the migration tasks are complete, verify that the MDisks are unmanaged by selecting **Pools** > **MDisks by Pool** and expand **Unassigned MDisk** to display all unmanaged MDisks.
   Ensure that all the MDisks that were removed are listed.
4. Access the external storage system and unmap the LUNs from the system ports.

   Note: If you no longer want to preserve the data that is on the LUNs, you can delete the LUNs.
5. Complete the following steps to have the system rescan the SAN:
   a) Click **Pools** > **MDisks by Pool**.
   b) Select **Discover MDisks** from the **Actions** menu.

   Note: If the unmanaged MDisk still exists in the view and is degraded or excluded, then include the MDisk and instigate discovery again to remove the MDisk.
6. Verify that the external storage system that you want to decommission has no MDisks.
7. Remove the external storage system from the SAN so that the ports can no longer access the external storage system.

External storage system zones

Ensure that you are familiar with the external storage system zoning details. Switch zones that contain storage system ports must not have more than 40 ports. A configuration that exceeds 40 ports is not supported.

Zones

The switch fabric must be zoned so that the nodes can detect the back-end storage systems and the front-end host HBAs. Typically, the front-end host HBAs and the back-end storage systems are not in the same zone. The exception to this is where split host and split storage system configuration is in use.

All nodes in a system must be able to detect the same ports on each back-end storage system. Operation in a mode where two nodes detect a different set of ports on the same storage system is degraded, and the system logs errors that request a repair action. This can occur if inappropriate zoning is applied to the fabric or if inappropriate LUN masking is used. This rule has important implications for back-end storage, such as IBM DS4000 storage systems, which impose exclusive rules for mappings between HBA worldwide node names (WWNNs) and storage partitions.

Each port must be zoned so that it can be used for internode communications. When configuring switch zoning, you can zone some node ports to a host or to back-end storage systems.

When configuring zones for communication between nodes in the same system, the minimum configuration requires that all Fibre Channel ports on a node detect at least one Fibre Channel port on each other node in the same system. You cannot reduce the configuration in this environment.

It is critical that you configure storage systems and the SAN so that a system cannot access logical units (LUs) that a host or another system can also access. You can achieve this configuration with storage system logical unit number (LUN) mapping and masking.

If a node can detect a storage system through multiple paths, use zoning to restrict communication to those paths that do not travel over ISLs.

However, to accommodate the limitations of some switch vendors on the number of ports or worldwide node names (WWNNs) that are allowed in a zone, you can further reduce the number of ports or WWNNs in a zone. Such a reduction can result in reduced redundancy and additional workload being placed on other system nodes and the Fibre Channel links between the nodes of a system.

If the round-trip latency between systems is greater than 80 milliseconds, stricter configuration requirements apply:

- Use SAN zoning and port masking to ensure that two Fibre Channel ports on each node that is used for replication are dedicated for replication traffic.
- Apply SAN zoning to provide separate intersystem zones for each local-to-remote I/O group pair that is used for replication.

The minimum configuration requirement is to zone both nodes in one I/O group to both nodes in one I/O group at the secondary site. The I/O group maintains fault tolerance of a node or port failure at either the local or remote site location. It does not matter which I/O groups at either site are zoned because I/O traffic can be routed through other nodes to get to the destination. However, if an I/O group that is doing the routing contains the nodes that are servicing the host I/O, there is no additional burden or latency for those I/O groups because the I/O group nodes are directly connected to the remote system.

Pool configuration guidelines

A storage pool provides the pool of storage from which volumes are created. For optimal performance and reliability, you must ensure that the MDisks that make up each tier of the storage pool have the same characteristics.

Notes:

1. Generally, the slowest MDisk in the storage pool governs the performance of that storage pool.

2. Generally, the weakest MDisk in the storage pool governs the reliability of that storage pool.
3. If a single MDisk in a pool fails, access to the entire pool is lost.

Use the following guidelines when you group similar disks:

- Group MDisks with equal performance in a single tier of a pool.
- Group similar arrays in a single tier. For example, configure all 6 + P RAID-5 arrays in one tier of a pool.
- Group MDisks from the same storage systems types in a single tier of a pool.
- Group MDisks that use the same type of underlying physical disk in a single tier of a pool. For example, group MDisks by Fibre Channel or SATA.
- Do not use single disks. Single disks do not provide redundancy. Failure of a single disk results in total data loss of the storage pool to which it is assigned.
- When possible, include similar sized MDisks in a storage pool tier to make it easier to balance the MDisks in that storage pool tier.

Important: When you group internal RAID MDisks into pools by using similar technology disks, the arrays must use a similar **slow_write_priority** setting to prevent creating a single array that cannot avoid writing to a slow component. Failure to use a similar setting results in an immediate critical performance bottleneck if a single drive in the pool became slow in the **redundancy** mode array. A simple exception to this rule is when you use the IBM Easy Tier function:

- The flash drive arrays are set to **redundancy** mode as the write response time technique is too coarse for typical flash drive latencies, and
- The spinning disk arrays are set to **latency** mode.

Managing pools

You can use the management GUI or the command line interface to work with your managed disks (MDisks) and storage pools. To work with MDisks in the management GUI, select **Pools** > **MDisks by Pool**.

CLI managed disk tasks

You can use the command-line interface to collect dump files from drives. Issue the **triggerdrivedump** command to collect internal log files from MDisks. Dump files are generated in the `/dumps/drive` directory. You can use **lsdumps** to list dump files, **cpdumps** to copy dump files and **cleardumps** to delete files from the dumps directory. Refer to the Command-line interface guide for more details.

You can use the command-line interface to create MDisks on a system.



Attention: If you add an MDisk to a system as an MDisk, any data on the MDisk is lost. If you want to keep the data on an MDisk (for example, because you want to import storage that was previously not managed), you must create image mode volumes instead.

Important: You can specify different extent sizes for different systems; however, you cannot migrate (volumes) between systems with different extent sizes. If possible, create all your systems with the same extent size.

Issue the **mkmdiskgrp** command to create a storage pool. You can specify MDisks at creation time, or add them later with the **addmdisk** command. Refer to the Command-line interface guide for more details.

You can use the command-line interface to remove MDisks from a system. When you remove MDisks, the managed disk objects that represent the MDisks might still exist. The system cannot access these MDisks because they were unconfigured or removed. You must remove the managed disk objects. Run the **includemdisk** command on each affected managed disk object. Run **rmmdisk** on each affected managed disk object to put it into the unmanaged mode. Run **detectmdisk** for the system to rescan the network and detect that the MDisks no longer exist. Refer to the Command-line interface guide for more details.

You can use the command-line interface to determine the relationship between volumes and managed disks (MDisks). To display a list of the IDs that correspond to the MDisks that comprise the volume, use

lsvdiskmember. To display a list of IDs that correspond to the volumes that are using this MDisk, use **lsmdiskmember**. To display a table of volume IDs and the corresponding number of extents that are being used by each volume, use **lsmdiskextent**. To display a table of MDisk IDs and the corresponding number of extents that each MDisk provides as storage for the specified volume, use **lsvdiskextent**.

You can use the command-line interface to determine the relationship between MDisks and controller LUNs. Each MDisk corresponds with a single RAID array, or with a single partition on a specified RAID array. Each RAID controller defines a LUN number for this disk. The LUN number and controller name or ID are needed to determine the relationship between MDisks and RAID arrays or partitions. Use **lsmdisk** to display the controller name or controller ID and the controller LUN number. Use **lscontroller** to display the vendor ID, product ID and WWNN of the controller. You can use this information to determine what is being presented to the MDisk. From the native user interface for the specified controller, list the LUNs it is presenting and match the LUN number with that found using **lsmdisk**. This provides the exact RAID array or partition that corresponds with the MDisk.

Managing throttles for pools

You can create, modify, and remove throttles for pools by using the management GUI or the command-line interface.

Throttling is a mechanism to control the amount of resources that are used when the system is processing I/Os on a specific pool. If a throttle is defined, the system either processes the I/O, or delays the processing of the I/O to free resources for more critical I/O. You can set throttles for both parent and child pools. Throttles work in a hierarchy when they are applied to both the parent and child pool. Throttles that are defined in the parent pool applies to volumes in both parent and child pools. However, throttles that are defined on the child pool only apply to the volumes in that child pool. For mirrored volumes that are in different pools, only the throttling that applies to the primary copy is considered, regardless of which copy the I/O is directed to. In this case, any throttles defined on the secondary copy of the volume are ignored. Only throttles on the primary copy are used, whether the pool is parent or child pool. When throttles are defined in both parent and child pools, the more restrictive throttle applies to the volume in the child pool, but the I/O is counted against both pools. Consider these examples:

- A parent pool is configured with a throttling limit of 5K IOPS and its single child pool had a limit at 3K IOPS. When I/O operations to a volume in the child pool reaches the throttling limit, the volume cannot receive any more I/O during that time. However, volumes in the parent pool can still use 2K of IOPS to process during that time.
- A parent pool is configured with a throttling limit of 4K IOPS . The parent pool also has two child pool and each child pool has a throttling limit of 3K of IOPS.
- A parent pool is configured with a throttling limit of 5K IOPS . The parent pool has also has two child pool without a throttle defined. All in all, the volumes in each pool cannot handle more than 4K IOPS but each of these pools can could serve its IOPS limit if the volumes in the other pools consume at a lower rate of IOPS In this example, the volumes in the parent and both child pools can reach the 5K IOPS together, regardless of which pool the volumes are located.

Using the management GUI

To create a new throttle for a specific pool, complete these steps:

1. In the management GUI, select **Pools** > **Pools**.
2. Right-click a pool and select **Edit Throttle...**.
3. On the **Edit Throttle** page, enter values for the following throttles:

   Bandwidth Limit
   Enter the maximum amount of bandwidth the pool can process before the system delays I/O processing for this pool. Click **Create**.

   IOPS Limit
   Enter the maximum I/O operations per second the pool can process before the system delays I/O processing for this pool. Click **Create**.

To edit an existing throttle for a specific pool, complete these steps:

1. In the management GUI, select **Pools** > **Pools**.
2. Right-click a pool and select **Edit Throttle...**.
3. On the **Edit Throttle** page, enter values for the following throttles:

   Bandwidth Limit
   Enter the maximum amount of bandwidth the pool can process before the system delays I/O processing for this pool. Click **Save**.

   IOPS Limit
   Enter the maximum I/O operations per second the pool can process before the system delays I/O processing for this pool. Click **Save**.

To remove throttles for a specific pool, complete these steps:

1. In the management GUI, select **Pools** > **Pools**.
2. Right-click a pool or multiple pools and select **Edit Throttle...**.
3. On the **Edit Throttle** page, click **Remove** by the throttle that you want to delete.

Using the command-line interface

To create a new throttle for bandwidth and IOPS for a pool with the command-line interface, enter the following command:

```
mkthrottle -type pool -bandwidth bandwidth_limit_in_mb -iops iops_limit -vdisk
vdisk_id_or_vdisk_name
```

Where *bandwidth_limit_in_mb* is the maximum amount of bandwidth that the system can process for the specified volume before the system delays I/O processing for that volume. The *iops_limit* is the maximum I/O operations per second that the system can process before the system delays I/O processing for this volume. The *vdisk_id_or_vdisk_name* is the ID or name of the volume that the throttle applies to.

To change an existing throttle for the volume, enter the following command:

```
chthrottle -type pool -bandwidth bandwidth_limit_in_mb -iops iops_limit throttle_name_or_id
```

Where *bandwidth_limit_in_mb* is the maximum amount of bandwidth that the system can process for the specified host before the system delays I/O processing for that volume. The *iops_limit* is the maximum I/O operations per second that the system can process before the system delays I/O processing for this volume. The *throttle_id_or_name* is the ID or name of the throttle that is being changed.

To remove a throttle for pool with the command-line interface, enter the following command:

```
rmthrottle throttle_name_or_id
```

Where *throttle_name_or_id* is the ID or name of the throttle that is being removed.

Creating data reduction pools

To use data reduction technologies on the system, you need to create a data reduction pool, create volumes with the data reduction pool, and map these volumes to hosts that support SCSI unmap commands.

Data reduction is a set of techniques that can be used to reduce the amount of usable capacity that is required to store data. An example of data reduction includes data deduplication. Data reduction can increase storage efficiency and performance and reduce storage costs, especially for flash storage. Data reduction reduces the amount of data that is stored on external storage systems and internal drives by reclaiming previously used capacity that are no longer needed by host systems. To estimate potential capacity savings that data reduction can provide on the system, use the Data Reduction Estimation Tool (DRET). This tool analyzes existing user workloads that are being migrated to a new system. The

tool scans target workloads on all attached storage arrays, consolidates these results, and generates an estimate of potential data reduction savings for the entire system.

For more information about DRET, see https://www.ibm.com/support/pages/node/6217841. For more information about Compretimator, see https://www.ibm.com/support/pages/node/6209688.

The system supports data reduction pools which can use different capacity savings methods simultaneously, increasing the capacity savings across the entire pool. Data reduction pools also support deduplication. When deduplication is specified for a volume, duplicate versions of data are eliminated and not written to storage, thus saving more usable capacity. Some models or software versions require specific hardware or software to use this function.

Support for the host SCSI unmap command is enabled by default.

Using the management GUI

To create data reduction on the system, complete these steps:

1. Create a data reduction pool by completing these steps:
   a. In the management GUI, select **Pools** > **Pools**.
   b. On the **Pools** page, click **Create**.
   c. On the **Create Pool** page, enter a name of the pool and select **Data Reduction**.
   d. Click **Create**.
2. Add storage to the data reduction pool by completing these steps:
   a. In the management GUI, select **Pools** > **Pools**.
   b. Right-click the data reduction pool that you created and select **Add Storage**.
   c. Select from the available storage and reserve capacity to the pool. Click **Assign**.

      Note: You cannot select from available storage when you are creating a data reduction child pool. A data reduction child pool has access to the complete available capacity of the parent pool without limit. A data reduction child pool cannot be created if the parent pool does not have available capacity.
3. Create volumes that use supported capacity savings methods in the data reduction pool and map them to hosts by completing these steps:
   a. In the management GUI, select **Volumes** > **Volumes**.
   b. On the **Volumes** page, click **Create Volumes**.
   c. On the **Create Volume** page, select the type of volume that you want to create.
   d. Enter the following information for the volume:

      Pool
      Select a data reduction pool from the list.

      Volume details
      Enter the quantity, capacity, and name for the volume or volumes that you are creating.

      Capacity savings
      Select one of the supported capacity savings methods. For any of these supported options, you can also select to use deduplication for the volume that you create. For example, you can create a thin-provisioned volume that also uses deduplication to remove duplicate data.

   e. Click **Create and Map**.

      Note: Select **Create** to create the volumes in the data reduction pool without mapping to hosts. If you want to map volumes to hosts later, select **Hosts** > **Hosts** > **Add Hosts**.

   f. On the **Create Mapping** page, select **Host** to display all hosts that are available for mapping. Hosts must support SCSI unmap commands. Verify that the selected host type supports SCSI unmap commands. Click **Next**.
   g. Verify the volume, and then click **Map Volumes**.

Using the command-line interface

Linked pools

High availability and disaster recovery replication require the storage pools that are used for creating replicated volumes to be linked between systems.

When a volume is created on the active management system for HA, or on the production system for DR, pool links define the pool that is used on the remote system to create the replicated volume. For information about linking pools for replication, see Pool links.

Easy Tier

The system includes IBM Easy Tier, which is a function that responds to the presence of high performance drives in a storage pool that also contains hard disk drives (HDDs). The system automatically and nondisruptively moves frequently accessed data from slower to faster tiers of storage.

A child pool inherits the Easy Tier settings from its parent pool. You cannot change the Easy Tier settings on a child pool. You can only change them on a parent pool.

If a pool contains one type of MDisk, Easy Tier goes into `balancing` mode. When the pool contains multiple types of MDisks, Easy Tier is automatically turned on. The local MDisks are classified as storage class memory, tier 0 flash, tier 1 flash, enterprise, or nearline drives. The system supports storage class memory, tier 0, and tier 1 flash drives. Storage class memory drives use Non-Volatile Memory Express (NVMe) technology that provides faster performance than the traditional SSDs but are slower than dynamic random-access memory (DRAM)-based drives. Tier 0 flash drives are higher-cost flash drives that provide high performance for read and write operations. Tier 1 flash drives are lower-cost flash drives that have large capacity but lower performance and write endurance characteristics.

Easy Tier modes of operation

Easy Tier offers several modes of operation to optimize the placement of data on flash drives and on HDDs in a storage pool. The system supports flash drives, which offer a number of potential benefits over magnetic hard disk drives (HDDs). The benefits can include faster data access and throughput, better performance, and less power consumption.

Determining the amount of data activity in an extent and when to move the extent to the proper storage tier is usually too complex a task to manage manually.

Easy Tier evaluation mode collects usage statistics for each storage extent for a storage pool where the capability of moving data from one tier to the other tier is not possible or is disabled. An example of such a storage pool is a pool of homogeneous MDisks, where all MDisks are typically HDDs. A summary file is created in the `/dumps` directory on the configuration node (`dpa_heat.`*`node_name.date.time`*`.data`), which can be offloaded and viewed by using the IBM Storage Tier Advisor Tool.

Easy Tier automatic data placement also measures the amount of data access. Easy Tier then acts on the measurements to automatically place the data into the appropriate tier of a storage pool that contains both MDisk tiers.

Dynamic data movement is not apparent to the host server and application users of the data, other than providing improved performance.

For a storage pool and volume to be automatically managed by Easy Tier, ensure that the following conditions are met:

- The volume must be striped.

Remember: Easy Tier is supported on compressed volumes, but it migrates data based only on read activity.

Volumes that are added to storage pools use extents from every type of MDisk initially. By default, the extents that are used are from the Enterprise tier, but Nearline and Flash tiers can also be used. Easy Tier then collects usage statistics to determine which extents to move to the MDisks.

Easy Tier evaluation mode

When IBM Easy Tier evaluation mode is enabled for a storage pool with a single tier of storage, Easy Tier collects usage statistics for all the volumes in the pool.

The system monitors the storage use at the volume extent level. Easy Tier constantly gathers and analyzes monitoring statistics to derive moving averages for the past 24 hours.

Volumes are not monitored when the `easytier` attribute of a storage pool is set to *off*. Volumes are monitored when the `easytier` attribute of a storage pool is set to *measure*.

You can control or view data placement settings by using the following command-line interface (CLI) commands:

chmdiskgrp
: Modifies the properties of the storage pool. Use this command to turn on evaluation mode on a storage pool with a single tier of storage and to turn off Easy Tier functions on a storage pool with more than one tier of storage.

lsmdiskgrp
: Lists storage pool information.

lsvdisk
: Lists volume information.

lsvdiskcopy
: Lists volume copy information.

mkmdiskgrp
: Creates a storage pool.

Other MDisk commands such as **addmdisk**, **chmdisk**, and **lsmdisk** can be used to view or set the tier an MDisk belongs to.

Automatic data placement

When IBM Easy Tier is active, the system completes several functions to automatically place data in the appropriate storage tier. Easy Tier measures the host access activity to the data on each storage extent. It also provides a mapping that identifies high activity extents, and then moves the high-activity data according to its relocation plan algorithms.

To automatically relocate the data, Easy Tier initiates the following processes:

1. Monitors volumes for host access to collect average usage statistics for each extent over a rolling 24-hour period of I/O activity.
2. Analyzes the amount of I/O activity for each extent to determine whether the extent is a candidate for migrating to or from the higher performing flash drive tier.
3. Develops an extent relocation plan for each storage pool to determine exact data relocations within that storage pool. Easy Tier then automatically relocates the data according to the plan.

While it relocates volume extents, Easy Tier follows these actions:

- Attempts to migrate the most active volume extents first.
- Refreshes the task list as the plan changes. The previous plan and any queued extents that are not yet relocated are abandoned.

Automatic data placement is enabled by default for storage pools with more than one tier of storage. When you enable automatic data placement, by default all striped volumes are candidates for automatic data placement. Image mode and sequential volumes are never candidates for automatic data placement. When automatic data placement is enabled, I/O monitoring is undertaken for all volumes whether or not the volume is a candidate for automatic data placement. When automatic data placement is enabled and sufficient activity warrants relocation, extents will be relocated within a day after enablement. You can control whether Easy Tier automatic data placement and I/O activity monitoring is enabled or disabled by using the settings for each storage pool and each volume. Each command in the following list that can create or change the settings for storage pools can enable or

disable both Easy Tier functions. Any command that can create or change the settings for volumes can enable or disable automatic data replacement, if automatic data replacement is enabled for the storage pool. You can control or view automatic data placement by using the following command-line interface (CLI) commands:

addvdiskcopy
Adds a copy to an existing volume by changing a nonmirrored volume into a mirrored volume.

chmdiskgrp
Modifies the properties of the storage pool. Use this command to turn on Easy Tier evaluation mode or I/O monitoring and to turn off Easy Tier functions on a storage pool with more than one tier of storage.

Note: When automatic data placement is active on a storage pool, set a warning threshold for the storage pool. Automatic data placement cannot function if the storage pool is used 100%.

chvdisk
Modifies the properties of a volume.

lsmdiskgrp
Lists storage pool information.

lsvdisk
Lists volume information.

lsvdiskcopy
Lists volume copy information.

mkmdiskgrp
Creates a storage pool.

mkvdisk
Creates sequential, striped, or image mode volumes.

If you want to disable automatic data placement for a volume or storage pool, set the `easytier` attribute to *off*.

Easy Tier acceleration mode

You can change the mode rates of extent migrations through the IBM Easy Tier function and pool balancing functions.

Normally, your system runs with the Easy Tier accelerated mode set to **off**. This system default rate ensures that migrations do not affect system performance.

To view the currently configured value, enter the **lssystem** command. The default view for `easy_tier_acceleration` is **off**. The maximum rate of extent migrations in normal mode (the system default) is 12 GB per 5 minutes. If `easy_tier_acceleration` is set to **on**, the maximum migration rate is 48 GB per 5 minutes. The migration rate of 48 GB per 5 minutes in accelerated mode cannot always be guaranteed.

You can use the **chsystem** command to turn on accelerated mode and extend the limit of the migration rate from 12 GB per 5 minutes to 48 GB per 5 minutes:

```
chsystem -easytieracceleration on
```

You can return to the rate of 12 GB per 5 minutes by turning off the accelerated mode:

```
chsystem -easytieracceleration off
```

You can also temporarily enable the accelerated mode to rapidly use new capacity, such as in the following examples:

- You added more capacity to an existing storage pool either by adding to an existing tier or by adding a tier to the pool. When you enable the accelerated mode, the system can quickly spread the existing volumes into the new capacity.

- You are migrating multiple volumes between pools, and the target pool has more tiers than the source pool. Volume migration between storage pools preserves the tier of the volume's extents. When you enable the accelerated mode, the system can quickly take advantage of the additional tiers in the target pool.

To avoid the possibility of performance issues that are caused by overloading the managed disks, use the accelerated mode during periods of reduced system activity.

Easy Tier automatic data placement requirements, recommendations, and limitations

Some limitations exist if you use the IBM Easy Tier function on your system.

- To ensure optimal performance, all MDisks in a storage pool tier must have the same technology and performance characteristics.
- Easy Tier automatic data placement is not supported on volume copies that are image mode or sequential. I/O monitoring for such volumes is supported, but you cannot migrate extents unless you convert the image or sequential volume copies to striped volumes.
- Automatic data placement and extent I/O activity monitors are supported on each copy of a mirrored volume. The Easy Tier function works with each copy independently of the other copy. For example, you can enable or disable Easy Tier automatic data placement for each copy independently of the other copy.
- When a volume is migrated out of a storage pool that is managed with the Easy Tier function, Easy Tier automatic data placement mode is no longer active on that volume. Automatic data placement is also turned off while a volume is being migrated between two pools that have Easy Tier automatic data placement enabled. Automatic data placement for the volume is reenabled when the migration completes.

Limitations when removing an MDisk by using the force parameter

When an MDisk is deleted from a storage pool with the **`force`** parameter, extents in use are migrated to MDisks in the same tier as the MDisk being removed, if possible. If insufficient extents exist in that tier, extents from the other tier are used.

Limitations when migrating extents

When Easy Tier automatic data placement is enabled for a volume, the **`migrateexts`** command-line interface (CLI) command cannot be used on that volume.

Limitations when migrating a volume to another storage pool

When the system migrates a volume to a new storage pool, Easy Tier automatic data placement between the generic flash drive tier and the Flash, Enterprise, or Nearline tier is temporarily suspended. After the volume is migrated to its new storage pool, Easy Tier automatic data placement between the generic flash drive tier and the Flash, Enterprise, or Nearline tier resumes for the newly moved volume, if appropriate.

When the system migrates a volume from one storage pool to another, it attempts to migrate each extent to an extent in the new storage pool from the same tier as the original extent. In some cases, for example when the target tier is unavailable, the other tier is used. For example, the generic flash drive tier might be unavailable in the new storage pool.

If the automatic data placement is enabled in the new storage pool, pending Easy Tier status changes are assigned after the volume completes its move to the new storage pool. Although the status changes are based on volume use in the old storage pool, the new status is honored in the new storage pool.

Limitations when migrating a volume to image mode

Easy Tier automatic data placement does not support image mode. No automatic data placement occurs in this situation. When a volume with Easy Tier automatic data placement mode active is migrated to image mode, Easy Tier automatic data placement mode is no longer active on that volume.

The Easy Tier function does support evaluation mode for image mode volumes.

Easy Tier reports

The management GUI supports monitoring Easy Tier data movement in graphical reports. The data in these reports helps you understand how Easy Tier manages data between the different tiers of storage, how tiers within pools are used, and the workloads among the different tiers. Charts for data movement, tier composition, and workload skew comparison can be downloaded as comma-separated value (CSV) files.

To view Easy Tier data and reports in the management GUI, select one of the following paths:

- From the management GUI, select **Monitoring** > **Easy Tier Reports**
- From the management GUI, select **Pools** > **View Easy Tier Reports**

Data movement statistics

The data movement chart displays the migration actions that are triggered by Easy Tier.

The data movement chart displays the data for the previous 24-hour period, in one-hour increments. You can change the span of time display, and the increments that you would prefer, such as hourly.

To view Easy Tier data and reports in the management GUI, select one of the following paths:

- From the management GUI, select **Monitoring** > **Easy Tier Reports**
- From the management GUI, select **Pools** > **View Easy Tier Reports**

Access the data movement statistics chart by selecting the **Data Movement** tab.

The first time that you open the Easy Tier reports page, the Data Movement chart is automatically displayed. After you modify this chart, or view a different chart, your last action is retained until a new set of data is acquired through the Easy Tier Statistics tool. (The tool processes the Easy Tier data and generates the HTML reports and CSV files).

The following list describes the type of data that is displayed in the chart:

Extent capacity
The Y-axis displays the number of moved capacity extents. Promoted (rebalanced) extents display above the line, and demoted (swapped) extents are displayed below the line.

Selected viewport time
Displays the time and date that is rendered in the viewport selection.

Viewport X-axis time selection
Displays the time selection in hours. Values that can be chosen are in hours, and the options are 24, 12, 6, 4, 3, 2, or 1 hour.

Restriction: Displaying large amounts of data, such as a 24-hour period in 5-minute increments, is limited to the display area. There is not enough display space to see all 288 data points in the viewport. Use the horizontal scroll bar to view all the data. You can also use the viewpoint shifter to move to the next set of data that is not being displayed in the current viewpoint.

X-axis increment selection
Displays the time that is selected in increments. You can modify the increments to choose the time that is displayed in the viewport. Values are 1 hour, 30 minutes, 15 minutes, 10 minutes, or 5 minutes, depending on the increments available for that section.

Data movement key
Displays all the data movement actions that take place. The types of actions are:

Promote
When active data is moved from a lower tier of storage to a higher tier to improve the overall system performance.

Warm Promote
When an MDisk becomes overloaded, active data is moved from a lower tier to a higher tier to reduce the workload on the MDisk.

Balancing Move
When data is moved within the same tier from an array with a higher workload, to one with a lower workload to balance the workload within the tier.

Balancing Swap
When data is moved within the same tier from an array with higher workload, to one with a lower workload to balance the workload within the tier.

Promote Swap
When active data is moved from a lower tier of storage to a higher tier, other less-active data must first be moved to make space.

Warm Demote
When an MDisk becomes overloaded, active data is moved from a higher tier to a lower tier to reduce the workload on the MDisk.

Cold Demote
When inactive, or less-active data is moved from a higher tier of storage to a lower tier to free up space on the higher tier.

Movement details
Displays a textual and visual description of the different data movement actions that are listed in the key.

The detailed format is described in the following table:

Table 17. Data move statistic descriptions

Data	Data format	Definition	Comments
index	Integer	Row number	
dss	String	Cluster ID	IBM.2145-169933
volume_id	Integer	Volume ID (VDisk ID)	Example: 0
copy_id	Integer	cdisk ID	Example: 0
extent	Integer	The ID of the extent in this volume	
time	ISO 8601 format	The time stamp when the movement happened	2016-10-26T21:37:05+05:00
type	Integer	Type of movement operation	Possible Value: 1~10

Table 17. Data move statistic descriptions (continued)			
Data	**Data format**	**Definition**	**Comments**
typedescription	String	Description of the type of data movement	Possible values: • PROMOTE(type=1) /* Move an extent to a more capable tier */ • SWAP(type=2) Move an extent to a less capable tier which frees space on a more capable tier. Also move an extent to the more capable tier. • DEMOTE(type=3) Move an extent to a less capable tier. • MOVE(type=4) Move an extent to another MDisk in the same tier • RB_SWAP(type=5) Swap extents between MDisks in the same tier • WARM_DEMOTE (type=6) Move an extent to a more capable MDisk in another tier. • WARM_PROMOTE(type=10)
source	Integer	The MDisk that extents are moved from	MDisk number
target	Integer	The MDisk that extents are moved to	MDisk number

Tier composition statistics

The Tier Composition chart displays the distributed workload between the top tier, middle tier, and bottom tier for the selected pool.

To view Easy Tier data and reports in the management GUI, select one of the following paths:

- From the management GUI, select **Monitoring** > **Easy Tier Reports**
- From the management GUI, select **Pools** > **View Easy Tier Reports**

Select the **Tier Composition** tab to display an interactive chart of the tiers that are within the selected pool.

The following list describes the type of data that is displayed in the chart:

Tier types
Displays the type of tiers that are in each category.

MDisk Mapping
Displays a table for the selected pool that describes which MDisks are mapped to which tier.

Extent Capacity
X-axis of the chart

Key
Displays the breakdown of the potential type of workload that might be in each tier category: The following values are available:

Active
Data with more than 0.1 IOPS / Extent access density for small IOPS

Inactive
Data with zero IOPS / Extent access density (no recent activity)

Active large
Data that is not classified

Low activity
Data with more than zero IOPS / Extent access density

Composition details
Displays more detailed information about the workload that is listed in the chart key.

The detailed format is described in the following table:

Table 18. Workload categorization descriptions

Data	Data format	Definition	Comments
index	Integer	Row number	
dss	String	Cluster ID	IBM.2145-169933
datetime	ISO 8601 format	The time stamp when the data is collected	2011-11-14T13:24:49+08:00
extpool	Integer	Pool ID	
tier	Integer	The tier of the MDisk	Possible tiers are in the range 0 - 2.
extentcount	Integer	Number of extents	
smallwps	Float	Small write IOPS targeted against the data	
smallrps	Float	Small read IOPS targeted against the data	
smallmbs	Float	Small I/O bandwidth targeted against the data	MBps
largewps	Float	Large write IOPS targeted against the data	
largerps	Float	Large read IOPS targeted against the data	
largembs	Float	Large I/O bandwidth targeted against the data	MBps

Table 18. Workload categorization descriptions (continued)			
Data	**Data format**	**Definition**	**Comments**
smallms	Float	Average response time on small I/O against the data	ms
largems	Float	Average response time on large I/O against the data	ms

Workload skew statistics

The Workload Skew Comparison chart displays the percentage of I/O workload compared to the total capacity.

To view Easy Tier data and reports in the management GUI, select one of the following paths:

- From the management GUI, select **Monitoring** > **Easy Tier Reports**
- From the management GUI, select **Pools** > **View Easy Tier Reports**

Access the Workload Skew Comparison chart by selecting the **Workload Skew Comparison** tab.

The following list describes the type of data that is displayed in the chart:

Tier Composition
Displays the usage of the different tiers in comparison to the I/O workload. The tiers are broken down to: Top Tier, Middle Tier, and Bottom Tier.

Key
Displays the sum of small I/O and the sum of MBps.

Skew Chart Details
Displays more details about the key and the make up of each tier.

Exporting Easy Tier report data by using the management GUI

You can export the Data Movement, Tier Composition, and Workload Skew Comparison data to your local system as a comma-separated value (CSV) file for viewing in other applications such as a spreadsheet.

To export the Easy Tier data by using the management GUI, complete these steps:

1. Access the Easy Tier statistics and charts by using one of two pathways:
   - From the management GUI, select **Monitoring** > **Easy Tier Reports**
   - From the management GUI, select **Pools** > **View Easy Tier Reports**

   Note: The first time that you open the page, the Data Movement chart is displayed. After you modify this chart, or view a different chart, your last action is retained until a new set of data is acquired through the STAT tool. (The STAT tool processes the Easy Tier data and generates CSV files).
2. Select the pool that contains the data that you want to export.
3. Click **Export** on the top of the GUI pane.
4. In the **Export Easy Tier Reports for Pool** page, select the type of data that you want to export for that pool. All data from the last 24 hours is exported. You can choose from the Data Movement, Tier Composition, and Workload Skew Comparison charts. You can select one, two or all three types of data.
5. Specify a destination that you want to save the file, and click **Save**.

   Note: A compressed file with the default name of easytierstats_*date_timestamp*.zip is downloaded. You can modify the file name to suit your needs.
6. Locate the saved file, and extract the data to view in a spreadsheet or text editor. The default file names are data movement.csv, skewcurve.csv, and workload.csv.

Extracting and viewing performance data with the IBM Storage Tier Advisor Tool

You can use the IBM Storage Tier Advisor Tool, which is also referred to as advisor tool, to view performance data that is collected by IBM Easy Tier over a 24-hour operational cycle. The advisor tool creates Hypertext Markup Language (HTML) files that you use to view the data when you point your browser to the file.

About this task

To obtain the Storage Tier Advisor Tool, search for "Storage Tier Advisor" at the www.ibm.com/support website and download the software from IBM Fix Central.

To extract the summary performance data, follow these steps by using the command-line interface (CLI).

Procedure

1. Enter the following command to find the current dpa_heat.*node_name.date.time*.data files in the clustered system.

   ```
   lsdumps -prefix /dumps/easytier
   ```

2. If necessary, enter the following command to copy the summary performance data files to the current configuration node.

   ```
   cpdumps -prefix /dumps/easytier/dpa_heat.node_name.date.time.data node_id | node_name
   ```

3. Use PuTTY scp (**pscp**) to copy the summary performance data in a binary format from the configuration node to a local directory.
4. From a Microsoft Windows command prompt, use the advisor tool to analyze the binary files. The results are stored in HTML files that are put into a local directory.
5. Point your browser to the HTML files in your local directory.

Results

The advisor tool displays the following statistical reports.

System summary report

- The number of storage pools that are monitored
- The number of volumes that are monitored
- The total capacity that is monitored
- The estimated total hot data capacity
- List of storage pools that are monitored, including pool ID, capacity, tier configuration, tier status, and data management status.

System recommendation report

- A sorted list of suggested flash MDisks to add to the storage pools (sorted by estimated performance improvement)
- A sorted list of suggested Enterprise MDisks to add to the storage pools (sorted by estimated IOPS improvement)
- A sorted list of suggested Nearline MDisks to add to the storage pools (sorted by estimated Bandwidth improvement)
- For each suggested MDisk: The estimated improvement

Storage pool recommendation report

- For each storage pool: list MDisk's current IOPS utilization and projected IOPS utilization

- For each storage pool: show the workload distribution on each tier
- For each storage pool: a list of suggested flash MDisks to add to the storage pools
- For each storage pool: a list of suggested Enterprise MDisks to add to the storage pools
- For each storage pool: a list of suggested Nearline MDisks to add to the storage pools
- For each suggested MDisk: The estimated improvement

Volume heat distribution report

- The distribution of hot data and cold data for each volume copy
- The configured capacity of a volume copy, along with the volume ID, copy ID
- The percentage of each volume copy's IO workload within the storage pool
- The capacity of each volume copy already on each tier

What to do next

You can view this information to analyze workload statistics and evaluate which logical volumes might be candidates for Easy Tier management. If you did not enable the Easy Tier function, the usage statistics that are gathered by the monitoring process can help you determine whether to use Easy Tier to enable potential performance improvements in your storage environment.

Easy Tier settings for storage pools and volumes

This table describes the storage pool Easy Tier settings.

Table 19. Easy Tier settings for storage pools and volumes

Storage pool Easy Tier setting	Number of tiers in the storage pool	Volume copy Easy Tier setting	Volume copy Easy Tier status
Off	One	Off	inactive (see note “1” on page 73)
Off	One	On	inactive (see note “1” on page 73)
Off	Two	Off	inactive (see note “1” on page 73)
Off	Two	On	inactive (see note “1” on page 73)
Measure	One	Off	measured (see note “2” on page 73)
Measure	One	On	measured (see note “2” on page 73)
Measure	Two	Off	measured (see note “2” on page 73)
Measure	Two	On	measured (see note “2” on page 73)
Auto	One	Off	measured (see note “2” on page 73)
Auto	One	On	balanced (see note “3” on page 73)
Auto	Two	Off	active (see note “4” on page 73)

Table 19. Easy Tier settings for storage pools and volumes (continued)

Storage pool Easy Tier setting	Number of tiers in the storage pool	Volume copy Easy Tier setting	Volume copy Easy Tier status
Auto	Two	On	measured (see note “2” on page 73)
On	One	Off	measured (see note “2” on page 73)
On	One	On	balanced (see note “3” on page 73)
On	Two	Off	measured (see note “2” on page 73)
On	Two	On	active (see note “4” on page 73)

Note:

1. When the volume copy status is `inactive`, no Easy Tier functions are enabled for that volume copy.
2. When the volume copy status is `measured`, the Easy Tier function collects usage statistics for the volume but automatic data placement is not active.
3. When the volume copy status is `balanced`, the Easy Tier function enables performance-based pool balancing for that volume copy.
4. When the volume copy status is `active`, the Easy Tier function operates in automatic data placement mode for that volume.

If the volume copy is in image or sequential mode or is being migrated, the volume copy Easy Tier status is `measured` instead of `active`.

The default Easy Tier setting for a storage pool is `auto`, and the default Easy Tier setting for a volume copy is on. These settings mean that Easy Tier functions except pool performance balancing are disabled for storage pools with a single tier, and that automatic data placement mode is enabled for all striped volume copies in a storage pool with two or more tiers.

Chapter 4. Volumes

A volume is a logical disk that the system presents to attached hosts.

Here are some of the relevant capacity-related terminologies:

Available capacity
The capacity that is available to provision new volumes, based on the MDisks that are added to the pool. The management GUI excludes pools that do not have available capacity.

Written capacity
The amount of data written to the volume by mapped hosts.

Real capacity
The total amount of usable capacity that is occupied by the volume.

Used capacity
The amount of usable capacity that contains data or metadata.

You can create different types of volumes, depending on the type of topology that is supported and configured on your system. All systems support standard topology, which is a single-site configuration. For systems with single-site configuration, you can create basic, mirrored, or custom volumes.

Volumes can be assigned to an ownership group. An ownership group defines a subset of users and objects within the system.

- The volume inherits the ownership group of the child pools that provide capacity for the volume and its copies. Volume copies can be created in different ownership groups for backup up scenarios. However, this value must be set intentionally by users that are not defined in ownership groups. When you create a volume copy or migrate volumes to other pools, you can specify child pools that are defined in different ownership groups in the management GUI, which establishes inconsistent ownership. However, it is not recommended to leave volumes or volume copies in different ownership groups. After the migration, the user with Security Administrator role needs to ensure all volumes or copies are within the same ownership group as the users who need access.
- With volume groups, the volume group and its volumes can belong to different ownership groups. However, the ownership of a volume group does not impact the ownership of the volumes that it contains.

Types

Each volume copy can be one of the following types:

Striped
A volume copy that was striped is at the extent level. One extent is allocated, in turn, from each MDisk that is in the storage pool. For example, a storage pool that has 10 MDisks takes one extent from each MDisk. The 11th extent is taken from the first MDisk, and so on. This procedure, which is known as a round-robin, is similar to RAID-0 striping.

You can also supply a list of MDisks to use as the stripe set. This list can contain two or more MDisks from the storage pool. The round-robin procedure is used across the specified stripe set.



Attention: By default, striped volume copies are striped across all MDisks in the storage pool. If some of the MDisks are smaller than others, the extents on the smaller MDisks are used up before the larger MDisks run out of extents. Manually specifying the stripe set in this case might result in the volume copy not being created.

If you are unsure if sufficient available capacity is available to create a striped volume copy, select one of the following options:

- Check the available capacity on each MDisk in the storage pool by using the **`lsfreeextents`** command.
- Let the system automatically create the volume copy by not supplying a specific stripe set.

© Copyright IBM Corp. 2024

This figure shows an example of a storage pool that contains three MDisks. This figure also shows a striped volume copy that is created from the extents that are available in the storage pool.

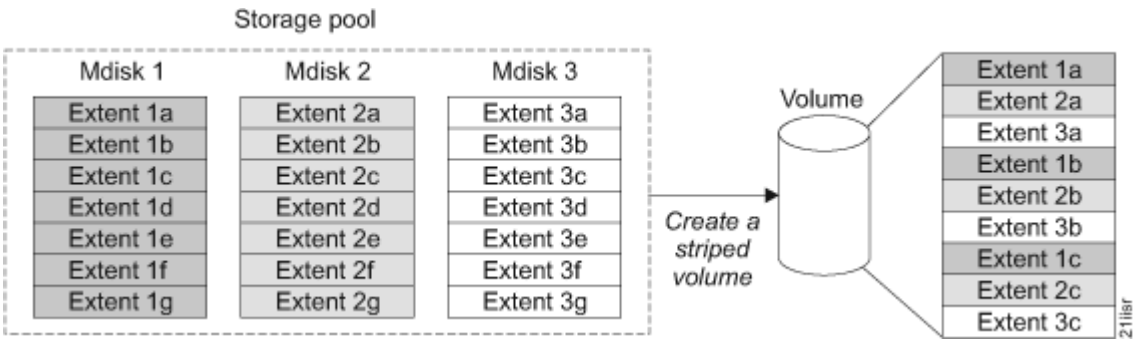


Figure 8. Storage pools and volumes

Sequential

When extents are selected, they are allocated sequentially on one MDisk to create the volume copy if enough consecutive free extents are available on the chosen MDisk.

Image

Image-mode volumes are special volumes that have a direct relationship with one MDisk. If you have an MDisk that contains data that you want to merge into the clustered system, you can create an image-mode volume. When you create an image-mode volume, a direct mapping is made between extents that are on the MDisk and extents that are on the volume. The MDisk is not virtualized. The logical block address (LBA) *x* on the MDisk is the same as LBA *x* on the volume.

When you create an image-mode volume copy, you must assign it to a storage pool. An image-mode volume copy must be at least one extent in size. The minimum size of an image-mode volume copy is the extent size of the storage pool to which it is assigned.

The extents are managed in the same way as other volume copies. When the extents are created, you can move the data onto other MDisks that are in the storage pool without losing access to the data. After you move one or more extents, the volume copy becomes a virtualized disk, and the mode of the MDisk changes from image to managed.

Attention: If you add a managed mode MDisk to a storage pool, any data on the MDisk is lost. Ensure that you create image-mode volumes from the MDisks that contain data before you start adding any MDisks to storage pools.

MDisks that contain existing data have an initial mode of unmanaged, and the clustered system cannot determine whether it contains partitions or data.

Image mode MDisks are members of a storage pool, but they do not contribute to free extents. Image mode volumes are not affected by the state of the storage pool because the storage pool controls image mode volumes through the association of the volume to an MDisk. Therefore, if an MDisk that is associated with an image mode volume is online and the storage pool of which they are members goes offline, the image mode volume remains online. Conversely, the state of a storage pool is not affected by the state of the image mode volumes in the storage pool.

You can use more sophisticated extent allocation policies to create volume copies. When you create a striped volume, you can specify the same MDisk more than once in the list of MDisks that are used as the stripe set. This allocation is useful if you have a storage pool in which not all the MDisks are of the same capacity. For example, if you have a storage pool that has two 18 GB MDisks and two 36 GB MDisks, you can create a striped volume copy by specifying each of the 36 GB MDisks twice in the stripe set so that two-thirds of the storage are allocated from the 36 GB disks.

If you delete a volume, you destroy access to the data that is on the volume. The extents that were used in the volume are returned to the pool of free extents that is in the storage pool. The deletion might fail if the volume is still mapped to hosts. If the deletion fails, you can specify the force-delete flag to delete both the volume and the associated mappings to hosts. Forcing the deletion deletes the Copy Services relationship and mappings.

States

This table describes the different possible states of a volume.

Table 20. Volume states

State	Description
Online	At least one synchronized copy of the volume is online and available if both nodes in the I/O group can access the volume. A single node can access a volume only if it can access all the MDisks in the storage pool that are associated with the volume.
Offline	The volume is offline and unavailable if both nodes in the I/O group are missing, or if none of the nodes in the I/O group that are present can access any synchronized copy of the volume. A thin-provisioned volume goes offline if a user attempts to write an amount of data that exceeds the available disk space.
Degraded	The status of the volume is degraded if one node in the I/O group is online and the other node is either missing or cannot access any synchronized copy of the volume. **Note:** If a volume is degraded and all of the associated nodes and MDisks are online, call your support center for assistance.
Deleting	For thin-provisioned or compressed volume copies in data reduction pools, the deleting status indicates that copies are being deleted. All volume copies, including fully allocated copies, are not accessible until the delete operation completes. In addition, several operations cannot be started until all copies are deleted. The following commands are restricted if one copy of a volume is in the process of being deleted: • **expandvdisksize** • **migratevdisk** • **rmvdiskcopy** • **rmvolumecopy** • **shrinkvdisksize** • **splitvdiskcopy**

Cache modes

You can select to have read and write operations that are stored in cache by specifying a cache mode. You can specify the cache mode when you create the volume. After the volume is created, you can change the cache mode.

This table describes the types of cache modes for a volume.

Table 21. Volume cache modes	
Cache mode	**Description**
readwrite	All read and write I/O operations that are performed by the volume are stored in cache. This is the default cache mode for all volumes. A volume or volume copy that is created from a data reduction pool must have a cache mode of readwrite. If you try to create a thin provisioned or compressed volume copy from a data reduction pool and the volume cache mode is not `readwrite`, the operation fails.
readonly	All read I/O operations that are complete by the volume are stored in the cache.
none	All read and write I/O operations that are complete by the volume are not stored in cache.

Basic volumes

A basic volume is the simplest type of volume, consisting of a copy in a single storage pool.

A basic volume is a volume that has only one physical copy. Basic volumes reside in a single pool on one site. In addition, basic volumes are supported in any system topology and are common to all configurations. A basic volume can be of any type of virtualization: striped, sequential, or image.

Mirrored volumes

By using volume mirroring, a volume can have two copies. Each volume copy can belong to a different pool, and each copy has the same provisioned capacity as the volume.

In the management GUI, an asterisk (*) indicates the primary copy of the mirrored volume. The primary copy indicates the preferred volume for read requests. When a server writes to a mirrored volume, the system writes the data to both copies. When a server reads a mirrored volume, the system picks one of the copies to read. If one of the mirrored volume copies is temporarily unavailable; for example, because the storage system that provides the pool is unavailable, the volume remains accessible to servers. The system remembers which areas of the volume are written and resynchronizes these areas when both copies are available.

You can create a volume with one or two copies, and you can convert a non-mirrored volume into a mirrored volume by adding a copy. When a copy is added in this way, the system synchronizes the new copy so that it is the same as the existing volume. Servers can access the volume during this synchronization process.

You can convert a mirrored volume into a non-mirrored volume by deleting one copy or by splitting one copy to create a new non-mirrored volume.

The volume copy can be any type: image, striped, or sequential. The volume copy can use thin-provisioning or compression to save capacity. If the copies are located in data reduction pools, you can also use deduplication to the volume copies to increase the capacity savings. If you are creating a new volume, the two copies can be of different types, but to use deduplication, both copies must reside in a data reduction pool. You can add a deduplicated volume copy in a data reduction pool to an existing volume with a copy in a standard pool. You can use this method to migrate existing volume copies to data migration pools.

You can use mirrored volumes for the following reasons:

- Improving availability of volumes by protecting them from a single storage system failure.
- Providing concurrent maintenance of a storage system that does not natively support concurrent maintenance.
- Providing an alternative method of data migration with better availability characteristics. While a volume is migrated by using the data migration feature, it is vulnerable to failures on both the source and

target pool. Volume mirroring provides an alternative because you can start with a non-mirrored volume in the source pool, and then add a copy to that volume in the destination pool. When the volume is synchronized, you can delete the original copy that is in the source pool. During the synchronization process, the volume remains available even if there is a problem with the destination pool.

- Converting standard-provisioned volumes to use data reduction technologies, such as thin-provisioning, compression, or deduplication.
- Converting compressed or thin-provisioned volumes in standard pools to data reduction pools to improve capacity savings.

When you use volume mirroring, consider how quorum candidate disks are allocated. Volume mirroring maintains some state data on the quorum disks. If a quorum disk is not accessible and volume mirroring is unable to update the state information, a mirrored volume might need to be taken offline to maintain data integrity. To ensure the high availability of the system, ensure that multiple quorum candidate disks are allocated and configured on different storage systems.

When a volume mirror is synchronized, a mirrored copy can become unsynchronized if it goes offline and write I/O requests need to be processed, or if a mirror fast failover occurs. The fast failover isolates the host systems from temporarily slow-performing mirrored copies, which affect the system with a short interruption to redundancy.

Note: If the capacity is standard-provisioned, the primary volume formats before synchronizing to the volume copies. The **-syncrate** parameter on the **mkvdisk** command controls the format and synchronization speed.

Write fast failovers

With write fast failovers, during processing of host write I/O, the system submits writes (with a timeout value of 10 seconds) to both copies. If one write succeeds and the other write takes longer than 10 seconds, the slower request times-out and ends. The duration of the ending sequence for the slow copy I/O depends on the backend from which the mirror copy is configured. For example, if the I/O occurs over the Fibre Channel network, the I/O ending sequence typically completes in 10 to 20 seconds. However, in rare cases, the sequence can take more than 20 seconds to complete. When the I/O ending sequence completes, the volume mirror configuration is updated to record that the slow copy is now no longer synchronized. When the configuration updates finish, the write I/O can be completed on the host system.

The volume mirror stops using the slow copy for 4 - 6 minutes; subsequent I/O requests are satisfied by the remaining synchronized copy. During this time, synchronization is suspended. Additionally, the volume's synchronization progress shows less than 100% and decreases if the volume receives more host writes. After the copy suspension completes, volume mirroring synchronization resumes and the slow copy starts synchronizing.

If another I/O request times out on the unsynchronized copy during the synchronization, volume mirroring again stops using that copy for 4 - 6 minutes. If a copy is always slow, volume mirroring attempts to synchronize the copy again every 4 - 6 minutes and another I/O timeout occurs. The copy is not used for another 4 - 6 minutes and becomes progressively unsynchronized. Synchronization progress gradually decreases as more regions of the volume are written.

If write fast failovers occur regularly, there can be an underlying performance problem within the storage system that is processing I/O data for the mirrored copy that became unsynchronized. If one copy is slow because of storage system performance, multiple copies on different volumes are affected. The copies might be configured from the storage pool that is associated with one or more storage systems. This situation indicates possible overloading or other back-end performance problems.

When you enter the **mkvdisk** command to create a new volume, the **mirror_write_priority** parameter is set to **latency** by default. Fast failover is enabled. However, fast failover can be controlled by changing the value of the **mirror_write_priority** parameter on the **chvdisk** command. If the **mirror_write_priority** is set to **redundancy**, fast failover is disabled. The system applies a full SCSI initiator-layer error recovery procedure (ERP) for all mirrored write I/O. If one copy is slow, the ERP can take up to 5 minutes. If the write operation is still unsuccessful, the copy is taken offline. Carefully

consider whether maintaining redundancy or fast failover and host response time (at the expense of a temporary loss of redundancy) is more important.



Attention: Mirrored volumes can be taken offline if no quorum disk is available. This behavior occurs because synchronization status for mirrored volumes is recorded on the quorum disk. To protect against mirrored volumes being taken offline, follow the guidelines for setting up quorum disks.

Read fast failovers

Read fast failovers affect how the system processes read I/O requests. A read fast failover determines which copy of a volume the system tries first for a read operation. The *primary-for-read copy* is the copy that the system tries first for read I/O; it is determined by a user implicated read algorithm.

The system submits host read I/O request to one copy of a volume at a time. If that request succeeds, then the system returns the data. If it is not successful, the system retries the request to the other copy volume.

With read fast failovers, when the primary-for-read copy goes slow for read I/O, the system fails over to the other copy. This means that the system tries the other copy first for read I/O during the following 4 - 6 minutes. After that, the system reverts to read the original primary-for-read copy. During this period, if read I/O to the other copy also goes slow, the system reverts immediately. Also, if the primary-for-read copy changes, the system reverts to try the new primary-for-read copy. This can happen when the system topology changes or when the primary or local copy changes. For example, in a standard topology, the system normally tries to read the primary copy first. If you change the volume's primary copy during a read fast failover period, the system reverts to read the newly set primary copy immediately.

The read fast failover function is always enabled on the system. During this process, the system does not suspend the volumes or make the copies out of sync.

Maintaining data integrity of mirrored volumes during storage system maintenance

Volume mirroring improves data availability by allowing hosts to continue I/O to a volume even if one of the backend storage systems failed. However, this mirroring does not affect data integrity. If either of the backend storage systems corrupts the data, the host is at risk of reading that corrupted data in the same way as for any other volume. Therefore, before you perform maintenance on a storage system that might affect the data integrity of one copy, it is important to check that both volume copies are synchronized. Then, remove that volume copy before you begin the maintenance. For example, the scenario would apply if you need to zero the data on the disks that the storage system is providing.

Adding a copy to a basic volume

The **addvdiskcopy** command adds a copy to an existing basic volume, which changes a nonmirrored volume into a mirrored volume. In the management GUI, select **Volumes** and right-click the basic volume and select **Add Volume Copies**.

Custom volumes

Custom volumes create volumes that are based on user-defined customization rather than taking the standard default settings for each of the options under quick volume creation.

Under most circumstances, the preset volume types work well in creating new volumes quickly and efficiently by setting the standard defaults for options. However, you can customize settings to create volumes that are specific to your environment.

Standard-provisioned volumes

A standard-provisioned volume completely uses storage at creation.

A volume's capacity can be considered in terms of its provisioned capacity and its real capacity. *Provisioned capacity* is the volume storage capacity that is available to a host. *Real capacity* is the storage capacity that is allocated to a volume from a pool. *Standard-provisioned volumes* are created with the real capacity equal to the provisioned capacity.

The extra metadata that describes the contents of thin-provisioned volumes is not required for standard-provisioned volumes. As a result, the I/O rates that are obtained from standard-provisioned volumes can be higher than the rates obtained from thin-provisioned volumes that are provisioned on the same MDisks.

By default, standard-provisioned volumes are automatically formatted by a background process after the volume is created. During this process, some system resources are used to make the volume available for immediate use and ensures the volume is clear of any prior data. The time taken to format is governed by the size of the volume and the mirror synchronization rate of the volume. Increasing the synchronization rate reduces the time taken to format the volume by using additional system resources. The progress and estimated completion time of each volume format process is reported by the system.

Each formatting process uses some system resources, so there is a limit on the number of volumes that can be formatted at the same time. In addition, some volume actions such as adding a volume copy, or moving, expanding, or shrinking the volume, are disabled until the system completes formatting the volume.

If it is not necessary for a standard-provisioned volume to be cleared of data when it is created, the formatting process can be skipped. Skipping formatting can also be useful if the volume will be the target of a Copy Services function because the Copy Services operation formats the volume.

Thin-provisioned volumes

A *thin-provisioned* volume presents a different capacity to mapped hosts than the capacity that the volume consumes in the storage pool.

The system supports thin-provisioned volumes in standard pools and in data reduction pools. In standard pools, thin-provisioned volumes are created as a specific volume type, that is based on capacity savings criteria. These properties are managed at the volume level. With data reduction pools, all the benefits of thin-provisioning are available to all the volumes that are assigned to the pool. Only standard-provisioned volumes do not gain these benefits. Data reduction pools enhance capacity efficiency for thin-provisioned volumes by monitoring the hosts use of capacity. When the host indicates that the capacity is no longer needed, the capacity is released and can be reclaimed by the data reduction pool to be redistributed automatically. Standard pools do not have the ability to reclaim capacity. For added capacity savings, you can specify compression and deduplication on volumes in data reduction pools.

Thin-provisioned volumes can also help simplify server administration. Instead of assigning a volume with some capacity to an application and increasing that capacity as the needs of the application change, you can configure a volume with a large virtual capacity for the application. You can then increase or shrink the real capacity as the application needs change, without disrupting the application or server.

Provisioned capacity is the volume storage capacity that is available to a host. *Real capacity* is the storage capacity that is reserved to a volume copy from a pool. In a standard-provisioned volume, the provisioned capacity and real capacity are the same. However, in a thin-provisioned volume, the provisioned capacity can be much larger than the real capacity.

The provisioned capacity of a thin-provisioned volume is typically significantly larger than its real capacity. Each system uses the real capacity to store data that is written to the volume, and metadata that describes the thin-provisioned configuration of the volume. As more information is written to the volume, more of the real capacity is used. The system identifies read operations to unwritten parts of the virtual capacity and returns zeros to the server without using any real capacity.

The system must maintain extra metadata that describes the contents of thin-provisioned volumes. As a result, the I/O rates that are obtained from thin-provisioned volumes can be lower than the rates obtained from standard-provisioned volumes that are allocated on the same MDisks.

When you configure a thin-provisioned volume in a standard pool, you can use the warning level attribute to generate a warning event when the used real capacity exceeds a specified amount or percentage of the total provisioned capacity. You can also use the warning event to trigger other actions, such as taking low-priority applications offline or migrating data into other storage pools. For thin-provisioned volumes in data reduction pools, the warning level value cannot be set because capacity reporting is handled at the pool level.

If a thin-provisioned volume in a standard pool does not have enough real capacity for a write operation, the volume is taken offline and an error is logged (error code 1865, event ID 060001). Access to the thin-provisioned volume is restored by either increasing the real capacity of the volume or increasing the size of the storage pool that it is allocated on.

When you create a thin-provisioned volume in standard pools, you can choose the grain size for allocating space in 32 KB, 64 KB, 128 KB, or 256 KB chunks. The grain size that you select affects the maximum provisioned capacity for the thin-provisioned volume in standard pools. The default grain size is 256 KB. If you select 32 KB for the grain size, the volume size cannot exceed 260,000 GB. The grain size cannot be changed after the thin-provisioned volume is created in a standard pool. Generally, smaller grain sizes save space but require more metadata access, which can adversely impact performance. If you are not going to use the thin-provisioned volume as a FlashCopy® source or target volume, use 256 KB to maximize performance. If you are going to use the thin-provisioned volume as a FlashCopy source or target volume, specify the same grain size for the volume and for the FlashCopy function. Grain size cannot be set on thin-provisioned volume copies in data reduction pools. The grain size of 8 KB is the default size for thin-provisioned volume copies in data reduction pools.

When you create a thin-provisioned volume, set the cache mode to `readwrite` to maximize performance. If the cache mode is set to `none`, the system cannot cache the thin-provisioned metadata, which decreases performance. A volume or volume copy created from a data reduction pool must have a cache mode of readwrite. If you try to create a thin provisioned or compressed volume copy from a data reduction pool and the volume cache mode is not `readwrite`, the operation fails.

The autoexpand feature prevents a thin-provisioned volume from using up its capacity and going offline. As a thin-provisioned volume uses capacity, the autoexpand feature maintains a fixed amount of unused real capacity, called the *contingency capacity*. For thin-provisioned volumes in data reduction pools, the autoexpand feature is always enabled to maintain contingency capacity. For thin-provisioned volumes in standard pools, the autoexpand feature is optional. However without this feature enabled, the contingency capacity can get used up, causing the volume to go offline. If you are using standard pools and want to determine whether an application requires the autoexpand feature, you can create a test thin-provisioned volume with the autoexpand feature turned off. If the application causes the volume to run out of capacity and go offline, you can then create a vdisk with the autoexpand feature turned on.

Converting thin-provisioned volumes

You can convert thin-provisioned volumes into standard-provisioned volumes.

You can non-disruptively convert a thin-provisioned volume into a standard-provisioned volume by using the following volume mirroring procedure:

1. Select a single-copy, thin-provisioned volume.
2. Add a standard-provisioned copy to the volume. You can choose to have the system automatically remove the thin-provisioned copy after the standard-provisioned volume is synchronized or you can manually remove it.
3. Wait while the volume mirroring feature synchronizes.
4. Remove the thin-provisioned copy from the standard-provisioned volume.

Converting standard-provisioned volumes

You can convert standard-provisioned volumes to thin-provisioned volumes.

You can non-disruptively convert a standard-provisioned volume into a thin-provisioned volume by following this procedure:

1. Select a single copy, standard-provisioned volume.
2. Add a thin-provisioned copy to the volume.

   Use a small real capacity and the autoexpand feature. You can also choose to have the system automatically remove the standard-provisioned copy after the thin-provisioned volume is synchronized or you can manually remove it.
3. Wait while the volume mirroring feature synchronizes the copies.
4. Remove the standard-provisioned copy from the thin-provisioned volume.

Any grains of the standard-provisioned volume that contain all zeros do not cause any real capacity to be allocated on the thin-provisioned copy. Before you create the mirrored copy, you can fill the available capacity on the volume with a file that contains all zeros.

Compressed volumes

When you create volumes, you can specify compression as a method to save capacity for the volume.

With compressed volumes, data is compressed as it is written to disk, saving more space. When data is read to hosts, the data is decompressed. Compression is available through data reduction support as part of the system. If you want volumes to use compression as part of data reduction support, compressed volumes must belong to data reduction pools. Data reduction pools also support reclaiming unused capacity automatically after mapped hosts no longer need the capacity for operations. These hosts issue SCSI unmap commands and the released capacity is reclaimed by the data reduction pool for redistribution. For compressed volumes in data reduction pools, the used capacity before compression indicates the total amount of data that is written to volume copies in the storage pool before data reduction occurs.

You can also monitor information on compression usage to determine the savings to your storage capacity when volumes are compressed. To monitor system-wide compression savings and capacity, select **Dashboard** to view the **Capacity Saving**. You can compare the amount of capacity that is used before compression is applied to the capacity that is used for all compressed volumes. In addition, you can view the total percentage of capacity savings when compression is used on the system. For systems that use multiple compression technologies, such as data reduction pools or self-compressing drives, the total percentage of capacity savings cannot be determined at a system level. To view each of the capacity savings for all the supported compression technologies for the system, select **View Compression Details**. You can also monitor compression savings across individual pools. You can also monitor compression savings across individual pools.

If your system currently does not use compression, the system automatically analyzes your configuration to determine the potential storage savings if compression is used. The management GUI incorporates the Comprestimator utility that uses mathematical and statistical algorithms to create potential compression savings for the system. The analysis for potential savings can be used to determine whether purchasing a compression license for the system is necessary to reduce cost of extra storage devices. To estimate compression savings in the management GUI, select **Volumes** > **Actions** > **Space Savings** > **Estimate Compression Savings**.

After the analysis completes, you can download a savings report that shows estimated savings for all the volumes with enough data to be analyzed. This report lists all currently configured volumes on the system and their potential compressions savings. To download a report, select **Volumes** > **Volumes** > **Actions** > **Capacity Savings** > **Download Savings Report**.

Deduplicated volumes

Deduplication can be configured with volumes that use different capacity saving methods, such as thin-provisioning.

Deduplicated volumes must be created in data reduction pools for added capacity savings. Deduplication is a type of data reduction that eliminates duplicate copies of data. Deduplication of user data occurs within a data reduction pool and only between volumes or volume copies that are marked as deduplicated. Some models or software versions require specific hardware or software to use this function. With deduplication, the system identifies unique chunks of data, called *signatures*, to determine whether new data is written to the storage. Deduplication is a hash-based solution, which means chunks of data are compared to their signatures rather than to the data itself. If the signature of the new data matches an existing signature that is stored on the system, then the new data is replaced with a reference. The reference points to the stored data, instead of writing the data to storage. This process saves capacity on the backend storage by not writing new data to storage and might improve performance on read operations to data with an existing signature. The same data pattern can occur many times and deduplication decreases the amount of data that needs to be stored on the system. A part of every hash-based deduplication solution is a repository that supports looking up matches for incoming data. The system contains a database that maps the signature of the data to the volume and its virtual address. If an incoming write operation does not have a signature that is stored in the database, then a duplicate is not detected and the incoming data is stored on backend storage. To maximize the space that is available for the database, the system distributes this repository between all nodes in the I/O groups that contain deduplicated volumes. Each node carries a distinct portion of the records that are stored in the database. If nodes are removed or added to the system, the database is redistributed between the nodes to ensure full use of available memory. Only certain models with specific hardware support deduplication. Verify your model and hardware components to use these functions.

When you create a volume, you can specify to include deduplication with other supported capacity savings methods. Deduplicated volumes must be created in data reduction pools. If you have existing volumes in standard pools, you can migrate them to data reduction pools to add deduplication to increase capacity savings for the volume.

Volume protection

Volume protection prevents active volumes or host mappings from being deleted inadvertently if the system detects recent I/O activity.

This global setting is enabled by default on new systems. You can either set this value to apply to all volumes that are configured on your system, or control whether the system-level volume protection is enabled or disabled on specific pools. To prevent an active volume from being deleted unintentionally, administrators can use the system-wide setting to enable volume protection. They can also specify a time period that the volume must be idle before it can be deleted.

The system-wide volume protection and the pool-level protection must both be enabled for protection to be active on a pool. The pool-level protection depends on the system-level setting to ensure that protection is applied consistently for volumes within that pool. If system-level protection is enabled, but pool-level protection is not enabled, any volumes in the pool can be deleted even when the setting is configured at the system level.

The following commands are affected by this setting:

- **rmvdisk**
- **rmvdiskcopy**
- **rmvvolume**
- **rmvdiskhostmap**
- **rmvolumehostclustermap**
- **rmmdiskgrp**
- **rmhostiogrp**

- **rmhost**
- **rmhostcluster**
- **rmhostport**

To change the volume protection setting and disable volume protection, complete the following steps:

1. To change the system-wide value and disable volume protection on the system, see **chsystem** command.
2. To change the pool-wide value and disable volume protection on the pools, see **chmdiskgrp** command.

Expand or shrink volumes

You can use the management GUI or the command-line interface (CLI) to expand or shrink the capacity of a volume.

Expand volumes

The system supports expanding the size of a volume concurrent with I/O operations, when supported by the host operating system.

You can expand volumes for the following reasons:

- To increase the available capacity on a particular volume that is already mapped to a host.
- To increase the size of a volume so that it matches the size of another volume so that it can be used in a FlashCopy mapping.

  Note:

  – For volumes in non-incremental FlashCopy mappings, you can increase the capacity regardless of whether the volume is the source or target volume. For volumes in incremental FlashCopy mappings, both the source and target volumes can be expanded, but the capacity of the target volume must be expanded first.

To expand a volume, see the **expandvdisksize** command.

Shrink volumes

You can reduce the capacity of a compressed or uncompressed volume by using the management GUI or the command-line interface (CLI).

Volumes can be reduced in capacity, if it is necessary. You can make a target or auxiliary volume the same capacity as the source or master volume when you create FlashCopy mappings,. However, if the volume contains data, do not shrink the size of the disk. The system disables shrinking a volume if the selected volume is performing quick initialization. After the quick initialization completes, you can shrink the volume.



Attention:

1. It is difficult to anticipate how an operating system or file system uses the capacity in a volume. When you shrink a volume, capacity is removed from the end of the disk, whether or not that capacity is in use. Even if a volume has available capacity, do not assume that only unused capacity is removed when you shrink a volume.
2. If the volume contains data that is being used, *do not attempt under any circumstances to shrink a volume without first backing up your data*.
3. For performance reasons, some operating systems or file systems use the outer edge of the disk.

You can use the **shrinkvdisksize** command or the management GUI to shrink the usable capacity that is provisioned to the particular volume by the specified amount. You can also shrink the provisioned capacity of a thin-provisioned volume without altering the usable capacity that is assigned to the volume.

You cannot shrink the following types of volumes:

- Volumes in a data reduction pool.
- A target volume in FlashCopy mappings cannot be shrunk. However, you can shrink a source volume, if the starting capacity of the FlashCopy mapping is not greater than the proposed capacity for the source volume. The management GUI filters target volumes that can be shrunk and disables the shrink action for those volumes that do not meet the criteria. If you are using the command-line interface, use the **lsfcmap** command to display the starting capacity (*start_capacity* parameter) of the FlashCopy mapping. This parameter indicates the capacity that was used the last time the FlashCopy operation was completed.

To shrink a volume, see the **shrinkvdisksize** command.

Dependent volumes

You can use the command-line interface (CLI) to list the volumes that depend on the status of a node.

Before you begin

If a node goes offline or is removed from a system, all volumes that depend on the node go offline. Before you take a node offline or remove a node from a system, run the **lsdependentvdisks** command to identify any node-dependent volumes.

About this task

By default, the **lsdependentvdisks** command also checks all available quorum disks. If the quorum disks are accessible only through the specified node, the command returns an error.

Various scenarios can produce node-dependent volumes. The following examples are common scenarios in which the **lsnodedependentvdisks** command returns node-dependent volumes:

1. The node contains flash drives with the only synchronized copy of a mirrored volume.
2. The node is the only node that can access an MDisk on the SAN fabric.
3. The other node in the I/O group is offline (all volumes in the I/O group are returned).
4. Pinned data in the cache is stopping the partner node from joining the I/O group.

To resolve (1), allow volume mirror synchronizations between Flash drive MDisks to complete. To resolve (2-4), bring any offline MDisks online and repair any degraded paths.

Note: The command lists the node-dependent volumes at the time the command is run; subsequent changes to a system require running the command again.

Procedure

1. Issue the **lsdependentvdisks** CLI command.

   The following example shows the CLI format for listing the volumes that depend on node01:

   ```
   lsdependentvdisks -enclosure -delim : 0:1
   ```

   The following example shows the output that is displayed:

   ```
   vdisk_id:vdisk_name
   4:vdisk4
   5:vdisk5
   ```

2. If the command **lsdependentvdisks** returns an error, you must move your quorum disks to MDisks that are accessible through all nodes.

   Rerun the command until no errors are returned.
3. Reissue the **lsdependentvdisks** command.

When the command returns no volumes, the system is free from any node-dependent volumes.

The following example shows the command syntax for listing the volumes that depend on node01:

```
lsdependentvdisks -delim : -node node01 :
```

If there are no node-dependent volumes in the system, the following example shows the command output:

```
vdisk_id       vdisk_name
```

Managing throttles for volumes

You can create, modify, and remove throttles for volumes by using the management GUI or the command-line interface.

Throttling is a mechanism to control the amount of resources that are used when the system is processing I/Os on a specific volume or volume group. If a throttle is defined, the system either processes the I/O, or delays the processing of the I/O to free resources for more critical I/O.

Managing throttles with the management GUI

To create a new throttle or edit an existing throttle for a specific volume, complete these steps:

1. In the management GUI, select **Volumes** > **Volumes**.
2. Right-click a volume and select **Edit Throttle...**.

   Note: You can select multiple volumes to apply throttles to more than one volume.
3. On the **Edit Throttle** page, enter values for the following throttles:

   Bandwidth Limit
   Enter the maximum amount of bandwidth the volume can process before the system delays the processing for this volume. Click **Create**.

   IOPS Limit
   Enter the maximum I/O operations per second the volume can process before the system delays the processing for this volume. Click **Create**.

To edit an existing throttle for a specific volume, complete these steps:

1. In the management GUI, select **Volumes** > **Volumes**.
2. Right-click a volume and select **Edit Throttle...**.
3. On the **Edit Throttle** page, enter values for the following throttles:

   Bandwidth Limit
   Enter the maximum amount of bandwidth the volume can process before the system delays the processing for this volume. Click **Save**.

   IOPS Limit
   Enter the maximum I/O operations per second the volume can process before the system delays the processing for this volume. Click **Save**.

To remove throttles for a specific volume, complete these steps:

1. In the management GUI, select **Volumes** > **Volumes**.
2. Right-click a volume or multiple volumes and select **Edit Throttle...**.
3. On the **Edit Throttle** page, click **Remove** by the throttle that you want to delete.

Managing throttles with the command-line interface

To create a new throttle or edit an existing throttle for a volume with the command-line interface, enter the following command:

```
mkthrottle -type vdisk -bandwidth bandwidth_limit_in_mb -iops iops_limit -vdisk
vdisk_id_or_vdisk_name
```

where *bandwidth_limit_in_mb* is the maximum amount of bandwidth that the system can process for the specified volume before the system delays processing of I/O for that volume. The *iops_limit* is the maximum I/O operations per second that the system can process before the system delays the processing for this volume. The *vdisk_id_or_vdisk_name* is the ID or name of the volume that the throttle applies to.

To change an existing throttle for the volume, enter the following command:

```
chthrottle -type vdisk -bandwidth bandwidth_limit_in_mb -iops iops_limit throttle_name_or_id
```

where *bandwidth_limit_in_mb* is the maximum amount of bandwidth that the system can process for the specified host before the system delays processing of I/O for that volume. The *iops_limit* is the maximum I/O operations per second that the system can process before the system delays the processing for this volume. The *throttle_id_or_name* is the ID or name of the throttle that is being changed.

To remove a throttle for volume with the command-line interface, enter the following command:

```
rmthrottle throttle_name_or_id
```

where *throttle_name_or_id* is the ID or name of the throttle that is being removed.

Volume groups

A volume group allows for many related volumes to be grouped so that they can be managed together. A volume group also can ensure consistency across all volumes in the group when used with snapshot or replication functions.

Volume groups can be used with the following functions:

Safeguarded Copy
Safeguarded Copy function is a cyber-resiliency feature that creates immutable copies of data that cannot be changed or manipulated.

A Safeguarded volume group describes a set of source volumes that are backed up collectively with the Safeguarded Copy function. Safeguarded snapshots can be configured through an internal scheduler that is defined in a snapshot policy or can be configured with an external snapshot scheduling application such as IBM Copy Services Manager.

Snapshot
Snapshots are point-in-time copies of a volume group that cannot be directly accessed by hosts. To access the snapshot contents, you can create a clone or thin clone. You can schedule snapshots using a snapshot policy or an external snapshot scheduling application. For more details, see <snapshots>

High availability and disaster recovery

Policy-based high availability provides an active-active high availability solution that uses two independent systems in different locations. The storage will automatically remain accessible to hosts if there is an event that makes one of the systems unavailable. When using policy-based high availability, everything in a storage partition is highly available; all volume groups in a partition are replicated for high availability.

Policy-based replication for disaster recovery provides asynchronous replication to another system for disaster recovery. Using this type of replication, a replication policy is assigned to individual volume groups to manage groups that are replicated.

Transparent cloud tiering

Transparent cloud tiering enables volume data to be copied and transferred to cloud storage.

The system supports creating connections to cloud service providers to store copies of volume data in private or public cloud storage. With transparent cloud tiering, administrators can move older data to cloud storage to free up capacity on the system. Point-in-time snapshots of data can be created on the system and then copied and stored on the cloud storage. An external cloud service provider manages the cloud storage, which reduces storage costs for the system. Before data can be copied to cloud storage, a connection to the cloud service provider must be created from the system.

A cloud account is an object on the system that represents a connection to a cloud service provider by using a particular set of credentials. These credentials differ depending on the type of cloud service provider that is being specified. Most cloud service providers require the host name of the cloud service provider and an associated password, and some cloud service providers also require certificates to authenticate users of the cloud storage. Public clouds use certificates that are signed by well-known certificate authorities. Private cloud service providers can use either a self-signed certificate or a certificate that is signed by a trusted certificate authority. These credentials are defined on the cloud service provider and passed to the system through the administrators of the cloud service provider.

Once the system is authenticated, it can then access cloud storage to either copy data to the cloud storage or restore data that is copied to cloud storage back to the system. The system supports one cloud account to a single cloud service provider. Migration between providers is not supported.

Each cloud service provider divides cloud storage into segments for each client that uses the cloud storage. These objects store only data specific to that client. The names of the objects begin with a prefix that you can specify when you create the account for the system. A prefix defines system-specific content that the object stores and supports multiple independent systems to store data to a single cloud account. Each cloud service provider uses different terminology for these storage objects.

The following requirements apply to Transparent Cloud Tiering:

1. When a cloud account is created, it must continue to use the same encryption type, throughout the life of the data in that cloud account. Even if the cloud account object is removed and remade on the system, the encryption type for that cloud account cannot be changed while back up data for that system exists in the cloud provider.
2. Performing rekey operations on a system with an encryption enabled cloud account, perform the commit operation immediately after the prepare operation. Remember to retain the previous system master key (on USB or in Key server) as this key can still be needed to retrieve your cloud backup data when performing a T4 recovery or an import.
3. Avoid the use of the **Restore_uid** option when backup is imported to a new system.

Supported cloud service providers

The system supports connections to various cloud service providers. Some cloud service providers require connections over external networks, and others can be created on a private network.

Each cloud service provider requires different configuration options. The system supports the following cloud service providers:

OpenStack Swift
OpenStack Swift is a standard cloud computing architecture from which administrators can manage storage and networking resources in a single private cloud environment. Standard application programmable interfaces (APIs) can be used to build customizable solutions for a private cloud solution.

Microsoft Azure
Microsoft Azure is an object storage solution that allows large volumes of unstructured data to be stored in the Azure public cloud. It provides an API and web-based interface to help you manage your cloud solution. Microsoft Azure is suitable for archiving, disaster recovery, and backup data storage.

Amazon S3
Amazon Simple Storage Service (Amazon S3) provides programmers and storage administrators with flexible and secure public cloud storage. Amazon S3 is also based on object storage standards and provides a web-based interface to manage, back up, and restore data over the web.

The following Amazon S3 (AWS) regions are supported by this code-level:

- us-east-2
- us-east-1
- us-west-1
- us-west-2
- af-south-1
- ap-east-1
- ap-south-2
- ap-southeast-3
- ap-southeast-4
- ap-south-1
- ap-northeast-3
- ap-northeast-2
- ap-southeast-1
- ap-southeast-2
- ap-northeast-1
- ca-central-1
- eu-central-1
- eu-west-1
- eu-west-2
- eu-south-1
- eu-west-3
- eu-north-1
- eu-south-2
- eu-central-2
- sa-east-1
- me-south-1
- me-central-1
- il-central-1

Security considerations for cloud accounts

Whenever the system accesses outside networks, the potential for unintentional or intentional exposure of sensitive data is a risk. When you are connecting the system to a cloud service provider over a public network, you can use encryption to protect data that is transferred to the cloud service provider.

The first level of encryption-based security provides secure communications between the system and the cloud service provider. The standard protocol, Transport Layer Security (TLS), protects these connections by encrypting data that is transferred between the system and the cloud service provider. Secure communications are mandatory for these connections and requires that public certificates are exchanged between the cloud service provider and the system. To configure certificates for secure communications in the management GUI, go to **Settings** > **Security** > **System Certificates**. You can also use the **chsystemcert** command to create system certificates. With secure communications, data is encrypted while it is transferred to the cloud, but might be stored on the cloud decrypted. Each cloud service

provider has its own security measures to protect data once it is located in cloud storage; however, breaches can still occur and data can be compromised. Clients that use cloud service providers can add extra encryption methods to protect their data after it is stored on the cloud.

Since the system supports encryption of at-rest data, you can optionally configure encryption key management to further protect data that is stored on the cloud storage. If key management is configured on the system, data is encrypted before it leaves the system and is stored on the cloud. The system supports key management through either a USB flash drive or an encryption key server. When encryption is configured, a master encryption key is created and is stored separately on either a USB flash drive or key server. When you create snapshots of data to send to the configured cloud service provider, each volume and each cloud account have separate encryption keys. The encryption key that is used by the cloud account protects encryption keys for the volumes. The master encryption key protects the encryption key that is used by the cloud account. Because the master encryption key is physically present on the USB flash drive or key server, you must ensure that security measures are implemented to protect the master encryption key from theft or loss. When the data is transmitted between the system and the cloud service provider, the data is also encrypted by certificates that are configured for secure communications. The master encryption key also protects the data in transit and the data remains encrypted while it is stored on the cloud storage. Data also remains encrypted with the encryption master key when it transferred back to the system from the cloud during restore operations. Finally, data can be decrypted when it arrives at the system or it can be stored on an encrypted volume on the system.

When a connection to a cloud service provider is configured, you must decide whether to encrypt data at rest in the cloud for this account. After you decide, the encryption setting for the account cannot be changed without restoring all data from the cloud, reconfiguring the account, and re-creating cloud snapshots for the data.

Encryption protects the data, but also requires careful configuration and management of keys. Depending on how key management is configured on the system, access to the keys is required when data is copied to or restored from the cloud service provider.

Configuring transparent cloud tiering

With transparent cloud tiering, you can configure a cloud account on the system to be able to create and restore cloud snapshots of system volumes.

Creating a DNS server

Before you create a connection to a cloud service provider or connect to cloud storage, ensure that you specify at least one DNS server to manage hostnames. You can have up to two DNS servers that are configured on the system. To configure DNS for the system, enter a valid IP address and name for each server. Both IPv4 and IPv6 address formats are supported.

To create a DNS server for host name resolution, use the **Settings** > **Network** > **DNS** panel in the management GUI.

To create a DNS server using the command-line interface, see **`mkdnsserver`** command.

Enabling a cloud connection

To configure transparent cloud tiering on the system, you must enable a cloud connection to a supported cloud service provider.

A cloud account is an object on the system that represents a connection to a cloud service provider by using a particular set of credentials. These credentials differ depending on the type of cloud service provider that is being specified. Most cloud service providers require the host name of the cloud service provider and an associated password, and some cloud service providers also require certificates to authenticate users of the cloud storage. Public clouds use certificates that are signed by well-known certificate authorities. Private cloud service providers can use either a self-signed certificate or a certificate that is signed by a trusted certificate authority. These credentials are defined on the cloud service provider and passed to the system through the administrators of the cloud service provider.

Enabling a cloud connection to OpenStack Swift

To enable a cloud connection between the system and OpenStack Swift, use the **Settings** > **System** > **Transparent Cloud Tiering** panel in the management GUI. Expand the **OpenStack account** and enter the following information:

Tenant
Enter the name of the OpenStack tenant that the system uses to access the cloud storage.

User name
Enter the user name that is associated with the account. The user name identifies clients that can access content that is stored by the cloud service provider.

Password
Enter the password that is associated with the user name for the cloud account.

Endpoint
Enter the URL that the system uses to access the cloud service provider.

Container prefix
Enter a short string that is added to the beginning of the container name to identify which containers are associated with the system. Prefixes can also be used to identify multiple systems that use the same cloud account to store snapshots of volume data.

SSL Certificate
Select the SSL certificate that is used to authenticate the system to the cloud service provider.

Keystone
Keystone authentication system is required for all OpenStack Swift cloud accounts. The management GUI automatically detects whether Keystone is enabled. If it is not enabled, a warning message displays. Click **Enable** to configure Keystone authentication for the account.

Encryption
If encryption is configured on the system, this value is preselected. If encryption is not enabled on the system, this option is unavailable.

Bandwidth
Determine how bandwidth is used for downloads to and uploads from the cloud service provider. Specify **No Limit** to allow the connection to use unlimited bandwidth, regardless of other operations that occur on the system. If you do not want to give unlimited bandwidth to these operations, you can specify the maximum amount of bandwidth that the connection can use for either downloads or uploads.

To enable a cloud connection between the system and OpenStack Swift interfaces using the command-line interface, see **`mkcloudaccountswift`** command.

Enabling a cloud connection to Amazon S3

To enable a cloud connection between the system and Amazon S3 , use the **Settings** > **System** > **Transparent Cloud Tiering** panel in the management GUI. Expand **Amazon S3 account** and enter the following information:

Access key ID
The public part of the access key credential for the Amazon user that the system uses to access the cloud storage.

Secret access key

Enter the private part of the Amazon credential that the system uses to authenticate the system to cloud storage.

Region
Select the Amazon region that is used to both access and store data on cloud storage.

Bucket prefix
Enter the prefix for the Amazon S3 bucket names that the system will use or create. Prefixes can also be used to identify multiple systems that use the same cloud account to store snapshots of volume

data. If you want to import or export snapshots between systems, use the same bucket prefix as the other system.

Encryption
If encryption is configured on the system, this value is preselected. If encryption is not enabled on the system, this option is unavailable.

Bandwidth
Determine how bandwidth is used for downloads to and uploads from the cloud service provider. Specify **No Limit** to allow the connection to use unlimited bandwidth, regardless of other operations that occur on the system. If you do not want to give unlimited bandwidth to these operations, you can specify the maximum amount of bandwidth that the connection can use for either downloads or uploads.

To enable a cloud connection between the system and Amazon S3 using the command-line interface, see **`mkcloudaccountawss3`** command.

Enabling a cloud connection to Microsoft Azure

To enable a cloud connection between the system and Microsoft Azure, use the **Settings** > **System** > **Transparent Cloud Tiering** panel in the management GUI. Expand **Cloud Provider Settings** page for Microsoft Azure, enter the following information:

Storage Account Name
Enter the name of the Microsoft Azure storage account that the system uses to authenticate to cloud storage.

Storage Account Key
Enter the account access key that the system uses to access the cloud storage.

Container Prefix
Enter the container prefix to identify the objects that the system creates and uses on the cloud storage.

Encryption
If encryption is configured on the system, this value is preselected. If encryption is not enabled on the system, this option is disabled by default.

Bandwidth
Determine how much bandwidth is used by download or upload operations for the cloud service provider. Specify **No Limit** to allow the connection to use unlimited bandwidth, regardless of other operations that occur on the system. If you do not want to give unlimited bandwidth to these operations, you can specify the maximum amount of bandwidth that the connection can use for either downloads or uploads. The maximum bandwidth limit for downloads and uploads is 10240 Mbps.

To enable a cloud connection between the system and Microsoft Azure using the command-line interface, see **`mkcloudaccountazure`** command.

Managing cloud accounts

A cloud account is an object on the system that represents a connection to a cloud service provider by using a particular set of credentials. These credentials differ depending on the type of cloud service provider that is being specified. Most cloud service providers require the host name of the cloud service provider and an associated password, and some cloud service providers also require certificates to authenticate users of the cloud storage. Public clouds use certificates that are signed by well-known certificate authorities. Private cloud service providers can use either a self-signed certificate or a certificate that is signed by a trusted certificate authority. These credentials are defined on the cloud service provider and passed to the system through the administrators of the cloud service provider. You can manage cloud accounts with the management GUI or the command-line interface.

Verify cloud account

To verify the cloud account state is online, use the **Settings** > **System** > **Transparent Cloud Tiering** panel in the management GUI. If the cloud account is not online, the connection to the cloud service provider is not available. Use the **Monitoring** > **Events** panel to investigate the cause of the problem. You might need to access the cloud service provider to determine the connection problem.

Changing settings on cloud account

Certain settings are controlled by the cloud service provider and cannot be changed. However, you can update credential information for a cloud account if necessary.

When settings are changed on the cloud account, any snapshot versions of volumes that are stored on the cloud becomes unavailable. Before you change any settings that are related to the cloud account, ensure that you restore data from the cloud storage or delete the cloud snapshots from cloud storage. After the changes are completed to the cloud account, new cloud snapshots can be created of volume data. Some situations might require that you update certain settings on the cloud account. For example, if a cloud service provider updates credential information, then changes to the cloud account information on the system are required. For some settings, the system restricts any updates.

OpenStack Swift account

For a cloud account with OpenStack Swift, the following settings cannot be changed:

Tenant
A tenant is a container that is used to store objects that are created for a particular account. Tenant names are defined by the cloud service provider and represent the underlying storage that is used by the account. Tenant names must remain consistent or data that is stored on the cloud cannot be accessed.

Container prefix
Container prefixes are added to the beginning of the tenant name to identify which tenants are associated with the system. Prefixes can also be used to identify multiple systems that use the same cloud account to store snapshots of volume data. Like with tenant names, these prefixes must remain consistent to access data on the cloud.

Encryption
After encryption is set for a cloud account, it cannot be changed.

Use the **Settings** > **System** > **Transparent Cloud Tiering** panel in management GUI to change settings for an OpenStack Swift account. Expand **Account information** to view the settings that are related to the account. The following settings can be updated:

User name
Enter the new user name that is associated with the account. The user name is used to identify clients that can access content that is stored by the cloud service provider.

Password
Enter the new password that gives access to the cloud storage for the specified user name. To update the password, click **Edit**.

Endpoint
Enter the new URL that the system uses to access cloud storage on the cloud service provider.

SSL Certificate
Select the new SSL certificate that is used to authenticate the system to the cloud service provider.

Keystone
Keystone authentication system is required for all OpenStack Swift cloud accounts. The management GUI automatically detects whether Keystone is enabled. If it is not enabled, a warning message displays. Click **Enable** to configure Keystone authentication for the account.

To change settings using the command-line interface for the OpenStack Swift storage account, see the **`chcloudaccountswift`** command.

Amazon S3 account

For a cloud account with Amazon S3, the following settings cannot be changed:

Region
Enter the user name that is associated with the account. The user name is used to identify clients that can access content that is stored by the cloud service provider.

Bucket Prefix
Enter short string that is added to the name of the bucket to identify which storage objects are associated with the system. Prefixes can also be used to identify multiple systems that use the same cloud account to store snapshots of volume data. A bucket is a container that Amazon S3 creates to hold objects that are associated with a particular client.

Encryption
If encryption is configured on the system, this value is preselected. If encryption is not enabled on the system, this option is unavailable.

Use the **Settings** > **System** > **Transparent Cloud Tiering** panel in management GUI to change settings for an Amazon S3 cloud account. Expand **Account information** to view the settings that are related to the account. The following settings can be updated:

Access key ID
Enter the new public part of the credential that the system uses to authenticate to cloud services. The access key ID is essentially a user name.

Secret access key
Enter the new private part of the credential that the system uses to authenticate the system to cloud services. To update the secret access key, click **Edit**.

To change settings using the command-line interface for the Amazon S3 storage account, see the **chcloudaccountawss3** command.

Microsoft Azure account

For a cloud account with Microsoft Azure, the following settings cannot be changed:

Container Prefix
Prefixes are added to the beginning of the container names. Container prefixes help to identify the objects that the system creates and uses on the cloud storage.

Encryption
If encryption is configured on the system, this value is preselected. If encryption is not enabled on the system, this option is not available.

Use the **Settings** > **System** > **Transparent Cloud Tiering** panel in management GUI to change settings for the Microsoft Azure storage account. Expand **Account information** to view the settings that are related to the account.

For Storage account key, you can add an access key or update the existing access key that the system uses to authenticate the cloud account.

To change settings using the command-line interface for the Microsoft Azure storage account, see the **chcloudaccountazure** command.

Testing a cloud account

The command-line interface supports diagnostics and status reports for network connectivity, authentication, and the ability to use cloud storage. You can use this information for troubleshooting and monitoring the status of your cloud account.

To test a cloud account, enter the command **testcloudaccount cloud_account_id_or_name**. For more information, see **testcloudaccount**.

Removing a cloud account

If a cloud account is no longer needed, it can be removed from the system. You can use the management GUI or the command-line interface to delete a cloud account.

If cloud snapshots are not currently in the cloud storage that is associated with the account, the cloud account settings and any containers and metadata are deleted from the cloud service provider. If cloud snapshots do exist on the cloud, only the cloud account information is deleted, but the cloud snapshots are retained on the cloud service provider. If at some point the data on the cloud is needed, you can re-create a new cloud account that uses the same settings and you can access the data that is stored on the cloud service provider. For the remove operation to succeed, no active cloud snapshots or restore operations can be in progress.

Use the **Settings** > **System** > **Transparent Cloud Tiering** panel in management GUI to remove the cloud account. Expand **Account information** to view the settings that are related to the account.

To remove a cloud account using the command-line interface, see the **`rmcloudaccount`** command.

Cloud volumes and snapshots

A cloud volume is any volume that is enabled for transparent cloud tiering. After transparent cloud tiering is enabled on a volume, point-in-time copies, or snapshots, can be created and copied to cloud storage that is provided by a cloud service provider. These snapshots can be restored to the system for disaster recovery purposes. Before you create cloud volumes, a valid connection to a supported cloud service provider must be configured.

With transparent cloud tiering, the system supports connections to cloud service providers and the creation of cloud snapshots of any volume or volume group on the system. Cloud snapshots are point-in-time copies of volumes that are created and transferred to cloud storage that is managed by a cloud service provider. A cloud account defines the connection between the system and a supported cloud service provider and must be configured before data can be transferred to or restored from the cloud storage. After a cloud account is configured with the cloud service provider, you determine which volumes you want to create cloud snapshots of and enable transparent cloud tiering on those volumes.

In the management GUI, volumes that do not have restrictions are enabled for cloud snapshots when a snapshot is created for that volume. If you are using the command-line interface, you must enable snapshots on the volume before you can create a snapshot of the volume.

A volume cannot be enabled for cloud snapshots if the volume is used in any of the following capacities:

- The volume is a VMware vSphere Virtual Volumes.
- A volume has a copy in a different storage pool.
- A volume is being migrated between different storage pools.
- A volume cannot be enabled for cloud snapshots if the cloud storage is set to import mode.
- A volume cannot be enabled for cloud snapshots if the maximum number of cloud volumes already exists. The maximum number of cloud volumes on the system is 1024.

  Note: If the system exceeds this limit, you can disable cloud snapshots on an existing cloud volume and delete its associated snapshots from the cloud storage to accommodate snapshots on new cloud volumes.
- The volume is associated with user-owned legacy FlashCopy mappings.

After the volume is enabled for the function, cloud snapshots of those volumes can be created and stored on cloud storage. Multiple snapshots can be created at different times and each version is timestamped and stored in the cloud storage. Each snapshot version on the cloud storage represents the data at a specific point in time, so different versions can be restored if necessary. The system supports both snapshots of local system data and also data on different systems. Some of these cloud snapshots can have corresponding cloud volumes on the local system, but other snapshots might not. You can restore data from these different snapshots of the cloud volume if necessary.

The system supports two types of cloud snapshots:

Full snapshots
When the cloud snapshot is created for the first time for a volume, all data is copied to the cloud storage. Any subsequent cloud snapshots of the volume include only the changed data from the initial snapshot. Each snapshot version of the volume is timestamped and stored on the cloud storage. Therefore, full plus incremental snapshots ensure that all data is copied to the cloud storage and is available if a restore operation is necessary. If the volume contains a large amount of data, full snapshots can be time-consuming.

Incremental snapshots
The version contains changed data from the last time the cloud snapshot was created for a selected volume. Incremental-only snapshots are quicker to complete than a full snapshot.

Individual volumes can be added to a single volume group for creating snapshots for all the volumes in that group. If volumes have similar content or are used by a specific host or application, they can be added to volume group to simplify snapshot operations. Separate snapshots are started for each volume in the group and the group must be created before the snapshots are created. Volume groups can be configured by using the **mkvolumegroup** command. The management GUI lists all cloud volumes and specifies the name of the volume group in which the volumes belong.

If you want to limit access of users to specific volume groups that are used in creating cloud snapshots, you can assign these volume groups to an ownership group. An ownership group defines a subset of users and objects within the system.Volumes that are within the volume group must be within a child pool.

Ownership can be defined explicitly or it can be inherited from the user, user group, or from other parent resources, depending on the type of resource. A user cannot change the ownership group of the resource, but can change the ownership group of the parent object. The following rules apply to volume groups that are defined in ownership groups:

- Volumes can be added to a volume group if both the volume and the volume group are within the same ownership group or if both are not in an ownership group. When migrating existing resources to ownership groups, the volume group and its volumes can belong to different ownership groups temporarily until the migration completes.
- The ownership of a volume group does not affect the ownership of the volumes it contains. If a volume group and its volumes are owned by different ownership groups, then the owner of the child pool that contains the volumes can change the volume directly. For example, the owner of the child pool can change the name of a volume within it. The owner of the volume group can change the volume group itself and indirectly change the volume, such as deleting a volume from the volume group. Neither the ownership group of the child pools nor the owner of the volume group can directly manipulate the resources that are not defined in their ownership group.

Individual cloud volumes and volume groups have different statuses to indicate whether the volume is available for snapshot operations. These statuses can be used to monitor in-progress snapshots and error states for the snapshot. Cloud volumes have the following snapshot statuses:

Off
Transparent cloud tiering is disabled. Snapshots are not enabled for the volume or volume group.

Ready
Transparent cloud tiering is enabled. A new snapshot can be started for the volume or volume group.

Copying
Transparent cloud tiering is enabled. A snapshot operation is in progress and data is being transferred to the cloud.

Copying error
Transparent cloud tiering is enabled. An existing snapshot operation is in-progress and cannot be completed because of an error. Causes for copying errors include losing the network connection to the cloud service provider or exceeding the amount of capacity on the cloud storage.

Not ready
Transparent cloud tiering is enabled. A new snapshot operation cannot be started because a restore is in progress for the volume.

In addition to creating snapshots of volume data to transfer to cloud storage, transparent cloud tiering also supports restoring volume data from cloud storage back to the system. Like with snapshots, versions of each copy operation are timestamped and stored on the cloud, which allows for specific point-in-time recovery operations to the original system or to another system. Older versions can be deleted and removed from the system and the cloud storage simultaneously to manage space. Snapshots can be restored only for individual volumes. You can restore multiple volumes simultaneously, but each restore operation needs to be started separately for each volume.

Certain recovery scenarios can require the restoration of an older version of the cloud snapshot. However, all newer versions on the cloud storage are deleted, including any in-progress snapshot operations. The management GUI verifies the version of cloud snapshot and issues a warning if the selected snapshot is not the most recent version. To proceed with the restore operation, you must confirm the deletion of these other versions.

You restore volume data with any snapshot version that was copied to the cloud, including restoring from a snapshot version that is still in progress. The snapshot version and volume that the snapshot version is being restored to must be the same size. A version cannot be deleted while it is being used for a restore. You can restore a snapshot version with following methods:

Restore to the production volume
: If the snapshot version is restored to the production volume, which is the original volume from which the snapshots were created. The snapshot version replaces the current data that exists on production volume with the data that is stored on the cloud storage. The production volume goes offline during the restore operations. Data is not fully restored to the production volume until the changes are committed.

Restore to a new volume
: When the snapshot version is restored to a new volume, you can use the restored data independently of the original volume from which the snapshot was created. If the new volume exists on the system, then the restore operation uses the unique identifier (UID) of the original volume. If the new volume does not exist on the system, you need to choose whether to use the UID from the original volume or create a new UID. If you plan on using the new volume on the same system, use the UID that is associated with the snapshot version that is being restored. You want a unique UID if you are restoring a version of a volume that exists on another system to the current system.

Individual cloud volumes have different statuses to indicate whether the volume is available for restore operations. These statuses can be used to monitor in-progress restore operations. Cloud volumes have the following restore statuses:

None
: No snapshot version for this volume available for restore.

Available
: Indicates that the selected volume has available snapshot versions on cloud storage that can be restored.

Restoring
: Indicates that the restore is in progress for this volume.

Restoring error
: Indicates that the restore operation is in progress, but an error occurred. Causes for restoring errors include losing the network connection to the cloud service provider or exceeding the amount of capacity on the pools in which the volume belongs.

Managing cloud volumes

Cloud volumes are any volume on the system that is enabled for cloud snapshots. You can use the management GUI or the command-line interface (CLI) to manage the cloud volumes and cloud snapshots.

Enable cloud snapshots on volumes

In the management GUI, when you create a snapshot, cloud snapshots are automatically enabled for the volume when the volume is not restricted from cloud snapshots. If you are using the command-line interface, you must enable cloud snapshots for the volume separately.

Cloud snapshot restrictions

A volume cannot be enabled for cloud snapshots if the volume is used in any of the following capacities:

- The volume is a VMware vSphere Virtual Volume.
- A volume has a copy in a different storage pool.
- A volume is being migrated between different storage pools.
- A volume cannot be enabled for cloud snapshots if the cloud storage is set to import mode.
- A volume cannot be enabled for cloud snapshots if the maximum number of cloud volumes already exists.

Note: If the system exceeds this limit, you can disable cloud snapshots on an existing cloud volume and delete its associated snapshots from the cloud storage to accommodate snapshots on new cloud volumes.

- Volume is associated with user-owned legacy FlashCopy® mappings.

Note: Cloud snapshots are supported with the system's new FlashCopy management model based on snapshot function. For more information about snapshot function, see Snapshots.

To enable snapshots on a volume using the command-line interface, see **chvdisk** command.

Disable cloud snapshots for a volume

If you no longer need to create snapshots for a cloud volume, you can disable the function for the volume with the management GUI or the command-line interface.

When cloud snapshots are disabled on a volume, the volume is disconnected from the cloud account. Existing snapshot versions of the volume that are located in cloud storage are not deleted. A full snapshot of the volume is required if a cloud snapshot is needed later. You cannot disable cloud snapshots on a volume if a current snapshot operation is in progress. You must either wait for the snapshot operation to complete or cancel the snapshot. The cloud snapshots can be used only with volumes that have a capacity greater than zero bytes.

To disable cloud snapshots for a volume, use the **Volumes** > **Cloud Volumes** panel in the management GUI. Right -click the volume and select **Disable cloud snapshots**.

To disable snapshots on a volume using the command-line interface, see **chvdisk** command.

Create volume groups

For cloud volumes, you can create a group of volumes of which you can create snapshots and transfer to cloud storage. Volumes groups can be created to easily manage cloud snapshots for volumes. It is convenient to provide a single action to take a snapshot of a preset group of volumes.

Volume groups provide consistent snapshot operations for cloud volumes. You can view cloud volumes and display whether the volumes are part of a group in the management GUI, but you must configure the volume group in the command-line interface.

To create volume group, see the **mkvolumegroup** command, and to create or add cloud volumes to the volume group, see the **chvdisk** command.

Create cloud snapshots

After you create a cloud account, you can specify an individual volume or groups of volumes to be copied and transferred to cloud storage by creating cloud snapshots.

Cloud snapshots are point-in-time copies of volumes that are created and transferred to cloud storage that is managed by a cloud service provider. You can select individual or multiple volumes to create cloud snapshots of the volume data. You can use the command-line interface to configure volume groups to create snapshots for multiple volumes for consistent management.

When a snapshot is created for multiple volumes or volumes in a group, a separate snapshot operation is started for each individual volume.

In the management GUI, volumes that do not have restrictions are enabled for cloud snapshots when a snapshot is created for that volume. If you are using the command-line interface, you must enable snapshots on the volume before you can create a snapshot of the volume.

To create a cloud volume, use the **Volumes** > **Cloud Volumes** panel in the management GUI and click **Add Volumes**. This enables cloud snapshots on the added volumes before you can create cloud snapshot of the volumes. User can choose the following options:

- **Let the system decide the type of snapshot**
- **Complete a full snapshot**

To enable snapshots on a volume using the command-line interface, see the **chvdisk** command and to create cloud snapshots on a volume, see the **backupvolume** command. If you want to complete a full snapshot where all data on the cloud volume is copied to the cloud storage use **backupvolume -full volumen_id_or_name**.

Restore cloud snapshots

With transparent cloud tiering enabled, snapshot versions of volumes are stored on the cloud storage that is provided by a supported cloud service provider. You can restore data from the cloud back to the system from any snapshot version to recover from data loss on the system.

A volume cannot be used for restoring data from the cloud if the volume is a VMware vSphere Virtual Volume.

To restore a snapshot generation from cloud storage to the system, use the **Volumes** > **Cloud Volumes** panel in the management GUI. Right-click the volume to display all snapshot versions that are available to restore and select **Restore** and snapshot version that you want to restore to the system.

Select the target volume from the following options:

Restore a production volume group with the contents of one of the volume groups snapshot
The contents of the production system is overwritten with a previously saved version of the data.

Restore to the production volume
The snapshot version is restored to the production volume, which is the original volume from which the snapshots were created. After the restore operation completes, the snapshot version completely replaces the current data that exists on production volume. During the restore operation, the production volume goes offline until it completes. If you do not want to have the production volume offline for the restore, you can restore a cloud snapshot to a new volume. The production volume remains online and host operations are not disrupted.

Restore to a new volume
When the snapshot version is restored to a new volume, you can use the restored data independent of the original volume from which the snapshot was created. If the new volume exists on the system, then the restore operation uses the unique identifier (UID) of the new volume. If the new volume does not exist on the system, you need to choose whether to use the UID from the original volume or create a new UID. If you plan on using the new volume on the same system, use the UID that is associated with the snapshot version that is being restored.

You can use the command-line interface to either restore a snapshot directly to the production volume or to a new volume. To determine the available snapshot versions, see the **lsvolumebackupgeneration** command.

Restore to a production volume
To restore the most recent or old version of the cloud snapshot to the production volume use **restorevolume**. To restore an older version of the cloud snapshot, you must specify the **-deletelatergenerations** parameter to confirm the deletion of all later versions of the snapshot.

Restore to a new volume
You can also restore a snapshot version for a volume that might not exist on the local system. In this case, the unique identifier of the volume snapshot data is used. You can also restore the entire snapshot by replacing the volume UID with the UID of the volume that was originally used for the snapshot. Ensure that the UID is not currently used on any other volume in the system. To restore a snapshot version to the selected volume use the **restorevolume** command. You must specify **-restoreuid** to retain the UID for that volume, otherwise it restores data on new volume without replacing volume UID of original volume.

Delete a version of a cloud snapshot

A cloud volume can have multiple versions of snapshots on the cloud storage. If capacity on the cloud becomes limited, you can delete versions of the volume that are no longer necessary.

Before you delete a version of a snapshot, ensure that no snapshot or restore operations are in progress for the selected volume. In addition, ensure that the connection to the cloud service provider is online, so the deletion can be validated and completed on the cloud storage. Wait until these operations complete before you delete any snapshots that are associated with the volume. Deleting a backup generation (as opposed to canceling an in-progress backup) requires the connection to the cloud account to be available when the task is submitted in order to validate user input before you start an asynchronous delete operation.

Before you delete a version of a cloud snapshot, ensure that the connection to the cloud service provider is online. To delete a version of a cloud snapshot from the cloud storage, use the **Volumes** > **Cloud Volumes** panel in the management GUI. Right-click the cloud volumes and select **Manage Cloud Snapshots** and select the version of the snapshot.

To delete a single or all snapshot version for a volume using the command-line interface, see the **rmvolumebackupgeneration** command.

Import cloud snapshots

You can restore a cloud snapshot for a volume that was backed up from another system to the same container by importing a snapshot version of that volume to the local system.

When you import a cloud snapshot of a volume that is not on the current system, you can choose to use the unique identifier of the snapshot to replace the existing UID on the volume. In this case, the UID of the snapshot replaces the existing UID for the volume that is selected to receive the restored data. Ensure that the new UID is not currently used on any other volume in the system. The management GUI does not support UID replacement when a snapshot is imported from another system.

Note: The cloud account on the target system must use the same login credentials and the same container prefix as the cloud account that was used to back up the volume initially.

To import a cloud snapshot from another system,

1. Determine the import system ID for the system that the cloud snapshot is being imported from by running CLI **lscloudaccountimportcandidate** on the target system.
2. To change the mode of the cloud account, you use different commands based on the type of cloud account that is configured, use CLI **chcloudaccountswift**, **chcloudaccountawss3**, **chcloudaccountazure**. You must specify -mode import and import system ID.
3. Use the **restorevolume** command to restore the cloud snapshot for the volume that was backed up.

4. To confirm that the cloud account is in import mode, use CLI **lscloudaccount**. The target system can now access the data that is in the cloud snapshot of the source system.
5. To return the cloud account to normal mode, you use different commands based on the type of cloud account. You must specify -mode normal.

Transparent cloud tiering restrictions

Transparent cloud tiering on the system is defined by configuration limitations and rules.

The system restrictions specific to transparent cloud tiering for this release follow.

- Cloud volumes cannot be expanded or shrunk.
- A volume cannot be configured for a cloud snapshot if any of the following conditions are valid:
  - The volume is part of a volume group that is replicated, or configured for high availability.
  - The volume is a VMware vSphere Virtual Volumes volume, including FlashCopy owned volumes that are used internally for Virtual Volumes restoration functions.
  - The volume is a mirrored volume with copies in different storage pools.
  - The volume is being migrated between storage pools.
  - The volume is associated with any user-owned legacy FlashCopy mappings.

    Note: Cloud snapshots are supported with the system's new FlashCopy management model based on snapshot function. For more information about snapshot function, see related information section.
- A volume cannot be enabled for cloud snapshots if the cloud storage is set to import mode.
- A volume cannot be enabled for cloud snapshots if the maximum number of cloud volumes exists. If the system exceeds this limit, you can disable cloud snapshots on an existing cloud volume and delete its associated snapshots from the cloud storage to accommodate snapshots on new cloud volumes.
- A volume cannot be used for a restore operation if the following conditions are valid:
  - A Virtual Volume, including FlashCopy volumes that are used internally for Virtual Volumes restoration functions.
- A volume that is configured for backup or is being used for restoration cannot be moved between I/O groups.
- Only one operation (cloud snapshot, restore, or snapshot deletion) is allowed at a time on a cloud volume.
- Cloud volume traffic is allowed only through management interfaces (1 Gbps or 10 Gbps).

Snapshots

A volume group snapshot is a consistent point-in-time copy of a volume group or a set of volumes.

A volume group snapshot can be used to restore the parent volume group or a subset of the parent volumes of the snapshot, or to create clones of the snapshot.

Snapshots can be added to a volume group by using the management GUI interface or crash consistent snapshots can be scheduled for a volume group by adding a snapshot policy to the volume group. This policy can be safeguarded, in which case the snapshots added are safeguarded.

These operations can also be performed by using the command-line interface, which in addition can add snapshots to a list of volumes rather than a volume group.

Cloning snapshots

A new volume group can be created and prepopulated with the contents of a snapshot. The new volumes and volume group inherit properties from the parent volumes and volume group of the snapshot.

The new volume groups can be of type thin-clone or clone. A thin-clone volume maintains space efficiency whereas a clone volume becomes an independent volume.

Administrators can use the management GUI or the command-line interfaces to create a clone or thin-clone volume group. In addition, the command-line interface supports the ability to single volume clones or thin-clone outside of a volume group.

Thin-clone

A thin-clone is a host-accessible copy of a snapshot. By default, it is created in the same storage pool and I/O group as the source volume. A thin-clone can be used like any other volume with a few exceptions:

- A thin-clone is always dependent on the snapshot from which it was created.
- The capacity of a thin-clone can be expanded or shrunk. However, it cannot be shrunk to a size that is smaller than the snapshot from which it was created.

Clone

A clone is a host-accessible copy of a snapshot. Once all the data from the source snapshot is copied to the clone volume, it becomes independent of the snapshot from which it was created.

During the copy process the volume displays as type 'clone', this changes once the copy is complete and the volume becomes a standard volume. By default, it is created in the same storage pool and I/O group as the source volume. Alternative pools or I/O groups can be specified.

A clone volume can be used like any other volume during the cloning process except it cannot be shrunk to a size smaller than the snapshot from which it was created.

Convert thin-clone to clone

A user can convert a thin-clone volume to a clone volume. When doing this any uncopied data is copied from the snapshot volume to the volume being converted and once that copy has completed the converted volume becomes a stand-alone volume. The system supports converting entire thin-clone volume groups and individual thin-clone volumes.

Refresh from a snapshot

Snapshot refresh can be used to refresh the data on a set of thin-clone volumes from a volume group snapshot.

Any existing data on the set of volumes is discarded and the volumes are populated with the data from the specified snapshot.

Snapshot refresh operates at the volume group level or to a user specified subset of volumes within a volume group.

Restoring from snapshots

Snapshot restore provides the user with the ability to restore a production volume group with the contents of one of the volume groups snapshots. This feature is sometimes referred to as *restore in place*.

This type of restore operation overwrites the contents of the production system with a previously saved version of the data.

Snapshot restore allows restore either at the volume group level or subsets of volumes, which include thin-clone and clone volumes.

Safeguarded snapshots

Safeguarded snapshots are snapshots with an expiration time and restrictions regarding their premature deletion, that is the deletion of a snapshot before its expiration time. Only a Security administrator can delete a safeguarded snapshot prematurely.

Like standard snapshots, safeguarded snapshots can be created with either an internal scheduler or an external scheduler.

To safeguard snapshots using the internal scheduler, you can assign a snapshot policy and specify that the snapshot schedule should be safeguarded. This schedule then adds safeguarded snapshots to the volume group.

Note: You cannot configure safeguarded snapshots with the external scheduler if the volume group has a safeguarded backup policy assigned to it.

Safeguarded snapshot user roles

The following section explains all the responsibilities and roles that are given to the safeguarded copy function users.

System administrator

The system administrator can do the following actions on the safeguarded copy objects:

- Create a safeguarded volume group.
- Assign volumes to the safeguarded volume group.
- Delete a safeguarded source volume.
- Delete a safeguarded volume group.
- Associate a safeguarded snapshot policy to a volume group.

The system administrator cannot do the following actions on the safeguarded copy function objects:

- Delete safeguarded snapshot.
- Remove safeguarded snapshot policy association from a volume group.

Security administrator

In addition to the system administrator, the security administrator can do the following actions on the safeguarded snapshots:

- Delete safeguarded snapshots.
- Remove safeguarded snapshot policy association from a volume group.

Snapshot properties

Properties include basic properties of a snapshot and specific requirements that must be met.

The basic properties of a snapshot are:

- Snapshots are immutable, meaning that their content is protected from manipulation by the host.
- Snapshots consume space in the parent volume's storage pool. By default, they are in the same pool as the parent volume, but they can be added to child pool of the parent volume's pool. Or, if the parent volume is in a child pool itself, then they can be added to a child pool that has the same parent pool as the parent volumes pool.

To add snapshots to a volume group or set of volumes, the following requirements must be met:

- Snapshots cannot be added to the volumes that are target volumes of legacy FlashCopy® mappings.
- On the SAN Volume Controller platform, volume group snapshots can be added to groups of volumes that are mirrored. On other platforms, volume mirroring is not permitted except in the migration use-case.
- Volumes of zero size cannot have snapshots.

To restore a volume group from snapshots, the following requirements must be met:

- Snapshots can only be used to restore to the parent volume group or volumes.
- The composition of the volume group must be the same at the time of the restore as it was at the time the snapshot was taken except for volumes that are deleted since the snapshot was taken.
- If volumes are added to the volume group, you can restore but you must use the CLI to specify the subset of volumes to be restored.
- The volumes being restored must be the same virtual capacity as when the snapshot was added.
- You cannot restore volumes if they are transparent cloud tiering (TCT) enabled, part of a 3 site configuration, using asynchronous policy-based replication and host attach.

Managing snapshots

You can use the management GUI or the command-line interface (CLI) to manage snapshots, safeguarded snapshots, and the volume groups.

Adding snapshots to volume groups

To add snapshots to a volume group, use the **Volumes** > **Volume Groups** panel in management GUI. Select the volume group that needs to be associated with the snapshots and under the **Snapshots** tab, select **Take Snapshot**.

Creating volume groups from snapshots

To create a volume group from snapshots, use the **Volumes** > **Volume Groups** > **Create Volume Group** panel in management GUI. Select **Choose existing snapshot from a volume group**, select the required snapshot and specify the copy type that is required.

Restoring a volume group from snapshots

To restore a volume group from snapshots, use the **Volumes** > **Volume Groups** panel in management GUI. Select the volume group that needs to be associated with the snapshots. Then, on the **Snapshots** tab, select the **Restore** option on the overflow menu for the snapshot being used for the restore. You can restore either an entire volume group or a subset of volumes within that group. Select the applicable option and the subset of volumes if applicable and select **Restore**.

In the management GUI, to see when a snapshot was last used for a restore, go to the **Snapshots** tab on the **Volumes** > **Volume Groups** panel. To see when a volume group was last restored from a snapshot, go to the `Properties` option on the **Volumes** > **Volume Groups** panel.

Deleting snapshots

To delete snapshots, use the **Volumes** > **Volume Groups** panel in management GUI. Select the volume group that needs to be associated with the snapshots and under the **Snapshots** tab, select the snapshot to delete and click overflow menu to select **Delete**.

Convert a thin-clone to clone

To convert a thin clone to a clone, use the **Volumes** > **Volume Groups** panel in management GUI. Select the volume group that needs to be converted and then select **Convert thin-clone to clone** menu option.

To convert a thin-clone to a clone by using the command-line interface, see **`converttoclone`** command.

Refresh a thin-clone from snapshots

To refresh a thin-clone volume group from a snapshot, use the **Volumes** > **Volume Groups** panel in management GUI. Select the volume group that needs to be converted and then select **Refresh from snapshot** menu option.

Chapter 5. Hosts

A host system is a server that is connected to the storage system through supported connection protocols.

The storage system supports host connections through Fibre Channel, Serial Attached SCSI (SAS), NVM Express (NVMe) over Fibre Channel (FC-NVMe), NVM Express (NVMe) over RDMA, NVM Express (NVMe) over TCP, or an IP network. See the System Overview page on the relevant Hardware Guide to understand which supported adapters can be used for each host attach protocol.

A host object is a logical object that represents a collection of host ports. Depending on host protocol, the host ports will be one of the following formats:

- Worldwide port names (WWPNs) for Fibre Channel and SAS hosts
- NVMe qualified names (NQNs) for FC-NVMe, NVM Express (NVMe) over RDMA and NVM Express (NVMe) over TCP hosts
- iSCSI qualified names (IQNs) or extended unique identifiers (EUIs) for iSCSI hosts

A typical configuration has one host object for each server that is attached to the storage system. If a cluster of hosts accesses the same storage, you can add the hosts into a Host Cluster object to make configuration simpler. For details, see .

The system can detect WWPNs that have been logged into the system but are not yet configured in host objects, such that these can be allocated to host objects on creation (or added to existing host objects). These are visible in the management GUI in **Hosts** > **Hosts** > **Unconfigured hosts**. Some HBA device drivers do not leave ports logged in if no disks are detected on the network.

The Fabric Device Management Interface (FDMI) enables any storage endpoint to register itself to the Fibre Channel (FC) fabric and query the HBA and port details of the entire fabric. For details, see “Fabric Device Management Interface” on page 305.

You can manually enter WWPNs, NQNs and iSCSI names when creating hosts or adding ports to host objects.

Ownership Groups

Hosts can be assigned to an ownership group. Users in an ownership group are restricted to viewing and managing objects within their ownership group. For details, see **Ownership groups**.

Host mapping

Host mapping is the process of controlling which hosts or host clusters have access to specific volumes within the system.

Host mapping is similar in concept to logical unit number (LUN) mapping or masking. LUN mapping is the process of controlling which hosts have access to specific logical units (LUs) within the disk controllers. LUN mapping is typically done at the storage system level. Host mapping is done at the software level.

The act of mapping a volume to a host makes the volume accessible to the WWPNs or iSCSI names, such as iSCSI qualified names (IQNs) or extended-unique identifiers (EUIs) that are configured in the host object.

Note: If a volume is mapped to more than one host then they must all use the same protocol. When a host must use a different protocol, you can take a snapshot of the volume and map the snapshot to the host with a different protocol.

Volumes and host mappings

Mappings can either be shared or private. Shared mappings are volume mappings that are shared among all the hosts that are in a host cluster. When a host cluster is created, any common volume mappings

© Copyright IBM Corp. 2024

become shared among all the hosts within the host cluster. If a mapping is not common, it remains a private mapping for that host only. Private mappings are mappings that are associated with an individual host. To view whether mappings are either private or shared mappings, select **Hosts** > **Mappings**. The Mappings view displays all private mappings by default. From the View menu, select either **Shared Mappings** to view shared mappings for hosts or **All Host Mappings** to display all configured private and shared mappings. Without a multipathing device driver, most operating systems present each path to a volume as a separate storage device. The multipathing software manages the many paths that are available to the volume and presents a single storage device to the operating system. If there are multiple paths, the system requires that the multipathing software run on the host or host cluster.

When you map a volume to a host, you can optionally specify a SCSI ID for the volume. This ID controls the sequence in which the volumes are presented to the host. Check the host software requirements for SCSI IDs because some require a contiguous set. For example, if you present three volumes to the host, and those volumes have SCSI IDs of 0, 1, and 3, the volume that has an ID of 3 might not be found because no disk is mapped with an ID of 2. The system automatically assigns the lowest available SCSI ID if none is specified.

Figure 9 on page 108 and Figure 10 on page 108 show two volumes, and the mappings that exist between the host objects and these volumes.

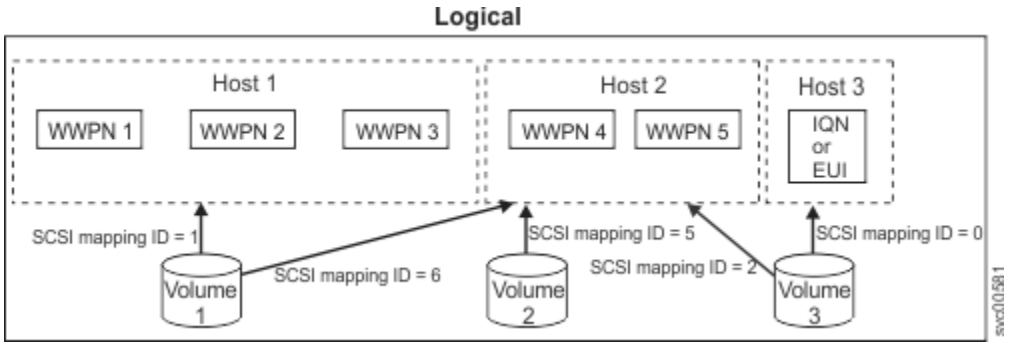


Figure 9. Hosts, WWPNs, IQNs or EUIs, and volumes

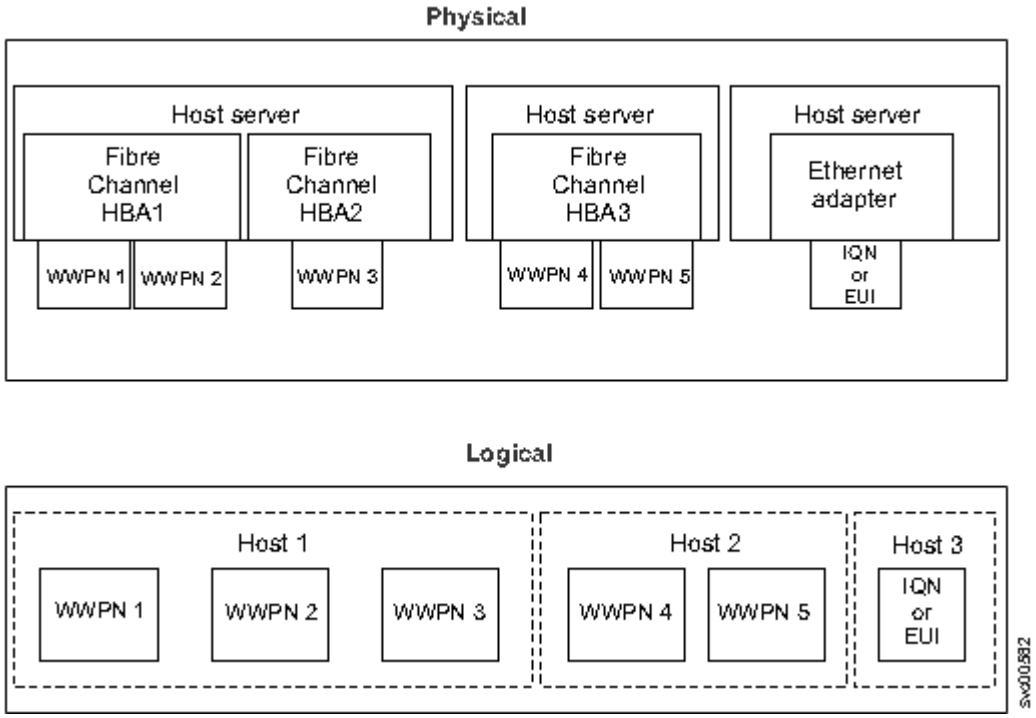


Figure 10. Hosts, WWPNs, IQNs or EUIs, volumes, and SCSI mappings

LUN masking is usually implemented in the device driver software on each host. The host has visibility of more LUNs than it is intended to use, and device driver software masks the LUNs that are not to be used by this host. After the masking is complete, only some disks are visible to the operating system. The system can support this type of configuration by mapping all volumes to every host object and by using operating system-specific LUN masking technology. However, the default, and recommended, system behavior is to map only those volumes that the host is required to access.

Host mappings can be assigned to an ownership group. An ownership group defines a subset of users and objects within the system.

Ownership can be defined explicitly or it can be inherited from the user, user group, or from other parent resources, depending on the type of resource. Host mappings inherit ownership from the parent resource. A user cannot change the ownership group of the resource, but can change the ownership group of the parent object. The following rule applies to volume-to-host mappings that are defined in ownership groups:

Volume-to-host mappings inherit the ownership group of the host or Host Cluster and volume in the mapping.

Determining the volume name from the device identifier on the host

You can use the command-line interface (CLI) to determine the volume name from the device identifier on the host.

About this task

Each volume that is exported by the system is assigned a unique device identifier. The device identifier uniquely identifies the volume and can be used to determine which volume corresponds to the volume that the host detects.

Complete the following steps to determine the volume name from the device identifier:

Procedure

1. Find the device identifier.
   For example, if you are using the native multipathing plug-in (NMP), the disk identifier is referred to as the virtual path (vpath) number. You can issue the following NMP command to find the vpath serial number:

   ```
   datapath query device
   ```

   For other multipathing drivers, refer to the documentation that is provided with your multipathing driver to determine the device identifier.
2. Find the host object that is defined to the system and corresponds with the host that you are working with.
   a) Find the worldwide port numbers (WWPNs) by looking at the device definitions that are stored by your operating system.
   For example, on AIX® the WWPNs are in the ODM and if you use Windows you must go into the HBA BIOS.
   b) Verify which host object is defined to the system for which these ports belong. The ports are stored as part of the detailed view so that you must list each host by issuing the following CLI command:

   ```
   lshost id | name
   ```

   Where *id* | *name* is the name or ID of the host.
   c) Check for matching WWPNs.
3. Enter the following command to list the host mappings:

   ```
   lshostvdiskmap hostname
   ```

Where *hostname* is the name of the host.

4. Find the volume UID that matches the device identifier and record the volume name or ID.

Volume reservations

The SCSI **Reserve** command and the SCSI **Persistent Reserve** command are specified by the SCSI standards. Servers can use these commands to prevent ports in other servers from accessing particular LUNs.

Reserving LUNs prevents accidental data corruption that is caused when a server overwrites data on another server. The **Reserve** and **Persistent Reserve** commands are often used by clustered-system software to control access to system volumes.

If a server is not shut down or removed from the server system in a controlled way, the standard and persistent reservations of the server are maintained. Reservations prevent other servers from accessing data that is no longer in use by the server that holds the reservation. In this situation, you might want to release the reservation to allow a new server to access the volume.

When possible, have the server that holds the reservation explicitly release the reservation to ensure that the server cache is flushed and that the server software is aware that access to the volume was lost. In circumstances where this action is not possible, you can use operating-system-specific tools to remove reservations. Consult the operating system documentation for details.

When you use the **rmvdiskhostmap** CLI command or the management GUI to remove host mappings, system nodes can remove the standard reservations and persistent reservations that the host has on the volume.

Host portsets

Portsets are groupings of logical addresses that are associated with the specific traffic types.

A host object can be configured to a single portset at a time. However, multiple host objects can be mapped to the same portset. A Host Cluster can consist of hosts that are configured to different portsets. However, it is not possible to view portset information by using a CLI command if the hosts in a cluster are configured to different portsets. You can configure a host to a specific portset by using the **mkhost** command. The **chhost** command can be used to change portset assignment to a host. The host loses its connectivity for all the applications until a new portset is assigned and contains an overlapping port with old portset.

Requirements for host portsets

In addition to portset 0, you can create more portsets for host traffic. These requirements are specific to host portsets:

- Portsets can have a maximum of 4 IP addresses per node.
- A single portset can either contain IPv4 or IPv6 or a mix of IPv4 and IPv6 addresses.
- For a host to log in to nodes on the system, the host must be mapped to a portset that contains at least one IP address from any of nodes on that system.

Requirements for Ethernet portsets

In general, Ethernet portsets have the following requirements:

Ethernet Portsets

- Each IP address in a portset must be configured on a separate Ethernet port.
- Same ports can share IP addresses across different portsets that allow the same IP address to be used for host, storage, and replication traffic. All shared IP addresses must use the same port and have the same VLAN, gateway, and prefix. When IP addresses are shared among multiple portsets, the system creates a logical copy of the IP address and its attributes, rather than a new IP address.

- Portsets that are owned by different ownership groups can share an IP address.
- The **Port Type** field value defines whether the portset is Ethernet.
- Each port can be configured with only one unique routable IP address (gateway specified). The routable IP can be shared among multiple portsets.
- Portset 0 is a default portset that is automatically configured when the system is updated or created. Portset 0 is a host portset by default and cannot be deleted even if it is empty. Portset 0 serves as the default portset for any IP addresses and host objects that are configured without a portset specified. Portset 0 allows administrators to continue with an original configuration that does not require multi-tenancy. After an update, all configured Ethernet-based host objects are automatically mapped to portset 0.

Requirements for Fibre Channel portsets

The Fibre Channel portsets have the following requirements:

- Many host objects can be mapped to a single Fibre Channel portset.
- If you do not map a host object to a portset, the host object is automatically mapped to the default portset to ensure operational continuity. The default portset for Fibre Channel portsets is **portset64**. You can create extra portsets based on the configurations and workload.
- Host objects can access volumes only through the Fibre Channel ports that are associated with the respective portset.
- The **Port Type** field value defines whether the portset is Fibre Channel. in the **port_type** parameter of the **mkportset** command.
- The Fibre Channel I/O ports are added to the portset by using Fibre Channel I/O port ID. It is applicable for all the nodes in the system.
- A Fibre Channel port can be associated with more than one portset.

A portset restricts a host to access only a specific set of IP addresses of a node. A host can access only those IP addresses that are configured on a portset and is mapped to that host. A portset object is a system-wide object and might contains IP addresses from every I/O group. To access multiple nodes in a system, a portset must be configured with the IP address of the nodes that the host wants to access.

Host connectivity

The system reports various information to help the user understand host connectivity and troubleshoot any issues.

For Fibre Channel, FC-NVMe, NVMe over RDMA, and NVMe over TCP-attached hosts, the system reports the node login count, which is the number of nodes that can detect each WWPN or NQN.

To view connectivity for Fiber Channel-attached hosts in the management GUI, select **Settings** > **Network** > **Fibre Channel Connectivity** or use the **lsfabric** command. To view connectivity issues for FC-NVMe, NVMe over RDMA, and NVMe over TCP-attached hosts in the management GUI, select **Settings** > **Network** > **NVMe Connectivity** or use the **lsnvmefabric** command.

For Ethernet-attached hosts, the number of logged-in nodes refers to sessions that are created between hosts and nodes. The number of logged-in nodes can be greater than the current number of nodes on the system.

- You can display the login information for each host in the management GUI to determine which ports for the logged-in nodes are connected to the Ethernet-attached hosts. To display this information, select **Hosts** > **Hosts** and right-click the host and select IP Login Information.
- To view login information in the command-line interface, use the **lshostiplogin** to list login session details for a specified host that is logged in to a configured Ethernet port on the system.
- You can also set a status policy and site status to determine how host statuses are displayed. For example, if you select Complete, the system displays an Online status only when all host ports are logged in to all nodes and hosts can access all mapped volumes. If a single host or node loses

connectivity, the status for the host displays as Degraded. If you select Redundant for the Status Policy, an Online status is displayed when a sufficient number host ports to node logins exist to retain access to all its mapped volumes, even when a host port or single node is removed or lost.

- If hosts are part of a multiple site configuration, each host needs to be assigned to a specific site and you can specify how host statuses are displayed for these hosts. In the management GUI, you can specify All for Site Status to indicate that host connectivity to the nodes on all sites determine the status of the host. If you select Local for Site Status, the host status relates only to connectivity to nodes on the local site. Host status for other nodes mapped to different sites is not included.

Host clusters

A *host cluster* is a group of logical host objects that can be managed together. For example, you can create a volume mapping that is shared by every host in the host cluster.

The system uses internal protocols to manage access to the volumes and ensure consistency of the data. Host objects that represent hosts can be grouped in a host cluster and share access to volumes. A host can only be in one host cluster at a time. New volumes can also be mapped to a host cluster, which simultaneously maps that volume to all hosts that are defined in the host cluster.

Each host cluster is identified by a unique name and ID, the names of the individual host objects within the cluster, and the status of the cluster. A host cluster can contain up to 128 hosts. However, a host can be a member of only one host cluster.

A host cluster can have one of the following statuses:

Online
All hosts in the host cluster are online.

Host degraded
All hosts in the host cluster are either online or degraded.

Host cluster degraded
At least one host is offline and at least one host is either online or degraded.

Offline
All hosts in the host cluster are offline (or the host cluster does not contain any hosts).

Host clusters can be assigned to an ownership group. An ownership group defines a subset of users and objects within the system.

Ownership can be defined explicitly or it can be inherited from the user, user group, or from other parent resources, depending on the type of resource. Host clusters can be owned if they are assigned an ownership group explicitly or by inheritance from the user creating them. New or existing hosts that are defined in a Host Cluster inherit the ownership group that is assigned to the Host Cluster.

Volume mapping

By default, hosts within a host cluster inherit all shared volume mappings from that host cluster, as if those volumes were mapped to each host in the host cluster individually. Hosts in a host cluster can also have their own private volume mappings that are not shared with other hosts in the host cluster.

With *shared mapping*, volumes are mapped on a host cluster basis. The volumes are shared by all of the hosts in the host cluster, if there are no Small Computer System Interface (SCSI) LUN conflicts among the hosts. Volumes that contain data that is needed by other hosts are examples of a shared mapping.

If a SCSI LUN conflict occurs, a shared mapping is not created. SCSI LUN conflicts are caused when multiple volumes are mapped to the same SCSI LUN ID. SCSI LUN conflicts can also be caused when a single volume is mapped to multiple SCSI LUN IDs.

With *private mapping*, individual volumes are directly mapped to individual hosts. These volumes are not shared with any other hosts in the host cluster. A host can maintain the private mapping of some volumes and share other volumes with hosts in the host cluster.

For example, Figure 11 on page 113 shows two hosts that are members of a host cluster. The volume that contains data is a shared mapping with Host A and Host B. However, Host A and Host B also have a private mapping to their respective boot volumes (LUN A and LUN B).

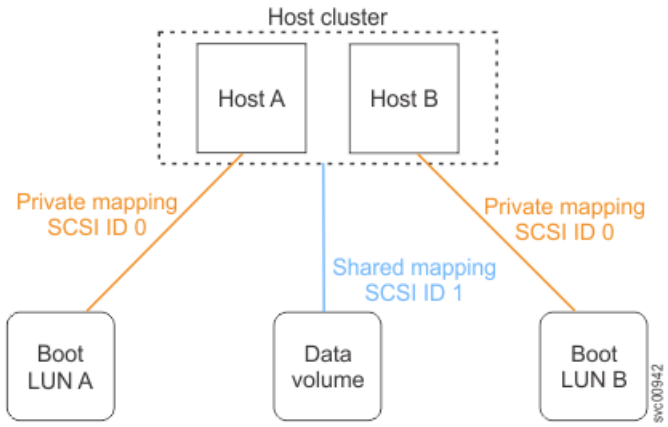


Figure 11. Host cluster that contains private and shared volumes

Managing host clusters

A *host cluster* is a group of logical host objects that can be managed together. For example, you can create a volume mapping that is shared by every host in the host cluster.

The system uses internal protocols to manage access to the volumes and ensure consistency of the data. Host objects that represent hosts can be grouped in a host cluster and share access to volumes. A host can only be in one host cluster at a time. New volumes can also be mapped to a host cluster, which simultaneously maps that volume to all hosts that are defined in the host cluster.

When you create a host cluster, you can specify the hosts to include in the cluster. You can also create an empty host cluster and add hosts later. If you select hosts to add to the host cluster, all of the common private volume mappings for the hosts become shared mappings for the host cluster. However, you can also exclude specific volumes from being shared. The private volume mappings are not shared with any other hosts in the host cluster. A host can be a member of only one host cluster at a time.

With *shared mapping*, volumes are mapped on a host cluster basis. The volumes are shared by all of the hosts in the host cluster, if there are no Small Computer System Interface (SCSI) LUN conflicts among the hosts. Volumes that contain data that is needed by other hosts are examples of a shared mapping.

If any mappings contain SCSI LUN conflicts, the operation fails. The management GUI displays all mappings with SCSI LUN ID conflicts. SCSI LUN conflicts can occur if multiple volumes are mapped with the same SCSI LUN ID or if same volume is mapped to multiple SCSI LUN IDs.

Note: When adding hosts to a host cluster, if you selected hosts with volumes that have SCSI ID conflicts, the system does not add these mappings to the host cluster. A SCSI LUN ID conflict occurs when multiple hosts are mapped to the same volume but with different SCSI IDs. In this case, a shared mapping is not created because the system does not allow a volume to be mapped more than once to the same host. The **Summary** page lists all volumes that contain conflicts and the system retains these mappings as private mappings to the original hosts.

If throttles are configured on any of the selected hosts, the throttles must be removed to be included in the host cluster. Throttles can be applied only to the host cluster and not individual hosts within the cluster. If you choose not to remove the throttles from the host, the host is excluded from the host cluster.

When a host is removed from a host cluster, you can convert the shared mappings of the host cluster to private mappings or remove all shared mappings from the host. If shared mappings are converted to private mappings, volume access is not removed since the volume remains mapped to the host and its SCSI ID remains unchanged. If you choose to remove the shared mappings when you remove a host, only the private mappings for that host are retained when it is removed from the host cluster. When you

remove the shared mapping, hosts cannot access those volumes that were mapped as shared mappings to the host cluster.

You can use several tools to manage host clusters, including the management GUI, and the command-line interface (CLI). To work with host clusters in the management GUI, select **Hosts** > **Host Clusters**.

Managing hosts

You can use several tools to manage hosts, including the management GUI, the command-line interface (CLI), and specialized utilities for working with host bus adapters. To work with hosts in the management GUI, select **Hosts**.

Configuring host objects

You can configure Fibre Channel, NVMe over Fibre Channel, NVMe over RDMA, NVMe over TCP, and Ethernet iSCSI-attached hosts to a system.

If you are configuring a host object on a Fibre Channel fabric-attached host, ensure that you complete all zone and switch configuration. If you are configuring a directly attached Fibre Channel host, there must be connections to both nodes in each I/O group that has volumes that are mapped to that host. Also, test the configuration to ensure that zoning was created correctly.

If you are configuring a host object on the system that uses iSCSI connections, ensure that you complete the necessary host-system configurations and configure the system for iSCSI connections.

If you are configuring a host, such as hosts that are running the VMware operating system, define each host in the system as a separate host object. Do not combine multiple worldwide port names (WWPNs) from multiple hosts as one host object.

If you are configuring a host to a portset, portset object is created first, and then the Fibre Channel port or IP address object, and then the host object is created and configured with a portset. A portset restricts a host to access only a specific set of logical addresses of a node. A host object can be configured to a single portset at a time. However, multiple host objects can be mapped to the same portset. There are two default portsets based on the portset type. *Portset 0* is used for Ethernet portsets and *portset 64* is the default portset for Fibre Channel portsets. If a host is not assigned to a portset, the host is added to the correct default portset based on the connection type of the host. The **`lsportset`** command shows the number of hosts count associated with the available portsets. For more information on portsets, refer to “Portsets” on page 329.

To create host objects in the management GUI, select **Hosts** > **Hosts** > **Add Host**.

You can edit host details from the **Hosts** > **Hosts** panel on the management GUI. To edit host details, open the **Actions** prompt on the host to edit or right-click on the host to edit and select **Edit and rename host**.

To create host objects by using the CLI, use the **`mkhost`** command. To add ports to an existing host object, use the **`addhostport`** command. To delete an existing host object, use the **`rmhost`** command.

Managing WWPN changes of replaced host HBAs

You can use the command-line interface (CLI) to record a change to a defined host object.

Before you begin

Sometimes it is necessary to replace the host-bus adapter (HBA) that connects the host to the SAN. You must inform the system of the new worldwide port names (WWPNs) that the replacement HBA contains.

Ensure that your switch is zoned correctly.

Procedure

To inform the system of a change to a defined host object, complete the following steps.

1. Enter the following CLI command to list the candidate HBA ports.

```
lsfcportcandidate
```

or

```
lssasportcandidate
```

You see a list of the HBA ports that are available for addition to host objects. One or more of these HBA ports corresponds with one or more WWPNs that belong to the new HBA port.

2. Locate the host object that corresponds with the host in which you replaced the HBA. The following CLI command lists all the defined host objects:

```
lshost
```

3. Enter the following CLI command to list the WWPNs that are currently assigned to the host object.

```
lshost hostobjectname
```

Where *hostobjectname* is the name of the host object.

4. Enter the following CLI command to add the new ports to the existing host object.

```
addhostport -fcwwpn one or more existing port names
 separated by : hostobjectname/ID
```

Where *one or more existing port names separated by :* is the WWPNs that are currently assigned to the host object and *hostobjectname/ID* is the name or ID of the host object.

5. Enter the following CLI command to remove the old ports from the host object.

```
rmhostport -fcwwpn one or more existing port names
 separated by : hostobjectname/ID
```

Where one or more existing WWPNs separated by a colon (`:`) are the WWPNs that are currently assigned to the host object and *hostobjectname/ID* is the name or ID of the host object.

Note: If the following conditions are met when volume protection is enabled for the system, the deletion of the specified host port fails.

- It is the last active port on the host.
- It is mapped to any volume that received I/O within the specified volume protection interval.

If volume protection is enabled, and the host port being deleted is the last port for a host, which is mapped to any volume that received I/O within the defined volume protection time period, then the command fails. If multiple hosts are mapped to the same active volume, the system deletes the port if the host is offline.

Results

Any mappings that exist between the host object and the volumes are automatically applied to the new WWPNs. Therefore, the host sees the volumes as the same SCSI LUNs as before.

What to do next

See the *IBM Multipath Subsystem Device Driver User's Guide* or the documentation that is provided with your multipathing driver for additional information about dynamic reconfiguration.

Managing throttles for hosts

You can create, modify, and remove throttles for hosts by using the management GUI or the command-line interface.

Throttles are a mechanism to control the amount of resources that are used when the system is processing I/Os on a specific host or host cluster. If a throttle is defined, the system either processes the I/O, or delays the processing of the I/O to free resources for more critical I/O.

Using the management GUI

To create a new throttle for a specific host, complete these steps:

1. In the management GUI, select **Hosts** > **Hosts**.
2. Right-click a host or multiple hosts and select **Edit Throttle...**.
3. On the **Edit Throttle** page, enter values for the following throttles:

   Bandwidth Limit
   Enter the maximum amount of bandwidth the host can process before the system delays processing for this host. Click **Create**.

   IOPS Limit
   Enter the maximum I/O operations per second the host can process before the system delays processing for this host. Click **Create**.

To change an existing throttle, complete these steps:

1. In the management GUI, select **Hosts** > **Hosts**.
2. Right-click a host cluster and select **Edit Throttle...**.
3. On the **Edit Throttle** page, enter values for the following throttles:

   Bandwidth Limit
   Enter the maximum amount of bandwidth the host can process before the system delays processing for this host. Click **Save**.

   IOPS Limit
   Enter the maximum I/O operations per second the host can process before the system delays processing for this host. Click **Save**.

To remove throttles for a specific host, complete these steps:

1. In the management GUI, select **Hosts** > **Hosts**.
2. Right-click a host or multiple hosts and select **Edit Throttle...**.
3. On the **Edit Throttle** page, click **Remove** by the throttle that you want to delete.

Using the command-line interface

To create a new throttle for bandwidth and IOPS for a host with the command-line interface, enter the following command:

```
mkthrottle -type host -bandwidth bandwidth_limit_in_mb -iops iops_limit -host
host_id_or_host_name
```

where *bandwidth_limit_in_mb* is the maximum amount of bandwidth that the system can process for the specified host before the system delays processing of I/O for that host. The *iops_limit* is the maximum I/O operations per second that the system can process before the system delays processing for this host. The *host_id_or_host_name* is the ID or name of the host that the throttle applies to.

To change an existing throttle for the host, enter the following command:

```
chthrottle -type host -bandwidth bandwidth_limit_in_mb -iops iops_limit throttle_name_or_id
```

where *bandwidth_limit_in_mb* is the maximum amount of bandwidth that the system can process for the specified host before the system delays processing of I/O for that host. The *iops_limit* is the maximum I/O operations per second that the system can process before the system delays processing for this host. The *throttle_id_or_name* is the ID or name of the throttle that is being changed.

To remove a throttle for a host with the command-line interface, enter the following command:

```
rmthrottle throttle_name_or_id
```

where *throttle_name_or_id* is the ID or name of the throttle that is being removed.

Chapter 6. Host attachment

This topic describes configuration activities to enable hosts to access volumes and use host-based solutions to manage the storage system.

You can attach hosts to the system by using the following protocols:

- Fibre Channel
- NVMe over Fibre Channel (FC-NVMe)
- NVMe over RDMA
- NVMe over TCP
- Serial-attached SCSI (SAS)
- Ethernet iSCSI
- Small Computer System Interface-Fibre Channel (SCSI-FCP)

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Hosts that use the Fibre Channel connections are attached to the system either directly or through a switched Fibre Channel fabric. Each port on a node is identified by a worldwide port name (WWPN).

Host connections through NVMe over RDMA use specialized network switches and require fewer resources than FC-NVMe. RDMA allows higher throughput and better performance with low latency. In addition, RDMA requires less expertise at the storage networking level than the Fibre Channel implementation, potentially reducing overall costs.

The advantage of using NVMe over TCP data transfer is that unlike RDMA data transfer, NVMe over TCP uses the existing Ethernet adapters on the host and network infrastructure.

Refer to the following pages for information about host configuration.

Using VMware

There are a number of integrations available between IBM Storage Virtualize and VMware that whilst not essential, will improve the overall functionality and provide a better user experience.

IBM Storage Virtualize plugin for vSphere

The IBM Storage Virtualize plugin for vSphere is based on VMware's Remote plugin architecture, and facilitates provisioning and management of IBM Storage from within the vCenter UI. The plugin and associated workflows are integrated into the vCenter user interface and allows a vSphere administrator to remotely perform administrative or maintenance actions across multiple storage systems. For more information on vSphere plugin, see https://www.ibm.com/docs/en/svpfv.

IBM Storage Virtualize Family - Storage Replication Adapter (SRA)

The IBM® Storage Virtualize Family Storage Replication Adapter (SRA) is a software add-on that integrates with the VMware Site Recovery Manager (SRM) solution to run Disaster Recovery failover scenarios with supported storage systems. For more information on SRA, see https://www.ibm.com/docs/en/spectrumvirtual-sra.

IBM VASA Provider for vSphere Virtual Volumes

The vSphere Virtual Volumes (vVol) framework is built upon vSphere APIs for Storage Awareness (VASA). These APIs provides tight integration, orchestration and automation between vSphere and the underlying storage system through a VASA provider which is embedded within supported storage platforms.

For more information on vVol, see the following page.

© Copyright IBM Corp. 2024

VMware Virtual Volumes (vVols)

The system provides native support for VMware vSphere APIs for Storage Awareness (VASA) through a VASA Provider (also known as a Storage Provider), which sends and receives information about storage that is used by VMware vSphere to the vCenter Server.

Through VASA, the system also supports VMware Virtual Volumes (also known as *vVols*), which allows VMware vCenter to automate the creation, deletion and mapping of volumes.

Before vSphere can be used to provision vVols, the storage system administrator must first enable vVols on the storage system. It is recommended that vVols are enabled by using the management GUI as this simplifies the setup process.

When vVols are enabled, the following objects are configured on the system:

- An ownership group
- A user account and user group
- A metadata volume
- A child pool with an associated provisioning policy

An ownership group is created that separates the resources that belong to VMware from the other resources in the system, and ensures that only these resources can be managed by VMware vCenter.

The storage system administrator delegates ownership of Virtual Volumes to VMware vCenter and the VASA Provider by creating a user account with the VASA Provider role. Although the storage system administrator can complete certain actions on volumes and pools that are owned by the VASA Provider, the VMware environment retains management responsibility for Virtual Volumes.

A metadata volume is created to store metadata for Virtual Volumes and vSphere storage policies. The system administrator selects a storage pool to provide capacity for the metadata volume. With each new volume created by the VASA Provider, VMware vCenter defines a small amount of metadata that is stored on the metadata volume. The metadata volume is exclusively used by the VASA Provider and cannot be deleted while vVols exist or mapped to hosts.

The storage system administrator decides what storage to allocate for Virtual Volumes by creating a child pool and assigning it to the VASA ownership group. Each vVols child pool is presented as a storage container in vSphere, from which a vVols datastore can be created. The storage system administrator must also associate a provisioning policy with the child pool, to specify how data Virtual Volumes, known as vmdk Virtual Volumes, are provisioned.

Storage Virtualize supports asynchronous replication for Virtual Volumes between two storage systems in different locations. The storage administrator preconfigures replication groups for vSphere by creating volume groups for vVols and assigning a replication policy to each group. The VMware administrator can use storage policies in vCenter to define the replication requirements for virtual machines, such as the second site where the VMs should be replicated and a recovery point objective. Virtual Volumes for multiple virtual machines can be grouped to manage them as a single unit for failover and disaster recovery.

Planning vVols

VMware Virtual Volumes (vVols) are not supported on all Storage Virtualize systems. Therefore, plan how to provision your vVols.

vVols is not supported on some products due to hardware limitation. If applicable, the limitations are documented in a Software feature limits page in the relevant Hardware Guide.

The system must be configured with the standard topology. Using stretched topology systems is not supported.

The vCenter Server and ESXi host management networks require TCP port 8440 to be accessible on the storage system's management IP address.

The storage system, vCenter Server and ESXi hosts must all be configured to use a Network Time Protocol (NTP) service. This ensures that the time settings between the storage system and the VMware environment remain consistent.

At least one standard storage pool is required. Data reduction pools are not supported for storing vVols or the vVols metadata volume.

Do not run vCenter Servers on vVols. This avoids the possibility of complicated failure scenarios with cyclic dependencies in the VMware environment.

Planning vVols replication

To implement vVols replication between two systems, each system must support vVols and asynchronous policy-based replication.

The following restrictions apply when you use vVols replication (For more information on asynchronous policy-based replication, see “Disaster recovery” on page 206.:

- A limit of one I/O group can be configured in each system.
- A limit of one partnership can be configured for vVols replication. Additional partnerships can be configured in each system but only a single partnership can be used for vVols replication.
- The size of a virtual volume cannot be changed while it is part of a vSphere replication group.

Each storage system is a separate fault domain for vVols replication. The storage systems must be sufficiently far apart such that anticipated disruptions affect only one location.

Use vCenter Server and ESXi versions that support vVols replication. vCenter Server and ESXi can discover replication groups, but do not manage their lifecycle. When you use vVols replication, the failover and disaster recovery operations are managed through vCenter Server integrations such as VMware Site Recovery Manager (SRM) and PowerCLI. A vCenter Server is required in each location.

Storage Virtualize supports the use of preconfigured replication groups only. The automatic creation of replication groups by using vSphere is not supported.

When you use replication with virtual volumes, you can apply a replication storage policy only to a configuration virtual volume and a data virtual volume. The swap virtual volume and any memory vVols are excluded from replication. Virtual machine snapshots are also excluded from replication.

When planning the VMware environment, if you intend to create virtual machine snapshots, you must check that the vCenter Server integrations you intend to use to manage vVol replication support replication without snapshots.

Implementing vVols

You can use the management GUI to enable Virtual Volumes.

Go to **Settings** > **System** > **VMware Virtual Volumes (vVols)** to enable vVols.

Before enabling vVols, ensure that the system certificate contains one of the following in the **Subject Alternative Name** field:

- IP address
- Fully qualified domain name (FQDN)
- Hostname

To specify a hostname or FQDN, a DNS server must be configured.

The properties that are defined in the **Subject Alternative Name** field are used by the vCenter Server to establish a secure connection to the storage system. The IP address, hostname, or FQDN as specified in the certificate must be used to register the storage provider in the vCenter Server.

For more information, see “System Certificates” on page 389.

Any hosts that are intended to access VMware vSphere Virtual Volume (vVol) datastores must be defined with the vVol or adminlun host type. By default, newly created hosts are configured with a generic host type, allowing only access to traditional storage volumes (for example, VMFS Datastores or Raw Device Mappings). However, hosts configured with vVol or adminlun host type allow access to both traditional volumes and vVols. Configure a host object for each ESXi host that accesses storage from the system. Protocol Endpoints (PEs) are automatically provisioned to hosts when the host type is configured as vVol or adminlun. It is a good practice to change the host type, and verify that all hosts are correctly detected the Protocol Endpoints (PEs) before registering the Storage Provider in vSphere. After the host type is set, rescan the ESXi Storage adapters and verify that SCSI devices are visible with SCSI IDs 768 & 769.

The storage system, vCenter Server, and ESXi hosts must all be configured to use a Network Time Protocol (NTP) service. This configuration helps ensure that the time settings between the storage system and the VMware environment are synchronized.

At least one standard storage pool is required. Data reduction pools are not supported for storing vVols or the vVols metadata volume.

When enabling vVols, the user account that is created for the vCenter Server is initially configured with a username and password. These credentials are only used when first registering the Storage Provider in vCenter. After successful registration, the password is removed and the user account is automatically reconfigured with a certificate for authentication.

After enabling vVols on the storage system, complete the remaining steps that are required in VMware vCenter to finish the configuration. If an internally or externally signed certificate is used, ensure that root or external CA certificates are imported into the vCenter trust store and pushed to ESXi hosts before registering the storage provider.

If registration of the storage provider encounters an error, see the troubleshooting topic Problem: Unable to add Storage Virtualize as a storage provider on VMware vCenter.

Each vVols child pool is presented as a storage container in vSphere, on which a vVols datastore can be created. The initial setup creates a single child pool for vVols and more pools can be created later if required.

The management GUI indicates when vVols are enabled on the storage system.

Implementing vVols Replication

After you enable vVols, you can implement vVols replication.

Implement vVols replication by completing the following steps:

1. Enable vVols on each system and register the storage system in each location in the corresponding vCenter Server. See Implementing vVols.
2. Create a 2-site partnership by using policy-based replication between the two storage systems.
3. Enable vVols replication on the system that will be used for creating production virtual machines.

Creating a partnership for policy-based replication

Use the management GUI to create a 2-site partnership between systems by using policy-based replication.

- Go to **Settings** > **System** > **Partnerships** to get started.
- If a partnership that uses policy-based replication exists between the two systems, no additional partnership configuration is needed to use vVols replication.
- The GUI indicates when the partnership is configured and ready for use with policy-based replication.

Enabling vVols replication

Use the management GUI to enable vVols replication.

- Go to **Settings** > **System** > **VMware Virtual Volumes (vVols)** to begin the setup.

- Configuring vVols replication links the VASA ownership groups and child pools between the two systems for replicated Virtual Volumes to be manageable by using the vCenter Server in the second site. The initial setup also creates a volume group with a replication policy. More volume groups and replication policies can be created later on either of the systems.
- After enabling vVols replication on the storage system, rescan the Storage Provider in vSphere to discover the new replication capabilities. If the storage system is already registered with vSphere on an earlier version, an extra step may be required to discover the new replication capabilities. See Updating from version 8.6.1 or earlier.
- You can now create VM storage policies that define the replication requirements for virtual machines.
- When a new virtual machine is created by using the policy, each of the compatible preconfigured volume groups on the storage system is available as a replication group choice in vSphere.
- Storage Virtualize supports the use of preconfigured replication groups only. The automatic creation of replication groups by using vSphere is not supported.

Updating from version 8.6.1 or earlier

If a system with vVols configured on version 8.6.1 or earlier is updated to version 8.6.2 or later, an extra step is required to discover new capabilities of the storage system in the vSphere environment.

After the storage system software is updated to version 8.6.2 or later, perform the following steps:

1. Using the vSphere Client, unregister the storage provider on the vCenter Server:
   a. Go to **VCSA** > **Configure** > **Storage Providers**.
   b. Select the storage provider and select **Remove**.
2. Using the storage system command-line interface, reset the user account that is used by the vCenter server:
   a. Use the following command to remove the stored certificate and specify a temporary password for the VASA user account: `chuser -nocertuid -password <new_password> <vCenter_user>`
3. Using the vSphere Client, reregister the storage provider on the vCenter Server:
   a. Go to **VCSA** > **Configure** > **Storage Providers**.
   b. Select **Add** and readd the storage provider.

Managing vVols

Create vVols child pools, resize the storage capacity of a vVols child pool, register the storage provider in vCenter, and perform other tasks to manage vVols.

Creating a vVols child pool

More child pools can be created for vVols by using the management GUI or the command-line interface. Each vVols child pool is presented as a storage container in vSphere, on which a vVols data store can be created.

To use the GUI to create a child pool for vVols, specify the VASA ownership group and provisioning policy when you create the child pool.

To use the command-line interface, use the following command:

```
mkmdiskgrp -name <vVol_child_pool_name> -owner vvol_child_pool -ownershipgroup VASA
-parentmdiskgrp <parent_pool> -provisioningpolicy <policy> -size <size> -unit <unit>
```

The parent pool must be a standard storage pool. Data reduction pools are not supported for vVols.

If you use the new child pool with vVols replication, add a pool link between the new pools on each system. Use the **Pools** window on the management GUI to add pool links for replication from either system.

After creating a new vVol child pool, the vCenter server administrator must rescan the storage provider in vSphere to refresh the available storage containers before creating a new vVols data store.

Resizing a vVols child pool

The storage administrator can increase or decrease the amount of storage capacity that is allocated to a vVols child pool by using the management GUI or the command-line interface.

After changing the capacity of a child pool, the vCenter server administrator must refresh the datastore capacity information in vSphere to detect the new capacity.

Considerations before updating the system software

The system operates with limited functions during the software update process. During this time, no configuration changes can be made to the system and vVols functions are reduced.

In vSphere, consider the following:

- Any virtual machines that are powered on continue to function.
- The storage provider might appear as offline.
- The vVols datastores might appear as inaccessible.
- Any powered off virtual machines on vVols datastores appear as inaccessible.
- Any attempt to perform VM-level management operations fail. These operations include, for example: power on and off, snapshots, vMotion, storage vMotion, and cloning.

Reregistering the storage provider in vCenter

If the storage provider needs to be removed and reregistered in vCenter, use the following command to remove the stored certificate and specify a temporary password for the VASA user account:

```
chuser -nocertuid -password <new_password> <VASA user ID>
```

These credentials are only used when registering the storage provider in vCenter. After successful registration, the password is removed and the user account is automatically reconfigured with a certificate for authentication.

Changing the system management IP address or DNS name

If the IP address, hostname or fully-qualified domain name (FQDN) specified in the subject alternative name (SAN) field of the system certificate is changed, a new system certificate must be generated by using either the internal or external certificate authority. After generating the certificate, the storage provider must be reregistered in vCenter (see Re-registering the storage provider in vCenter).

User accounts

The user account created when enabling vVols is initially configured with a username and password. These credentials are only used when first registering the storage provider in vCenter. After successful registration, the password is removed and the user account is instead automatically reconfigured with a certificate.

If any manual vVols-related service or recovery tasks are required on the storage system, the storage administrator must log in using a user account with the VASA provider role. This can either be a new user in the same user group, or the same user account that the storage provider was registered with.

To create a new user account, use the following command:

```
mkuser –usergrp VASAProvider –name <new user> -password <password>
```

To configure a password for the existing user account, use the following command:

```
chuser -password <password> <vvol_user_name>
```

Managing vVols Replication

Create and remove a replication group, modify the existing storage policy or assign a different policy in vSphere, and manage failover operations.

Creating a Replication Group

Storage Virtualize supports the use of preconfigured replication groups only. The automatic creation of replication groups using vSphere is not supported.

To create a replication group for use with vSphere Virtual Volumes, you must first create a volume group for vVols and assign a replication policy on the storage system. To create a new volume group for vVols replication using the management GUI, go to the **Volumes** > **Volume groups** page.

To use the command-line interace, use the following command:

```
mkvolumegroup -owner vvol -ownershipgroup VASA -replicationpolicy <policy> -name <name>
```

After creating volume groups for vVols, rescan the Storage Provider in vSphere to refresh the available replication groups.

Removing a Replication Group

When a replication group is no longer required in vSphere, the corresponding volume group can be removed on the storage system. The volume group must be empty before it can be removed. When the volume group is removed on the production system, the volume group is automatically removed on the recovery system.

After removing volume groups for vVols, rescan the Storage Provider in vSphere to refresh the available replication groups.

Assigning or Changing the Replication Group for a Virtual Machine using vSphere

If the storage requirements for a virtual machine change, you can modify the existing storage policy or assign a different policy in vSphere. If the new storage policy will use a different replication group, the existing storage policy must first be removed from the virtual machine, or replaced with a policy that does not include replication, before the new policy can be assigned.

When editing the settings for a virtual machine with a storage policy that includes replication, the replication group selection is mandatory. If the replication group drop-down is initially not displayed in the vSphere Client GUI, you must wait for the GUI to update to select a replication group before applying the changes to the VM.

Changing the size of a Virtual Volume using vSphere

The size of a virtual volume cannot be changed if the volume is in a replication group. You must remove the storage policy from the virtual machine, or assign a different policy that does not include replication, before resizing the VMDK for a virtual machine using vSphere. After applying the new size, the original storage policy can be reassigned to the virtual machine.

Managing Failover Operations

When using vVols replication, the failover and disaster recovery operations are managed through the vCenter Server integrations such as VMware Site Recovery Manager (SRM) and PowerCLI.

Volume groups used for vVols replication can be monitored using the storage system GUI and command-line interface. However, the management of the volume groups is delegatated to the VMware environment.

vCenter Server and ESXi can discover replication groups, but do not manage their life cycle. When using vVols replication, the failover and disaster recovery operations are managed through vCenter Server integrations such as VMware Site Recovery Manager (SRM) and PowerCLI.

VMware can perform three types of failovers on vVol-based VMs:

Planned Failover
Movement of a VM from one site to another for a planned migration. The systems at both sites are required to be accessible throughout the failover process. Once a planned failover is complete, replication can optionally be reversed so that the failed over VM can be reprotected.

Unplanned Failover
Movement of a VM in response to a failure in the production site. Only the recovery site is required to be accessible to perform an unplanned failover. If the original production datacenter recovers after the failover, replication can optionally be reversed so that the failed over VM can be reprotected. In the event that the original production environment cannot be recovered, a new storage policy with alternative replication settings can be configured.

Test Failover
Allows the recovery copy of a VM to be brought up for testing, without taking down the production VM. Test failover enables temporary access to the recovery environment to allow a failover plan to be verified before an actual disaster or planned migration. When the test failover is stopped, any changes made to the recovery copy during the test are discarded.

Performing a Test Failover

When performing a failover test in the VMware environment, the following restrictions apply to the replication groups being tested on the recovery site:

- Creating and deleting vVols is not supported in the test copy of the replication group. These operations are still supported on the production copy of the replication group and any changes made to the production copy while the test is active will be replicated to the recovery copy when the test is stopped.
- Any snapshots created from virtual machines in the replication group must be removed before stopping the test.
- Promoting a replication group directly from test to production is not supported. The test must first be stopped before performing a failover.

Using OpenStack

The IBM Storage Virtualize family volume management driver (formerly known as the Storwize® Cinder driver) offers various block storage services.

For further details about OpenStack, see the following links:

- IBM Storage Virtualize family volume driver documentation
- Cinder driver support matrix
- Cinder release notes

The following OpenStack versions are supported with this release:

- Wallaby
- Xena
- Yoga
- Zed
- Antelope
- Bobcat
- Caracal
- Dalmatian

- Epoxy
- Flamingo

Using Kubernetes and OpenShift

Red Hat® OpenShift® Data Foundation (ODF) uses the ODF FlashSystem Operator and ODF FlashSystem Driver to integrate with your IBM Storage FlashSystem. When using the OpenShift Data Foundation with your storage system ensure that all configuration needs are met by following the Red Hat ODF documentation set.

For more information on supported Red Hat OpenShift Data Foundation (ODF) version support ODF FlashSystem Operator and ODF FlashSystem Driver and FlashSystem storage systems, see ODF FlashSystem driver support summary.

- For more information about Red Hat OpenShift Data Foundation, a software-defined storage for containers, see the OpenShift Data Foundation page.

The ODF FlashSystem Operator and ODF FlashSystem Driver enable IBM Storage Virtualize family products to be used with Red Hat OpenShift Data Foundation.

For more information and documentation, see the following:

- User information and release notes documentation for the Red Hat OpenShift Data Foundation: OpenShift Data Foundation documentation.
- General information about the OpenShift Data Foundation, a software-defined storage for containers: OpenShift Data Foundation

Copy offload by using ODX on Microsoft Windows

Offloaded data transfer (ODX) is a feature in Microsoft Windows Server. ODX function frees up hosts and speeds the copy process by offloading the entire orchestration to a storage array.

For virtual machine (VM) operations and large data transfers, Microsoft ODX offloads the heavy lifting of data movement to the host storage array instead of doing normal read and write operations. Examples include copy operations such as:

- VM creation
- VM migration
- VM cloning
- Microsoft Windows Virtual Hard Disk (VHD) creation
- VHD conversion
- VM backup and recovery
- File copy

ODX is most relevant in a Hyper-V environment for VM heavy lifting operations. Standard buffered copy is run by reading the data from the storage controller into the host, buffering it, and then writing it to another volume. The ODX function frees up hosts and speeds the copy process by offloading the entire orchestration to a storage array. This offloading is done by using tokenization for read operations and write operations, and it avoids buffering, which can ultimately cut down on processor cycles.

Planning for offloaded data transfer (ODX)

For a copy workload in a Microsoft Windows environment, you can use Microsoft offloaded data transfer (ODX) to offload the copy workload to the storage controller to speed job completion. When you decide on an ODX implementation, review the planning considerations to help you determine whether ODX is appropriate for your environment.

It is important if you have a Windows environment to know whether existing Windows workload consists of copy workload, and if it does, what is its extent. By default, Windows offloads any copy greater than 256 K to the controller if the volumes are published as ODX capable. Two important points exist.

1. With ODX enabled, Windows offloads any copy greater than 256 K to the system controllers. The offload frees up the hosts CPU and network. Depending upon the type of copy, the system can even experience faster copy in certain situations. However, for multiple Windows hosts in the environment, all Windows hosts start pumping copy offload jobs to the storage controllers. If you use ODX, plan for this impact.
2. Workloads other than copy might experience higher latencies if the copy workload is high. Since the offloaded jobs tend to "Copy offload by using ODX on Microsoft Windows" on page 125 faster, they require more controller resources in a specified time. The result might be higher latencies with non-copy workloads.

If you have a Hyper-V environment, the benefit is immense through copy offload ODX technology. With ODX enabled, all Windows jobs like file copy, VM migration, cloning, and VHD creation would be offloaded to the storage controller. It is advisable to check the impact of copy offload on other types of workloads in such an environment.

If you decide to implement ODX, it is advisable that you review existing workloads and make sure that capacity planning is completed to meet the demand. If you are not interested in the benefits of copy offload, you can disable system-wide ODX by using the CLI. Not using ODX prevents existing workloads from experiencing extra latency due to copy offload.

SCSI Fibre Channel host attachments

Small Computer System Interface (SCSI) Fibre Channel hosts with different operating systems can be attached to the system.

The system supports configurations that use N_port ID virtualization (NPIV) in the host bus adapter when they are connected to a SAN switch. However, these configurations are not supported when they are directly attached to the Fibre Channel ports on the system.

IBM Power Systems AIX

You can attach IBM Power Systems AIX hosts via Fibre Channel.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Refer to AIX Documentation for further configuration information.

Configuring the AIX operating system

You must configure the AIX operating system to use IBM Power Systems hosts with the system.

Configuring for fast fail and dynamic tracking

You must enable the fast fail and dynamic tracking attributes for hosts systems that run an AIX 5.2 or later operating system.

Before configuring your host system to use these attributes, ensure that the host is running the AIX operating system version 5.2 or later.

To configure your host system to use the fast fail and dynamic tracking attributes follow these steps.

Note: These steps need to be taken only on systems running AIX 6.1 or 5.2. The commands require the fscsi*X* device to be in the defined state.

1. Issue the `chdev -l fscsi0 -a fc_err_recov=fast_fail` command to set the Fibre Channel SCSI I/O Controller Protocol Device event error recovery policy to fast_fail for each Fibre Channel adapter. This example command is for adapter fscsi0.

   Note: By default, dynamic tracking is enabled on all systems running AIX 7.1.

2. Issue the `chdev -l fscsi0 -a dyntrk=yes` command to enable dynamic tracking for each Fibre Channel device. This example command is for adapter fscsi0.

Multipath support for IBM Power Systems hosts

You must install multipathing software on all IBM Power Systems AIX hosts that are attached to the system.

On IBM Power Systems hosts, the AIX Path Control Module (AIXPCM) provides multipathing support.

For more information about the multipath drivers, see www.ibm.com/support.

Configuring system devices with multiple paths per LUN

- The system supports multiple LUNs on an IBM Power Systems AIX host.
- The system supports multiple path configurations for a LUN. You can have multiple hdisks (logical hard disks) available on the host for each physical LUN.
- To configure multiple paths for all LUNs, run the **cfgmgr** command to scan for new devices.
  - If you are using the AIX Path Control Module (AIXPCM) and want to check whether the multipath device is available, run the **lsdev** command and search for the `hdisk#` line with the `MPIO IBM #### FC Disk` label against it.

For more information about which multipath driver to use, see http://www.ibm.com/support.

Multipathing configuration maximums for IBM Power Systems hosts

When you configure, keep in mind the maximum configuration for IBM Power Systems AIX hosts.

Table 22 on page 127 provides the maximum volumes and paths per volume for the Native Multipathing Plug-In (NMP).

Table 22. Configuration maximums for NMP on IBM Power Systems AIX hosts

Object	NMP maximum	Description
Volume (hdisk)	2048	The maximum number of volumes that can be supported by the NMP (per host object). The maximum number of volumes is enforced by the system. **Note:** The system supports 512 volumes by using iSCSI.
Paths per volume	8	The maximum number of paths to each volume. The number of paths directly corresponds with the resulting path-failover time. Although the maximum number of supported paths is eight, do not use more than two paths per adapter port.

SAN boot support for IBM Power Systems hosts

You can reload the operating system from a boot disk on the SAN over Fibre Channel. You can initiate this process from the system volume if your IBM Power Systems hosts use AIX operating system version 5.3 or later.

You must use the AIX Path Control Module (AIXPCM) to boot from the SAN. Create an installation volume and map it to the host. Proceed with the installation per the AIX installation instructions and, when you are prompted, select the previously defined volume as the target installation disk.

Dynamically increasing volume size

If your IBM Power Systems AIX hosts use AIX 5.2 or a later AIX operating system version, the system supports the ability to dynamically increase volume size.

The **chvg** command options make it possible to increase the size of a physical volume that the Logical Volume Manager (LVM) uses, without interruptions to the use or availability of the system. See the AIX publication *System Management Guide: Operating System and Devices* for more information.

Virtual input/output for IBM Power Systems hosts

The system provides both single and dual Virtual input/output (VIO) server configurations on IBM Power Systems hosts that support VIO.

You can present the system volumes to the VIO server host bus adapters (HBAs) by using the same method as a standard AIX installation. For single VIO server configurations, volumes can be split up into logical volumes by the VIO server and mapped to the VIO clients. For dual VIO server configurations, volumes cannot be split into logical volumes, and must instead be mapped intact through both servers to the VIO clients.

Known AIX issues and limitations

The restriction of Direct Memory Access (DMA) resources is one of the known issues and limitations with the system and an IBM Power Systems AIX host.

On a heavily loaded system, the following symptoms can indicate that the host is low on direct memory access (DMA) resources:

- You might see errors that indicate that the host bus adapter (HBA) was unable to activate an I/O request on the first attempt.
- You might see lower-than-expected performance with no errors being logged.

To reduce the incidence of these messages, you can increase the resources by modifying the maximum transfer size attribute for the adapter as follows:

1. Type the `lsattr -El` *HBA* `-a max_xfer_size` command to view the current setting, where *HBA* is the name of the adapter that is logging the error. For this example, the HBA is fcs0.
2. Type the `chdev -l fcs0 -P -a max_xfer_size=0x1000000` command to increase the size of the setting.

   Note: To view the range of allowable values for the attribute, type `lsattr -Rl fcs0 -a max_xfer_size`.
3. Restart the host to put these changes into effect.

IBM Power Systems with Virtual I/O Server

Successful Fibre Channel attachments for IBM Power Systems with Virtual I/O Server require knowledge of related configuration details and known issues and limitations with the server.

Configuring the IBM i operating system

You must configure the Virtual I/O Server and the IBM i client operating system before you can use IBM i as a host.

Before you configure the host operating systems, the following tasks must be completed:

- Your system must be installed.
- You must install the appropriate host bus adapter and driver on your host.

After the prerequisite tasks are complete, use the following general steps to configure your host operating system:

1. Zone the Virtual I/O Server to the system on the Fibre Channel network.

2. Install the AIX Path Control Module (AIXPCM) on the Virtual I/O Server to enable the management of multiple paths to the system volumes.
3. Create the host on the system by using the worldwide port names (WWPNs). Map the volumes to the host as required.

   Note: The IBM i virtual I/O client supports boot from a SAN with a volume for the load source of 20 GB or larger.
4. Run the **cfgdev** command on the Virtual I/O Server to configure the new disks. Use the **lspath** command to ensure that disks are listed for all paths to the system.
5. Map the system LUNs (hdisks on the Virtual I/O Server) uniquely to the IBM i virtual SCSI client adapters by using the **mkvdev** command.
6. Use the Hardware Management Console (HMC) to verify that the I/O tagging for the IBM i virtual I/O client partition is correctly set for the load source and the alternate restart device.
7. Install IBM i version 6 release 1 or later.

IBM z Systems that run the Linux operating system

You can attach IBM z Systems® that run the Linux operating system via Fibre Channel.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Refer to IBM z Systems Linux documentation and the following pages for further configuration information.

Configuring the Linux operating system for IBM z Systems® hosts

You must configure the Linux operating system before you can use z Systems hosts with the system.

Multipath support for IBM z Systems hosts

You must install multipathing software on all z Systems hosts that are attached to the system.

On z Systems hosts that are running the Linux operating system, the following software provides multipathing support:

Linux Kernel 2.4
The logical volume manager.

Linux Kernel 2.6

- The multipath-tools package on SUSE Linux Enterprise Server.
- The device-mapper-multipath package on Red Hat Enterprise Linux.

For more information about using the logical volume manager or the multipathing tools package with z Systems hosts that run Linux kernel 2.4 or Linux kernel 2.6, see the documentation at the following website:

www.ibm.com/developerworks/linux/linux390/index.html.

Installing and configuring the host bus adapter on an IBM Z® host system

The host bus adapters (HBAs) for an IBM Z® host system must be ordered as features. These features are either factory-installed when you order a new system or installed into an existing system by a service representative.

To configure the HBA:

- Ensure that the appropriate features (FICON®, FICON Express, FICON Express2, or FICON Express4) are installed on your IBM Z® host system.
- Configure the HBA to run in FCP mode.

For more information about FCP connectivity, see www.ibm.com/support.

Multipathing configuration maximums for LVM and mp-tools

When you configure, keep in mind the maximum configuration for the logical volume manager (specifically, LVM1) for the Linux operating system.

Logical volume manager

Table 23 on page 130 provides the maximum number of physical volumes, logical volumes, and paths for LVM1.

Table 23. Configuration maximums for LVM1

Object	Maximum	Description
Number of volume groups	99	The number of volume groups that can be defined per host
Physical volumes per volume group	256	The maximum number of physical volumes that can be assigned to one volume group
Paths per physical volume	16	The maximum number of paths to each PV
Logical volumes	256	The total number of logical volumes that are supported by LVM1 (because of the 256 minor number limit of the kernel)

Multipath-tools

The number of paths per physical volume is limited to eight paths with Kernel 2.6.

SAN boot support on IBM z Systems hosts

On a SCSI LUN, you can complete an initial program load (IPL) process.

IPL processes on SCSI LUNs can fail because there is no path failover capability during the boot process. See the latest *How to use FC-attached SCSI devices with Linux on z Systems* publication for more information about using IPL processes with z Systems hosts.

Defining the number of disks on IBM z Systems hosts

When you define the number of disks on z Systems hosts that are running the Linux operating system, you are provisioning capacity for configured disks.

On the Linux operating system, a device node is used for each path to the disk and a second device node is used for the path grouped device. The maximum number of devices depends on your Linux configuration.

System configuration for IBM z Systems hosts

Each attached system LUN has a set of device files in the `/dev` Linux directory.

These device files are intended to be used only by the multipath tools. Allowing them to be used directly by your system or applications can produce data inconsistencies. Instead, configure your system and applications to use the device nodes that are created by the multipath tools.

Known issues and limitations for IBM z Systems hosts

There are some restrictions for z Systems hosts that are running the Linux operating system.

The following website provides currently known restrictions for the latest Linux on z Systems releases:

www.ibm.com/developerworks/linux/linux390/index.html.

IPL restrictions

Device Mapper Multipath (DM-Multipath) is not available on either the root or boot devices.

For information about DM-Multipath configuration and administration, see Device Mapping Multipath. System re-IPL (`shutdown -r`) is supported on z/VM® guests only. It is not supported in LPAR mode.

Changing the timeout value for the FC transport class

You can change the timeout value for the Fibre Channel (FC) transport class when the I/O operation stalls.

The FC transport class keeps all outstanding I/O for a certain time until it determines that the remote port is not responding. This time period is controlled by the timeout value. The default time is 60 seconds. This situation causes a temporary I/O stall in the case of FC link incidents or when paths are varied offline.

Change the timeout value.

- Set the timeout value for each remote port by using the `sysfs` file system.

```
/sys/class/fc_remote_ports/rport-0:0-0/dev_loss_tmo
```

- Preset the timeout value for all remote ports when you load the `scsi_transport_fc` module by using the `dev_loss_tmo` parameter.

SCSI device state changes from running to offline

In some instances, the Small Computer System Interface (SCSI) mid layer changes the SCSI device state from **running** to **offline**.

I/O that is later sent to that SCSI device is rejected by the mid layer and never reaches the SCSI lower-level device drivers for processing. To solve this problem, you must increase the SCSI timeout value to at least 60 seconds as shown in the following example:

```
SUBSYSTEM=="scsi", ACTION=="add", ATTR{vendor}=="IBM",
ATTR{model}==", ATTR{timeout}="60"
```

IBM z Systems running the IBM z/VSE operating system

You can attach IBM z Systems servers (hosts) that are running the IBM z/VSE® operating system. The z/VSE operating system can operate in a logical partition (LPAR) and in an IBM z/VM guest environment.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

The following IBM website provides general z/VSE information about supported z/VSE releases, and supported IBM z Systems servers: www-03.ibm.com/systems/z/os/zvse/about/status.html.

You need a FICON Express (FICON Express, FICON Express2 or higher) adapter (also named host bus adapter (HBA)) on your z Systems server.

You must configure the FICON® Express adapter or HBA in FCP mode. FCP mode means that the HBA is configured with channel path ID (CHPID) type FCP in the input/output control data set (IOCDS) of your z Systems server.

Note: With N_Port ID Virtualization (NPIV), multiple virtual FCP adapters can be defined. Each virtual FCP adapter can have its own unique WWPN. You can use the WWPN of a virtual FCP adapter to map volumes to the host. This action can be used to define multiple hosts that share the same physical FCP adapter. The NPIV support was introduced with System z9®.

Configure your z/VSE system to use FCP-attached SCSI disks. After you configure a path to the SCSI disk, the SCSI disks are recognized by z/VSE user programs and system programs as fixed block architecture (FBA) disks only. You can use the SCSI disks with existing FBA interfaces. The z/VSE operating system does not provide SCSI command support for applications. See the IBM z/VSE Administration publication for additional information.

SAN boot support for IBM z Systems servers running the z/VSE operating system

If you have installed the z/VSE operating system on SCSI disks, you can complete the initial program load (IPL) process from the SCSI disks.

See the *IBM z/VSE Administration* publication for information regarding what specifications are required to initiate the IPL process in an LPAR environment and in a z/VM guest environment.

Hosts that run the Linux operating system

You can attach hosts running the Linux operating system via Fibre Channel.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Refer to Linux Documentation for further configuration information.

Requirements for attaching systems to hosts running the Linux

Ensure that your system meets the requirements for attaching to a host that is running the Linux operating system.

The following list provides the requirements for attaching the system to the host that is running the Linux operating system:

- Check the LUN limitations for your host system.
- Ensure that you have the documentation for your host and access to the hardware installation information for the correct model of your system. All system publications are available from the following website:www.ibm.com/support.
- Ensure that you install the correct operating systems and are running a supported kernel of Linux.
- When you attach the system to a BladeCenter blade server, see the BladeCenter documentation for SAN configuration details.

You must ensure the following configuration settings to restore the path if they are lost during an upgrade or other hardware maintenance that results in a node being offline:

- SCSI INQUIRY TIMEOUT
- SCSI COMMAND TIMEOUT
- Multipath settings in multipath.conf

Note: If you configure your system to use NPIV, there are less chances that the paths will get lost.

To set SCSI INQUIRY TIMEOUT use these steps:

- Add **`scsi_mod.inq_timeout=70`** to the kernel boot command line through *GRUB* configuration. By adding the **`scsi_mod.inq_timeout=70`** parameter, the change in the parameter is persistent from a server reboot. Linux hosts can also regain system node paths when lost. This change can be done by completing the following steps.

  For SLES12 and SLES15 servers, follow these steps:

  1. To make the permanent change, edit `/etc/default/grub` and add to the `GRUB_CMDLINE_LINUX_DEFAULT` line:

     ```
     scsi_mod.inq_timeout=70
     ```

  2. Run the following command to rewrite the boot record:

     ```
     grub2-mkconfig -o /boot/grub2/grub.cfg
     ```

- For RHEL6, RHEL7 and RHEL8 servers, follow these steps:

1. To make the change permanent, edit `/etc/sysconfig/grub` and add to the GRUB_CMDLINE_LINUX line:

```
scsi_mod.inq_timeout=70
```

2. Run the following command to rewrite the boot record:

```
grub2-mkconfig -o /etc/grub2.cfg
```

The previous steps will be effective after rebooting. In RHEL6, RHEL7, RHEL8, SLES12, and SLES15, you can choose to change the **inq_timeout** parameter temporarily without rebooting. This method will not keep the parameter value persistent if the system ever reboots in the future and if you do not edit the GRUB configuration by following these steps. .

Use the following command to change the **inq_timeout** parameter temporarily without rebooting:

```
echo 70 > /sys/module/scsi_mod/parameters/inq_timeout
```

In RHEL6, RHEL7, RHEL8, SLES12, and SLES15 enter the following command to view that the change was made:

```
systool -m scsi_mod -A inq_timeout
```

The output of the command shows that the value is changed to 70:

```
Module = "scsi_mod"
```

```
inq_timeout         = "70"
```

Note: It is best to perform both tasks in case the server is rebooted in the future.

To set the SCSI COMMAND TIMEOUT

- Set the *udev* rules for SCSI command timeout to 120s. This is the recommended setting for all versions of Linux.

Udev rules file creation

1. To increase the SCSI command timeout for the system, create the following udev rule:

```
# Set SCSI command timeout to 120s (default == 30 or 60) for IBM 2145 devices
SUBSYSTEM=="block", ACTION=="add", ENV{ID_VENDOR}=="IBM",ENV{ID_MODEL}=="2145", RUN+="/bin/sh -c 'echo 120 >/sys/block/%k/device/timeout'"
```

2. After you set up your volumes, confirm that they are set for 120 seconds. Locate the block device paths by running `multipath -ll | grep sd` from the command prompt. Then, run `cat /sys/block/sdX/device/timeout` (where X is each 2145 block device path).
3. To reload the udev rules without rebooting (or dynamically) you can run the following commands:

```
udevadm control -R
/sbin/udevadm trigger --type=devices --action=add
```

Settings for Linux hosts

To ensure path recovery in failover scenarios, certain Device Mapper Multipath (DMMP) settings and udev rules for the attachment of Linux hosts to the system are recommended. These settings are valid for IBM System x, all Intel or AMD-based servers, and Power® platforms.

You must restart your host after you complete the following two steps:

- Editing the multipath settings in `/etc/multipath.conf`
- Editing the udev rules for SCSI command timeout

For each Linux distribution and releases within a distribution, refer to the default settings under `[/usr/share/doc/device-mapper-multipath.*]` for Red Hat and `[/usr/share/doc/packages/multipath-tools]` for Novell SuSE. Ensure that the entries added to `multipath.conf`

match the format and syntax for the required Linux distribution. Use the `multipath.conf` only from your related distribution and release. Do not copy the `multipath.conf` file from one distribution or release to another.

For some operating system levels, the `polling_interval` needs to be located under defaults instead of under **device settings**. If `polling_interval` is present in the device section, comment out `polling_interval` by using a # key.

For example:

```
Under Device Section
#         polling_interval 5,

Under Defaults Section
defaults {
        user_friendly_names yes
        polling_interval 5
}
```

Multipath settings for specific Linux distributions and releases

Edit `/etc/multipath.conf` with the following parameters and confirm the changes by entering:

```
multipathd -k
multipathd> show config
```

Note: Oracle Linux versions are same as Red Hat Linux versions.

Red Hat Linux version 7.x, 8.x, and 9.x

```
     vendor "IBM"
     product "2145"
     path_grouping_policy "group_by_prio"
     path_selector "service-time 0" # Used by Red Hat 7.x
     prio "alua"
     path_checker "tur"
     failback "immediate"
     no_path_retry 5
     rr_weight uniform
     rr_min_io_rq "1"
     dev_loss_tmo 120
```

SUSE Linux Versions 12 and 15

```
     vendor "IBM"
     product "2145"
     path_grouping_policy "group_by_prio"
     path_selector "service-time 0"
     prio "alua"
     path_checker "tur"
     failback "immediate"
     retain_attached_hw_handler "yes"
     no_path_retry 5 # or no_path_retry "fail"
     fast_io_fail_tmo 5
     rr_min_io 1000
     rr_min_io_rq 1
     rr_weight "uniform"
```

Ubuntu

```
     vendor "IBM"
     product "2145"
     path_grouping_policy "group_by_prio"
     path_selector "service-time 0"
     prio "alua"
     path_checker "tur"
     failback "immediate"
     no_path_retry 5 # or no_path_retry "fail"
     retain_attached_hw_handler "yes"
     fast_io_fail_tmo 5
     rr_min_io 1000
```

```
    rr_min_io_rq 1
    rr_weight "uniform"
```

DM-MPIO for dev_loss_tmo

After a problem is detected on an FC port and it set to infinity, the SCSI layer can wait until 2147483647 seconds (68 years) before removing it from the system. The default value is determined by the OS.

All Linux hosts should have a **dev_loss_tmo** setting, but the value in seconds is how long to wait for the device/paths to be pruned. The suggested duration is 120-150 seconds, but extended duration is also supported.

Care needs to be taken if it is too low since if paths are pruned, then they also need to be rediscovered and if too low, that may require manual rescan later. If inquiry timeout is right, the host should be able to re-add the paths when the SVC nodes are restored.

If the inquiry is too short such as 20 seconds then the inquiry may timeout before the paths are ready.

Multipathing driver

If you lose paths and are not automatically restored, you can manually get them back with the following process.

- If you are using Linux *dm-multipath* as the `multipathing` software on the host, and have NPIV disabled on the SVC, it may be necessary to rescan ports after each node restores its paths.
- If you have NPIV disabled on the SVC, additional configuration is required. During the upgrade, an SVC with NPIV disabled will shut down the ports for an extended period of time, which may cause Linux to remove the ports.
- This setting may be applied to a running system, however it must also be applied to the GRUB configuration in order to ensure it persists over reboots.
- It is possible that, even with this setting, an SVC upgrade may keep the ports down longer than the timeout setting allows. It may be necessary to rescan the ports once the SVC node has restored operation.
- Use the **multipath** command to check path status once the SVC node has completed the upgrade, before updating the next node:

```
# multipath -ll /dev/mapper/mpatha
(360050768028211d8b000000000000061) dm-11 IBM,2145
size=10G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=enabled
| |- 4:0:1:0  sder    129:48  failed faulty running
| `- 5:0:1:0  sdfb    129:208 failed faulty running
`-+- policy='service-time 0' prio=10 status=active
  |- 4:0:0:0  sdem    128:224 active ready running
  `- 5:0:0:0  sdew    129:128 active ready running#
```

Choose a device name, instead of mpatha in the above example, to match a device being utilized by the SVC that is performing the upgrade. The failed and faulty status, indicates that paths are still down for Linux. You can use **multipath -ll** command to list all devices name, and then scan the output for failed paths, confirming that failed paths are on devices associated with the SVC upgrade.

To rescan all SCSI targets and the ports on Red Hat Enterprise Linux or SUSE Linux Enterprise Server, use the **rescan-scsi-bus.sh** command which is part of the `sg3_utils` package.

If your Linux distribution does not include the **rescan-scsi-bus.sh** command, use the **SCSI-rescan** command to rescans all the SCSI targets.

Hosts that run the Microsoft Windows Server operating system

You can attach hosts running the Microsoft Windows Server operating system via Fibre Channel.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Use the Microsoft Device Specific Module (MSDSM) multipathing driver for the Windows Server operating system. MSDSM is included in the Windows Server operating system and Hyper-V.

Refer to Windows documentation and the following pages for further configuration information.

Changing the disk timeout on Microsoft Windows Server

To change the disk timeout value on Windows Server operating systems, follow steps in the Windows Server host and in the registry browsing tool.

On your Windows Server hosts, change the **disk I/O timeout** value to 60 in the Windows registry, as follows:

1. In Windows, click **Start** > **Run**.
2. In the dialog text box, type `regedit` and click **Enter**.
3. In the registry browsing tool, locate the `HKEY_LOCAL_MACHINE\System\CurrentControlSet\Services\Disk\TimeOutValue` key.
4. Confirm that the value for the key is 60 (decimal value). If necessary, change the value to 60.

Configuring the Brocade HBA for hosts that run a Windows Server operating system

After you install the Brocade host bus adapter (HBA) and the driver on a Microsoft Windows Server operating system host, you must configure the HBA.

For more information about supported Brocade models, see the following website:

www.ibm.com/support

To configure the HBA BIOS, either use the Brocade Host Connectivity Manager (HCM) software or reboot into the Brocade configuration tool. Configure the following settings:

- Host Adapter BIOS: Disabled (unless the machine is configured for SAN Boot)
- Queue depth: 4

You can disable the BIOS by using the command-line configuration tool:

- `bcu bios –disable 1/0`
- `bcu bios –disable 1/0`

You can set the queue depth by using the command-line configuration tool:

- `bcu fcpim –qdepth 1/04`
- `bcu fcpim –qdepth 1/14`

You can query the settings by using the command-line configuration tool:

- `bcu port –query 1/0`
- `bcu port –query 1/1`

Configuring the QLogic HBA for hosts that run the Windows Server operating system

After you install the QLogic HBA and the device driver on hosts that run a Microsoft Windows Server operating system, you must configure the HBA.

About this task

For more information about supported QLogic models, see the following IBM website:

www.ibm.com/support

To configure the HBA BIOS, either use the QLogic HBA manager software or reboot into the Fast!UTIL tool. Configure the following settings:

- Host Adapter BIOS: Disabled (unless the machine is configured for SAN Boot)
- Adapter Hard Loop ID: Disabled
- Connection Options: 1 - point to point only
- LUNs Per Target: 0
- Port Down Retry Count: 15

Set the execution throttle to a suitable queue depth for your environment, for example, a value of 100. For more information about the execution throttle and queue depth, see QLogic documentation.

If you are using Native Multipathing Plug-In (NMP), set Enable Target Reset to No. See Table 24 on page 137 to include the required parameters for the registry key.

Table 24. Registry key parameters for QLogic models

Key	Required parameters
HKEY_LOCAL_MACHINE > SYSTEM > CurrentControlSet > Services > ql2*xxx* > Parameters > Device > DriverParameters	**`Buschange=0;FixupInquiry=1`** **Note:** If you are using QLogic driver version 9.1.2.11 or later, **Buschange** cannot be set to zero. See your device driver documentation for details.

Configuring the Emulex HBA for hosts that run a Windows Server operating system

After you install the Emulex host bus adapter (HBA) and the driver on hosts that run a Microsoft Windows Server operating system, you must configure the HBA.

For the Emulex HBA StorPort driver, accept the default settings and set topology to 1 (1=F_Port Fabric). For the Emulex HBA FC Port driver, use the default settings and change the parameters that are given in Table 25 on page 137.

Note: The parameters that are shown in parentheses correspond to the parameters in HBAnywhere.

Table 25. Configuration file parameters for the Emulex HBA

Parameters	Recommended Settings
Query name server for all N-ports (BrokenRSCN)	Enabled
LUN mapping (MapLuns)	Enabled (1)
Automatic LUN mapping (MapLuns)	Enabled (1)
Allow multiple paths to SCSI targets (MultipleSCSIClaims).	Enabled
Scan in device ID order (ScanDeviceIDOrder)	Disabled
Translate queue full to busy (TranslateQueueFull).	Enabled
Retry timer (RetryTimer)	2000 milliseconds
Maximum number of LUNs (MaximumLun)	Equal to or greater than the number of the system LUNs that are available to the HBA.

Configuring Microsoft Device Specific Module (MSDSM)

Use the Microsoft Device Specific Module (MSDSM) multipathing driver for the Windows Server operating system. MSDSM is included in the Windows Server operating system and Hyper-V.

The Microsoft Multipath I/O (MPIO) feature must also be enabled when using MSDSM .

Complete the following steps to install MSDSM on the Windows Server operating system.

1. To use MSDSM, install the Microsoft Multipath I/O (MPIO) feature by using the **Add Roles and Features** wizard in the Windows Server Manager.
2. Open the **MPIO Properties** window from the MPIO application in Windows Administrative Tools.
3. Select **IBM 2145** on the **Discover Multi-Paths** tab in the **MPIO Properties** window and click **Add**.
4. Click **OK** to restart the host when prompted. Wait for the host to restart.
5. Open Windows PowerShell and enter the **`mpclaim -s -d`** command.

   Ensure that MSDSM is installed correctly and can claim the volumes correctly. Use the default MSDSM settings.

Migrating existing SAN boot images

If you have a host that runs the Microsoft Windows Server operating system and existing SAN boot images that are controlled by storage controllers, you can migrate these images to image-mode volumes that are controlled by the system.

About this task

To migrate your existing SAN boot images, complete the following steps:

Procedure

1. Shut down the host.
2. Complete the following configuration changes on the storage controller:
   a) Remove all the image-to-host mappings from the storage controller.
   b) Map the existing SAN boot image and any other disks to the system.
3. Zone one port of each host bus adapter (HBA) to one of the system ports that is associated with the I/O group for the target image-mode volume.
4. Complete the following configuration changes on the system:
   a) Create an image-mode volume for the managed disk (MDisk) that contains the SAN boot image.
      Use the MDisk unique identifier to specify the correct MDisk.
   b) Create a host object and assign it to the HBA port that you zoned to the system port in step “3” on page 138.
   c) Map the image mode volume to the host.
      For example, you might map the boot disk to the host with SCSI LUN ID 0.
   d) Map the swap disk to the host, if required.
      For example, you might map the swap disk to the host with SCSI LUN ID 1.
5. Change the boot address of the host by completing the following steps:
   a) Restart the host and open the BIOS utility of the host during the booting process.
   b) Set the BIOS settings on the host to find the boot image at the worldwide port name (WWPN) of the node that is zoned to the HBA port.
6. Boot the host in single-path mode.
7. Uninstall any multipathing driver that is not supported for system hosts that run the applicable Windows Server operating system.
8. Install a supported multipathing driver.

9. Restart the host in single-path mode to ensure that the supported multipath driver was properly installed.
10. Zone each HBA port to one port on each system node.
11. Add HBA ports to the host object that you created in step “4.b” on page 138.
12. Configure the HBA settings on the host by using the following steps:
    a) Restart the host and open the host's BIOS utility during the booting process.
    b) Ensure that all HBA ports are boot-enabled and can see both nodes in the I/O group that contains the SAN boot image. Configure the HBA ports for redundant paths.
    c) Exit the BIOS utility and finish booting the host.
13. Map any additional volume to the host as required.

Hosts that run the VMware ESXi operating system

Fibre Channel attachments between the system and various guest host operating systems that run on the ESXi operating system have specific requirements and other guidelines that need to be followed.

Note: The system supports VMware vSphere Virtual Volumes that allow VMware vCenter to manage system objects like volumes and pools. Virtual Volumes provide storage that is used by specific virtual servers, and also correspond to single volumes on the system. The system administrator can assign ownership to VMware administrators to create the pools and volumes that are used exclusively for Virtual Volumes. For more information, see the information about configuring Virtual Volumes.

Configuring the ESXi operating system

You must configure the VMware ESXi operating system before you can connect hosts that run VMware ESXi with the system.

Before you begin

Before you configure the host operating system, the following tasks must be completed:

- Your system must be installed.
- You must install the appropriate host bus adapters and ensure that the latest drivers and firmware are applied.

Procedure

After the prerequisite tasks are complete, use the following general steps to configure your host.

1. Define the host object by using the worldwide port name identifiers (WWPNs) for each physical ESXi host in the VMware vSphere cluster.
2. If more than one physical ESXi host is present in the vSphere cluster, create a host cluster and add each ESXi host object as a member.

   When you map volumes to ESXi hosts, ensure that you create shared host-cluster mappings for all shared storage to ensure consistent SCSI IDs across the ESXi hosts for the same volume. If the ESXi host is SAN booted from a volume on the system, use private mappings for these volumes.
3. Review the multipath settings by running the following CLI command:

   ```
   # esxcli storage nmp device list -d naa.<UUID of volume>
   ```

   The output is similar to the following example:

   ```
   naa.600507640084011070000000000008c
      Device Display Name: IBM Fibre Channel Disk (naa.600507640084011070000000000008c)
      Storage Array Type: VMW_SATP_ALUA
      Storage Array Type Device Config: {implicit_support=on; explicit_support=off;
   explicit_allow=on; alua_followover=on; action_OnRetryErrors=on; {TPG_id=16,TPG_state=AO}
   {TPG_id=17,TPG_state=ANO}}
      Path Selection Policy: VMW_PSP_RR
   ```

```
   Path Selection Policy Device Config: {policy=rr,iops=1,bytes=10485760,useANO=0;
lastPathIndex=1: NumIOsPending=0,numBytesPending=0}
   Path Selection Policy Device Custom Config:
   Working Paths: vmhba4:C0:T0:L250, vmhba3:C0:T0:L250
```

Verify the following settings:

a. The Storage Array type plug-in for IBM volumes is set to VMW_SATP_ALUA.

b. The path selection policy is set to RoundRobin.

c. The Round Robin IOPS is set to 1 to evenly distribute I/Os across as many ports on the system as possible. For information about how to make this change, see the VMware information at this website: Adjusting Round Robin IOPS limit from default 1000 to 1 (2069356) .

4. Enable ATS hardware accelerated locking and set to ATS-only public.To verify the current settings, use the following CLI command:

```
# vmkfstools -Ph -v1 /vmfs/volumes/<datastore_name> | grep Mode
Mode: public ATS-only
```

If the datastore Mode is not reporting public ATS-only, consider either of the following two options:

a. Creating a LUN and creating a VMFS5 or VMFS6 datastore and use Storage vMotion to migrate virtual machines to the new datastore.

b. Enforcing ATS-only on all affected datastores. This process requires stopping all I/O to the datastore first. For more information, see Change VMFS Locking to ATS-Only. This CLI command provides the UUID of the device:

```
# vmkfstools -Ph -v1 /vmfs/volumes/<datastore_name>  | grep -A 1 Partitions
Partitions spanned (on "lvm"):        naa.60050764008401107000000000000008b:1
```

Using the UUID of the device, the following CLI commands can be used to change the Lock Mode to ATS-only:

```
# vmkfstools --configATSOnly 1 /vmfs/devices/disks/naa.60050764008401107000000000000008b:1
```

5. Set the XCopy transfer size to 4096.

```
# Get-VMHost  | Get-AdvancedSetting -Name DataMover.MaxHWTransferSize | select Entity,
name, value


Entity                          Name                          Value
------                          ----                          -----
vmlab11c2.ssd.hursley.ibm.com DataMover.MaxHWTransferSize  4096
```

6. Use VAAI, which is a set of SCSI commands that are used by VMware ESXi hosts to offload operations to the storage array. Enable **HardwareAcceleratedMove**, **HardwareAcceleratedLocking**, and **HardwareAcceleratedInit**.

```
# Get-VMHost  | Get-AdvancedSetting -name *HardwareAccelerated* | select Name, value
Name                               Value
----                               -----
DataMover.HardwareAcceleratedMove     1
VMFS3.HardwareAcceleratedLocking      1
DataMover.HardwareAcceleratedInit     1
```

7. Wherever possible, set the guest virtual machines to have SCSI timeouts set to at least 60 seconds although 120 seconds is recommended. Refer to the host vendor documentation for instructions on which settings to change.

8. The number of paths per volume is recommended not to exceed 8.

However, during operations such as a nondisruptive volume move, if the Fibre Channel target port mode is transitional for the I/O group, up to 16 paths per volume is supported. See the VMware documentation for the maximum number of devices and paths for the version of VMware ESXi that you are using.

9. If you use Datastore Clusters where vSphere Storage vMotion is used to migrate virtual machines based on I/O latency, it is important to understand the system requirements.
    a. Ensure that each LUN has independent and isolated performance characteristics:
        i) Where possible, present each volume from a different pool, especially when you use arrays or managed disks that are created from traditional spinning (non-SSD or flash) media.
        ii) Review the preferred node on each LUN in the Datastore Cluster to ensure an even distribution between the nodes within the I/O group.
    b. Given VMs can migrate between datastores within the Datastore Cluster. Be aware of dynamic I/O workloads and how Easy Tier will respond. Do one of the following tasks:
        i) Disable automatic storage migrations when you use Easy Tier, or
        ii) Use a single tier of storage between all LUNs that are used in a Datastore Cluster.
10. If you have dual-site topologies, consider the following requirements:
    - If the system topology is set to `stretched`, consider using Host and Virtual Machine DRS Groups to restrict a given VM to move only to other hosts within the same site.
    - VMware ESXi host objects should have a site set to avoid unnecessary intersite traffic.

VMware ATS heartbeating

VMware ESXi uses the SCSI Compare and Write command (VMware refers to this command as Atomic Test and Set - ATS) to "heartbeat" periodically to datastores.

Note: The use of ATS heartbeating is not supported on ESXi hosts that run version 5.5 update 2 or later or ESXi version 6.0 before update 3.

Due to the low timeout value for heartbeat I/O that uses ATS, host disconnects can occur if delays of 8 seconds or longer are experienced in completing individual heartbeat I/Os on backend storage systems or the SAN infrastructure. Under certain circumstances, such as if the ATS heartbeats are ended due to delays of 8 seconds or more, the ESXi host might retry the ATS heartbeat command and it might fail with a miscompare or lost access to datastore message that is logged in the vmkernel.log on the host. In many cases, ATS heartbeat miscomparisons during verification and lost access to datastore messages are results of ATS heartbeats that work as expected when I/O delays are present.

ATS heartbeating increases the load on the system and can lead to access issues on busy systems, particularly during maintenance procedures. To reduce this load, ATS heartbeats can be disabled.

For VMFS5 and VMFS6 datastores, run these commands:

- To disable ATS heartbeats, run the following CLI command:

```
# esxcli system settings advanced set -i 0 -o /VMFS3/UseATSForHBOnVMFS5
```

- To enable ATS heartbeats, run the following CLI command:

```
# esxcli system settings advanced set -i 1 -o /VMFS3/UseATSForHBOnVMFS5
```

For VMFS3 datastores, run the following command to disable ATS heartbeats:

```
# esxcli system settings advanced set -i 0 -o /VMFS3/useATSForHBOnVMFS3
```

SAN boot support for hosts running ESXi operating systems

The system can be used as a boot device for the ESXi operating system.

For SAN boot support for the ESXi operating system, you must meet the following requirement:

- The ESXi operating system must be on a SAN disk.

See the software restrictions page on the following IBM support website for any other restrictions for SAN boot support: www.ibm.com/support.

Hosts that run the Microsoft Hyper-V operating system

You can attach hosts running the Microsoft Hyper-V operating system via Fibre Channel. This allows you to run multiple guest operating systems on the server.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Refer to the Windows Documentation and the following pages for further configuration information.

Configuring the QLogic HBA for a host that runs the Microsoft Hyper-V operating system

After you install the QLogic host bus adapter (HBA) and the device driver on hosts that are running the Microsoft Hyper-V operating system, you must configure the HBA.

About this task

To configure the HBA BIOS, either use the QLogic HBA manager software or reboot into the Fast!UTIL tool. Configure the following settings:

- **Host Adapter BIOS:** Disabled (unless the machine is configured for SAN Boot)
- **Adapter Hard Loop ID:** Disabled
- **Connection Options:** 1 - point to point only
- **LUNs Per Target:** 0
- **Port Down Retry Count:** 15

Set the execution throttle to a suitable queue depth for your environment, for example, a value of 100. For more information about the execution throttle and queue depth, see QLogic documentation.

If you are using Native Multipathing Plug-In (NMP), set **Enable Target Reset** to No. See Table 26 on page 142 to include the required parameters for the registry key.

Table 26. Registry key parameters for QLogic models

Key	Required parameters
HKEY_LOCAL_MACHINE > SYSTEM > CurrentControlSet > Services > ql2*xxx* > Parameters > Device > DriverParameters	**Buschange=0;FixupInquiry=1** **Note:** If you are using QLogic driver version 9.1.2.11 or higher, **Buschange** cannot be set to zero. See your device driver documentation for details.

Configuring the Emulex HBA for hosts that run the Microsoft Hyper-V operating system

After you install the Emulex host bus adapter (HBA) and the driver on hosts that are running the Microsoft Hyper-V operating system, you must configure the host bus adapter.

Before you begin

For the Emulex HBA StorPort driver, accept the default settings and set topology to 1 (1=F_Port Fabric).

Known issues and limitations of using Microsoft ODX on Microsoft Hyper-V hosts

Ensure that you are familiar with the Microsoft Hyper-V host attachment requirements to verify and troubleshoot your configuration.

ODX is disabled by default so that you can enable the MPIO feature. After enabling the MPIO feature, you can enable the ODX by entering the **chsystem -odx on** CLI command.

Although Microsoft ODX is supported, there is a limitation from the Windows side for supporting ODX online. The volumes that are mapped to the host are not *ODX capable* in Windows until they are unmapped and then mapped again.

To unmap and map the volumes, follow these steps:

1. Bring down the applications by using the volumes so that the applications stop using the volumes.
2. Disconnect or logout the host from the storage.
3. Rediscover the volumes from the Server Manager (the volumes go away).
4. Reconnect or log in the host to the storage.
5. Rediscover the volumes from the Server Manager (the disks reappear).
6. Start the applications.

Note: If these LUNs are formatted with an allocation unit size that is less than 4 KB, ODX might not work on these LUNs. A subpage-aligned I/O (less than 4 KB) is not allowed.

You can add new LUNs to a Windows host for consumption of an ODX-aware application. The system array works best with ODX when LUNs that are exported to the Windows host are formatted with 32 KB NTFS cluster size. Therefore, ensure that the allocation unit size for the New Technology File System (NTFS) is greater than or equal to 32 KB. The default on the Windows Server 2012 release 2 is 4 KB, which might not result in the performance that the system can deliver.

Citrix XenServer hosts

Fibre Channel attachments between the system and a Citrix XenServer host require that you follow certain instructions and guidelines.

Attachment requirements for Citrix XenServer hosts

Before you attach the system to a Citrix XenServer host, you must be familiar with the requirements of the host.

To successfully attach the host, the following requirements must be met:

- Ensure that you have the documentation for your host and that you have access to the Installing section in this product information. All system publications are available from the following website: www.ibm.com/support.
- Ensure that a supported Citrix XenServer version is installed.

The following website provides the system requirement information for the Citrix XenServer host: www.citrix.com.

Note: From the **Citrix** home page, follow this path: **Products & Solutions** > **XenServer/Essentials** > **Essentials for XenServer** > **Specifications**.

Host bus adapters for Citrix XenServer hosts

The Citrix XenServer hosts must use the correct host bus adapters (HBAs).

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Drivers and firmware for Citrix XenServer hosts

You must use the correct HBA device driver and firmware levels for your Citrix XenServer hosts.

The HBA drivers are installed when you install the Citrix XenServer software. No external installation of the driver or firmware is required for an HBA.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Configuring the Citrix XenServer host

You must configure the Citrix XenServer host before you can use the host with the system.

The following Citrix XenServer website provides the most current information about configuring the Citrix XenServer:

www.citrix.com

Linux operating system

Linux virtual machines (VMs) are always supported in a paravirtualized mode.

Windows operating system

Windows VMs are always supported as hardware-based virtual machines (HVMs) with paravirtualized drivers for storage and network. The drivers are installed from the Citrix XenServer installation CDs.

If you plan to use Windows VMs running under Citrix XenServer, ensure that the server processors have hardware virtualization support, such as Intel Virtualization Technology (VT)-based processors or AMD-V-based processors.

Multipath support for Citrix XenServer hosts

The multipath driver is installed as part of the Citrix XenServer software.

The following website provides the most current information about the IBM configuration file (`multipath.conf`) for the system:

```
www.ibm.com/support
```

To find the appropriate configuration file, complete the following steps:

1. Search for the following term: `subsystem device driver for Linux`
2. Scroll down the list until you find the device-mapper multipath configuration file for the system.
3. Select and download the file for Red Hat Enterprise Linux 5.

Clustering support on Citrix XenServer hosts

Clustering for guest operating systems either within or across Citrix XenServer hosts, or clustering of Citrix XenServer hosts is not supported.

Zoning and portsets

Ensure that you are familiar with these zoning details and host zones. More details are included in the SAN configuration and zoning rules summary.

Paths to hosts

The number of paths through the network from the node canisters to a host must not exceed eight. Configurations in which this number is exceeded are not supported.

Portset avoids wrong zoning configuration and limits the number of paths though which hosts can access external storage devices. Zoning host ports to too many ports of external system can create redundant paths. Also, unnecessary host login to too many Fibre Channel ports of external system consumes resources and leads to non-uniform distribution of host login on ports. Portsets limit the number of paths that the host can access to external storage system and distributes the host login evenly across multiple ports. For more information on Fibre Channel portsets, refer to the **Fibre channel portset**.

Fibre Channel portset is mapped with Fibre Channel port and host object. The host port WWPNs must be zoned only to external system Fibre Channel ports that are associated with the portset. If host ports are zoned to any other Fibre Channel port, the system notifies non-porset member login. Such logins can be removed by switching zones. An administrator can identify such logins by using the **lsfabric** command. The **mkhost** command shows host login counts on ports and enable the administrator to identify less loaded portsets. The command also shows the WWPNs that an administrator can use with host ports for zoning.

Adding WWPN to portsets provides greater flexibility and can be used for specific functional use cases. For example, SCSI host WWPN on a Fibre Channel port can be added to Portset1, whereas NVMeFC host WWPN on that port can be added to different portset.

To find the worldwide port names (WWPNs) that are required to set up Fibre Channel zoning with hosts, use the **lstargetportfc** command. This command also displays the current failover status of host I/O ports.

Portset avoids wrong zoning configuration and limits the number of paths though which hosts can access external storage devices. Zoning host ports to too many ports of external system can create redundant paths. Also, unnecessary host login to too many Fibre Channel ports of external system consumes resources and leads to non-uniform distribution of host login on ports. Portsets limit the number of paths that the host can access to external storage system and distributes the host login evenly across multiple ports. For more information on Fibre Channel portsets, refer to the **Fibre channel portset**.

Host zones

The configuration rules for host zones are different depending upon the number of hosts that access the system. For configurations of fewer than 64 hosts per system, the system supports a simple set of zoning rules that enable a small set of host zones to be created for different environments. For configurations of more than 64 hosts per system, the system supports a more restrictive set of host zoning rules.

Zoning that contains host HBAs must ensure host HBAs in dissimilar hosts or dissimilar HBAs are in separate zones. Dissimilar hosts means that the hosts are running different operating systems or are different hardware products; thus different levels of the same operating system are regarded as similar.

You can map each volume that you create on the system to multiple HBA Fibre Channel ports in a specific host. There can also be multiple paths across the SAN. For these reasons, each host must run multipathing software, such as the Microsoft Device Specific Module (MSDSM). The multipathing software manages the many paths that are available to the volume and presents a single storage device to the operating system. The system supports various multipathing software packages. The specific multipathing software that the system supports depends on the host operating system that you use the software with.

The number of paths through the network from the system nodes to a host must not exceed eight. Configurations that exceed this number are unsupported.

- Each system node has multiple ports and each I/O group has two system nodes. Therefore, without any zoning, the number of paths to a volume is the number of system ports in the I/O group times the number of host ports.
- This rule exists to limit the number of paths that must be resolved by the multipathing device driver.

If you want to restrict the number of paths to a host, zone the switches so that each HBA port is zoned with one system port for each node in the clustered system. If a host has multiple HBA ports, zone each port to a different set of system ports to maximize performance and redundancy.

To obtain the best overall performance of the system and to prevent overloading, the workload to each port must be equal. This can typically involve zoning approximately the same number of host Fibre Channel ports to each Fibre Channel port.

Systems with fewer than 64 hosts:

For systems with fewer than 64 hosts that are attached, zones that contain host HBAs must contain no more than 40 initiators, including the ports that act as initiators. A configuration that exceeds 40 initiators is not supported. A valid zone can be 32 host ports plus 8 ports. When it is possible, place each HBA port in a host that connects to a node into a separate zone. Include exactly one port from each node in the I/O groups that are associated with this host. This type of host zoning is not mandatory, but is preferred for smaller configurations.

Note: If the switch vendor recommends fewer ports per zone for a particular SAN, the rules that are imposed by the vendor take precedence over system rules.

Systems with more than 64 hosts:

Each HBA port must be in a separate zone and each zone must contain exactly one port from each node in each I/O group that the host accesses.

Fibre Channel port-name identification

The locations of port names are identified for Fibre Channel connections.

About Fibre Channel port-name identification

The format and content of the Fibre Channel (FC) port identifier are determined by the manufacturer of the link control facility for the applicable Fibre Channel port. The identifier is an eight-byte field, which the Fibre Channel protocols use to uniquely identify the Fibre Channel port.

The worldwide port name (WWPN) consists of 16 hexadecimal characters (0 - 9 and A - F). The system uses the WWPN to uniquely identify the Fibre Channel host interface adapter that is installed in your host. The system automatically finds the WWPN for your host Fibre Channel host interface adapter when you attach your host to the system.

Note: Each Fibre Channel port has a unique WWPN. There are 1 or more ports per host interface adapter, and therefore 1 or more WWPNs per host interface adapter. Each host interface adapter port has a unique WWPN that must be added to the host object.

Locating the WWPN for an IBM System p, eServer, or an RS/6000 AIX host

You can locate the WWPN for an IBM System p, eServer, or an RS/6000 AIX host by completing the following steps:

1. Log in as root.
2. Type `lscfg -vl fcs`*x*, where *x* is the adapter number.

   The network address is the Fibre Channel adapter port WWPN value.

Locating the WWPN for a host that is running the Linux operating system

You can locate the WWPN for a host that is running the Linux operating system with a QLogic adapter by completing the following steps.

Procedure

1. Restart the server.
2. Press Alt+Q to get the **FAST!Util** menu.

   If you have more than one Fibre Channel host bus adapter (HBA) installed, all the Fibre Channel HBAs are displayed. Scroll down to the adapter you want. Press Enter.
3. From the **FAST!Util** menu, scroll down and select **Select Host Adapter**.

4. Scroll up and highlight **Configuration Settings**. Press Enter.
5. From the **Configuration Settings** menu, click **Host Adapter Settings**.
6. Write down the 16-digit alphanumeric string that is displayed.

Locating the WWPN for a host running the Microsoft Windows operating system

Determining the WWPN of a host that runs a Windows operating system depends on the type of HBA in your host server.

For QLogic, you can use the SANsurfer GUI/IBM FAStT Management Suite Java™ (MSJ) if you have it, or restart the host and enter Ctrl+Q to open the QLogic BIOS, where you can find the HBA WWPNs.

For Emulex hosts, use the elxcfg tool that is packaged with the firmware. This tool opens in the Windows operating system and does not require a restart.

Locating the WWPN for an Oracle SPARC host

You can locate the worldwide port name (WWPN) for an Oracle SPARC host by following the steps in this topic.

Procedure

1. After you install the adapter and you restart the host system, view the `/var/adm/messages` file.
2. Search for the line that contains the applicable phrase for your host bus adapter (HBA):
   a. For the JNI SBUS HBA, search for `fcaw`*x*`: Fibre Channel WWNN`, where *x* is the adapter number (0, 1, and so on). You can find the WWPN on the same line immediately after the worldwide node name (WWNN).
   b. For the JNI PCI HBA, search for `fca-pci`*x*`: Fibre Channel WWNN`, where *x* is the adapter number (0, 1, and so on). You can find the WWPN on the same line as the WWNN.
   c. For the QLogic QLA2200F HBA, search for `qla2200-hba`*x*`-adapter-port-name` where *x* is the adapter number (0, 1, and so on).

Locating the WWPNs for a host that is running a VMware operating system

You can locate the worldwide port names (WWPNs) for a host that is running a VMware operating system.

Procedure

To locate the WWPNs for the host, complete the following steps:

1. Open the VMware Management Interface and click the **Options** tab.
2. Select **Storage Management**.
3. In the new window, click the **Adapter bindings** tab.

   The WWPN is listed at the end of each port heading line, which is the lines that start with `vmhba`. For example, in the following line, **`21:00:00:E0:8B:1A:E4:C6`** is the WWPN of the HBA port:

```
vmhba0: QLogic Corp QLA231x/2340 (rev 02) (21:00:00:E0:8B:1A:E4:C6)
```

Locating the WWPN for an SGI Origin host

You can locate the WWPN for an SGI Origin host that is running the IRIX operating system with a QLogic adapter by following the steps.

1. Restart the server.
2. Type the `scsiha -w [bus_number | device]` command.

For example, type `scsiha -w 6 7 8 9`. shows an example of the command output.

```
# scsiha -w 6 7 8 9
6 Portname: 210000e08b05d207
7 Portname: 210000e08b04d539
8 Portname: 210000e08b050808
9 Portname: 210000e08b038fe6
#
```

Queue depth in Fibre Channel hosts

In Fibre Channel networks, the queue depth is the number of I/O operations (SCSI commands) that can be run in parallel on a device.

Usually a higher queue depth can lead to better performance. However, when too many concurrent I/Os are sent to a storage device, the device responds with an I/O failure message of queue full. This message is intended to cause the host to try the I/O again a short time later. However, not all operating systems correctly handle the queue-full failure, which can cause unnecessary I/O failures or delays to applications.

To avoid these delays or failures, configure your hosts so that they send only a finite number of I/Os to the storage system to avoid exhausting the resources of the storage system.

In configurations with multiple hosts, configure each host with a similar queue depth to maintain fairness between the hosts.

- For most hosts, set the HBA queue depth to 32.
- For hosts that are significantly busier or where not many hosts are configured on the storage system, set the HBA queue depth to 128.

Ensure that the queue depth total for all hosts that are connected to a single physical Fibre Channel port on the storage system does not exceed 2048.

If host I/O queue full failures regularly occur, consider reducing the queue depths or distributing the host zoning across more physical Fibre Channel ports on the storage system.

iSCSI Ethernet host attachment

iSCSI Ethernet hosts with different operating systems can be attached to the system.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

iSCSI overview

iSCSI is an IP-based standard for transferring data that supports host access by carrying SCSI commands over IP networks. The iSCSI standard is defined by RFC 3720.

Connections from iSCSI-attached hosts to system nodes are supported. iSCSI connections from system nodes to Storwize family, IBM Spectrum Accelerate, and Dell EqualLogic systems that are used as external storage systems are also supported.

Table 27 on page 148 shows that iSCSI and Fibre Channel terms have analogous components.

Table 27. Comparison of iSCSI and Fibre Channel components

iSCSI components	**Fibre Channel components**
iSCSI host bus adapter	Fibre Channel host bus adapter
Network interface controller (NIC) and iSCSI software initiator	Fibre Channel host bus adapter
IP switch	Fibre Channel switch
IP router	–

Table 27. Comparison of iSCSI and Fibre Channel components (continued)	
iSCSI components	**Fibre Channel components**
iSCSI name, such as IQN (iSCSI qualified name) or EUI (extended-unique identifier)	WWNN (worldwide node name)

iSCSI initiators and targets

In an iSCSI configuration, the iSCSI host or server sends requests to a node. The host contains one or more initiators that attach to an IP network to initiate requests to and receive responses from an iSCSI target. Each initiator and target are given a unique iSCSI name such as an iSCSI qualified name (IQN) or an extended-unique identifier (EUI). An IQN is a 223-byte ASCII name. An EUI is a 64-bit identifier. An iSCSI name represents a worldwide unique naming scheme. This scheme identifies each initiator or target in the same way that worldwide node names (WWNNs) are used to identify devices in a Fibre Channel fabric.

iSCSI targets are devices that respond to iSCSI commands. An iSCSI device can be an end node such as a storage device, or it can be an intermediate device such as a bridge between IP and Fibre Channel devices. Each iSCSI target is identified by a unique iSCSI name. The system can be configured as one or more iSCSI targets.

To transport SCSI commands over the IP network, an iSCSI driver must be installed on the iSCSI host and target. The driver is used to send iSCSI commands and responses through a network interface controller (NIC) or an iSCSI HBA in the host or target hardware.

For maximum performance, use a gigabit Ethernet adapter that transmits 1000 megabits per second (Mbps) for the connection between the iSCSI host and the iSCSI target.

iSCSI host connection options

Figure 12 on page 149 shows an iSCSI host that connects to the system over an Ethernet network.

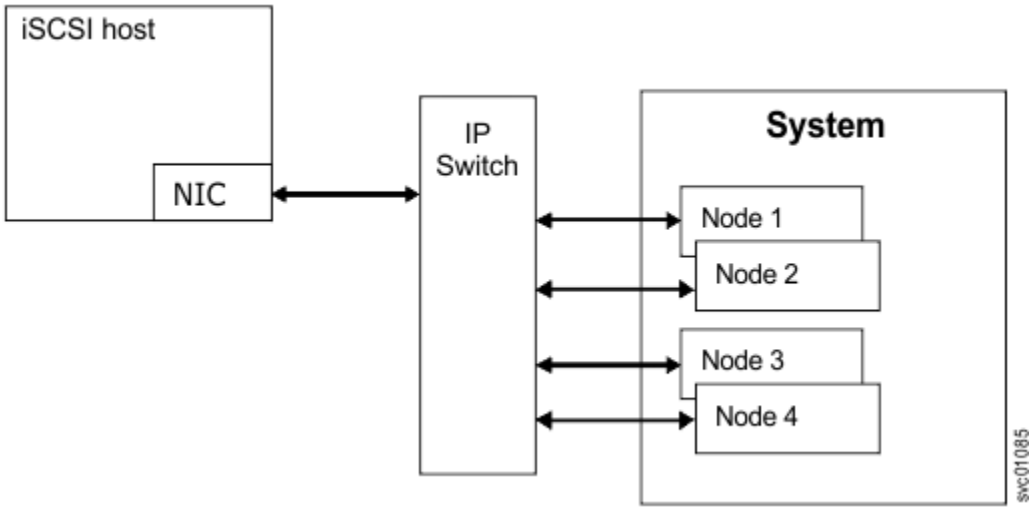


Figure 12. Transmitting SCSI over TCP/IP

iSCSI multisession support

Multisession support, or host-level multipath support, provides multiple paths between iSCSI initiators and targets, which is useful for high availability and load balancing.

Some products require that an iSCSI initiator login with a specific IP address of the target versus logging in with a target name. Logging in with a target name in those products causes the iSCSI initiator to log in to all of the IP addresses in the target and causes session reinstantiation that results in session loss of the previous login.

The system offers single subnet multi-session configurations and multiple and dual subnet multisession configurations that remove the restriction against logging in to a target name. iSCSI initiators use Internet storage name service (iSNS) servers to log in, which enables the discovery of the iSCSI targets and their IP addresses.

- An iSCSI session is a TCP relationship between an iSCSI initiator node port and an iSCSI target node port. When established, iSCSI control, data, and status messages are communicated over the session.
- Each session is identified with multiple iSCSI initiator names and the target name plus the iSCSI initiator session ID (ISID) plus the target portal group tag (target-side ID).
- The system multi-session approach is to have relationships between multiple initiator instances (separate ISIDs) to a single target portal group in an iSCSI target. This approach is similar to having multiple SCSI initiator ports from the same iSCSI initiator device have relationships with a single SCSI target port.

iSCSI storage support

The system supports iSCSI connections to Storwize family and Dell EqualLogic systems that are used as external storage systems. For specific configuration information, see "iSCSI configuration details for host connections" on page 152.

Discovery and path configuration for iSCSI controllers is done as follows:

- Specify only a target IP and issue a discovery request to the iSCSI target controller
- The target controller returns a list of all configured iSCSI qualified names (IQNs) on that controller
- Establish connections to the discovered target IQNs

Relationship of WWPNs and iSCSI names for host objects

A host can be created with worldwide port names (WWPNs) or iSCSI names. The WWPN name space and the iSCSI name space within the system share the same internal system resources.

The iSCSI name in a host object can take up to a maximum equivalent of four WWPNs depending on the number of I/O groups in which the host participates. This affects the maximum number of hosts that you can configure in a system.

- Creating an iSCSI host object with one iSCSI name in one I/O group is equivalent to creating a Fibre Channel host with one WWPN.
- Creating an iSCSI host object with two iSCSI names in one I/O group is equivalent to creating a Fibre Channel host object with two WWPNs.

Example iSCSI configurations

Configure your system in a network that uses Internet Small Computer System Interface (iSCSI).

This figure shows a two-node clustered system that is connected to a single subnet. Each node has two Ethernet ports, each of which is used for iSCSI data transfers. One node in the system also acts as the system configuration node. In this example, port 1 on the configuration node provides the system management IP interface.

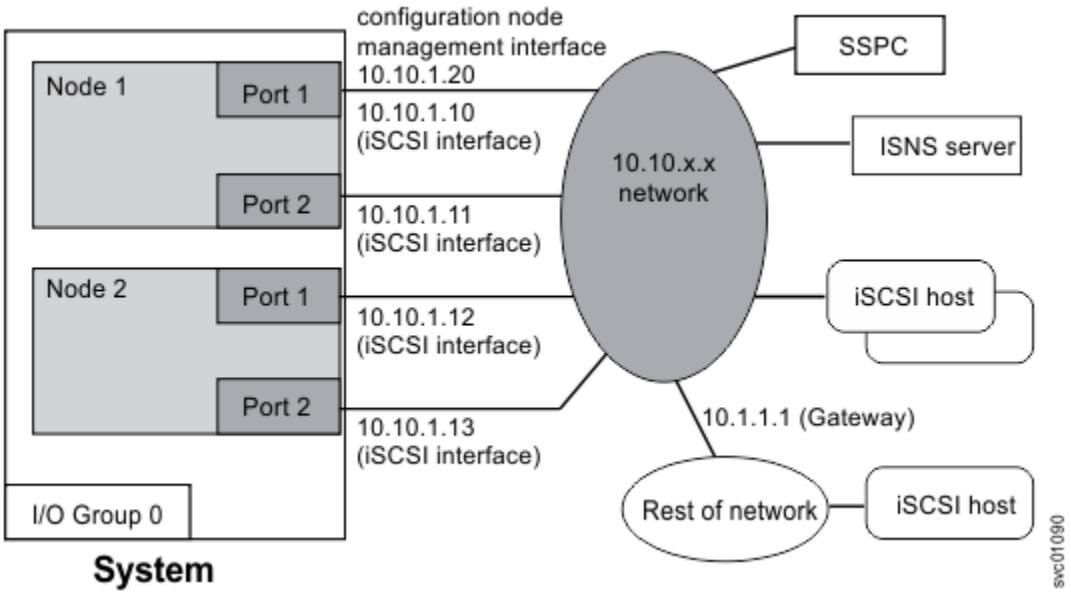


Figure 13. iSCSI configuration with a single subnet

This figure shows a two-node system that is connected to multiple subnets. Each node has two Ethernet ports (port 1 and port 2) that are connected to different IP subnets. One node in the system also acts as the configuration node. This configuration provides alternate IP interfaces on different subnets for the system management interface.

Note: With a multiple subnet configuration, if one subnet fails and the iSNS server is part of that subnet, the nodes and hosts are not able to access the iSNS server. The I/O is not affected and the other subnet is still active.

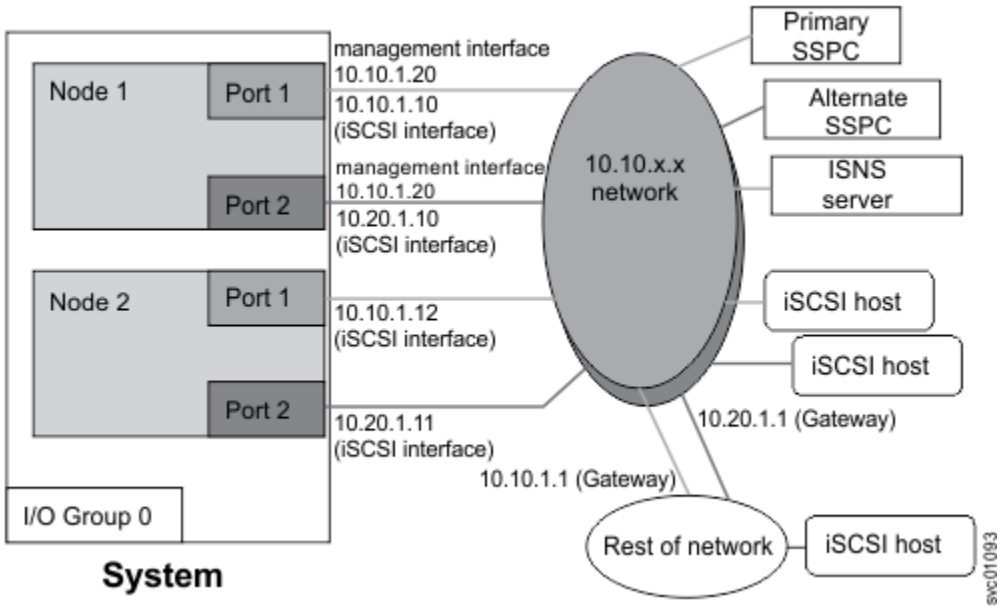


Figure 14. iSCSI configuration that uses multiple subnets and provides alternative configuration interfaces

This figure shows the same two-node system as Figure 14 on page 151. However, it illustrates the benefits of using multipathing and redundant host interface adapters in the host systems.

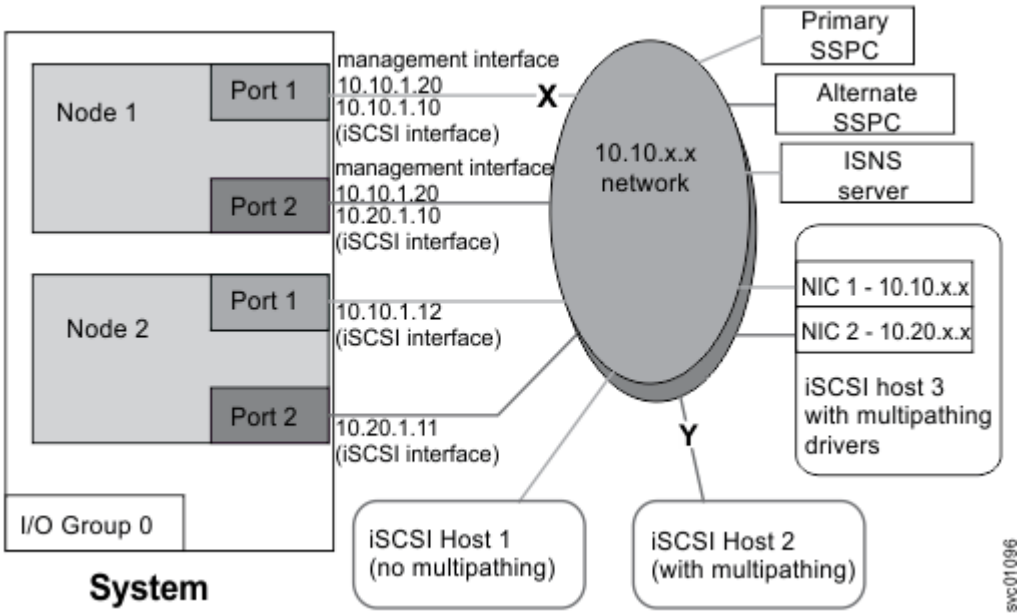


Figure 15. iSCSI configuration with host multipathing

In this example, host 1 does not use multipathing. A volume in the system's I/O group appears as four separate devices in host 1. The host selects one device to perform I/O to the volume, which corresponds to a particular IP address at a system node port, 10.10.1.10. If a connection between the host and this system port is broken (the link at X is broken), an I/O error is recorded on host 1 for that volume if I/O is in progress. No system state changes or IP failover occurs.

Host 2 uses multipathing. A volume in the system I/O group appears as a single device to the applications on host 2, even though the multipathing driver can detect four separate devices for each volume. The multipathing driver selects one or more of these devices during I/O. If the connection between the host and one system node port is lost, the multipathing driver can select an alternative path to the system I/O group. The I/O between the host and system continue without error. However, host 2 has only one host interface adapter. Therefore, host 2 reports I/O errors (such as the link at Y is broken) if the connection between that host interface adapter and the network is lost.

Host 3 uses multipathing and redundant host interface adapters. If a host interface adapter fails, the multipathing driver can still find paths from the host to a volume in the system I/O group, and the application I/O can continue without error. Because the host interface adapters are connected to different IP networks, the overall configuration can tolerate a single network failure without I/O errors occurring on host 3.

Multipathing drivers are not required to do system maintenance when system nodes are removed or replaced in an I/O group. However, multipathing host drivers are required for load balancing and for surviving host interface adapter, link, or network failures.

iSCSI configuration details for host connections

Follow these configuration details for iSCSI host connections.

Attach the system to Internet Small Computer Systems Interface (iSCSI) hosts by using the Ethernet ports on the systems.

Note: The system supports SAN devices that bridge iSCSI connections into a *Fibre Channel* network.

iSCSI connections route from hosts to the systems over the LAN. Follow these configuration rules:

- There is no SLP support for discovery.
- Header and data digest support is provided only if the initiator is configured to negotiate.
- Only one connection per session is supported.

- Only ErrorRecoveryLevel 0 (session restart) is supported.

iSCSI hosts connect to the system through the node-port IP addresses. During node failure, in addition to node-port IP addresses, the iSCSI name, and iSCSI alias for the failed node are also transferred to the partner node. After the failed node recovers, the node-port IP address and the iSCSI name and alias are returned to the original node.

Note: The 100 Gbps adapter supports iSCSI. However, the performance is limited 25 Gbps per port.

IP requirements for iSCSI

Note: For general IP requirements, see "iSCSI Ethernet host attachment" on page 148.

IP addresses are used to discover storage volumes and to use I/O to access the volumes. Node iSCSI IP addresses are also used to access a remote Internet Storage Name Service (iSNS) server, if configured.

- Each node Ethernet port can be configured on the same subnet with the same gateway, or you can have each Ethernet port on separate subnets and use different gateways.
- If you are configuring a system to use node Ethernet ports 1 and 2 for iSCSI I/O, ensure that the overall configuration also meets the system IP requirements that are listed previously.
- To help ensure iSCSI IP failover operations, nodes in the same I/O group must be connected to the same physical segments on the same node ports. However, you can configure node Ethernet ports in different I/O groups to use different subnets and different gateways.

You can configure iSNS to facilitate scalable configuration and management of iSCSI storage devices. Currently, you can have only one type of protocol that is used by the iSNS server at a time: either IPv4 or IPv6. For example, if you try to configure an IPv6 iSNS IP address when you already configured an IPv4 iSNS IP address, the new IPv6 IP address becomes the iSNS IP address. The old IP address can no longer be used for iSNS function.

iSCSI authentication

iSCSI supports two types of authentication through the Challenge Handshake Authentication Protocol (CHAP):

1. **One-way authentication**: The iSCSI target authenticates the iSCSI initiators.
2. **Two-way (mutual) authentication**: The iSCSI target and the iSCSI initiators authenticate each other.

Attention: With the iSCSI initiator, you can set two passwords: one for discovery and another for iSCSI session I/O. However, the system requires that both passwords for each type of authentication are the same. That is, two identical passwords for one-way CHAP, and two identical passwords for two-way CHAP that are different from those passwords for one-way CHAP.

Host mapping

Use this information for specific iSCSI host mapping information. For additional general host mapping information, see "Ethernet connectivity" on page 297.

iSCSI hosts that have a multipath driver that is installed can be moved in a nondisruptive manner.

For more information about host mapping, see "Host mapping" on page 107.

iSCSI protocol supported session parameters

The following Ethernet session parameters are supported:

```
initial_r2t = 1
immediate_data = 0
max_connections = 1
Max_recv_segment_data_length = 32k
max_xmit_data_length = 32k
max_burst_length = 32k
first_burst_length = 32k
default_wait_time = 2
```

```
default_retain_time = 20
max_outstanding_r2t = 1
data_pdu_inorder = 1
data_sequence_inorder = 1
error_recovery_level = 0
header_digest = CRC32C,None
data_digest = CRC32C,None
ofmarker = 0
ifmarker = 0
ofmarkint = 2048
ifmarkint = 2048
```

Configuring the system for iSCSI hosts

You need to complete several tasks to configure the system to work with iSCSI-attached hosts. The tasks include general tasks on the host system before you configure a system.

Because the system supports various host machines, consult the documentation for specific instructions and requirements for a particular host. For a list of supported hosts, see this website: www.ibm.com/support

To configure a system for supported Ethernet connections, follow these general tasks on the host system:

Port IP addresses are the IP addresses that are used by iSCSI-attached hosts to process I/O. Host port group IDs are automatically assigned to the port. Host port grouping groups the ports that have the same maximum port speed and helps ensure that no more than four ports are discovered by a host.

1. To configure a new port IP address to a specified Ethernet port of a node with an IPv4 or IPv6 address, enter the following command-line interface (CLI) command:

   ```
   mkip -node node_name | node_id -ip ipaddr
   -gw gateway -prefix prefix -vlan vlan_id -port port_id -portset portset_id
   ```

   where *node_name* | *node_id* specifies the name or ID of the node that is being configured, *ipaddr* is the IPv4 or IPv6 address for the Ethernet port, *gateway* is the IPv4 or IPv6 gateway IP address, *prefix* is subnet prefix for IPv4 or IPv6, *port_id* specifies the Ethernet port ID (1 or 2) and *portset_id* specifies the portset ID or name. To view a list of ports, use the **lsportethernet** command. Also, **mkip** has *-shareip* option to share IP address between multiple portset.

   The optional **-vlan** parameter sets the virtual local area network (VLAN) ID for an IP address that is configured for iSCSI host attachment.

   For more information about the **-vlan** parameter, see the information about configuring VLAN for iSCSI using the CLI.

2. To remove an IP address from a node Ethernet port, enter either of these CLI commands.

   The following command deletes an IP configuration for the specified Ethernet port:

   ```
   rmip ip_address_id
   ```

   To list the valid values for the IP address ID, enter the **lsip** command.

After you configure your IP addresses, you can optionally create or configure several iSCSI items.

Configuring or modifying an iSCSI alias by using the CLI

You can use the command-line interface (CLI) to optionally create or change the iSCSI alias for the selected node. An iSCSI alias is a user-assigned name that identifies a node to the iSCSI-attached host.

To configure or modify an iSCSI alias, follow these steps:

1. To specify that the name or iSCSI alias that is being set is the name or alias of the node in the I/O group, enter the following CLI command:

   ```
   svctask chnode -iscsialias alias node_name | node_id
   ```

   where *alias node_name* | *node_id* specifies the name or ID of the node.

2. To specify that the name or iSCSI alias that is being set is the name or alias of the partner node in the I/O group, enter the following CLI command.

   When a system does not have a partner node, the values set are applied to the partner node when it is added to the system. If this parameter is used when there is a partner node, the name or alias of that node changes.

   ```
   svctask chnode -iscsialias alias -failover node_name | node_id
   ```

   where *alias* specifies the iSCSI name of the node and *node_name* | *node_id* specifies the node to be modified.

After you create iSCSI aliases, you can optionally configure the address for the internet Storage Name Service (iSNS) server for the system.

Configuring the iSNS server address with the CLI

If you are using Ethernet-attached hosts with a system, you can use the command-line interface (CLI) to optionally configure the address for the internet Storage Name Service (iSNS) server for the system. Host systems use the iSNS server to manage targets and for discovery.

1. To specify an IPv4 address for the iSCSI storage name service (SNS), enter the following CLI command:

   ```
   chsystem -isnsip sns_server_address
   ```

   where *sns_server_address* is the IP address of the iSCSI storage name service in IPv4 format.
2. To specify an IPv6 address for the iSCSI storage name service (SNS), enter the following CLI command:

   ```
   chsystem -isnsip_6 ipv6_sns_server_address
   ```

   where *ipv6_sns_server_address* is the IP address of the iSCSI storage name service in IPv6 format.

Configuring system iSCSI authentication with the CLI

You can use the command-line interface (CLI) to configure the system to authenticate with iSCSI-attached hosts by using the Challenge-Handshake Authentication Protocol (CHAP). After the CHAP is set for the system, all attached hosts must be configured to authenticate. When you are troubleshooting a problem, you can delay your configuration of the CHAP authentication until after you configure the first one or two hosts and test their connectivity.

To configure authentication between the system and the iSCSI-attached hosts, follow these steps:

1. To configure one-way CHAP authentication for an iSCSI host, enter the following CLI command:

   ```
   chhost -iscsiusername iscsi_username -chapsecret chap_secret host_name
   ```

   where *iscsi_username* is the username, *chap_secret* is the CHAP secret to be used to authenticate the system, and *host_name* is the name of the host. The *chap_secret* value must be 12 characters. If you do not specify the username, the initiator's IQN is taken as the username for one-way CHAP authentication.
2. To set the two-way CHAP authentication method for the Ethernet-based communications of the system, enter the following CLI command:

   ```
   chsystem -iscsiauthmethod chap -chapsecret chap_secret
   ```

   where *chap* specifies that CHAP is the authentication method and *chap_secret* is the CHAP secret to be used. The specified CHAP secret cannot begin or end with a space.

   Note: The two-way CHAP authentication username must match with `clustername` which can be found by **`lssystem`** command.

3. To clear all CHAP secrets for authentication that were previously set, enter the following CLI command:

```
chsystem -nochapsecret
```

   if the **chapsecret** parameter is specified, the **nochapsecret** parameter is not allowed.
4. Run the **lsiscsiauth** command to display the Challenge Handshake Authentication Protocol (CHAP) secret that you configured.

After you configure the CHAP secret for the system, ensure that the system CHAP secret is added to each iSCSI-attached host. On all iSCSI-attached hosts, specify a CHAP secret that the hosts use to authenticate to the system.

After you configure the iSCSI host object, use the **lshost** CLI to display the status of the configured host. An iSCSI host can have one of the following statuses:

Online
The host is logged in to both FlashSystem nodes and has a vdisk mapped, which indicates full connectivity and vdisk access.

Degraded
The host is logged in to one or both FlashSystem nodes without a vdisk mapping, which limits access to vdisks and results in degraded functionality.

Offline
The host has no active sessions with either node, which results in no connectivity. The vdisk mapping does not apply in this state.

iSCSI support (considerations)

Be aware of the following general iSCSI support considerations.

The general iSCSI parameters are applicable regardless of the operating system.

- The system supports:
  - The standard iSCSI port (3260)
  - For more information on hosts per I/O group and sessions per I/O group, see Configuration limits.
  - I/O from Fibre Channel and iSCSI initiators in different hosts to the same volumes.

    Note: The system does not support I/O from Fibre Channel and iSCSI initiators in the same hosts to the same volumes.
  - Multiple sessions from iSCSI hosts, with a maximum of four sessions from one iSCSI host to each system iSCSI target.
  - One-way CHAP authentication, where you can specify an authentication username with a maximum of 31 ASCII characters, or you can use the initiator's IQN as the username .
  - BI-CHAP authentication with initiators that accept a blank username field.
  - CHAP algorithms MD5, SHA1, SHA2, and SHA3-256 for one-way and two-way CHAP authentication.
  - A large number of sessions per host can cause multiple issues such as iscsid soft lockup or OOM Killer on the host side. To avoid these issues, it is recommended to have a maximum 512 sessions per host.
- iSCSI uses either iSCSI qualified name (IQN) (223 bytes) or extended unique identifier (EUI) (64-bit) names.

  Note: Ensure that the IP takeover facility in an I/O group is enabled. Then, if the node that is acting as the iSCSI target fails, the partner node takes over the IP addresses of the failed node, thus continuing operations. During takeover, the iSCSI initiator is logged out from the failed node. A new session or login is reestablished with the partner (working) node that uses the IP address of the failed node.
- Each iSCSI target can support both IPv4 and IPv6 concurrently.

- Host operating systems such as Oracle Linux that have many Ethernet interfaces are subject to an Address Resolution Protocol (ARP) Flux problem. This problem might occur when a host replies to ARP requests for interfaces on the same or different subnet from any interface on that same or different subnet. Usually, this behavior is not a problem. However, in specific cases, ARP Flux generates unexpected behavior of applications due to an incorrect mapping between IPv4 addresses and MAC addresses.

  To avoid ARP Flux on Linux, use the following setting on the host:

  ```
  echo 1 > /proc/sys/net/ipv4/conf/all/arp_filter
  ```

  To make this behavior persistent, add a `net.ipv4.conf.all.arp_filter=1` entry in `/etc/sysctl.conf`.

  If you use other operating systems such as VMware ESX, check your settings to avoid ARP Flux on those hosts as well.
- Volumes are mapped to a host that uses the same host mapping mechanism as Fibre Channel attachment. A volume can be mapped to a Fibre Channel host or an iSCSI host. Mapping a volume through both iSCSI and Fibre Channel to the same host is not supported.
- All IP addresses (service and configuration) associated with a clustered-system Ethernet port must be on the same subnet. However, IP addresses associated with a node Ethernet port that is used for iSCSI traffic can be configured to belong to different subnets.
- A host can discover and login to only those IP addresses that are mapped to it via portsets.
- For a host to login to 'n' nodes of SVC, it must be mapped to a portset that contains IP addresses from all 'n' nodes.
- Unconfigured logins will now be rejected.
- A host that is associated with portset x, if tries to discover through portset y, discovery succeeds and will give IP addresses associated with portset x.
- Modifying IP addresses mapped to a portset results in login disruptions from hosts mapped to that portset.
- Enter the following command to disable **iWARP Port Mapper** (iwpmd) service on RHEL 7.5 and higher host machine:

  ```
  $ systemctl stop iwpmd
  ```

  Note: With the IBM Storage Virtualize software, the **iWARP Port Mapper** (iwpmd) service results in host system hang or restart.

Setting up the host server as an iSCSI initiator

Follow this procedure to set up a host server for use as an iSCSI initiator with the system. The specific steps vary depending on the particular host type and operating system that is involved.

To configure a host, select a software-based iSCSI initiator, such as the Microsoft Windows iSCSI software initiator.

You can use either the management GUI or the command-line interface to set up a host server. To set up your host server for use as an iSCSI software-based initiator with the system, complete the following steps:

Using the Management GUI

1. To set up your system for iSCSI, complete these steps:
   a. In the management GUI, select **Settings** > **Network** > **Ethernet Ports**.
   b. Expand the I/O group, right-click the node that you want to change, and select **Modify IP Settings**.
   c. On the **Modify IP Settings** page, enter the IP address for the node. You can select either IPv4 or IPv6 addresses. Click **Modify**.

d. Verify that you configured the node and the system Ethernet ports correctly. The **Ethernet Ports** page displays all the changed nodes as configured and **Host Attach** is set to **Yes**.

e. To create hosts, select **Hosts** > **Hosts** and select **Add Host**. Select **iSCSI Hosts** and enter an iSCSI initiator name in the iSCSI port field. Check the host server to determine the iSCSI initiator name. Enter more details about the host and click **Add Host**.

f. To create volumes, select **Volumes** > **Volumes** and select **Create Volumes**. Select the type of volume that you want to create and select the pool, size, and other details for the volume. Click **Create and Map to Host**.

g. After the volume is created, click **Continue** to map the host to the volume. Select the host that you create and click **Map Volumes**.

2. To set up your host server, complete these steps:

a. Ensure that you configured your IP interfaces on the server.

b. Install the software for the iSCSI software-based initiator on the server.

c. On the host server, run the configuration methods for iSCSI so that the host server iSCSI initiator logs in to the system and discovers the volumes. The host then creates host devices for the volumes

3. After the host devices are created, you can use them with your host applications.

Using the CLI

1. To set up your system for iSCSI, complete these steps:

a. Select a set of IPv4 or IPv6 addresses for the Ethernet ports on the nodes that are in the I/O groups that use the iSCSI volumes.

b. Verify that for those node Ethernet ports Host flag is enabled through the **lsportethernet** command. If they are not enabled, then enable it using **chportethernet** command.

c. Configure the node Ethernet ports on each node in the system with the **mkip** command.

d. Verify that you configured the node and the system Ethernet ports correctly by reviewing the output of the **lsip**, **lsportethernet** and **lssystemip** commands.

e. Use the **mkvdisk** command to create volumes on the system.

f. Use the **mkhost** command to create a host object on the system that describes the iSCSI server initiator to which the volumes are to be mapped.

g. Use the **mkvdiskhostmap** command to map the volume to the host object in the system.

2. To set up your host server, complete these steps:

a. Ensure that you configured your IP interfaces on the server.

b. Install the software for the iSCSI software-based initiator on the server.

c. On the host server, run the configuration methods for iSCSI so that the host server iSCSI initiator logs in to the system and discovers the volumes. The host then creates host devices for the volumes

3. After the host devices are created, you can use them with your host applications.

What to do next

Be aware of the following considerations:

- AIX does not support iSCSI with IPv6.
- Each connection between a host initiator and a system Ethernet port can use IPv4 or IPv6 (where supported). Concurrent use of both IPv4 and IPv6 for the same initiator to the system Ethernet port is not supported.
- When the host object on the system side is created or re-created for an iSCSI host, you must log out of any existing sessions from that host and then log back in.

- Host objects can be associated to different host attach portsets by using **mkhost** and **chhost** commands. Similarly, IP addresses can also be configured on different portsets by using **mkip** command. A host object can login through IP addresses belonging to same portset. It is recommended to read the limits and considerations for the portset.
- Host performance is reduced if you configure RTC with the iSCSI host attachment on the system because some CPU cores are dedicated to RTC processing, thus reducing the number of CPU cores available for iSCSI processing.

Linux iSCSI host attachment

You can use iSCSI Ethernet to attach a Linux host to the system.

Before you begin host attachment procedures:

- Read the information about setting up the host server.
- Ensure that you are familiar with the command-line interface (CLI) commands. Information about the CLI is available in this product information.

Installing the Linux software iSCSI initiator

To install the Linux software iSCSI initiator for iSCSI host attachment, set the iSCSI qualified name, target discovery, and authentication. Also, enable multipathing for the Linux hosts.

Before you begin

The system supports the use of the iSCSI-initiator software that is provided by your Linux distribution vendor. Ensure that you have the latest release of the initiator software that is installed on your host before you continue with the instructions.

The initiator software on RHEL systems is packaged as *iscsi-initiator-utils*. The initiator software on SLES systems is packaged as *open-iscsi*.

Procedure

1. Configure the iSCSI-initiator daemon by editing the `/etc/iscsi` file.

```
[root@R64SRV6 ~]# cd /etc/iscsi/

[root@R64SRV6 iscsi]# ls

initiatorname.iscsi  iscsid.conf

[root@R64SRV6 iscsi]#
```

2. Manage the software by using the **chkconfig** command and the **service** command. The service on RHEL systems is named *iscsid*. The service on SLES systems is named *open-iscsi*. Consider the examples for RHEL and SLES systems.

   For RHEL systems:

```
[root@R64SRV6 iscsi]# chkconfig --list | grep "iscsid"
iscsid          0:off   1:off   2:off   3:off   4:off   5:off   6:off

[root@R64SRV6 iscsi]# chkconfig iscsid on

[root@R64SRV6 iscsi]# chkconfig --list | grep "iscsid"
iscsid          0:off   1:off   2:on    3:on    4:on    5:on    6:off
[root@R64SRV6 iscsi]#
```

   For SLES systems:

   Yas1 and Yast2 provide an appropriate method to properly configure iSCSI services and initiator. From a command line for iSCSI services in SLES:

```
#/etc/init.d/open-iscsi start

#chkconfig open-iscsi on
```

3. Use the default settings for the iSCSI initiator. These settings are described in `/etc/iscsi/iscsid.conf`.

Setting the iSCSI qualified name for the Linux host

To correctly specify the iSCSI qualified name (IQN) for the Linux host, the initiator name that you choose must match the iSCSI name that is given to the host object for this system.

The iSCSI initiator IQN that you configure on the host side should never be the same as the target IQN on the system node.

1. Edit the `/etc/iscsi/initiatorname.iscsi` file to specify the IQN name. The following example shows a sample name:

```
InitiatorName= iqn.1986-03.com.ibm:iscsihost.0
```

2. After you change the initiator name, you must start the iSCSI-initiator service again.

Configuring more than one Linux host Ethernet port for iSCSI

If you want to configure more than one Linux host Ethernet port for iSCSI, methods are available to improve routing and maintain optimal redundancy in cases of dynamic discovery.

About this task

You can use one of the following two methods:

- Configure different subnet IP addresses for the host ports and the system ports. Each target IP of a system node is accessible only from one host IP of the corresponding subnet. As a result, pairs of host and system ports are in separate subnets.
- In the `/etc/iscsi/ifaces` subdirectory, create a separate *iface* configuration file for each network object that you want to bind a session to. Each iface config file must have a unique name that is less than or equal to 64 characters in length.

Procedure

The following *iface* example shows a simple case with two network devices (eth0 and eth1):

1. Edit the `/etc/iscsi/ifaces/eth0` file to specify the iface.net_ifacename:

```
iface.net_ifacename = eth0
```

2. Similarly, edit the `/etc/iscsi/ifaces/eth1` file:

```
iface.net_ifacename = eth1
```

iSCSI target discovery for Linux hosts

Two possible methods can be used to discover iSCSI target portals. The following list includes the methods:

- Send Targets
- Internet storage name service (iSNS)

After discovery, you use the **iscsiadm** command to log in to or log out of a specific portal on a discovered target.

Discovering iSCSI targets by using send targets

You can discover iSCSI target portals by using the send targets method. The **iscsiadm** command returns the IQN and target portals that are associated with the system node.

Run the discovery command:

```
iscsiadm --mode discovery --type sendtargets --portal x.x.x.x
```

where *x.x.x.x* is the IP address of a node Ethernet port on the system.

Tip: You can substitute `-m` for `--mode`, `-t` for `--type`, and `-p` for `--portal` in the discovery command.

Command output similar to the following example is displayed:

```
x.x.x.x:3260, 1 iqn.1986-03.com.ibm:2145.china5.hlnc111874
```

Discovering iSCSI targets on your Linux host by using iSNS

You can discover iSCSI target portals by using the internet storage name service (iSNS) method. Before you use the iSNS discovery, you must do the following tasks:

- Configure the system to use iSCSI.
- Configure the system to use your iSNS server.
- Verify that the system nodes registered their target portals with the iSNS server. If the node iSCSI addresses are visible to the iSNS server, this registration occurs automatically.

To discover iSCSI targets by using iSNS, complete the following steps:

1. Modify the iSNS server settings in the `/etc/iscsi/iscsid.conf` file by entering the IP address and the port number that is used by your iSNS server:

   ```
   isns.address = <iSNS server IP address>
   isns.port = <iSNS server port>
   ```

2. Make the iSNS configuration changes effective by restarting the iSCSI initiator service.
3. Run the iSNS discovery command to generate a list of all iSCSI target portals that are registered with the iSNS server:

   ```
   iscsiadm --mode discovery --type isns
   ```

Logging in to discovered targets

You can connect to a specific portal on a discovered target by using the **`iscsiadm`** command. The system supports only one iSCSI session between an initiator and a target. Ensure that you do not attempt to connect to the same target node more than once.

Log in the iSCSI initiator to your selected system targets by using the following appropriate command examples:

For SLES:

```
iscsiadm --mode node --target target_IQN --portal x.x.x.x --login
```

For example: `iscsiadm --mode node --target iqn.1986-03.com.ibm:2145.cluster10g1.node1 --portal 192.168.64.191 --login`.

For RHEL:

```
iscsiadm --mode node --targetname target_IQN --portal x.x.x.x --login
```

For example: `iscsiadm --mode node --targetname iqn.1986-03.com.ibm:2145.cluster10g1.node1 --portal 192.168.64.191 --login`.

When the login command completes, the host can access the volumes that are mapped to the host from the target after the SCSI devices are rescanned.

Logging out from targets

You can log out of a specific system target or you can log out of all established sessions.

Note: After you log out of a system target, the iSCSI volumes that it serves are no longer available to the host.

1. To log out of a specific system target, enter the following command:

```
iscsiadm --mode node --target <IQN> --portal x.x.x.x --logout
```

2. To log out of all your established sessions, enter the following command:

```
iscsiadm --mode node --logoutall=all
```

iSCSI timeout configuration settings for the system

You can set the timeout value for disks. The disk timeout is set in the `/sys/block/<device>/device/timeout` file. Depending on operating system and OS version, increase the default value of 30 or 60 seconds to 120 in most situations.

For example, if the device is `/dev/sdc`, you would use:

```
[root@R64SRV6 ~]#echo -n 60 > /sys/block/sdc/device/timeout
```

To increase the SCSI command timeout for the system, create the following *udev* rule so that it is not necessary to set the timeout manually for each mapped disk:

```
udev rules file
cat /etc/udev/rules.d/99-ibm-2145.rules

# Set SCSI command timeout to 120s (default == 30 or 60) for IBM 2145 devices
SUBSYSTEM=="block", ACTION=="add", ENV{ID_VENDOR}=="IBM",
ENV{ID_MODEL}=="2145", RUN+="/bin/sh -c 'echo 120 > /sys/block/%k/device/timeout'"
```

Reconfirm the settings after your restart the system. After you set up your volumes, use `multipath -ll | grep sd` to locate the block device paths. Also, check with `cat /sys/block/sd<x>/device/timeout` for 120 seconds, where <x> is a 2145 path. For example:

```
edam:~ # multipath -ll | grep sd | tail -1
  `- 10:0:0:4 sdl 8:176 active ready running
edam:~ # cat /sys/class/block/sdl/device/timeout
120
edam:~ #
```

Setting up authentication for Linux hosts

You can set up one-way CHAP authentication for Linux hosts. After you configure one-way authentication that is working for your host, you can optionally set up two-way authentication.

Before you begin

The system supports two Challenge Handshake Authentication Protocol (CHAP) methods:

- One-way CHAP authentication (the system authenticates the host iSCSI initiator).
- Two-way CHAP authentication (both the system and the initiator authenticate each other).

Note: CHAP secrets that you select for one-way authentication and two-way authentication must be different.

Procedure

To set up authentication for a Linux host, follow these steps:

1. Open `/etc/iscsi/iscsid.conf` or `/etc/iscsid.conf` by using an appropriate editor.
2. Go to the CHAP settings paragraph.

   The following example shows the output of CHAP settings for a Linux host:

```
#*************
#CHAP Settings
#*************

#To enable CHAP authentication set node.session.auth.authmethod
```

```
#to CHAP. The default is None.
#node.session.auth.authmethod = CHAP

#To set a CHAP username and password for initiator
#authentication by the target(s), uncomment the following lines:
#node.session.auth.username = username
#node.session.auth.password = password
node.session.auth.username = rhel_username
node.session.auth.password = xxxxxxxxxxxxx
#To set a CHAP username and password for target(s)
#authentication by the initiator, uncomment the following lines:
#node.session.auth.username_in = username_in
#node.session.auth.password_in = password_in
node.session.auth.username_in = clustername
node.session.auth.password_in = yyyyyyyyyyyy
#node.session.auth.chap_algs = SHA3-256,SHA256,SHA1,MD5
#To enable CHAP authentication for a discovery session to the target
#set discovery.sendtargets.auth.authmethod to CHAP. The default is None.
#discovery.sendtargets.auth.authmethod = CHAP
#To set a discovery session CHAP username and password for the initiator
#authentication by the target(s), uncomment the following lines:
#discovery.sendtargets.auth.username = username
#discovery.sendtargets.auth.password = password

#To set a discovery session CHAP username and password for target(s)
#authentication by the initiator, uncomment the following lines:
#discovery.sendtargets.auth.username_in = username_in
#discovery.sendtargets.auth.password_in = password_in
```

3. Set up authentication.
   - Set up one-way authentication:
     a. Set a CHAP username and password to your initiator name.
        i) `node.session.auth.authmethod = CHAP`
        ii) `node.session.auth.username = <initiator's user name>`
        iii) `node.session.auth.password = <CHAP secret for host>`
        iv) (Optionally) `node.session.auth.chap_algs = SHA3-256,SHA256,SHA1,MD5`
     b. Set a discovery session CHAP username and password to your initiator name.
        i) `discovery.sendtargets.auth.authmethod = CHAP`
        ii) `discovery.sendtargets.auth.username = <initiator's user name>`
        iii) `discovery.sendtargets.auth.password = <CHAP secret for host>`
     c. Save these settings. Log out of any current sessions and rediscover the system iSCSI target for the CHAP secret to be effective.

     Note: In the previous example, *xxxxxxxxxxxx* is the CHAP secret for the host, and the *rhel_username* is the IQN name of the initiator. This username must be the same value that you set with the **chhost** command (*iscsiusername* field) for this host.
   - Set up two-way authentication.

     Note: It is not mandatory to set up two-way authentication. Before you configure for two-way authentication, ensure that one-way authentication is configured and is working for your host.
     a. Edit the `password_in` to CHAP secret that you set up with the **chsystem** command on the system. Set the CHAP username as `clustername`. The `clustername` can be found by **lssystem** command on the system.
        i) Set a CHAP username and password for the target or targets.
           – `node.session.auth.username_in = clustername`
           – `node.session.auth.password_in = <CHAP secret for clustered system>`
        ii) Set a discovery session CHAP username and password for the target or targets.
           – `discovery.sendtargets.auth.username_in = clustername`

- discovery.sendtargets.auth.password_in = *<CHAP secret for clustered system>*

b. Save these settings. Log out of any current sessions and rediscover the system iSCSI target for the CHAP secret to be effective.

Note: Before upgrading to 8.5.3.0, if the customer has configured two-way chap authentication, they must first switch to one-way chap and then back to two-way chap once the upgrade is complete. This is required because `clustername` as a username is not supported for two-way chap secret in earlier releases.

Enabling multipathing for Linux hosts

You can enable the system to support multipathing for Linux hosts.

Before you begin

The system supports the Linux device-mapper multipathing driver for the iSCSI host attachment. After the iSCSI devices are discovered, you can enable multipathing.

Procedure

To enable multipathing for a Linux host, complete the following steps:

1. If the daemon is already running, stop the multipathing daemon with the `service multipathd stop` command.
2. Add the entry for the system by editing the multipath configuration file, `/etc/multipath.conf`. In the DMMP configuration file, `/etc/multipath.conf`, ensure that the default settings include the values that are described in Settings for Linux hosts.

   Note: Prevent blacklisting of the IBM 2145 devices by either disabling blacklisting of devices entirely or by adding an appropriate entry in the `blacklist_exceptions` section.
3. Save the file.
4. Restart the multipathing daemon with the `service multipathd start` command.

   If you are using SUSE Linux Enterprise Server 10 Service Pack 2 or later, replace the line `prio "/sbin/mpath_prio/alua /dev/%n"` with `prio "alua"` to disable the `Using deprecated prio` message.

   Note: Linux usually enables multipathing automatically. However, in some cases such as the RH 6.3 SAN boot installation, the multipathing is not configured properly. If this situation occurs, you must manually edit the `/etc/multipath.conf` file. For information about DM-multipath, see the Linux *Storage Administration Guide* or *DM-Multipathing Installation and Configuration Guide* for your kernel from the Linux provider. For more information about Linux system configuration, see the following website:

   https://www.ibm.com/docs/en/linux-on-systems

Configuring multisession on the Linux operating system

Hosts with Linux operating systems support multisession connections between the host and target.

About this task

Discover the target, and then complete the following steps to configure a session between the host and the target:

Procedure

1. To create a session between the host and the target, run one of the following commands:
   - `iscsiadm –m node -l`

This command establishes a session on all of the portals reported by the iSCSI target at the time of discovery.

- `iscsiadm –m node –T iSCSI target name -l`

  This command establishes a single session between the host and the target at the specified target portal.

2. To remove a session between the host and the target, run one of the following commands:

   - `iscsiadm –m session –u`

     This command removes sessions on all of the portals reported by the iSCSI target at the time of discovery.

   - `iscsiadm –m node –T iSCSI target name -u`

     This command removes a single session between the host and the target at the specified target portal.

Microsoft Windows iSCSI host attachment

You can use iSCSI Ethernet to attach a Microsoft Windows host to the system.

Before you begin host attachment procedures:

- Read the information about setting up the host server.
- Ensure that you are familiar with the command-line interface commands.

Setting the iSCSI qualified name for the Microsoft Windows host

The iSCSI qualified name (IQN) can be viewed or changed on the Windows host.

To view or change the iSCSI IQN, click the **Configuration** tab on the Windows Server operating system. The initiator name that you choose must match the iSCSI name that is given to the host object for the system.

iSCSI target discovery for Windows hosts

The supported methods for discovering iSCSI target portals are:

- Send Targets
- Internet storage name service (iSNS)

Discovering iSCSI targets using Send Targets

You can discover iSCSI target portals by using the Send Targets method. The Send Targets method is configured from the **Discovery** tab of the iSCSI Initiator control panel.

This form of discovery requires that you to type the address of one or more of your system node Ethernet ports. For each node that you want to complete discovery on, follow these steps:

1. Click the **Add** button (click **Discover Portal** on Windows Server 2008) that is associated with the list of target portals.
2. Type the IP address or DNS name of the node Ethernet port on which you want to complete discovery. Leave the IP port name at the default value of 3260. If CHAP authentication is configured for this host on the system, see “Authentication for Windows hosts” on page 167.
3. Click **OK**. The system node port address is added to the list of target discovery portals.

Discovering iSCSI targets on your Windows host by using iSNS

You can discover iSCSI target portals by using the internet storage name service (iSNS) method. Before you use the iSNS discovery, complete the following tasks:

- Configure the clustered system to use your iSNS server.
- Verify that the system nodes registered their target portals with it.

The iSNS method is configured from the **Discovery** tab of the iSCSI Initiator control panel.

1. Click the **Add** button that is associated with the list of iSNS servers.
2. Type the IP address or DNS host name of your iSNS server.

Connecting to discovered targets

After you discover the iSCSI targets, you connect to a specific portal on a target.

The system supports only one iSCSI session between an initiator and a target. Ensure that you do not attempt to connect to the same target node more than once. Open the **Targets** tab of the **iSCSI Initiator** control panel. The list of discovered targets shows an entry for each discovered node.

To connect to a specific portal on a discovered target, complete the following steps:

1. Select the node that you want to connect to from the list and click **Log on...** (click **Connect** on Windows Server 2008). You can select whether to restore the connection automatically at boot or to enable multipathing for the iSCSI target from the **Connect to Target** window.
2. To connect to a specific portal on the target, such as to perform load balancing between the node Ethernet ports or to configure CHAP authentication, click **Advanced...** to access these settings.
3. Click **OK** to change the status for the selected targets from **Inactive** to **Connected**.
4. Select the node to disconnect from the list and click **Log Off** to stop the I/Os to the iSCSI disks. Use this option to log off from the iSCSI target.

Viewing and managing the discovered disks

You can use the Windows host to view the discovered disks and bind the devices to the **Microsoft iSCSI Initiator Service**. The system volumes that are mapped to the iSCSI host are now visible to the Windows disk management services. The system volumes can be initialized, formatted, and mounted.

Ensure that the iSCSI devices are available before dependent applications or services are started. You can bind these devices to the **Microsoft iSCSI Initiator Service** from the **Bound Volumes/Devices** tab on the **iSCSI Initiator** control panel.

If the system is rebuilt or restored from backups, you must log off and log back in to any previously connected iSCSI targets that are supplied from the system.

From the Windows host, you can view information to help you differentiate each of the discovered disks. To do so, complete the following steps.

1. From the Windows **Command Prompt**, type `diskpart`.
2. Next, type `list disk`.
3. Next, select the disk for which you want to view the details. Type `select disk` *x* where *x* is the number of the disk that you want to view.
4. Type `detail disk` to see the type of disk and other information. The following screen shows a sample output.

```
DISKPART> list disk
  Disk ###  Status      Size     Free     Dyn  Gpt
  --------  ----------  -------  -------  ---  ---
  Disk 0    Online       149 GB    78 GB   *
  Disk 1    Online       149 GB    78 GB   *
  Disk 2    Online       565 MB   565 MB
  Disk 3    Online       337 MB   337 MB

DISKPART> select disk 2

Disk 2 is now the selected disk.

DISKPART> detail disk

IBM      2145             SCSI Disk Device
Disk ID: 00000000
Type   : iSCSI
Bus    : 0
Target : 2
LUN ID : 0

There are no volumes.

DISKPART>
```

Changes to system registry to optimize the initiator

You can change the system registry to optimize the iSCSI operation.

For reliable iSCSI operation during system maintenance, you can update these values in the system registry. See Table 28 on page 167. In addition, you can also update the system registry to change the delayed acknowledgment settings to address slow performance with iSCSI operations between systems. .

Table 28. System registry

Registry key	Type	Value (decimal)
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\Class\{4D36E97B-E325-11CE-BFC1-08002BE10318}*bus id*\Parameters\LinkDownTime	DWORD	120
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\Class\{4D36E97B-E325-11CE-BFC1-08002BE10318}*bus id*\Parameters\MaxRequestHoldTime	DWORD	120
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Disk\TimeOutValue	DWORD	60
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\Class\{4D36E97B-E325-11CE-BFC1-08002BE10318}*bus id*\Parameters\ MaxPendingRequests	DWORD	2048

Note: You must restart the computer for these changes to take effect.

Authentication for Windows hosts

The system supports both one-way and two-way Challenge Handshake Authentication Protocol (CHAP) methods for Microsoft Windows host attachments.

Discovery sessions or normal sessions use one-way CHAP authentication where the system authenticates the host iSCSI initiator. In a two-way CHAP authentication, the system and the initiator authenticate each other.

Setting up authentication for discovery sessions for Windows hosts

You can set up authentication for discovery sessions for Microsoft Windows hosts at the same time that you are connecting to the system volumes. The CHAP authentication information is on the same **Advanced Settings** panel.

To set up one-way authentication for a Windows host, complete the following steps:

1. From the Control Panel, select the **iSCSI Initiator** option.
2. From the **iSCSI Initiator Properties** panel, click the **Discovery** tab.
3. After you click the **Discovery** tab, click **Add** under the **Target Portals** section to see the **Add Target Portal** dialog box.

   Note: For Windows Server 2012, select **Discover Portal** and then enter the IP address under **IP address or DNS name**.
4. Click **Advanced...** to see the **Advanced Settings** panel.
5. Select **CHAP logon information**.

   Note: For Windows Server 2012, select **Enable CHAP logon**.
6. Type in a value for the **User name**. The user name must be the same value that you set with the **chhost** CLI command (*iscsiusername* field) for this host.
7. Type in a value for the **Target secret**. The target secret must be a value of 12 characters and is the same value that you set with the **chhost** command on the system for this host. Click **OK**.

Tasks in the management GUI:

1. Select **Hosts** > **Hosts**.
2. Right-click the host to configure CHAP authentication and select **Properties**.
3. On the **Properties** page, ensure that **Show Details** is selected and click **Edit**.
4. In the **iSCSI CHAP secret** field, enter the CHAP secret that you use for the host.
5. In the **iSCSI user name** field, enter the host user name that you use for the host. Click **Save**.

Setting up authentication for normal sessions for Windows hosts

You can set up authentication for normal sessions for Microsoft Windows hosts at the same time that you are connecting to a target or volume. The CHAP authentication information is on the same **Advanced Settings** panel.

To set up one-way authentication for a Windows host, complete the following steps:

1. From the Control Panel, select the **iSCSI Initiator** option.
2. From the **iSCSI Initiator Properties** panel, click the **Discovery** tab.
3. After you click the **Targets** tab, click **Log On...** under the **Targets** section to see the **Log On to Target** panel.

   Note: For Windows Server 2012, highlight your target, select **Connect**, and then select **Advanced** from the pop-up screen.
4. Click **Advanced...** to see the **Advanced Settings** panel.
5. Select **CHAP logon information**.

   Note: For Windows Server 2012, select **Enable CHAP logon**.
6. Type in a value for the **User name**. The user name must be the same value that you set with the **chhost** CLI command (*iscsiusername* field) for this host.
7. Type in a value for the **Target secret**. The target secret must be a value of 12 characters and is the same value that you set with the **chhost** command on the system for this host. Click **OK**.

For tasks in the management GUI, refer to the Setting up authentication for discovery sessions for Windows hosts section.

Setting up two-way authentication for Windows hosts

You can set up two-way authentication for Microsoft Windows hosts.

Verify that you have a working one-way authentication and that you configured a system iSCSI Challenge Handshake Authentication Protocol (CHAP) secret on the system.

To set up two-way authentication for a Windows host, complete the following steps:

1. From the Control Panel, select the **iSCSI Initiator** option.
2. From the **iSCSI Initiator Properties** panel, click the **General** tab.

   Note: For Windows Server 2012, select **Configuration**.
3. Click **Secret** to see an **iSCSI Initiator** panel from which you can type your CHAP secret. Enter the system CHAP secret.

   Note: For Windows Server 2012, click **CHAP...** and then enter the system CHAP secret.

   Click **OK**.

   Notes:

   a. This setting applies to both the discovery session and normal session.

   b. The same length restrictions for CHAP secrets that apply to one-way authentication apply to two-way authentication.

   c. The CHAP secrets for the system and the host initiator cannot be the same.
4. To set up the two-way authentication, repeat the previous steps, but in this instance, select **Perform mutual authentication** from the **Advanced Settings** panel.
5. Select **iSCSI Initiator Properties** > **iSCSI Initiator Mutual CHAP secret**. The initiator username value must be *clustername* and chap secret must match with what is configured with **`chsystem`**. The `clustername` can be found by **`lssystem`** command.

Tasks in the management GUI:

1. Select **Settings** > **Network** > **iSCSI**.
2. On the **iSCSI Configuration** page, select **Modify CHAP Configuration**.
3. On the **Modify CHAP Configuration** page, enter a system-wide CHAP secret and select **Use for iSCSI-attached hosts**. Click **Modify**.

   Note: Before upgrading to 8.5.3.0, if the customer has configured two way chap authentication, they must first switch to one way chap and then back to two way chap once the upgrade is complete. This is required because `clustername` as a user name is not supported for two way chap secret in the earlier releases.

Configuring multisession on the Microsoft Windows operating system

The Microsoft Windows operating system supports multisession connections.

About this task

Discover the iSCSI targets, and then complete the following steps to configure a multisession connection.

Note: The following Microsoft Multipath I/O (MPIO) timer key values are recommended when you use multisession on Microsoft Windows and the Microsoft iSCSI Multipath DSM:

- PathVerifyEnabled:

  This flag enables path verification by MPIO on all paths every *N* seconds (where *N* depends on the value set in the PathVerificationPeriod panel). The type is boolean and must be filled out with either 0 (disable) or 1 (enable). The value is set to Disabled by default. You must change the value to Enabled.
- PathVerificationPeriod:

This setting indicates the period, in seconds, within which MPIO is expected to complete path verification. This field is only honored if PathVerifyEnabled is True. This default value is 30 seconds. The maximum value that is allowed is MAXULONG. You must set this value to 60 seconds.

- PDORemovePeriod:

  This setting controls the amount of time, in seconds, that the multipath pseudo-LUN will remain in system memory, even after losing all paths to the device. The default is 20 seconds. The maximum value that is allowed is MAXULONG. You must set this value to 60 seconds.

  Detailed information about the procedure to set these values is available by searching for "Microsoft iSCSI Software Initiator Version 2.08" from the following website:

  www.microsoft.com/downloads

To create a multisession connection, complete the following steps.

Procedure

1. Select the **Targets** tab from **iSCSI Initiator Properties**, and click **Log On**.
2. From **Log On to Target**, select **Enable multi-path**, and then select one of the following two options:
   - Click **OK**.

     Repeat this procedure to create multiple sessions. Each time that you repeat this procedure, a new session is established between the initiator and the target.
   - Click **Advanced**, then make the following selections:
     - At **Local Adapter**, click **Microsoft iSCSI Initiator**.
     - At **Source IP**, click the host IP from which your target portal can be pinged.
     -
     - At **Target Portal**, click **Default**.

     Click **OK**.

What to do next

To delete a multisession session, select the **Targets** tab from **iSCSI Initiator Properties**, and click **Details**. Place a check mark in the box of each identifier connection to be terminated, and then click **Log off** and **OK**.

Note: Multiple sessions are not supported for Microsoft Clustering with iSCSI.

IBM AIX iSCSI host attachment

You can use iSCSI Ethernet to attach an IBM AIX host to the system.

Before you begin host attachment procedures:

- Read the information about setting up the host server.
- Ensure that you are familiar with the command-line interface commands.

Configuring the AIX iSCSI software initiator

Ensure that you follow all the steps to set up the host server and configure the AIX iSCSI software initiator.

Install the AIX software initiator on your AIX host if it is not already installed.

1. To verify that the file set `devices.iscsi_sw.rte` is installed, use the **`lslpp -l`** command.
2. If necessary, install the interim fix for the system Object Data Manager (ODM) stanzas.

The software initiator is configured by using the System Management Interface Tool (SMIT). To configure the software initiator, follow these steps:

1. Select **Devices**.
2. Select **iSCSI** and select **iSCSI Protocol Device**.
3. Select **Change / Show Characteristics of an iSCSI Protocol Device**. Select the device that is associated with the software iSCSI initiator, such as `iscsi0`.
4. Verify that the **Initiator Name** value is correct. The initiator name value is used by the iSCSI target during login. The name must match the iSCSI name that was set for the system host object that is associated with this host.

   Note: A default initiator name is assigned when the software is installed. You can change the initiator name to match the local network naming conventions.
5. The **Maximum Targets Allowed** field corresponds to the maximum number of iSCSI targets that can be configured. If you reduce this number, you also reduce the amount of network memory that is preallocated for the iSCSI protocol driver during configuration.

Adding the iSCSI targets for AIX hosts

Add the iSCSI targets for AIX hosts so that they can be discovered through a properly configured network interface.

To add the iSCSI targets during the device configuration, follow these steps:

Note:

The system does not support Ethernet port aggregation. Therefore, host-based multipath driver support for system LUNs is necessary to provide active-active controller use. Operating systems such as Linux, Windows, and VMware can configure iSCSI with multiple paths to system LUNs by using iSCSI interconnects. However, AIX MPIO does not support multiple paths to system LUNs by using iSCSI interconnects.

This statement does not mean that LUN redundancy across controller failure is not supported over iSCSI for AIX hosts. If a controller fails, the surviving controller takes over all the IP addresses and the iSCSI target that is configured on the failing controller. The AIX host then reconnects with the failed over iSCSI targets on the surviving controller and continues with its IO operations. Therefore, controller failure does not disrupt operations when AIX hosts communicate to the system through iSCSI.

1. Edit the `/etc/iscsi/targets` file to add the iSCSI target portal details for all system nodes that the host must log on to.

   Each uncommented line in the file represents an iSCSI target and is in the following format.

   ```
   192.168.1.7     3260      iqn.1986-03.com.ibm:2145.sahyadri.node1
   ```

   Note: As part of the "no iSCSI MPIO" restriction, it is not valid to have the same target IQN multiple times in the targets file, even with different IP addresses. If the same target exists multiple times, comment out the duplicate targets in order for the process to run.
2. After you edit the `/etc/iscsi/targets` file, type the following command:`cfgmgr -v -l iscsi0`

   This command causes the driver to attempt to log on to each of the targets that are listed in the `/etc/iscsi/targets` file. It also defines a new hard disk (hdisk) for each LUN on the targets that are found.
3. Review the discovered hard disks by entering the following command:

   ```
   lsdev -c disk
   ```

   The system volumes appear in the output in the following format:

   ```
   hdisk2   Available               IBM 2145 iSCSI Disk Drive
   ```

   Note: If the appropriate disks are not discovered, review the configuration of the initiator, the target, and any iSCSI gateways to ensure that they are correct. Run the **`cfgmgr`** command again.

Configuring iSCSI host disk timeouts

To configure the iSCSI host disk timeouts for the system, use the **chdev** CLI command to set the value and then verify your changes.

Use the value of 60 seconds for the rw_timeout attribute for the system iSCSI host disks. To set the value, complete the following steps:

1. Set the value with the **chdev** command as shown in the following example for each iSCSI device:

```
chdev -l hdisk10 -a rw_timeout=60
```

2. Verify the changes by using the **lsattr** command:

```
lsattr -E -l hdisk10 -a rw_timeout
```

Setting up authentication for AIX hosts

AIX hosts can be correctly set up for authentication on the system by following certain guidelines and tasks.

Although the system supports both one-way authentication and two-way authentication for iSCSI, the AIX software initiator currently supports only one-way authentication. The system target authenticates the initiator.

CHAP settings are defined in the `/etc/iscsi/targets` file on the host. The AIX initiator or host bus adapter (HBA) always uses its iSCSI qualified name (IQN) as the CHAP user name.

To set up authentication on an AIX host, complete the following steps:

1. Open the `/etc/iscsi/targets` file with any editor.
2. For each line that contains a target definition, append the CHAP secret of the initiator in quotation marks:

```
192.168.1.7      3260     iqn.1986-03.com.ibm:2145.sahyadri.node1 "secret"
```

   The CHAP secret value that you set here must match the value that was configured on the system for the host object that is associated with this host. Because the system authenticates on a per-initiator basis, the CHAP secret is the same for all the targets on a particular clustered system.

   Note: When setting up authentication for two way chap username and password, the `node.session.auth.username_in` value must be `clustername`. The `clustername` can be found by **lssystem** command.

   An example of CHAP settings for an AIX host for the `/etc/iscsi/targets` file is shown below.

```
#ChapSecret             = %x22*( any character ) %x22
#                       ;   "                      "
#                       ; ChapSecret is a string enclosed in double quotes. The
#                       ; quotes are required, but are not part of the secret.
#
#EXAMPLE 1: iSCSI Target without CHAP(MD5) authentication
#      Assume the target is at address 192.168.3.2,
#      the valid port is 5003
#      the name of the target is iqn.com.ibm-4125-23WTT26
#The target line would look like:
#192.168.3.2 5003 iqn.com.ibm-4125-23WWT26
#
#EXAMPLE 2: iSCSI Target with CHAP(MD5) authentication
#      Assume the target is at address 10.2.1.105,
#      the valid port is 3260
#      the name of the target is iqn.com.ibm-K167-42.fc1a
#      the CHAP secret is "This is my password."
#The target line would look like:
#10.2.1.105 3260 iqn.com.ibm-K167-42.fc1a "This is my password."
#
#EXAMPLE 3: iSCSI Target with CHAP(MD5) authentication and line continuation
#      Assume the target is at address 10.2.1.106,
#      the valid port is 3260
#      the name of the target is iqn.com.ibm:00.fcd0ab21.shark128
```

```
#      the CHAP secret is "123ismysecretpassword.fc1b"
#The target line would look like:
#10.2.1.105 3260 iqn.2003-01.com.ibm:00.fcd0ab21.shark128


192.168.1.41 3260 iqn.1986-03.com.ibm:2145.pahar.dvt110702
192.168.2.43 3260 iqn.1986-03.com.ibm:2145.moscow.dvt110706 "svcchapsecret"
```

The two targets in the previous example are members of different clustered systems. One target is configured to authenticate the initiator, and the other target is not configured to authenticate the initiator.

Note: Before upgrading to 8.5.3.0, if the customer has configured two-way chap authentication, they must first switch to one-way chap and then back to two-way chap once the upgrade is complete. This is required because `clustername` as a user name is not supported for two-way chap secret in earlier releases .

Updating ODM stanzas for system iSCSI devices

Updating the Object Data Manager (ODM) stanzas for the system iSCSI devices requires that you follow certain guidelines.

An interim fix is available to update the AIX ODM stanzas to recognize iSCSI system volumes. Ensure that you have installed the official PTFs. The following website provides information about the available PTFs: www.ibm.com/support

When you install the interim fix package, the AIX host can recognize iSCSI system volumes. The system volumes are shown in the below example with the following attributes for `lsdev -C -1 hdisk1`.

```
lsdev -C -l hdisk1

hdisk1     Available              IBM 2145 iSCSI Disk Drive


# lsattr -E -l hdisk1
clr_q          no                            Device CLEARS its Queue on error True
host_addr      9.71.43.106                   Hostname or IP Address           False
location                                     Location Label                   True
lun_id         0x0                           Logical Unit Number ID           False
max_transfer   0x40000                       Maximum TRANSFER Size            True
port_num       0xcbc                         PORT Number                      False
pvid           none                          Physical volume identifier       False
q_err          yes                           Use QERR bit                     True
q_type         simple                        Queuing TYPE                     True
queue_depth    8                             Queue DEPTH                      True
reassign_to    120                           REASSIGN time out value          True
reserve_policy no_reserve                    Reserve Policy                   True
rw_timeout     60                            READ/WRITE time out value        True
start_timeout  60                            START unit time out value        True
target_name    iqn.1986-03.com.ibm:2145.china6.hlcn111890   Target NAME      False
unique_id      3521360050768019102968800000000000000204214503IBMiscsi Unique device
identifier      False

lscfg -v -l hdisk1


  hdisk1             IBM 2145 iSCSI Disk Drive

        Manufacturer.................IBM
        Machine Type and Model......2145
        ROS Level and ID............30303030
        Serial Number...............
        Device Specific.(Z0)........0000043268101002
        Device Specific.(Z1)........
        Device Specific.(Z2)........
        Device Specific.(Z3)........
```

VMware iSCSI host attachment

iSCSI Ethernet hosts for the VMware operating system can be attached to the system.

Before you begin host attachment procedures:

- Read the information about setting up the host server.
- Ensure that you are familiar with the command-line interface (CLI) commands.

Installing the VMware initiator software

Attach a VMware host to the system using iSCSI. The system supports the use of the iSCSI-initiator software on VMware Version 4.0 ESX servers. Ensure that you have the latest release of the initiator software that is installed on your host before you continue with the instructions.

Complete the following steps to install the VMware initiator software:

1. Click the VMware vSphere client to access the ESX 4.0 client.
2. Enter the IP address of the ESX server with your user ID and password to log in to the system. A list of available operating systems is displayed.
3. Configure the vSwitches.

Note: Using the default initiator SCSI settings is recommended.

The following functions are not currently supported:

- Guest OS support
- SAN boot support

Virtual Switches

A virtual switch (or vSwitch) is a software program that enables one virtual machine to communicate with another.

Similar to a physical Ethernet switch, a vSwitch directs communication on the network by inspecting packets before passing them on. VMware uses a vSwitch to work with the system.

The vSwitch contains the VMkernel (the kernel that is used by VMware to access the iSCSI I/O) and the service console (the console through which management activities are performed), and uses a network port group. For the system to work with VMware, the IPs configured on the VMkernel must ping to and from the nodes in the clustered system to enable communication between VMware and the system.

About VMkernels

The VMkernel is the operating system of your ESX or ESXi server and has complete control over all hardware including CPU, memory, network interface adapters, and host bus adapters.

The VMkernel completes the virtualization and is also referred as the *hypervisor*.

VM consoles

The VM service console is a specialized virtual machine that enables management of the ESX host.

Note: There is no service console under ESXi because this management function resides on the VMkernel.

VM port groups

The network port group connects all guest operating systems to the network.

Use the service console to manage a Guest OS.

Configuring vSwitches

You must configure the vSwitches on the VMware host before creating any virtual machines on the host. Follow these steps to configure vSwitches:

1. From the vSphere Client that you opened in "Installing the VMware initiator software" section, select the **Configuration** tab.

2. Click **Networking** > **Add Networking** > **Virtual Machine** and click **Next**.
3. Click **Create a Virtual Switch** and click **Next**.
4. Select a network interface that is running and is not in use by a vSwitch, and then click **Finish**.

   A vSwitch is created within a network port group. All hosts that you create are attached to this network port group.

You must add a service console to this vSwitch to access the ESX server and provide the hosts with access to the external environment.

Adding a service console to the vSwitch

You must add a service console to this vSwitch to access the ESX server and provide the hosts with access to the external environment. Then, to access the internal environment, you must add a second vSwitch and service console.

Perform the following steps to add a service console to a vSwitch:

1. Open the vSphere Client and click the **Configuration** tab.
2. Click **Networking**, and then select the vSwitch.
3. Click **Properties** > **Add** and select **Service Console**.
4. Select **Use the following IP settings**, and enter the external IP values.

   This IP setting is the management IP for the VMware vSwitch.

   Important: Be sure to enter the correct values for the service console gateway. If the wrong gateway values are entered, vSphere is unable to access the VMware machine.
5. To set internal IP values, create a second vSwitch by following the steps at "Configuring vSwitches" section.
6. From the vSphere Client, select the guest OS.
7. Click **Add** > **Ethernet Adapter**, and click **Next**.
8. Add the next Ethernet adapter, which is the VMware network in the new switch, to this guest OS. Perform this procedure for each guest OS.
9. Add a service console to this vSwitch by completing steps “1” on page 175 through “4” on page 175, entering the internal IP values.

Add a VMkernel to this vSwitch to assign I/O traffic to it.

Creating the VMkernel

You must create a VMkernel to assign I/O traffic to a vSwitch. Complete the following steps to create a VMkernel:

1. From the vSphere Client, select the v Switch and click **Properties**.
2. Click **Add** > **VMkernel** and click **Next**.
3. Select **Use the following IP settings** and enter the IP values that you want to assign to the VMkernel, and then click **Next**.
4. Click **Finish** to complete the procedure. This VMkernel is used for the iSCSI traffic.

To test the configuration, access the target node and ping the internal IPs that belong to the service console of the VMkernel. You can also test the configuration from the guest OS by pinging the target IPs.

Setting the initiator name

The iSCSI IQN can be viewed or changed from the Configuration tab of the iSCSI software adapter control panel. To view or change the iSCSI qualified name (IQN), complete the following steps.

1. Click the **General** tab and click **Configure**. The status of the Software Initiator Properties section shows Enabled.
2. Click **OK**.

The initiator name that you choose here must match the iSCSI name that is given to the system host object for this machine.

Discovering targets on VMware

Two methods are supported for iSCSI target portal discovery: iSNS and Send Targets discovery.

Both methods are configured from the Discovery tab of the iSCSI Initiator control panel module. To use the Send Targets method, complete the following steps for each node on which you want to perform discovery:

1. From the iSCSI Initiator control panel, click **Select iSCSI Software Adapter**.
2. Click **Properties** > **Dynamic Discovery** > **Add**.
3. Enter the target IP address. The target IP address is the IP address of a node in the I/O group from which you are mapping the iSCSI volume.
4. Leave the IP port number at the default value of 3260, and click **OK**. The connection between the initiator and target is established after clicking **OK**.
5. Click **iSCSI Software Adapter**.
6. Right-click this tab and rescan the devices for discovery.

   The maximum number of volumes that can be mapped to the VMware iSCSI initiator is 256.

   Note: iSNS discovery is not supported in VMware ESX 4.

Mapping VMware iSCSI volumes to a guest OS

You can map VMware iSCSI volumes to a guest OS. After you first create the virtual machine, complete the following steps to map the volumes:

1. From the vSphere Client, select the type of virtual machine from the list on the left side of the inventory. Right-click your selection and power on the virtual machine.
2. Select the **Console** tab and install the guest OS from your installation CD or your network.
3. When the OS installation process is complete, power off the virtual machine. From the vSphere Client, right-click the virtual machine and click **Edit Settings**.
4. Right-click the guest OS and click **Edit Settings**.
5. Click **Add** to add the iSCSI volumes and power off the guest OS.
6. Click **Hard Disk** and click **Next**.
7. Click **Raw Device Mapping** and click **Next**.
8. Click the volume to be mapped and click **Next**.
9. Click **Store With Virtual Machine** and click **Next**.
10. Click **Physical** and click **Next**.
11. Click **Virtual Device Node** as the default and click **Next**.
12. Click **Finish**, and then click **OK**.
13. Power on the virtual machine and verify that the mapping process was successful:
    - On Linux, issue the `fdisk-l` command.
    - On Microsoft Windows, start the Device Manager and scan for iSCSI volumes.

When the volumes are mapped to the guest OS, I/O functions are available on the new iSCSI volume.

Setting up authentication for VMware hosts

One-way and two-way CHAP authentication for VMware hosts can be set up.

To set up one-way and two-way CHAP authentication on a VMware host, complete the following steps:

1. From the vSphere Client screen, click the **Inventory** tab and click **Storage Adapters**.
2. Select the iSCSI storage adapter, and then right-click and select **Properties**.
3. Click **CHAP authentication**.
4. Select the preferred settings and enter the passwords:
   a. From the CHAP (target authenticates host) menu, click **Use CHAP**.
   b. Click **Use initiator name** to use *initiator's IQN* as the user name for CHAP authentication or type the user name that you set with the **chhost** CLI command (*iscsiusername* field) on the system. Enter the CHAP secret.
5. To set up the two-way authentication, repeat the previous steps and then complete these steps:
   a. From the Mutual CHAP authentication menu, click **Use CHAP**.
   b. Click **Name** and enter the value as `clustername`. The `clustername` can be found by **lssystem** command.
   c. Enter the CHAP secret. The chap secret must match with what is configured with the **chsystem**.

   Note: Before upgrading to 8.5.3.0, if the customer has configured two way chap authentication, they must first switch to one way chap and then back to two way chap once the upgrade is complete. This is required because `clustername` as a username is not supported for two way chap secret in earlier releases.

Host multipathing

VMware supports different types of multiple paths.

Fixed path
The I/O initiates access from the path that is initially selected, and switches to a different path if problems are encountered.

Most Recently Used Path
VMware uses the most recently used path, and switches to a different path if problems are encountered.

Round Robin
VMware accesses the different paths available at fixed intervals.

Viewing available paths

You can view the available multiple paths in VMware. To view the available paths in VMware, complete the following steps.

1. From the vSphere Client screen, click the **Configuration** tab.
2. Click **Storage Adapters** and select the iSCSI storage adapter.
3. Right-click on a volume and select **Manage Paths**. The available paths are listed.

To create multiple paths to the volume, log in to multiple targets in the same I/O group.

Enabling IPv6

You can enable IPv6 support on a host in VMware 4.0 or later. Complete the following steps to enable IPv6 support in VMware 4.0 or later:

1. From the vSphere Client screen, select the **Configuration** tab.
2. Click **Networking** > **Properties**.

3. Click **Enable IPv6 support on this host**, and then click **OK**. IPv6 support is enabled when you restart VMware.
4. To complete IPv6 discovery and login functions, open the vSphere Client.
5. Select the **Configuration** tab and click **Networking**, and then click the vSwitch.
6. Click **Properties** > **Service Console** and click **Edit**.
7. Click the **IP Settings** tab and enter your IP values, and then click **OK**.

iSCSI guest OS initiator

Configuration of a VMware host includes understanding several requirements that relate to the iSCSI guest OS initiator.

VMware recognizes the guest OS as an application and provides a link between the guest OS and the target. The VMware iSCSI initiator itself does not process any actions.

Before you use the iSCSI initiator of the guest OS, ensure that the following items are true:

- The target network setup in VMware and the target IPs ping from VMware are configured.
- The IPs on the guest OS are configured, and they ping the target IPs.
- The internal IPs on the guest OS ping from the target (node).

When the network is configured, you can start iSCSI discovery and logging activities by using the iSCSI initiator of the guest OS. You can also use a multipathing driver from the guest OS to handle IP path failure.

For installation and use of iSCSI initiator, see the iSCSI installation topic for your operating system. There is no change in the commands and setup concerning the guest OS iSCSI initiator; it uses the commands that are used by the operating system.

Note: Only Red Hat Enterprise Linux (RHEL) V5.4 is supported as a guest OS initiator with VMware.

Multisession on the VMware system

This section provides information about multisession capabilities on the VMware system.

Only Version 4.0 of VMware supports multisession.

The VMware initiator automatically discovers all of the portals and logs in to them.

For a Guest OS initiator, follow the steps under that OS type for multisession connection for that operating system type.

Note: Removing the target IP on the Dynamic Discovery tab removes all of the target portals.

Queue depth in iSCSI hosts

In iSCSI networks, the queue depth is the number of I/O operations that can be run in parallel on a device.

Configure your servers to limit the queue depth on all of the paths to the volumes in configurations that contain many servers or volumes.

Where a number of servers in the configuration are idle or do not initiate the calculated quantity of I/O operations, queue depth might not need to be limited.

Queue depth limits in iSCSI networks

The queue depth is the number of I/O operations that can be run in parallel on a device. If you are designing a configuration for an iSCSI network, you must estimate the queue depth for each node to avoid application failures.

If a node reaches the maximum number of queued commands, it returns error codes to the host such as `Resource unavailable`. Many operating systems cannot recover if the situation persists for more than

15 seconds. This inability to recover can result in one or more servers that present errors to applications and application failures on the servers.

The formula for queue depth calculation considers the following factors:

- The maximum number of queued commands is calculated on a per-node basis. An I/O group contains two nodes. The system must continue to function when one of the nodes in an I/O group is not available.
- If a volume is mapped to be seen by more than one server, each of the servers can send commands to it.
- If a device driver times out of a command, it immediately reissues the command. The commands are stored in the command queue.

After you calculate the queue depth limit, you must apply it. Each operating system has a particular method of limiting the queue depth on a per-volume basis; see your host operating system documentation to change queue depth.

Homogeneous queue depth calculation in iSCSI hosts

Set the homogeneous queue depth to a calculated value in iSCSI hosts.

Set the queue depth for each volume on the servers by using the calculation:

```
q = (n/(v * c)) * 0.75
```

where

- n = the allowed command pool per I/O group; this value is fixed at 4000
- v = number of volumes per server for the particular I/O group under consideration
- c = number of servers in an I/O group

Therefore, `q = (4000/(v * c)) * 0.75`.

Example

Consider a system with 20 host servers `(c = 20)`, each with 20 volumes mapped `(v = 20)`. For this configuration, the value of the queue depth is `q = (4000/(20 * 20)) * 0.75 = 7.5` (or the lower boundary limit of 7).

Use the minimum value of 2 if the value of derived queue depth exceeds the maximum allowed value on a particular host operating system.

Some common host server configurations are summarized in Table 29 on page 179, considering 20 volumes mapped to each host server.

Table 29. Common queue depth configurations

Number of iSCSI hosts per I/O group	**<20**	**20-40**	**64**
Recommended queue depth	Seven (or the maximum number that is allowed by the host OS, whichever is smaller)	Three to four (or the max number that is allowed by the host OS, whichever is smaller)	Two

Nonhomogeneous queue depth calculation in iSCSI hosts

After you understand certain criteria for nonhomogeneous queues, you can calculate their queue depth in iSCSI hosts.

Nonhomogeneous queues meet one of the following criteria:

- One or more servers must be allocated more resources so that they can queue more commands.
- Volumes must not be distributed evenly among the I/O groups in the clustered system.

For each volume, consider each server to which that volume has a mapping in an I/O group. This mapping results in a set of server/volume pairs. If the sum of the server and volume queue depth for all of the pairs is less than 4000, the server does not experience problems due to a full queue.

Multisession

Multisessions for iSCSI hosts can be configured by using different operating systems. Several characteristics must be considered when you configure an iSCSI multisession.

The following points apply to multisession configuration:

- The AIX iSCSI initiator does not support multisession.
- The use of the supported iSCSI multipathing is recommended when running the I/O with multisession.
- If you use CHAP authentication under the Microsoft Windows operating system, use the same secret across each session that is initiated from that host to a particular target.
- If the Microsoft Windows clustered system is rebuilt or restored from a backup, hosts must log off and then log back on to any previously connected iSCSI targets that are supplied from the system.
- If the host application uses SCSI 3 Persistent Reservation, you must run only a single iSCSI session per node.
- Running with multiple iSCSI sessions is not supported for applications that use Microsoft 2008 Failover clustering over an iSCSI interface. You must run only a single session per node in these situations.

For more information on session limit maximums for an iSCSI target, an I/O group, and a system, see Configuration limits.

Single subnet configurations

Some multisession configurations can support a single subnet.

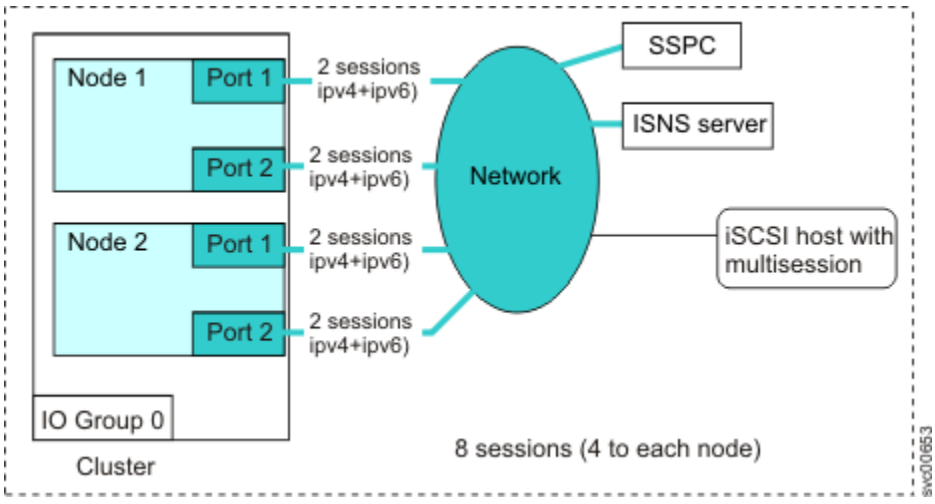


Figure 16. Four two-session iSCSI multisession configurations for a single subnet

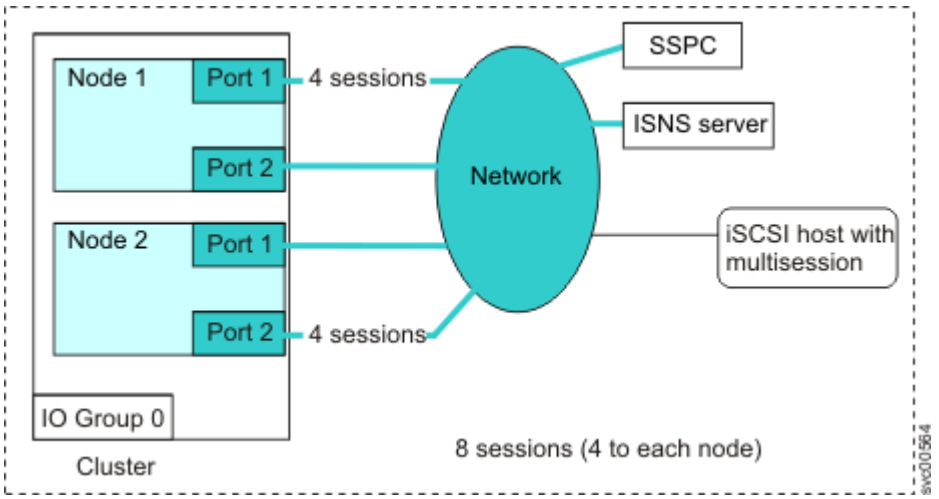


Figure 17. Two four-session iSCSI multisession configurations for a single subnet

Multiple and dual subnet configurations

Some multisession configurations support multiple and dual subnets.

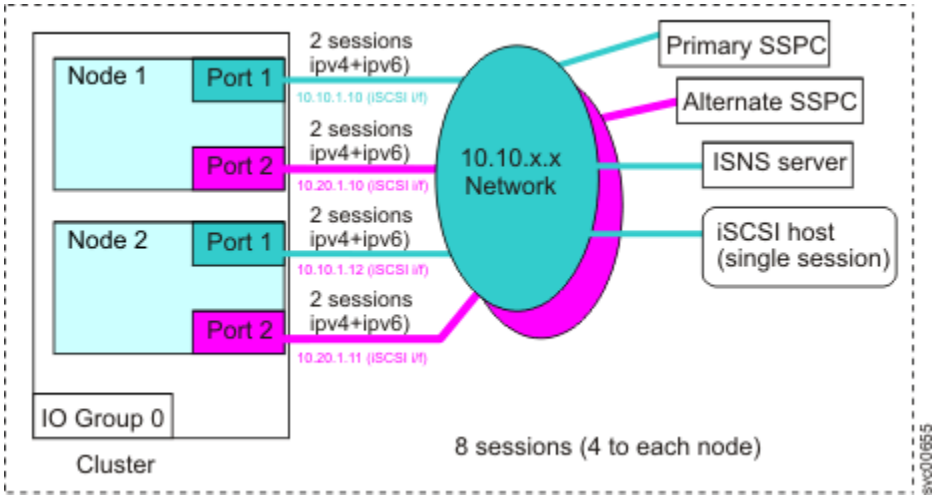


Figure 18. Four two-session iSCSI multisession configurations for multiple subnets and dual configuration

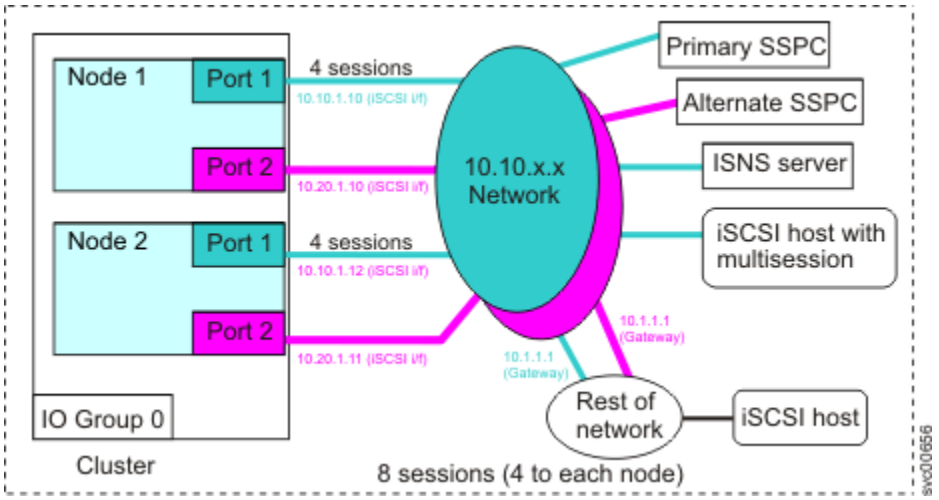


Figure 19. Two four-session iSCSI multisession configurations for a multiple subnets and dual configuration

NVMe over Fibre Channel host attachments

An NVM Express (NVMe) over Fibre Channel (FC) host can be attached to the system.

FC-NVMe configuration guidelines

Consider NVMe over Fibre Channel target guidelines when you plan and configure the hosts.

Include the following details in your plan:

- An NVMe host can connect to a maximum of four NVMe controllers on each system node and with another four if there is a failover.

  Zone up to four ports in a single host to detect up to four ports on a node. To allow failover and avoid outages, zone the same or extra host ports to detect an extra four ports on the second node in the I/O group.

NVMe hosts that run the Linux operating system

When you install and configure attachments between the system and a host that runs the Linux operating system, follow specific guidelines.

For other specific information about NVMe over Fibre Channel (FC-NVMe) such as interoperability requirements, see Configuration limits.

Attachment requirements for hosts that are running the Linux operating system

Ensure that your system meets the requirements for attaching to a host that is running the Linux operating system.

1. Follow the HBA vendor instructions to update to the right firmware and driver level.
2. Install the required NVM-Express user space tools, by using one of the following commands:
   - For SLES

     ```
     zypper install nvme-cli
     ```

   - For RHEL

     ```
     yum install nvme-cli
     ```

3. If you are working with an Emulex adapter, verify that Emulex auto-connect is installed. To verify, run the following command:

   ```
   rpm -q nvmefc-connect
   ```

   Note: For SLES 15, RHEL 8.3 and latest kernel levels, the command **`nvmefc-connect`** is no longer applicable.

   Note: When using FC-NVMe on RHEL 9.0 , use nvme-cli version 1.16.

Configuring the Linux operating system for FC-NVMe hosts

After you ensure that your system meets the requirements for attaching to a Linux host, configure the Linux operating system.

You must install the appropriate host bus adapters with the correct firmware and driver levels that support FC-NVMe.

1. Zone the host ports to the NVMe ports on the system.

   For more information about how to identify NVMe ports on the system, see the CLI host commands.
2. Find the host NVMe Qualified Name (NQN) address (under `/etc/nvme/hostnqn`).

3. On the system, create the NVMe host object by using the host NQN.

```
svctask mkhost -nqn nqn.2014-08.org.nvmexpress:uuid:449f8291-9c1e-446c-95c1-0942f55fa208
 -protocol nvme -name mcr-host-246
```

4. Map relevant volumes to the NVMe host. The same volumes cannot be mapped concurrently to both NVMe and SCSI.
5. To discover and connect to NVMe targets, enter the following commands.
   a. The **NVMe Discover** command.

```
nvme discover --transport=fc --traddr=nn-0x$twwnn:pn-0x$twwpn
--host-traddr=nn-0x$wwnn:pn-0x$wwpn
```

   The return of this command is the **NVMe Discover** log page, which consists of the target's subsystem NQN.

```
Discovery Log Number of Records 1, Generation counter 0
=====Discovery Log Entry 0======
trtype:  fibre-channel
adrfam:  fibre-channel
subtype: nvme subsystem
treq:    not required
portid:  <>
trsvcid: none
subnqn:  nqn.1986-03.com.ibm:nvme:2145.<>.iogroup<>
traddr:  nn-$twwnn:pn-0x$twwpn
```

   b. The **NVMe Connect** command, which has the same syntax with the addition of the subnqn.

```
nvme connect --transport=fc --traddr=nn-0x$twwnn:pn-0x$twwpn
--host-traddr=nn-0x$wwnn:pn-0x$wwpn -n $subnqn
```

   c. **NVMe connect-all** includes the **Discover** and **Connect** commands within the same command. This command connects automatically to the subnqn that is presented in the discover command:

```
nvme connect-all --transport=fc --traddr=nn-0x$twwnn:pn-0x$twwpn
--host-traddr=nn-0x$wwnn:pn-0x$wwpn
```

   d. Initiate, the **NVME disconnect** command, to disconnect any NVMe device:

```
nvme disconnect -d /dev/nvme-$x
```

   e. Initiate the **NVME disconnect-all** command, to disconnect all NVMe devices:

```
nvme disconnect-all
```

Removing a LUN

To remove a LUN from the storage system, perform the following steps:



Attention: Before removing a LUN from the storage system, ensure that all processes and applications that access the LUN are stopped and nothing is accessing one or more storage systems. If any process is still accessing the removed LUN, an emergency shutdown may occur, resulting in loss of access.

1. Stop all files access to the LUN, which needs to be removed and unmount the files system, if applicable.

   If LVM was used on one or more storage systems from which the LUN is being removed, make sure that no Logical Volumes, Volume Groups or Physical Volumes are still present or in use of one or more relevant storage systems. To ensure the above function, run the following commands:

   - **man lvremove**
   - **man remove**
   - **man pvremove**

2. To verify that all files are closed, enter the **lsof/dev/nvme*n*** commands. The output must be empty.
3. Place the LUNs on the storage side. From the storage system, unmap the volume from the host that uses the GUI.
   a. From the left-side panel, open the **Hosts** tab.
   b. Right-click on the hostname and select **Modify Volume Mappings**.
   c. Select the volume to be unmapped and click **Remove Volume Mapping**.

Multipath configuration for FC-NVMe hosts

Follow FC-NVMe multipath configuration recommendations for successful attachment of Linux hosts to the system.

Turning Native NVMe Multipath on or off

Hosts can be configured to work with the traditional Device Mapper or with the Native NVMe Multipath as follows:

- On SLES15, Native NVMe Multipath is turned on by default. On SLES12SP4, Native NVMe Multipath is not turned on by default.

  To check whether Native NVMe Multipath is on, enter:

  ```
  # systool -m nvme_core -A multipath
  Module = "nvme_core"

      multipath           = "Y"
  ```

  If the multipath = "N", and you want to turn it on, enter:

  ```
  echo "options nvme_core multipath=Y" > /etc/modprobe.d/50-nvme_core.conf
  dracut -f
  reboot
  ```

  Note: On SLES12SP4, if the multipath is not turned on even after reboot, then follow these steps:

  1. Edit the file `/etc/default/grub`.
  2. Append **nvme-core.multipath=on** at the end of the GRUB_CMDLINE_LINUX_DEFAULT variable.
  3. Save the file.
  4. Run `grub2-mkconfig -o /boot/grub2/grub.cfg` to apply the new configuration.
  5. Reboot.

  If you choose to work with Device Mapper, enter:

  ```
  echo "options nvme_core multipath=N" > /etc/modprobe.d/50-nvme_core.conf
  dracut -f
  reboot
  ```

- On Red Hat Enterprise Linux 8.0 or later, Native NVMe Multipath is not turned on by default.

  To check whether Native NVMe Multipath is enabled in the command-line, enter:

  ```
  # cat /sys/module/nvme_core/parameters/multipath

      Y
  ```

  If the output is "N", and you want to enable it, follow these steps:

  – Using command-line:

1. Add the `nvme_core.multipath=Y` option on the command-line:

```
# grubby --update-kernel=ALL --args="nvme_core.multipath=Y"
```

2. On the 64-bit IBM Z architecture, update the boot menu:

```
# zipl
```

3. Reboot the system.

– Using module configuration file:

1. Create the `/etc/modprobe.d/nvme_core.conf` configuration file with the following content:

```
options nvme_core multipath=Y
```

2. Back up the `initramfs` file system:

```
# cp /boot/initramfs-$(uname -r).img \
     /boot/initramfs-$(uname -r).bak.$(date +%m-%d-%H%M%S).img
```

3. Rebuild the `initramfs` file system:

```
# dracut --force --verbose
```

4. Reboot the system.

If you choose to work with Device Mapper, enter:

1. Remove the `nvme_core.multipath=Y` option from the command-line:

```
# grubby --update-kernel=ALL --remove-args="nvme_core.multipath=Y"
```

2. On the 64-bit IBM Z architecture, update the boot menu:

```
# zipl
```

3. Remove the `options nvme_core multipath=Y` line from the `/etc/modprobe.d/nvme_core.conf` file, if it is present.
4. Reboot the system.

For more details to enable native NVMe and Device Mapper on Red Hat Enterprise Linux 8.0 or later, see Enabling multipathing on NVMe devices.

Device Mapper or Native NVMe Multipath configuration

Two framework options are available to configure the multipath configuration file: Device Mapper or Native NVMe Multipath.

To configure Device Mapper, complete the following steps:

1. Edit the `/etc/multipath.conf` file to include the following code:

```
devices {
    device {
        vendor "NVME"
        product "IBM     2145"
        path_grouping_policy "multibus"
        path_selector "round-robin 0"
        prio "ANA"
        path_checker "none"
        failback "immediate"
        no_path_retry "queue"
        rr_weight uniform
        rr_min_io_rq "1"
        fast_io_fail_tmo 15
        dev_loss_tmo 600
    }
}
defaults {
```

```
    user_friendly_names yes
    path_grouping_policy    group_by_prio
}
```

2. Run the following commands to validate that the multipath daemon is running:

```
systemctl enable multipathd.service
systemctl start multipathd.service

# ps -ef | grep -v grep | grep multipath
root      1616     1  0 Nov21 ?        00:01:14 /sbin/multipathd -d -s
```

   After you enable the multipath service for SLES12SP4, rebuild `initrd` with multipath support:

```
dracut --force --add multipath
```

3. Run the following commands to apply configurations:

```
multipath -F
multipath
multipath -ll
```

If you use Native NVMe Multipath, add the following code to the multipath configuration file:

```
blacklist {
    device {
        vendor  "NVME"
        product "IBM\s+2145"
    }
}
```

Device Mapper performance

Testing shows that performance is superior when you work with multiqueue I/O scheduling as described in the following steps.

Note: Before you apply these steps, contact a SUSE representative to verify that this option is supported.

For performance improvements, apply multiqueue I/O scheduling with blk-mq (in SLES):



Warning: Do not change **`nvme-core.multipath`** value when Native NVMe is in use.

1. Edit the `/etc/default/grub` file and the following text:

```
GRUB_CMDLINE_LINUX_DEFAULT="BOOTPTimeout=20 BootpWait=20
biosdevname=0 powersaved=off resume=/dev/system/swap splash=silent
quiet showopts crashkernel=175M,high dm_mod.use_blk_mq=y
scsi_mod.use_blk_mq=1 transparent_hugepage=never"
```

2. Apply the new configuration:

```
swfc178:~ # grub2-mkconfig -o /boot/grub2/grub.cfg
Generating grub configuration file ...
Found theme: /boot/grub2/themes/SLE/theme.txt
Found linux image: /boot/vmlinuz-4.12.14-25.19-default
Found initrd image: /boot/initrd-4.12.14-25.19-default
Found linux image: /boot/vmlinuz-4.12.14-23-default
Found initrd image: /boot/initrd-4.12.14-23-default
done
```

3. Reboot.
4. Validate that the multiqueue feature is enabled in multipath:

```
mpatho (eui.880000000000000b0050760071c60044) dm-3 NVME,IBM     2145
size=1.5G features='3 queue_if_no_path queue_mode mq' hwhandler='0' wp=rw
```

VMware ESXi installation and configuration for NVMe over Fibre Channel hosts

VMware ESXi contains two levels of connectivity.

The first level of VMware ESXi is the host (ESXi machine), which is the actual server that includes the host bus adapter and the NVMe SCSI support. The host manages both hardware resource and virtualized components by using the GUI and CLI (esxcli).

The other level includes one or more guest virtual machines with an operating system. The virtual machines detect the defined volumes as VMware datastores and can read or write to them.

Configuring the ESXi operating system

You must configure the VMware ESXi 8.0 and up operating system before you can connect hosts that run VMware ESXi datastores with the system.

Before you configure the host operating system, the following tasks must be completed:

- Your system must be installed.
- You must install the appropriate host bus adapters and ensure that the latest drivers and firmware are applied.

After the prerequisite tasks are complete, use the following general steps to configure your host.

1. Make sure you use Emulex Gen6 HBA minimal supported versions.
   - Driver version: 12.4.293.2
   - Firmware version: 12.4.243.17

   Enter the following command to verify:

   ```
   esxcli storage san fc list
   ```

2. Make sure that FC-NVMe is enabled by setting lpfc_enable_fc4_type to 3.

   To change the value, enter the following command:

   ```
   esxcli system module parameters set -m lpfc -p lpfc_enable_fc4_type=3
   ```

3. Make sure all port of the ESXi hosts and system ports are zoned correctly.
   To see the WWPNs of FC-NVMe, enter the following command:

   ```
   esxcli storage core adapter list
   ```

   The NVMe WWPNs have the driver name `lpfc`.
4. Define the host object by using the NQN identifiers for each physical ESXi host in the VMware vSphere cluster.
   You can find the NQN by entering the following command:

   ```
   esxcli nvme info get
   ```

5. If more than one physical ESXi host is present in the vSphere cluster, create a host cluster and add each ESXi host object as a member.
6. After you create volumes on the system and map them to the host, do VMware storage rescan.
   If rescan does not work, try resetting the Fibre Channel ports. To reset the ports, enter the following command:

   ```
   esxcli storage san fc reset -A <HBA name>
   ```

7. Create VMware datastores based on the system mapped volumes.

NVMe hosts that run the Microsoft Windows operating system

Ensure that your system meets the requirements for attaching to a host that is running the Microsoft Windows operating system.

The following list provides the requirements for attaching the system to the host that is running the Windows operating system:

- Before you install the Emulex HBA Manager application, verify the correct level with Broadcom. At the time of publication, `OneInstall-Setup-12.6.240.40.exe`, which includes driver level 12.6.165.0, was used.
- Install the Emulex HBA Manager application from Broadcom. Under **Installation options**, select the **FC Driver** checkbox to include the driver in the installation. Also, select the correct **Emulex HBA Manager application** kit checkbox.

Configuring the Windows operating system for FC-NVMe hosts

After you ensure that your system meets the requirements for attaching to a Microsoft Windows host, configure the Windows operating system.

You must install the appropriate host bus adapters with the correct firmware and driver levels that support FC-NVMe.

1. Zone the host ports to the NVMe ports on the system.

   For more information about how to identify NVMe ports on the system, see the CLI host commands.
2. Find the Windows host NVMe Qualfied Names (NQNs).

   The NQNs of the Windows host displays as a string composed from a constant and the WWPN address of the host port. The constant is `nqn.2017-01.com.broadcom:ecd:nvmf:fc:` plus add the port WWPN.
3. On the system, create the Windows host by using the Windows ports NQN. The system supports up to 2 NQN addresses per host object, which are separated by a comma. For instructions about creating a host, refer to the host commands information.
4. On the system, map relevant volumes to the Windows NVMe host. The same volumes cannot be mapped concurrently to both NVMe and SCSI.
5. Open Emulex HBA Manager application.

   Select a port to configure and click **Driver Parameters**. Change the following defaults and repeat for each port:

   a. For **`EnableNVMe`**, set to 1 to enable NVMe support.

   b. For **`NumNVMENode`**, set to the maximum value of 512.

   c. For **`NumNVMENS`**, set to the number of volumes that are mapped or to the maximum value of 255.
6. Apply the settings and restart the Windows host.
7. Either validate the correct firmware level or select **Download Firmware** to update.

Multipath configuration for FC-NVMe Windows hosts

Follow FC-NVMe multipath configuration recommendations for successful attachment of Microsoft Windows hosts to the system.

1. On **Windows server manager** > **Tools**, MPIO might need to be added. You can add MPIO either through **Manage** > **Add Roles/Features** or through the **`powershell`** command **Add-WindowsFeature -Name 'Multipath-IO'**.
2. In the **MPIO** window, select **Discover Multi-Paths**. Under **Others**, add `IBM NVMe 2145`. This process might take a while.
3. Restart the Windows host.
4. You can view the multipath device by entering the **`powershell`** command `mpclaim -s -d`.

IBM Power Systems AIX

You can attach an NVMe host to an IBM Power Systems AIX host by following specific requirements and instructions.

Attaching to IBM Power Systems AIX hosts

For interoperability information about attachments between the system and an IBM Power Systems AIX host, check the IBM support website: http://www.ibm.com/support

Note: The IBM Power Systems information applies to all AIX hosts that are listed on the interoperability support site.

Attachment requirements for IBM Power Systems hosts

Ensure that your system meets the requirements for attaching to a host that is running the AIX operating system.

Before you attach an IBM Power Systems host, ensure that you meet the following prerequisites:

- You installed the correct operating systems and version levels on your host, including any updates and APARs (Authorized Program Analysis Reports) for the operating system.
- You have the documentation for your host system and you have access to the Installing section in this product information.

  All system publications are available from the following website:

  IBM

  https://www.ibm.com/docs/en/aix

At the time of writing, the supported servers are POWER9 servers with minimum firmware version of 950.

AIX environments for IBM Power Systems hosts

Ensure that each IBM Power Systems host uses a supported operating system and level.

The following website provides current interoperability information about the supported operating system levels for IBM Power Systems hosts:

https://www.ibm.com/docs/en/aix

At the time of writing, the supported operating system is AIX 7.2 TL 5 from 2035_72V and above.

Drivers and firmware for IBM Power Systems hosts

Ensure that you use the correct host bus adapter device driver and firmware levels for yourIBM Power Systems AIX hosts.

For POWER9 hosts, the minimum firmware level is 950.

The following website provides current interoperability information about device driver and firmware levels:

https://www.ibm.com/docs/en/aix

Configuring the AIX operating system

You must configure the AIX operating system before you can use IBM Power Systems hosts with the system.

Before you configure the AIX host operating systems, the following tasks must be completed:

- The system must be installed.
- The appropriate host bus adapters must be installed.

After the prerequisite tasks are complete, use the following general steps to configure your AIX host system.

1. Zone the host to the system nodes on the Fibre Channel (FC) SAN.
2. Verify that the FC ports appear in AIX. Each FC port appears through the **lsdev** command with a name starts with *fcs*.

```
bash-4.4# lsdev | grep fcs
fcs0       Available 00-00       PCIe3 2-Port 32Gb FC Adapter (df1000e314101506)
fcs1       Available 00-01       PCIe3 2-Port 32Gb FC Adapter (df1000e314101506)
```

   Note: If *fcs* device is not found, you can run the **cfgmgr** command.
3. Check for FC-NVMe device availability.

   Note: AIX creates children devices of the *fcs* device for each protocol that AIX supports.

   a. The FC-NVMe protocol appears through the **lsdev** command with a name starts with *fcnvme*.

```
bash-4.4# lsdev | grep fcnvme
fcnvme0    Available 00-00-02    FC-NVMe Protocol Device
fcnvme1    Available 00-01-02    FC-NVMe Protocol Device
```

   b. If the FC-NVMe device is not detected, you can enable it by using a **smitty** command for each port.

```
smitty -> Devices -> FC Adapter -> FC-NVMe Device -> Enable a FC-NVMe Protocol Device
```

4. Find the host NVMe Qualified Name (NQN) from each FC-NVMe port.

```
bash-4.4# lsattr -El fcnvme0
attach     switch      How this adapter is connected  False
autoconfig available   Configuration State            True
host_nqn   nqn.2018-04.com.ibm:aix.fc.pn.0x100000109ba9d1c9.nn.0x0x200000109ba9d1c9   Host
NQN (NVMe Qualified Name) True
```

5. Create FC-NVMe host type in target.

   Note: In AIX, each port has its own NQN. You can configure two NQNs to add a host to the target. If you have more ports, you can work on the host cluster, where each pair is defined as `host`.
6. Create a volume and map to the FC-NVMe host object.
7. Run the **cfgmgr** command without any parameters.
8. To check the FC-NVMe paths and disks, run the following command:

```
lsdev | grep -i nvme
```

   The NVMe Discovery Controller and the NVMe Dynamic Controller appear for each path. The following example shows the four paths between the host and the target for the NVMe Discovery Controllers and the NVMe IO dynamic Controllers:

```
nvme0 Available 00-00-02 NVMe Discovery Controller
nvme1 Available 00-00-02 NVMe Discovery Controller
nvme2 Available 00-01-02 NVMe Discovery Controller
nvme3 Available 00-01-02 NVMe Discovery Controller
nvme4 Available 00-00-02 NVMe Dynamic Controller
nvme5 Available 00-00-02 NVMe Dynamic Controller
nvme6 Available 00-01-02 NVMe Dynamic Controller
nvme7 Available 00-01-02 NVMe Dynamic Controller
```

   The following example shows the mapped volumes as hdisk:

```
hdisk1 Available 00-00-02 NVMe Flash Disk
hdisk2 Available 00-00-02 NVMe Flash Disk
hdisk3 Available 00-00-02 NVMe Flash Disk
hdisk4 Available 00-00-02 NVMe Flash Disk
hdisk5 Available 00-00-02 NVMe Flash Disk
hdisk6 Available 00-00-02 NVMe Flash Disk
hdisk7 Available 00-00-02 NVMe Flash Disk
hdisk8 Available 00-00-02 NVMe Flash Disk
```

```
hdisk9 Available 00-00-02 NVMe Flash Disk
hdisk10 Available 00-00-02 NVMe Flash Disk
```

NVMe over RDMA and NVMe over TCP host attachments

You can attach NVM Express (NVMe) over RDMA or NVMe over TCP hosts to the system.

Note: NVMe over RDMA or NVMe over TCP is not supported for direct attach host connection.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Linux

You can attach NVM Express (NVMe) over RDMA or NVMe over TCP hosts that run the Linux operating system.

Attachment requirements for hosts that are running Linux

Ensure that the host is running a supported operating system and adapter cards. Refer to vendor documentation if updates are needed.

Note: For more information about host supported level, see IBM System Storage Interoperation Center (SSIC).

NVM-Express user space tools for Linux are required. Install these tools by using the following commands:

- SLES: **zypper install nvme-cli**
- RHEL: **yum install nvme-cli**

Use the following version of the nvme-cli tool:

- On SLES, only nvme-cli version 1.13 is supported.
- On RHEL 8.4, use nvme-cli version 1.12.
- On RHEL 8.6 or later, use nvme-cli version 1.16.

Configuring the host attachment

1. For connectivity and performance benefits, the following sysctl settings on host are recommended:

   net.ipv4.conf.all.arp_filter is enabled
   net.ipv4.conf.all.arp_ignore is set to 2
   TCP Delayed ACK is enabled

   Refer to Operating System documentation for configuration instructions.
2. Use the following command to identify the host NVMe Qualified Name (NQN) address:

   ```
   cat /etc/nvme/hostnqn
   ```
3. On the storage system, create the NVMe host object using the **Hosts** panel on the GUI, or the **mkhost** command on the CLI.
4. On the storage system, map volumes to the NVMe host using the **Volumes** panel on the GUI, or the **mkvdiskhostmap** command on the CLI.
5. To discover and connect to NVMe targets on the host, use the **nvme discover** and **nvme connect** commands. Refer to your operating system documentation for more information.

VMware ESXi

You can attach NVM Express (NVMe) over RDMA or NVMe over TCP hosts that run the VMware ESXi operating system.

VMware ESXi contains two levels of connectivity. The first level of VMware ESXi is the host (ESXi machine), which is the actual server that includes the host bus adapter and the NVMe support. The host manages both hardware resource and virtualized components by using the GUI and CLI (esxcli).

The other level includes one or more guest virtual machines with an operating system. The virtual machines detect the defined volumes as VMware data stores and can read or write to them.

Configuring the VMware ESXi

Ensure that the host is running a supported operating system and adapter cards. Refer to vendor documentation if updates are needed.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Table 30. NVMe over RDMA minimum supported Ethernet cards and firmware versions for VMware ESXi 7.0 UP2

	CX4	**CX5**	**CX6**	**N2100G**
Firmware version	14.26.1040	16.32.1010	14.26.1040	219.0.144.0
Ethernet Card version	4.21.71.101-1OEM.702.0.0.17630552	4.21.71.101-1OEM.702.0.0.17630552	4.21.71.101-1OEM.702.0.0.17630552	219.0.29.0-1OEM.700.1.0.15843807 219.0.13.0-1OEM.700.1.0.15843807

For more information about Ethernet cards and latest firmware levels, see ConnectX Ethernet Cards for VMware ESXi Server and Broadcom Ethernet Network Adapters.

Configure the VMware ESXi 7.0 or ESXi 8.0 operating system before connecting hosts that run VMware ESXi data stores with the system.

Any single host NQN is only permitted to have a single connection to any target IP address. If a host is using multiple initiators, they should be configured on separate subnets with their intended target IP addresses.

Configure the adapter software

Before configuring the host operating system, the following tasks must be completed:

1. If you're using RDMA or TCP with this ESXi for the first time, enable the adapter software from the VMware ESXi GUI.
   a. Verify that Adapter MTUs are consistently set end-to-end.
   b. Validate that RDMA vSphere Installation Bundles (VIBs) are loaded with the correct vSphere version (OS dependent).
   c. Create a standard virtual switch per RDMA or TCP adapter.
      - For more information about configuring RDMA adapters, see **Configure Adapters for NVMe over RDMA (RoCE v2)** and **Configure VMkernel Binding with a vSphere Standard Switch** in VMware vSphere Storage product documentation.
      - For more information about configuring TCP adapters, see **Configure Adapters for NVMe over TCP Adapter Storage** and **Configure VMkernel Binding for a TCP Adapter with a vSphere Standard Switch** in VMware vSphere Storage product documentation.
   d. It is recommended that delayed ack be enabled. Refer to VMware ESXi vendor guides for more details.

Configure the host software

After the prerequisite tasks are complete, use the following general steps to configure your host.

1. Define the host object by using the NQN identifiers for each physical ESXi host in the VMware vSphere cluster.

   Find the NQN using the following command:

   ```
   esxcli nvme info get
   ```

2. Discover the vmhba value to identify each RDMA or TCP adapter.

   ```
   esxcfg-scsidevs -a
   ```

3. Discover and connect to the storage controller.

   - Identify the discovered controller through command:

     ```
     esxcli nvme controller list
     ```

   - To discover and connect use commands:

     a. `esxcli nvme fabrics discover -i <ip_address_of_target_rdma/tcp> -a <vmhba> -p 4420`

     b. `esxcli nvme fabrics connect -i <ip_address_of_target_rdma/tcp> -a <vmhba> -p 4420 -s <subsystem-nqn> -c <controller-id> -t <keep-alive-timeout>`

     where <controller-id> is an optional parameter.

4. Run the following command to validate that the multipath and device mapper has loaded successfully:

   ```
   esxcli storage core plugin list
   ```

   Validating with High Performance Path (HPP) plug-in for NVMe:

   ```
   esxcli storage core claimrule list -c MP
   Rule Class   Rule  Class    Type       Plugin
   Matches                                                                  XCOPY Use Array
   Reported Values  XCOPY Use Multiple Segments  XCOPY Max Transfer Size KiB  Config String
   ----------  -----  -------  ---------  ---------
   ------------------------------------------------------------------------
   -----------------------------  --------------------------
   ---------------------------  -------------
   MP             50  runtime  transport  NMP
   transport=usb
          false                          false                            0
   MP             51  runtime  transport  HPP
   transport=sata
          false                          false                            0  pss=FIXED
   MP             52  runtime  transport  NMP
   transport=ide
          false                          false                            0
   MP             53  runtime  transport  NMP
   transport=block
          false                          false                            0
   MP             54  runtime  transport  NMP
   transport=unknown
          false                          false                            0
   MP            101  runtime  vendor     MASK_PATH  vendor=DELL model=Universal Xport
   pci_vendor_id=* pci_sub_vendor_id=*
   false                         false                          0
   MP            101  file     vendor     MASK_PATH  vendor=DELL model=Universal Xport
   pci_vendor_id=* pci_sub_vendor_id=*
   false                         false                          0
   MP          65531  runtime  transport  HPP
   transport=sas
          false                          false                            0  pss=FIXED
   MP          65532  runtime  transport  HPP
   transport=parallel
          false                          false                            0  pss=FIXED
   MP          65533  runtime  transport  HPP
   transport=pcie
   ```

```
        false                           false                               0  pss=FIXED
MP          65534  runtime  vendor     HPP        nvme_controller_model=*
pci_vendor_id=* pci_sub_vendor_id=*
false                          false                          0
MP          65535  runtime  vendor     NMP        vendor=* model=* pci_vendor_id=*
pci_sub_vendor_id=*
false                          false                          0
```

Configure the attachment

To validate that the namespaces are retrieved, complete the following steps:

1. Run the following command:

   ```
   esxcli nvme namespace list
   ```

   This provides an empty output.
2. Map all volumes to the host by using the **mkvdiskhostmap** command or the management GUI. For more information about **mkvdiskhostmap** command, see **mkvdiskhostmap**.
3. Rerun the **esxcli nvme namespace list** command.

Chapter 7. Policies

A policy defines a set of rules that can be used to simplify storage management. The system supports several different types of policies.

Provisioning policies

A provisioning policy is an entity that defines a set of rules for allocating capacity from a pool. Any volumes that are created in the pool adopt the capacity-saving methods that are defined in the provisioning policy.

An administrator or security administrator can use a provisioning policy to define a set of rules that are applied when volumes are created within a storage pool. When a storage pool is assigned with a provisioning policy, all volumes that are created in that pool get automatically created with the capacity savings that are defined by the policy. Provisioning policies allow a simpler and consistent provisioning process that can be used with external automation software. The capacity savings for existing volumes do not change if the provisioning policy is removed from the storage pool or a different policy is applied. A provisioning policy defines only the capacity savings that are used when you create new volumes in the pool.

Provisioning policies support the following capacity-saving options:

None
Use the **none** option when you don't want volumes to use any additional capacity savings. With this option, volumes use only the built-in capacity saving functions of the storage pool. For example, pools containing certain drive types, such as FlashCore Module drives, are self-compressing.

Thin-provisioning
Use the thin-provisioning option to indicate that all volumes in a pool use thin-provisioning to save capacity. Thin-provisioning allocates capacity only when data is written to storage.

Compression
Use the compression option to indicate that all volumes in a pool use compression to save capacity. Compression reduces the capacity that data requires by removing unnecessary content from the data.

Deduplication
Use deduplication when you create a new provisioning policy to add deduplication to your capacity saving for a data reduction pool. Both compression and thin-provisioning capacity saving can be used with deduplication to increase capacity savings for data reduction pools. Standard pools do not support deduplication.

A provisioning policy can be renamed after the policy is created. However, the capacity savings option of a provisioning policy cannot be changed. To change the capacity savings to be used for new volumes, create and assign a new provisioning policy to the storage pool, replacing any existing policy. The old policy can be deleted if it is no longer required. A provisioning policy can be associated with multiple pools. However, a pool can only have one policy.

The capacity savings for a volume can be changed after a volume is created. A provisioning policy for a storage pool is only applied when new volumes or volume copies are created.

After adding a provisioning policy to a pool, or assigning a different provisioning policy, it is possible to use the **addvolumecopy** command to convert existing volumes to use the capacity savings of the new provisioning policy. The **-autodelete** parameter can be used to automatically delete the original copy of the volume after the new copy is synchronized.

Configuring provisioning policies

You can configure provisioning policies to simplify the management of capacity saving for volumes and pools.

To create a provisioning policy by using management GUI, complete these steps:

© Copyright IBM Corp. 2024

1. In the management GUI, select **Policies** > **Provisioning Policies**.
2. On the **Provisioning Policies** page, click **Create Provisioning Policy**.
3. On the **Create Provisioning Policy** page, specify the **Policy Name**.
4. Select the **Capacity Type** from the **Additional Capacity Savings**. The following capacity types can be selected:
   - **None**
   - **Thin-provisioned**
   - **Thin-provisioned and deduplicated**
   - **Compressed**
   - **Compressed and deduplicated**
5. Click **Create**.

You can use the command-line interface (CLI) to configure provisioning policies. To configure provisioning policies, complete the following steps:

1. To create a new provisioning policy that uses no additional capacity savings and uses built-in capacity saving functions, enter the following command.

   ```
   mkprovisioningpolicy -name <name> -capacitysaving none
   ```

   where *name* is the name of the new provisioning policy.
2. To create a new provisioning policy that uses thin provisioning, enter the following command.

   ```
   mkprovisioningpolicy -name <name> -capacitysaving thin
   ```

   where *name* is the name of the new provisioning policy.
3. To create a new provisioning policy that uses compression, enter the following command.

   ```
   mkprovisioningpolicy -name <name> -capacitysaving compressed
   ```

   where *name* is the name of the new provisioning policy.
4. To create a new provisioning policy that uses compression and deduplication, enter the following command.

   ```
   mkprovisioningpolicy -name <name> -capacitysaving compressed -deduplicated
   ```

   where *name* is the name of the new provisioning policy.

After you create provisioning policies, assign them to pools with the **mkmdiskgrp** or **chmdiskgrp** command.

To assign a provisioning policy to a new pool, you can use the **mkmdiskgrp** command with the following syntax:

- ```
  mkmdiskgrp -name <pool_name> -ext <extent_size> -provisioningpolicy <provisioning_policy_name
  | provisioning_policy_id >
  ```

  where the *pool_name* variable indicates the name of the pool. The *extent_size* variable indicates the size of extents. The *provisioning_policy_name* variable indicates the name of the provisioning policy.
- To assign a provisioning policy to an existing pool, you can use the **chmdiskgrp** command with the following syntax:

  ```
  chmdiskgrp -provisioningpolicy <provisioning_policy_name | provisioning_policy_id> pool_name
  ```

  where the *pool_name* variable specifies the name of the pool to apply the provisioning policy to.
  ```

After you create the pool, use the **mkvolume** command to add volumes to the pools. These volumes adopt the capacity saving method that is configured in the provisioning policy that is assigned to the pool.

Creating provisioning policy and assigning to pools

A provisioning policy simplifies the volume provisioning process by specifying capacity savings methods per pool.

Any volumes created in the pool are created using the capacity savings defined by the provisioning policy. A provisioning policy can be associated with multiple pools. However, each pool or child pool can be assigned only with a single provisioning policy. You can use management GUI and command-line interface to create provisioning policies.

If you are using policy-based replication, provisioning policy ensures that capacity savings for volumes are consistent between the production system and the recovery system.

Note: Provisioning policies is optional when using policy-based replication. However, if provisioning policies are not used, any replication-target volumes created automatically by the system will be created with no additional capacity savings.

Using management GUI

To create a provisioning policy by using management GUI, complete these steps:

1. In the management GUI, select **Policies** > **Provisioning Policies**.
2. On the **Provisioning Policies** page, click **Create Provisioning Policy**.
3. On the **Create Provisioning Policy** page, specify the **Policy Name**.
4. Select the **Capacity Type** from the **Additional Capacity Savings**. The following capacity types can be selected:
   - **None**
   - **Thin-provisioned**
   - **Thin-provisioned and deduplicated**
   - **Compressed**
   - **Compressed and deduplicated**
5. Click **Create**.

To assign a provisioning policy to a parent pool or a child pool, complete these steps:

1. In the management GUI, select **Pools** > **Pools**.
2. On the **Pools** page, right-click on the selected pool and click **Assign Provisioning Policy**.
3. On the **Assign Provisioning Policy** page, select the policy to assign to the pool. When you assign a provisioning policy to a storage pool, the management GUI displays provisioning policies with compatible capacity savings types for the selected pool. For example, for standard pools, only the none and thin-provisioned options are compatible. For data reduction pools that do not contain compressing drives, all five capacity saving options are available. For data reduction pools that contain compressing drives, only the none or compressed and deduplicated capacity savings options are available.
4. Click **Assign**.

Using command-line interface

To create a provisioning policy from command-line interface, use the following command:

```
mkprovisioningpolicy -name <name> -capacitysaving <none|thin|deduplicated|compressed>
```

where the **-name** command specifies the provisioning policy name. The **-capacitysaving** specifies the capacity-saving methods, such as *thin*, *deduplicated*, *none*, and *compressed* that you can use on creating volumes in a pool.

Managing provisioning policies

A provisioning policy of a pool defines how to provision new volume capacity. Use the management GUI interface or command-line interface to manage provisioning policies.

With a provisioning policy, an administrator or security administrator can define the new volume capacity to be provisioned within a pool.

For existing pools, you can assign new provisioning policies. Any new volumes that are created and assigned to the pool automatically adopt the capacity savings method that is defined in the provisioning policy. Older volumes in the pool retain their capacity savings method that was defined on the volume when it was created.

You can migrate existing volumes in the pool to use a new provisioning policy by creating a copy of the volume within the pool. When the volume is migrated, you can delete the original copy of the volume from the pool.

Migrating volumes to use provisioning policies

A provisioning policy defines how new volume capacity must be provisioned within a pool. You can migrate volumes to use the provisioning policy of a pool with the **addvolumecopy** command. The volume migrates to take on the characteristics of the provisioning policy associated with the pool. The old volume copy can be deleted automatically.

You can use the command-line interface (CLI) to migrate volumes to use a provisioning policy. To migrate a volume to use a provisioning policy, complete the following steps:

1. To view provisioning policies that are configured on the system, enter the following command:

   ```
   lsprovisioningpolicy
   ```

   Verify the capacity savings method and whether the deduplication flag is assigned. Deduplication is only supported on data reduction pools and can be added to both capacity savings methods. Standard pools support only compression and thin-provisioning.
2. To create a new pool that uses the specified provisioning policy, enter the following command:

   Standard pool
   To create a standard pool, enter the following command:

   ```
   mkmdiskgrp -name pool_name -ext extent_size -provisioningpolicy policy_name | policy_id
   ```

   where *policy_name* | *policy_id* is the name of the provisioning policy to assign to the standard pool.

   Data reduction pool
   To create a data reduction pool, enter the following command:

   ```
   mkmdiskgrp -name pool_name -ext extent_size -datareduction yes -provisioningpolicy
   policy_name | policy_id
   ```

   where *policy_name* | *policy_id* is the name of the provisioning policy to assign to the data reduction pool.
3. To create a copy of the original volume in the new pool that uses the provisioning policy, enter the following command:

   ```
   addvolumecopy -pool pool_name -autodelete volume_name
   ```

   where *-pool name* is the new pool name where the provisioning policy is assigned. A volume copy is created in the new pool and the newly created volume copy takes all the characteristics of the policy that is associated with that pool.

Note: The preceding example uses the **-autodelete** parameter to indicate that the system automatically deletes the source volume of the copy. If you do not specify this parameter, you can delete the original volume copy manually after the migration process has completed.

Snapshot policies

A *snapshot policy* is a set of rules that controls the creation and expiration of snapshots.

A snapshot policy has the following properties:

Backup interval
This is the interval between snapshot backups. The minimum value is 60 minutes.

Retention period
The length of time which snapshots are retained. The maximum length is 365 days.

With snapshot policies, administrators can schedule the creation of snapshots for volumes in a volume group at specific intervals and retain based on their security and recovery point objectives (RPO). A snapshot policy has following properties:

- It can be assigned to one or more volume groups.
- Only one *snapshot policy* can be scheduled to one volume group.
- The system supports a maximum number of 40 snapshot policies.

The system supports an internal scheduler to manage and create snapshot policies on the system. The system provides predefined polices that contain specific retention and frequency values for common use-cases. The system also allows creation of the user-defined policies.

The following predefined policies are supported:

Predefinedsspolicy0
Select this policy for the most frequent copies and retention. For this policy, snapshots are created every six hours and retained for a week. Use this policy for volume data that requires the highest recovery point objective (RPO). For example, volume data that is frequently updated and critical to your business can benefit from frequent copies and retention. Customer accounts, orders, or proprietary information are examples of data that can need more frequent backups. For more information, refer to your organization's business continuity plan.

Predefinedsspolicy1
Select this policy for less frequent copies and medium retention. For this policy, snapshots are created weekly and retained for a month. Use this policy for application data that is updated frequently and requires a high RPO, but might not contain business-critical data.

Predefinedsspolicy2
Select this policy for less frequent copies and longer retention. For this policy, snapshots are created monthly and retained for a year. Use this policy for older data that is not updated frequently but still requires retention, such as past customer accounts or employee records.

Predefinedsspolicy38
Select this policy to create cloud snapshots only. For this policy, cloud snapshots are created daily and retained for a month. Use this policy for storing snapshots of the volumes to cloud storage.

Predefinedsspolicy39
Select this hybrid policy to create schedules for local snapshots and cloud snapshots. Select this policy for creating both, local snapshots and cloud snapshots. For this policy, local snapshots are created every six hours and retained for a week. It also creates cloud snapshots daily and are retained for a month. Use this policy for volume data that requires the highest recovery point objective (RPO) and for storing snapshots of the volumes to cloud storage.

Note: Snapshots can also be added and managed with external schedulers like IBM Copy Services Manager.

Managing snapshot policies

An administrator or security administrator can create, define, and assign a snapshot policy to a volume group.

Assigning snapshot policies

Policies can be created and assigned to volume groups from the **Policies** > **Snapshot Policies** tab of the management GUI.

If the backup start time is in the past, it will be advanced according to the backup interval until a time in the future is reached. A monthly scheduled snapshot creates snapshot on the same day of each month for days 1-28 based on the start time. If the scheduled day is not available in that month, the last day of the month is considered.

For example, if snapshot policy is specified for every month and the start date is 31 May, the next snapshot date would be 30 June, followed by 31 August. Daylight savings do not affect the backup and retention intervals that are specified by the schedule. For example, if snapshots are triggered every 24 hours and occur every day at 2:00 AM before daylight savings start, they are created at 1:00 AM when daylight savings pushes the time ahead by an hour. The backup interval remains constant at 24 hours in this example.

The command-line interface provides the following command to assign snapshot policies:

- **mksnapshotpolicy** creates a snapshot policy.
- **mkvolumegroup** assigns a snapshot policy to a volume group at the time of creation.
- **chvolumegroup** changes or assigns the snapshot policies on an existing volume group.

If you assign a policy to a volume group that already has a policy associated with it, the newly assigned policy replaces the existing policy.

Deleting snapshot policies

Use the **rmsnapshotpolicy** command in the command-line interface to delete the user-defined snapshot policies.

The predefined policies cannot be deleted.

Security administration privileges are required to delete policies that are associated with volume groups.

Changing snapshot policies

Use the **chsnapshotpolicy** command in the command-line interface to change the name of the user-defined snapshot policy. The name can also be changed from the **Policies** > **Snapshot Policies** tab in the management GUI.

Suspending snapshot policies

If the source volumes are compromised during a cyberattack or during a disaster recovery scenario, any snapshot taken during an attack can contain malicious data or software. The system supports suspending snapshot policies until the compromised data can be determined. This can be done at a volume group level by using the **Volumes** > **Volume Groups** panel on the GUI, or the **chvolumegroupsnapshotpolicy** command.

Alternatively, all snapshot policies on the system can be suspended using the **chsystem** command. Snapshots can be used for cloning or restoring while any snapshot policies are suspended.

Replication policy

A replication policy defines how replication is configured between systems.

A replication policy is defined between two systems in a partnership that supports policy-based replication. For more information, see “Replication policies” on page 215.

Chapter 8. Replication

Policy-based replication can be used to implement high availability and disaster recovery solutions between systems in different locations.

The possible configurations include:

High availability
Active-active high availability between two independent systems.

Disaster recovery
Asynchronous replication to another system for disaster recovery.

High availability and disaster recovery
Active-active high availability between two independent systems, with asynchronous replication to a third system for disaster recovery.

Note: On some platforms, software features, or combinations of software features are limited by available hardware. If applicable, the limitations are documented in a *Software feature limits* page in the relevant Hardware Guide.

High availability

Policy-based high availability provides an active-active high availability (HA) solution using two independent systems in different locations. The storage will automatically remain accessible to hosts if there is an event that makes one of the systems unavailable.

Storage partitions are used to implement policy-based high availability and disaster recovery solutions. Partitions contain volumes, volume groups, hosts, and host-to-volume mappings. For more information, see Storage partitions.

When an HA replication policy is associated with a partition, all the objects contained in the partition are automatically configured across both systems that are defined in the replication policy. Hosts that are created in a partition will be able to discover paths to the mapped volumes through both systems. Hosts that are configured with a location prioritize the use of a co-located system for I/O.

Using policy-based high availability, everything in a storage partition is highly available. When required, both management and I/O failover between systems to maintain access.

Each partition that is associated with an HA replication policy has two properties:

- *preferred management system*
- *active management system*

The active management system is the system from which all configuration tasks must be performed. In the event of a failure that affects the current active management system, the active management system role will automatically failover to the other system in the partnership.

The preferred management system is the system that you would like to be the active management system under ideal conditions and is the system that is preferred to keep the partition accessible if the two HA systems are disconnected. The preferred management system can be changed from the current active management system. Different partitions can be configured with different preferred management systems.

If the active management system and the preferred management system are not the same system, the system will automatically failover the active management system role to the preferred management system when it is able to.

As part of the high availability solution, a partnership IP quorum application must be deployed on a server that is separate from the storage systems. The partnership IP quorum application is used to arbitrate between systems in a partnership, to determine which system will operate as the active management system in the event of a failure.

© Copyright IBM Corp. 2024

Related concepts

"IP quorum application " on page 493
The IP quorum application is a Java™ application that runs on a server that is separate from the storage system.

Related information

"Partnerships" on page 213
Partnerships are used to connect systems together to enable migration, data replication, and high-availability solutions.

"Policies" on page 195
A policy defines a set of rules that can be used to simplify storage management. The system supports several different types of policies.

"Storage partitions" on page 243
Storage partitions are used to implement policy-based high availability and disaster recovery solutions. Partitions on different systems can be linked to configure replication for disaster recovery between the systems. Replication policies can be assigned to individual volume groups, enabling different policies to be used for different groups within the same partition.

Planning high availability

When you plan for replication for high availability, check that your environment supports it and that you understand the limits of configuration.

The following configuration limits apply to policy-based high availability:

- A limit of one I/O group can be configured in the system.
- A maximum of one partnership can be used for high availability.
- A partnership cannot be used for replication for both high availability and disaster recovery.
- Only Fibre Channel SCSI hosts can be configured in a storage partition that is configured for high availability.
- Host clusters cannot be used for hosts that are configured in a storage partition that is configured for high availability.
- Only the *standard* system topology is supported.
- SCSI Persistent Reserve commands are not supported.
- Volumes configured for high availability cannot be mapped to other IBM Storage Virtualize systems.

Policy-based high availability requires a partnership to be configured between the systems using either Fibre Channel or Ethernet (short distances using RDMA) connectivity.

The systems should be sufficiently far apart such that anticipated disruptions affect only one location. The systems must be sufficiently close together to not cause the response time for I/O to exceed the maximum that an application can tolerate because of latency in replication link communications. For policy-based high availability, a maximum of 1 millisecond round-trip time (RTT) is supported for the inter-site link between systems.

IP connectivity is required between systems for management. The following TCP ports are required to be accessible on the system management IP address:

- Port 7443 for REST API access between systems.
- Replication setup using the management GUI requires access to both systems from the host where the web browser is running. Ensure that any firewalls between the web browser and storage systems allow inbound traffic to port 7443 and 443 on the system management IP address.

In addition, both systems must have IP network connectivity to an external partnership IP quorum application to arbitrate if a site or system failure occurs. Multiple quorum applications can be deployed for redundancy.

Hosts must use an ALUA-compliant multipath policy. Policy-based high availability supports a reduced set of the operating systems supported by IBM Storage Virtualize systems. For more information, see Supported host operating systems for high availability.

The following restrictions apply to policy-based high availability:

- The name of a volume group cannot be changed while the volume group is in a storage partition that has a replication policy assigned.
- Policy-based high availability cannot be used with volumes that are:
  - Image mode
  - Cache disabled
  - Configured to use Transparent Cloud Tiering (TCT)
- A volume in a storage partition with a replication policy assigned cannot be:
  - Renamed
  - Resized (expanded or shrunk)
  - Migrated to image mode, or have an image mode copy added

Implementing high availability

Policy-based high availability can be configured starting from the **Storage partitions** panel in the management GUI, navigating to the partition to be modified.

The GUI guides you through the required steps to:

- Configure a partnership between systems, or select an existing partnership.
- Link storage pools between systems.
- Configure a partnership IP quorum application.

Configuration and monitoring can also be performed by using the REST API, or the CLI.

Creating a new high availability configuration by using the CLI

It is recommended that policy-based high availability (HA) is configured and managed by using the GUI. Configuration can also be performed by using the REST API or the CLI.

Prerequisite to configure your systems to support the High Availability solution:

- If not already configured, follow the link to set up a partnership: “Partnerships” on page 334.
- If not already configured, follow the link to set up a linked storage pool: “Pool links” on page 214.
- If not already configured, follow the link to set up a partnership IP quorum application: “IP quorum application ” on page 493.

Follow the steps to create a new policy-based high availability (HA) configuration by using command-line interface:

1. To create a replication policy with the *2-site-ha* topology, enter the following command.

   ```
   mkreplicationpolicy -topology 2-site-ha -location1system <name or ID of one system>
   -location1iogrp 0 -location2system <name or ID of other system> -location2iogrp 0
   ```

   if necessary, use **lspartnership** to get the name or ID of each system. For more information, see **mkreplicationpolicy**.
2. To create a Storage Partition and associate it with the replication policy, enter the following command.

   ```
   # Creating a single-site storage partition that is later configured for HA
   mkpartition -name my_partition
   ...
   chpartition -replicationpolicy <HA policy ID or name> my_partition
   ```

```
# Creating an HA storage partition
mkpartition -name ha_partition -replicationpolicy <HA policy ID or name>
```

For more information, see **mkpartition**.

3. To create a logical host object that is maintained across both systems by specifying that the host should be associated with the Storage Partition, enter the following command.

```
mkhost -partition <storage partition ID or name>...
```

For more information, see **mkhost**.

Note: To prioritize a colocated system for I/O, specify a location for the host.

4. To create a volume group that is maintained across both systems by specifying that the volume group should be associated with the Storage Partition, enter the following command.

```
mkvolumegroup -partition <storage partition ID or name>...
```

For more information, see **mkvolumegroup**.

5. To create a volume in this volume group, enter the following command.

```
mkvolume -volumegroup <volume group ID or name>...
```

For more information, see **mkvolume**.

6. To map the created volume to the created host, enter the following command.

```
mkvdiskhostmap -host <host ID or name> <volume ID or name>...
```

For more information, see **mkvdiskhostmap**.

7. Hosts can now discover paths to volumes at both locations. You can monitor and change the active and preferred management systems using the **lspartition** and **chpartition** commands.

Adding high availability to an existing configuration by using the CLI

It is recommended that policy-based high availability (HA) is configured and managed by using the GUI. The GUI helps you create a new configuration or modify an existing configuration. You can also configure policy-based HA by using the REST API or the CLI.

Prerequisites to configure your systems to support the HA solution:

- If not already configured, follow “Partnerships” on page 334 to set up a partnership.
- If not already configured, follow “Pool links” on page 214 to set up a linked storage pool.
- If not already configured, follow “IP quorum application ” on page 493 to set up an IP quorum application.

Follow these steps to add policy-based high availability to an existing configuration by using the CLI:

1. Volumes must be in a volume group to add high availability. If necessary, to create a new volume group, enter the following command:

```
mkvolumegroup
```

For more information, see **mkvoluemgroup command**.

2. If needed, use **lsvolumegroup** to see the ID of the volume group. For more information, see **lsvolumegroup** command.
3. To add the volumes that you want to be HA to the volume group, enter the following command:

```
chvdisk –volumegroup <volume group ID or name> <volume ID or name>
```

For more information, see **chvdisk command**.

4. To create a draft partition, enter the following command:

```
mkpartion –draft –name my_partition
```

For more information, see **mkpartition** command.

Note: lspartition shows **draft=yes** for a draft partition. A draft partition can be used to construct a storage partition, but has no effect until it is published.

5. To add a volume group to the draft partition, enter the following command:

```
chvolumegroup –draftpartition my_partition <volume group ID or name>
```

For more information, see **chvolumegroup command**. Adding the volume group to the draft partition automatically adds any dependent hosts and other volume groups to the draft partition. It does not automatically add volumes that are not in volume groups.

6. The draft partition must be published to make it into a storage partition that is eligible for use in a high availability configuration. The draft partition can be published only when all the required volumes are in volume groups. To publish a draft partition, enter the following command:

```
chpartition –publish my_partition
```

For more information, see **chpartition command**.

7. If required, you can create another draft partition and repeat the process. Only one draft partition is allowed for each system.
8. If required, partitions can be merged. To merge a second partition into the first partition, enter the following command:

```
mergepartition –targetpartition <name or ID of other storage partition> my_partition
```

For more information, see **mergepartition command**.

9. To create a replication policy with the *2-site-ha* topology, enter the following command:

```
mkreplicationpolicy –topology 2-site-ha –location1system <name or ID of one system>
-location2system <name or ID of other system> -location1iogrp 0 -location2iogrp 0
```

If necessary, use **lspartnership** to get the name or ID of each system. For more information, see **mkreplicationpolicy command**.

10. To associate the published storage partition with the replication policy, enter the following command:

```
chpartition –replicationpolicy <HA policy ID or name> my_partition
```

Note: As long as the partition is associated with the replication policy, the volumes, volume groups, hosts and host-to-volume mappings that are contained in the partition are maintained across both systems.

11. Hosts can now discover paths to volumes at both locations. Use **lspartition** to monitor the health of the HA solution. Use **lspartnership** to monitor connectivity between the systems. Use **chpartition** to change the active and preferred management systems.

Note: To prioritize a colocated system for I/O, use **chhost** to specify a location for the host.

Supported host operating systems for high availability

Policy-based high availability supports a reduced set of the operating systems that are supported by IBM Storage Virtualize.

The following operating systems are supported with policy-based high availability when connected using Fibre Channel SCSI:

- Red Hat Enterprise Linux, including with KVM (all supported versions).
- SUSE Linux Enterprise (all supported versions).
- IBM AIX (version 7.2 and later).

- Linux on IBM Z (all supported versions and distributions).
- VMware ESXi (all supported versions, excluding vVols).
- IBM i 7.4 TR11 or later, IBM i 7.5 TR5 or later, IBM i 7.6 or later. Fabric-attached Fibre Channel only.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Hosts configured for high availability must have the generic host type. Existing hosts with deprecated host types should be changed to the generic host type.

Volumes configured for high availability cannot be mapped to other IBM Storage Virtualize systems.

Disaster recovery

Policy-based replication can be used to implement asynchronous replication between systems for disaster recovery.

In the event of a disaster that prevents access to the production storage, host access can be enabled to the recovery copy. When the original production system is accessible again, replication can be restarted in either direction to restore redundancy.

Replication for disaster recovery can be configured with or without the use of storage partitions. Using storage partitions allows high availability to be easily added later and will also allow the use of future FlashSystem grid functionality. Storage partitions can be used where the host environment meets the supported configurations. For more information, see Storage partitions.

Existing configurations using replication for disaster recovery without storage partitions can be migrated into a storage partition by using the CLI. This allows an existing 2-site disaster recovery (DR) configuration to be extended to a 3-site HA and DR configuration while continuing to support disaster recovery operations and without requiring resynchronization of the recovery copy.

Replication for disaster recovery offers the flexibility to use different replication policies for different volume groups to match business requirements for different recovery point objectives, even on the same partnership or within the same storage partition.

Planning replication for disaster recovery

Before implementing replication for disaster recovery, you must determine your recovery point objectives, application requirements, and consider the connectivity available between locations.

To implement replication for disaster recovery, consider the following:

- Policy-based replication for disaster recovery supports only non-zero RPO and RTO. For zero RPO and RTO requirements, consider using policy-based high availability.
- Determine the round-trip time (RTT) between the systems. As the RTT increases, the maximum throughput decreases.
- Determine the change rate of data for applications storing data on the system. Environments with high change rates and low bandwidth may be unable to achieve low RPOs.
- The use of replication protects data from a large-scale disaster on the site or storage systems. Consider using safeguarded snapshots in addition to replication to also provide protection against data corruption and cyber attacks.

Replication for disaster recovery requires a partnership to be configured between the systems. Partnerships for asynchronous replication can use any of the supported connectivity types. Partnerships can be configured with up to 3 other systems for disaster recovery.

IP connectivity is required between systems for management. The following TCP ports are required to be accessible on the system management IP address:

- Port 7443 for REST API access between systems.

- Replication setup using the management GUI requires access to both systems from the host where the web browser is running. Ensure that any firewalls between the web browser and storage systems allow inbound traffic to port 7443 and 443 on the system management IP address.
- Additionally, IP replication requires inbound and outbound access using port 3260 on the system management IP and port 3265 on the IP addresses for the data path connections.

Policy-based replication for disaster recovery can be used with all host operating systems that are supported by IBM Storage Virtualize systems. The list of supported operating systems is on IBM System Storage Interoperation Center (SSIC).

For the best performance, configure the host multipath drivers to use SCSI ALUA or NVMe ANA.

With policy-based replication, you can replicate thin-clones and FlashCopy target volumes to the recovery system. You cannot replicate Safeguarded snapshots.

The following restrictions apply to policy-based replication for disaster recovery:

- The name of a volume group cannot be changed while a replication policy is assigned.
- Policy-based replication cannot be used with volumes that are:
  - Image mode
  - Cache disabled
  - Configured to use Transparent Cloud Tiering (TCT)
- A volume in a volume group with a replication policy assigned cannot be:
  - Renamed
  - Resized (expanded or shrunk)
  - Migrated to image mode, or have an image mode copy added
  - Moved to a different I/O group
- All volumes in a volume group must be in the same I/O group.

Implementing replication for disaster recovery

You can implement replication for disaster recovery with policy-based replication.

Policy-based replication uses volume groups and replication policies to deploy and manage replication for disaster recovery.

Configuring disaster recovery using storage partitions

From the **Storage partitions** panel, navigate to the partition to be configured.

The GUI guides you through the required steps to:

- Configure a partnership between systems, or select an existing partnership.
- Link storage pools between systems.
- Create and link a partition on the remote system for disaster recovery.

After configuring the partition, navigate to the **Volume group** view within the partition to create and assign replication policies on individual volume groups as required.

Configuring disaster recovery without using storage partitions

If not configured already, navigate to the **Settings** > **System** > **Partnerships** page to create a partnership between systems for replication.

From the **Volumes** > **Volume groups** panel, navigate to the volume group to be configured and select the **Policies** tab.

The GUI guides you through the required steps to:

- Select an existing partnership.
- Link storage pools between systems.
- Create a new replication policy or select an existing policy to assign to the volume group.

Managing replication for disaster recovery

You can manage replication for disaster recovery by using the management GUI, CLI, or REST API.

When a replication policy for disaster recovery is assigned to a volume group, the volume group operates in one of three modes:

Production
The volumes in a production volume group are accessible for host I/O and configuration changes are allowed to the volume group. This system acts as the source for replication to the target volumes on the recovery system. The production volume group contains the most up-to-date copy of data for host access.

Recovery
The volumes in a recovery volume group are the target for replication and cannot be used for host I/O in this state. Configuration changes are not allowed on the recovery copy, but changes made to the production copy are copied to the recovery copy.

Independent
If access is enabled for the recovery copy of a volume group, such as during a disaster, each copy of the volume group is independently accessible for I/O and configuration changes are allowed on both. Replication is suspended in this state. While independent, any data or configuration changes that are made on either system are not copied to the other system. When replication is restarted, the system automatically updates the data and the configuration on the recovery volume group to match the production volume group.

Replication policies do not define a direction for replication. The initial direction of replication is determined when a replication policy is associated with a volume group or a volume group is created specifying a replication policy. The system where the replication policy is assigned starts as the production system.

Removing a replication policy from the production copy of a volume group causes the recovery volume group and its volumes to be deleted. Independent access must be enabled on the recovery copy first to keep the copy, and the replication policy can then be removed from the volume group without affecting the other copy.

The replication policy assigned to a volume group can be replaced with a different replication policy without requiring any resynchronization, for example, to use a policy with a different recovery point objective alert for the volume group. The new policy must specify the same systems in the same locations.

Managing and monitoring disaster recovery for a volume group

Using the management GUI
To configure disaster recovery, assign a replication policy to the volume group. Using the management GUI, navigate to the **Volume groups** page and select the desired volume group, then select the **Policies** tab. When a replication policy is assigned to the volume group, this page is used for managing and monitoring replication.

Using the command-line interface
To assign a replication policy to a volume group using the CLI, use the following command:

```
chvolumegroup -replicationpolicy <policy> <volumegroup>
```

To unassign a replication policy from a volume group, use the following command:

```
chvolumegroup -noreplicationpolicy <volumegroup>
```

To view a list of all the volume groups that are configured for replication, use the following command:

```
lsvolumegroupreplication
```

To view the detailed status of replication for an individual volume group, use the following command:

```
lsvolumegroupreplication <volumegroup>
```

Accessing the recovery copy of a volume group

Using the management GUI

1. Log in to the system with the recovery copy.
2. Navigate to the **Volume groups** page and select the desired volume group.
3. Select the **Policies** tab and review the recovery point for the copy.
4. If the recovery point is acceptable, click **Enable access** to enable independent access to the volume group.
5. If the production system is inaccessible, the production copy is made independent when connectivity is restored. The recovery copy will always become independent and available for use immediately.

Using the command-line interface

You can use the **chvolumegroupreplication** command to enable independent access when logged in to the system with the recovery copy:

```
chvolumegroupreplication -mode independent <volumegroup>
```

Restarting replication from the independent state

When the replication mode for a volume group is in the independent state, replication can be restarted in either direction.

The replication policy assigned to the volume group on the system you want to use as the production copy will be used to define the replication requirements, such as the RPO alert for the group. If the volume group was previously the recovery copy in a storage partition, it might not yet have a replication policy assigned. You must assign a replication policy to the volume group before you can restart replication from this system. If you use the management GUI to restart replication, the GUI guides you through the required steps to create a new replication policy or select an existing policy to assign to the volume group as part of the restart task.

Using the management GUI

1. Log in to the system with the volume group you want to use as the new production copy.
2. Navigate to the **Volume groups** page and select the desired volume group.
3. Select the **Policies** tab.
4. Click the **Restart replication** action to restart replication using the volume group on the local system as the production copy. Any changes made to the copy on the remote system will be overwritten when replication is restarted.

Using the command-line interface

You can use the **chvolumegroupreplication** command to restart replication when logged in to the system with the copy you want to use as the production copy:

```
chvolumegroupreplication -mode production <volumegroup>
```

Migrating replication for disaster recovery to use partitions

Existing configurations that use replication for disaster recovery without storage partitions can be migrated into a storage partition by using the CLI. This allows an existing 2-site DR configuration to

be extended to a 3-site HA+DR configuration while continuing to support disaster recovery operations and without requiring resynchronization of the recovery copy.

This procedure requires all the volume groups that are to be migrated to the same partition to replicate in the same direction.

Using the command-line interface

1. On each system:
   a. Use **mkpartition -draft -name <partition_name>** to create a new draft partition.

      Note: If appropriate for your environment, consider naming the partitions differently to allow easier identification of which partition is typically used for disaster recovery during normal running.

   b. Use **chvolumegroup -draftpartition <partition> <volumegroup>** to move the volume groups to be migrated into the new draft partition.

      Note: Moving an object into a draft partition may automatically add associated objects to the partition too. Review the complete set of objects in the draft partition before you publish it.

   c. Publish the partition by using: **chpartition -publish <partition>**
2. On the system with the recovery volume groups:
   a. Identify the **uuid** value of the partition that will be used for recovery by using: **lspartition <recovery_partition>**
3. On the system with the production volume groups:
   a. Configure a link for disaster recovery between the partitions by using: **chpartition -makedrlink -remotedrlinkedpartitionuuid <recovery_partition_uuid> -remotesystem <recovery_system_id/name> <production_partition_id/name>**

      Note: The remote partition being linked to must contain all of the associated remote volume groups for all disaster recovery volume groups in the local partition.

   b. For each 2-site-async-dr policy that needs to be replaced, create and assign a new replication policy with topology ‘async-dr’ in the partition:
      i) Use the **lsvolumegroupreplication** command to identify the current replication policy assigned to each volume group to be migrated.
      ii) Use the **lsreplicationpolicy** command to identify the RPO alert value for each policy.
      iii) Create the new replication policy by using: **mkreplicationpolicy -name <name> -topology async-dr -partition <production_partition_id/name> -rpoalert <rpoalert>**
      iv) Assign the appropriate new replication policy to each volume group by using: **chvolumegroup -replicationpolicy <policy> <volumegroup>**
      v) Optionally, any unused 2-site-async-dr policies can now be removed by using: **rmreplicationpolicy <policy>**

After the CLI procedure above has been completed, further management and monitoring can be performed using the GUI, the CLI, or the REST API.

High availability and disaster recovery (3-site replication)

Combining policy-based high availability and asynchronous policy-based replication for disaster recovery enables a 3-site replication solution for the most mission-critical workloads.

Policy-based HA provides a zero recovery point objective (RPO) solution by using two independent storage systems with synchronous replication across metro-area distances and active-active access for hosts at both sites with seamless failover for zero recovery time objective (RTO). Asynchronous replication (RPO > 0) to a third system provides disaster recovery capability (RTO > 0) when a metro-area disaster.

Three-site replication for a storage partition can be configured in any of the following ways:

- Starting with an existing 2-site high availability configuration, then adding replication for disaster recovery at a third side.
- Starting with an existing 2-site configuration for disaster recovery using storage partitions, then adding high availability to either of the partitions.
- Configuring 3-site replication for high availability and disaster recovery in a single operation.

The planning, implementation, management, and monitoring information for high availability and disaster recovery, when used individually, also applies when used in a 3-site configuration.

- Planning high availability
- Planning replication for disaster recovery

Partnerships must be configured between all three systems. High availability requires either a Fibre Channel or an RDMA-based Ethernet partnership. The partnership for disaster recovery also supports the use of a long-distance TCP partnership. The different partnerships between systems can use different types of connectivity.

The use of separate node ports is recommended to isolate the different types of traffic for:

- High availability partnerships
- Disaster recovery partnerships
- Hosts and external storage, which can also be separated further if required

Example error scenarios

This table lists examples of possible error scenarios and the expected behavior in a 3-site solution, starting from an initial healthy state.

Table 31. Example error scenarios

Scenario	High availability	Disaster recovery
Loss of connectivity between the non-preferred HA system and the DR system	No impact	No impact
Loss of connectivity between the preferred HA system and the DR system	The active management system will change to the non-preferred system to maintain replication to the DR system. The storage partition remains active-active with no interruption to host access. The preferred system will automatically become the active management system again when connectivity to the DR system is restored.	No impact
Loss of connectivity between the HA systems	If the connectivity issue lasts more than 3 seconds, HA will be suspended. The volumes within the storage partition will remain accessible through one of the two HA systems.	No impact. Replication will continue between the active management system and the DR system.

Table 31. Example error scenarios (continued)

Scenario	High availability	Disaster recovery
Both HA systems lose connectivity to the DR system	No impact	Replication is suspended. Replication will resume automatically when connectivity is restored.
One or both of the HA systems lose connectivity to all quorum applications	No impact. However, the loss of quorum will prevent the non-preferred system from taking over the partition if the HA systems lose connectivity, or if the preferred system becomes unavailable.	No impact
The preferred HA system loses connectivity to the non-preferred system and all quorum applications at the same time. The non-preferred system still has connectivity to the active quorum application.	The storage partition will continue to be accessible on the non-preferred system. HA will automatically re-establish when connectivity is restored.	Replication will continue between the active management system and the DR system.
The preferred HA system is unavailable (for example, power failure, hardware failure)	The storage partition will continue to be accessible on the non-preferred system. HA will automatically re-establish when the preferred system is available.	Replication will continue between the active management system and the DR system.
The non-preferred HA system is unavailable (for example, power failure, hardware failure)	The storage partition will continue to be accessible on the preferred system. HA will automatically re-establish when the non-preferred system is available.	Replication will continue between the active management system and the DR system.
The DR system is unavailable (for example, power failure, hardware failure)	No impact	Replication is suspended. Replication will resume automatically when the DR system is available.
Loss of management IP connectivity only between the HA systems	No impact on host I/O. If high availability is established for the partition, configuration changes that would result in the loss of high availability are blocked. If high availability is not established, configuration changes are permitted and HA will re-establish automatically when connectivity is restored.	No impact.

Table 31. Example error scenarios (continued)		
Scenario	**High availability**	**Disaster recovery**
Loss of management IP connectivity only between the active management system and the DR system	No impact.	Configuration changes made to the HA storage partition cannot be replicated to the DR system. Replication will be suspended if configuration changes are required on the DR system. If replication is suspended, it will resume automatically when connectivity is restored.
Access is enabled to the recovery copy of one or more volume groups in the storage partition	No impact.	Replication is suspended for the independent volume groups. Host I/O and configuration changes are allowed on the HA and DR partitions independently.
Replication is restarted to synchronize changes from the DR system to the HA systems	HA is suspended while replication is running from the DR system to the active management system. HA will re-establish automatically when there are no volume groups replicating from the DR system.	No impact.

Partnerships

Partnerships are used to connect systems together to enable migration, data replication, and high-availability solutions.

You can create a partnership in the following ways:

- By using Fibre Channel connectivity.
- By using IP connectivity.

In addition to the partnership for data transfer between systems, IP connectivity is also required between the management IP addresses of each system to access the REST API for management. Management traffic uses authentication certificates to ensure secure communications between systems.

When you use the management GUI to configure high availability or disaster recovery by using storage partitions, the GUI guides through the steps that are required to configure suitable partnerships between systems. It is not necessary to create partnerships between systems beforehand.

Using the management GUI

When you create a partnership between systems for high availability or disaster recovery by using GUI, certificates are automatically exchanged between systems to create a truststore on each system. No additional steps are required.

Using the command-line interface

When using the command-line interface, you must exchange certificates and create a truststore on each system as a separate activity after you create the partnership between systems.

1. On the first system, use the following command to export the certificate:

```
chsystemcert -export
```

The certificate is exported to a file named */dumps/certificate.pem*.

2. Copy the certificate file from the system to your local workstation by using an SCP client. Then, use SCP to upload the certificate file to the `/tmp/` directory on the second system.
3. On the second system, create the truststore by using the following command:

```
mktruststore -file /tmp/certificate.pem -restapi on -name <remote_cluster_name>
```

   Naming the truststore is optional. But, in environments with multiple partnerships or multiple truststores, it helps to identify which truststores are used for which purpose.
4. Repeat the preceding steps to export the certificate from the second system and create a truststore on the first system.

You can use the **lssystemcert** command to view information about the current system certificate and the **lstruststore** command to list the truststores that exist on the system.

Pool links

High availability and disaster recovery replication require the storage pools that are used for creating replicated volumes to be linked between systems. When a volume is created on the active management system for HA, or on the production system for DR, pool links define the pool that is used on the remote system to create the replicated volume.

For SAN Volume Controller systems that use a stretched system topology with mirrored volumes, only the pools that are in site 1 need to be linked.

Use of the management GUI to configure pool links simplifies the process by allowing pools to be linked from a single system without needing to login to the other system. Configuration can also be performed by using the CLI or the REST API, connecting to each system with pools to be linked.

Additionally, if child pools currently exist on one system only, you can use the management GUI on the partnered system to create and link a child pool in a single step. The management GUI simplifies the process of creating a linked pool on the second system, displaying the properties like name, capacity, and provisioning policy from the first system.

When you configure pool links by using the GUI, you can optionally assign a provisioning policy to the pools that you are linking. Provisioning policies control how capacity is provisioned on new volumes that are created in the linked pool. Pools on each partnered system can be assigned provisioning policies with different capacity savings methods. The system creates two default provisioning policies when the first parent pool is created. You can create more user-defined policies to specify alternative capacity savings if desired. The system creates fully provisioned volumes if a provisioning policy is not configured. For more information, see Provisioning policy.

Using the management GUI

When adding high availability or disaster recovery, the management GUI guides you through the required steps to link storage pools between systems, and optionally assign provisioning policies on each linked pool.

To link extra pools later, or remove the links between pools, go to the **Pools** > **Pools** page in the management GUI and select the pool that you want to modify. Right-click on the pool and select the action-prompt to **Add Pool Link for Replication** or **Modify Pool Links for Replication** as required.

Using the command-line interface

Pools on partnered systems are linked if they have the same *replication_pool_link_uid* value. To identify pools that are linked, use the **lsmdiskgrp** command on each system.

To modify or remove the links between pools by using the command-line interface, use the **chmdiskgrp** command.

Replication policies

A replication policy defines how replication is configured between systems.

The topology of a replication policy specifies the type of replication to be used. The following topologies are available:

High availability

- This type of replication policy is used with storage partitions.
- The policy specifies the systems that are used for high availability.
  - The locations of the systems in the policy define the layout of the systems seen in the management GUI. The system in location 1 is shown on the left, and the system in location 2 is shown on the right. This layout remains fixed and is not affected by the direction of replication between systems, simplifying monitoring and management from either of the systems.
- This type of policy can be used for 2-site high availability and 3-site (high availability with disaster recovery) configurations.
- You can start from the **Storage partitions** panel in the GUI to add or remove high availability for a partition.

Asynchronous disaster recovery

- This type of replication policy is used with volume groups that are in a storage partition.
- The storage partitions are linked to define the systems that are used for disaster recovery replication.
- This type of policy can be used for 2-site disaster recovery and 3-site (high availability with disaster recovery) configurations.
- You can use the **Volume groups** panel within a storage partition to add or remove this type of replication for a volume group.

2-site asynchronous disaster recovery

- This type of replication policy is used with volume groups that are not in a storage partition.
- This type of policy is suitable for environments that cannot use storage partitions for disaster recovery. For example, replication for VMware Virtual Volumes (vVols), or for hosts that are not currently supported in a partition. For more information, see Storage partitions.
- The policy specifies the systems and the I/O groups in those systems that are used for replication.
  - The locations of the systems in the policy define the layout of the systems seen in the management GUI. The system in location 1 is shown on the left, and the system in location 2 is shown on the right. This layout remains fixed and is not affected by the direction of replication between systems, simplifying monitoring and management from either of the systems.
- This type of policy can be used for 2-site replication only.
- You can use the **Policies** tab of the **Volumes** > **Volume Groups** panel in the management GUI to add or remove this type of replication for a volume group.

Both types of replication policies for asynchronous replication define a maximum acceptable recovery point for the recovery copy. An alert is generated if the recovery point exceeds this value.

Replication policies cannot be changed after they are created. However, a new policy can be used to replace the existing one if changes are required.

Replication statistics

You can use the metrics that the system collects for policy-based replication to analyze performance issues. Statistics for policy-based replication help you troubleshoot and improve system performance.

Volume group statistics

Policy-based replication operates on volume groups rather than individual volumes. Therefore, most of the replication statistics are recorded against a volume group. The following statistics are logged for volume groups:

Table 32. Statistics collection for volume group

Statistic name	Description
rarp	Indicates the average recovery point in seconds for the volume group since the last time statistics were collected.
rnrw	Indicates the cumulative number of replication writes to the target sent by this node for this volume group.
rnrb	Indicates the cumulative number of blocks that this node has replicated for this volume group.
rwrp	Indicates the worst recovery point in seconds for the volume group since the last time statistics were collected.

Node statistics

For node-level values, policy-based replication adds to the existing node XML statistics file.

Table 33. Statistics collection for node statistics

Statistic name	Description
rjru	Indicates the percentage of journal resources that are used for replicating data, averaged across all threads. Value out of 10000 representing a percentage to two decimal places.
rjrw	Indicates the highest percentage of journal resources that are used across all threads. Value out of 10000 representing a percentage to two decimal places.

Volume statistics

Generally, all statistics regarding replication are reported on the volume groups but many existing tools focus on displaying statistics on the individual volumes.

Table 34. Statistics collection for volume statistics

Statistic name	Description
varp	Indicates the average recovery point in seconds for the volume group for this volume since the last time statistics were collected.
vwrp	Indicates the worst recovery point in seconds for the volume group for this volume since the last time statistics were collected.

Chapter 9. Access

The Access menu in the GUI maintains who can log in to the system, defines the access rights to the user, and tracks what was done by each privileged user to the system.

Ownership groups

An ownership group defines a subset of users and objects within the system.

Ownership groups restrict access for users in the ownership group to only those objects that are defined within that ownership group. An *owned object* can belong to one ownership group. Users in an ownership group are restricted to viewing and managing objects within their ownership group. These restricted users can access only this subset of objects. Unrestricted users, who are not in an ownership group, can continue to view or manage all the objects on the system based on their defined user role, including objects within ownership groups. The following objects can be assigned to ownership groups:

- Child pools
- Volumes
- Volume groups
- Hosts
- Host clusters
- Host mappings
- FlashCopy mappings
- FlashCopy consistency groups
- User groups
- Portsets

Two basic use cases can be applied to using ownership groups on the system. The first use case for using ownership groups is on a system where new objects are created within the ownership group. There can also be other existing objects on the system that are not in the ownership group. The second use case for using ownership groups is on a system where these supported objects are already configured and you want to migrate these objects to use ownership groups.

When a user group is assigned to an ownership group, the users in that user group retain their role, but are restricted to only those resources within the same ownership group. The role that is associated with a user group can define the operations on the system and the ownership group can further limit access to individual resources. For example, you can configure a user group with the Copy Operator role, which limits user access to FlashCopy operations. Access to individual resources, such as a specific FlashCopy consistency group, can be further restricted by assigning it to an ownership group.

Inheriting ownership

Generally two basic inheritance rules that apply to all objects that can be associated to an ownership group:

1. Owned objects inherit the ownership group of the owner who creates those objects.
2. Only users with Security Administrator role can create, manage, or assign ownership group to new objects or existing objects.

Depending on the type of resource, ownership groups can be defined explicitly or inherited from these explicitly defined objects.

The following graphic shows how different objects inherit ownership from ownership groups:

© Copyright IBM Corp. 2024

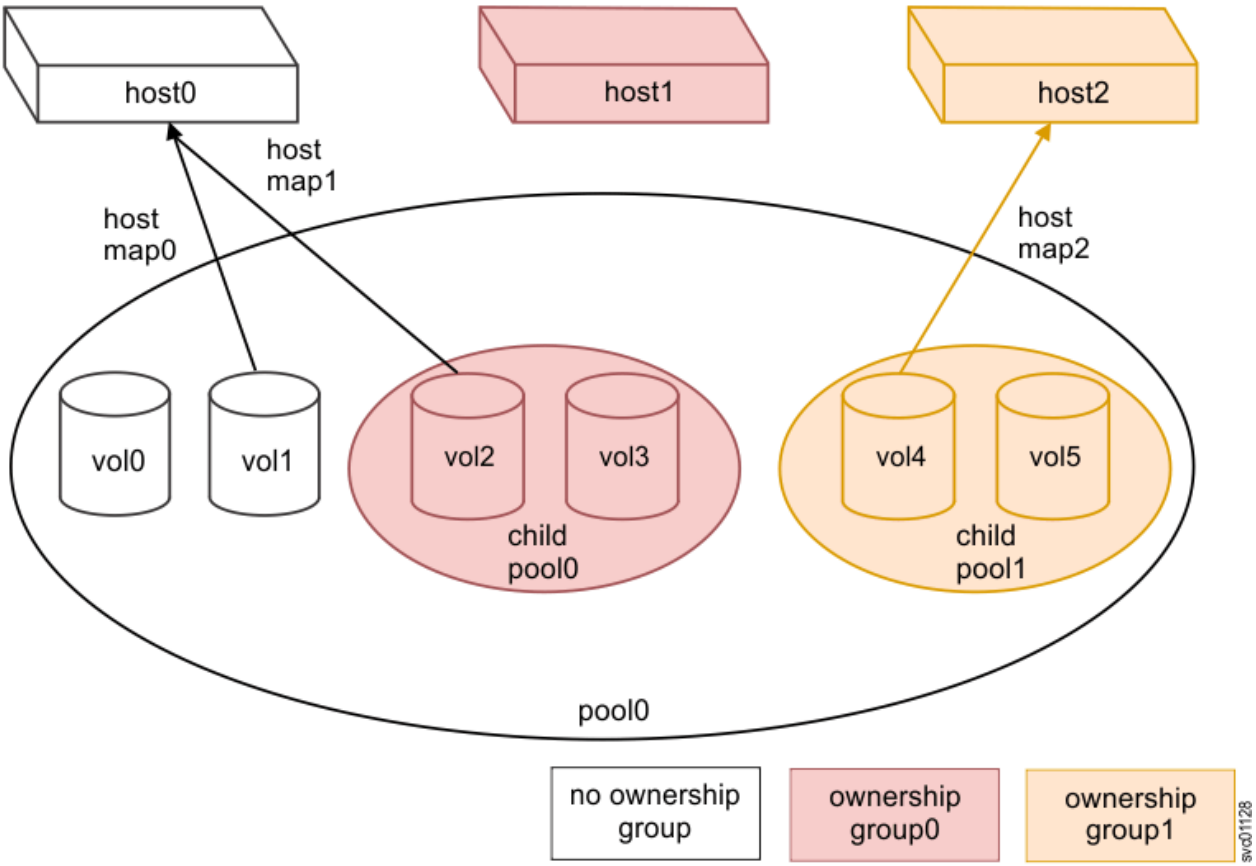


Figure 20. Ownership group inheritance

You can explicitly define ownership groups for the following objects:

Child pools
New or existing volumes that are defined in the child pool inherit the ownership group that is assigned for the child pool.

Host clusters
New or existing hosts that are defined in a Host Cluster inherit the ownership group that is assigned to the Host Cluster.

Hosts that are not part of a Host Cluster
If a host is not part of a Host Cluster, an ownership group can be associated with the host when the host object is created or changed.

Volume groups
Volume groups can be created to manage multiple volumes that are used with transparent cloud tiering support. The following rules apply to volume groups that are defined in ownership groups:

- Volumes can be added to a volume group if both the volume and the volume group are within the same ownership group or if both are not in an ownership group. When migrating existing resources to ownership groups, the volume group and its volumes can belong to different ownership groups temporarily until the migration completes.
- The ownership of a volume group does not affect the ownership of the volumes it contains. If a volume group and its volumes are owned by different ownership groups, then the owner of the child pool that contains the volumes can change the volume directly. For example, the owner of the child pool can change the name of a volume within it. The owner of the volume group can change the volume group itself and indirectly change the volume, such as deleting a volume from the volume group. Neither the ownership group of the child pools nor the owner of the volume group can directly manipulate the resources that are not defined in their ownership group.

Note: Volume groups can only be created by using the CLI.

FlashCopy consistency groups

FlashCopy consistency groups can be created to manage multiple FlashCopy mappings. The following rules apply to FlashCopy consistency groups that are defined in ownership groups:

- Users that are defined in ownership groups can add FlashCopy mappings to a consistency group if the volumes in the mapping and in the consistency group are within the same ownership group. For users that are not defined in an ownership group, no restriction for adding mappings to a FlashCopy consistency group. In this case, FlashCopy consistency groups are containers only and do not influence the ownership of the contents.
- When you are migrating existing resources to ownership groups, a FlashCopy consistency group and its resources can belong to different ownership groups temporarily until the migration completes.
- As with volume groups and volumes, the ownership of the consistency group has no impact in the ownership of the mappings it contains.

User groups

The following rules apply to user groups:

- Only users with Security Administrator role can create or manage ownership groups.
- Users with Security Administrator role cannot be assigned to an ownership group.

Portsets

The following rules apply to portsets:

- Users who are defined in an ownership group can view and create IP addresses or assign portset to host that are in their ownership group.
- Users who are defined in an ownership group can assign portset to host that are not assigned to an ownership group.
- When a user who are defined in an ownership group creates a portset, it is automatically assigned to the same ownership group as that restricted user.
- Users who are defined in an ownership group can view all IP addresses in portsets owned by them and portsets not assigned to an ownership group.
- Users who are defined in an ownership group cannot modify portsets associated with other restricted users.
- Users who are defined in an ownership group cannot view portsets that have been assigned to a different ownership group.
- Users who are defined in an ownership group cannot view IP addresses that are associated with a portset assigned to a different ownership group.
- Users who are defined in an ownership group cannot create or delete IP addresses from a portset that is part of a different ownership group.
- Users who are defined in an ownership group cannot own **storage** portsets.
- Users with the Security Administrator role can view, create, and manage all portsets on the system.
- Users with the Security Administrator role can view and create IP addresses to any portsets on the system.
- Users with the Security Administrator role can assign portsets that are not assigned to an ownership group to hosts in an ownership group.
- Users with the Security Administrator role can assign portset to hosts that are assigned the same ownership group.
- Users with the Security Administrator role cannot assign portset to hosts if the host and portsets are in different ownership groups.
- Users with the Security Administrator role cannot change the ownership of **storage** portsets.

Other objects inherit the ownership group from these explicitly assigned objects. Usually the ownership group is defined when an object is created. Ownership can also be changed when the explicitly defined

object from which the ownership is established is assigned to a different ownership group. The following objects inherit the ownership from the explicitly defined object:

Hosts that are a part of a Host Cluster

The following rules apply to hosts that are defined in ownership groups:

- The host inherits the ownership group of the Host Cluster that it belongs to.
- If a host is removed from a Host Cluster within an ownership group, the host inherits the ownership group of the Host Cluster it used to belong to.
- If a host is removed from a Host Cluster not within an ownership group, the host inherits no ownership group.
- Hosts can be added to a Host Cluster if the host and Host Cluster have the same ownership group.
- Changing the ownership group of a Host Cluster automatically changes the ownership group of all the hosts inside the Host Cluster.

Volumes

The following rules apply to volumes that are defined in ownership groups:

- The volume inherits the ownership group of the child pools that provide capacity for the volume and its copies. Volume copies can be created in different ownership groups for backup up scenarios. However, this value must be set intentionally by users that are not defined in ownership groups. When you create a volume copy or migrate volumes to other pools, you can specify child pools that are defined in different ownership groups in the management GUI, which establishes inconsistent ownership. However, it is not recommended to leave volumes or volume copies in different ownership groups. After the migration, the user with Security Administrator role needs to ensure all volumes or copies are within the same ownership group as the users who need access.
- With volume groups, the volume group and its volumes can belong to different ownership groups. However, the ownership of a volume group does not impact the ownership of the volumes that it contains.

Users

The following rules apply to users that are defined in ownership groups:

- A user inherits the ownership group of the user group it belongs to.
- Users that use LDAP for remote authentication can belong to multiple user groups, but belong to just one ownership group that is associated with one of the user groups.

Volume-to-host mappings

The following rule applies to volume-to-host mappings that are defined in ownership groups:

Volume-to-host mappings inherit the ownership group of the host or Host Cluster and volume in the mapping.

FlashCopy mappings

The following rules apply to FlashCopy mappings that are defined in ownership groups:

- FlashCopy mappings inherit the ownership group of both the volumes that are defined in the mapping.
- Like with FlashCopy consistency groups, it is possible for a consistency group and its mappings to belong to different ownership groups. However, the ownership of the consistency group has no impact in the ownership of the mappings that it contains.

Configuring ownership groups

You can configure ownership groups to manage access to resources on the system.

Before you create ownership groups and assign resources and users, it important to understand the following guidelines:

- Users within an ownership group can view or change resources within the ownership group in which they belong. Only users outside of an ownership group, usually users with Security Administrator roles, can view and manage all resources on the system, including users within an ownership group.
- Users within an ownership group cannot change any objects outside of their ownership group. This restriction includes global resources that are related to resources within the ownership group. Global resources cannot be assigned to ownership groups. For example, a user can change a volume in the ownership group, but not the MDisk or drive that provides the storage for that volume. The MDisk or drives are considered global resources and cannot be assigned to ownership groups.
- Users within an ownership group cannot view or change resources if those resources are assigned to another ownership group or are not assigned to any ownership group. However, users within ownership groups can view and display global resources. For example, if a volume is assigned to an ownership group, a user outside of that ownership group cannot display information for that volume. However, the same user can display information on drives on the system, since drives are a global resource that cannot be assigned to any ownership group.

Using the management GUI

These steps are for new systems where owned objects, such as child pools, hosts, and volumes are not configured yet.

To configure supported objects to use ownership groups on system, complete these steps:

1. In the management GUI, select **Access** > **Ownership Groups**.
2. On the **Ownership Groups** page, click **Create Ownership Group**. A new ownership group is created and displays on the **Ownership Groups** page.
3. After you have created an ownership group, create a user group to include users that are the owners of the ownership group. Select **Access** > **Users by Group** and click **Create User Group**.
4. On the **Create User Group** page, enter the following information:

   Name
   Enter the name of the user group.

   Role
   Select the role for all users within the user group. User groups that are assigned in ownership groups cannot use the Security Administrator role.

   Ownership Group
   Select the ownership group that you created earlier and assign it to this user group.
5. On the **Access** > **Users by Group** page, select existing users or create new users to assign to the user group. These users automatically inherit the ownership group that is assigned to the user group.
6. On the **Pools** > **MDisks by Pools** page, if a parent pool does not exist, click **Create Pool** and assign storage to that pool. Right-click the new parent pool and select **Create Child Pool**.
7. On the **Create Child Pool** page, enter the following information:

   Name
   Enter the name of the child pool.

   Capacity
   Select the amount of capacity that you are allocating from this child pool.

   Note: The **Capacity** field is not available when you are creating a data reduction child pool. A data reduction child pool has access to the complete free capacity of the parent pool without limit. A data reduction child pool cannot be created if the parent pool does not have free capacity.

   Ownership Group
   Select the ownership group that you created earlier and assign it to this child pool.
8. After child pools are configured and assigned to the ownership groups, users that you created in Step“5” on page 221 can create objects like hosts, volumes, and FlashCopy mappings.

Managing ownership groups

You can use the management GUI or command line interface to manage ownership groups.

Migrating to ownership groups

If you want to use ownership groups for objects that exist on your system, you need to reconfigure certain resources if you want to configure ownership groups.

An ownership group defines a subset of users and objects within the system. If you want to use ownership groups, you might need to reconfigure the following objects to implement ownership groups on the system:

- Child pools
- Volumes
- Volume groups
- Hosts
- Host clusters
- Host mappings
- FlashCopy mappings
- FlashCopy consistency groups
- User groups
- Portsets

If these objects are configured on your system and you want to use ownership groups to manage access, identify the objects that you want to assign to ownership groups and determine any dependencies these objects have to related objects outside of the ownership group. For example, if you have a volumes that are used in a FlashCopy mapping ensure that both volumes are assigned to the same ownership group, even if different child pools provide capacity. In this case, two separate child pools are configured with the same ownership group and their volumes inherit the ownership group from the pool. Objects like FlashCopy mappings and host mappings cannot be created with inconsistent ownership. Inconsistent ownership means that either an object is owned by different owners or dependent objects do not belong to the same ownership group. Some of these inconsistencies are can be temporary and resolve when all related objects are assigned to the correct ownership group. However some actions, such as adding a volume copy or migrating a volume can be created with inconsistent ownership intentionally. For example, when adding a volume copy or migrating a volume, you can have these objects be in different ownership groups. In the management GUI, this is done automatically when a volume is copied or migrated to a child pool in a different ownership group and warnings are generated. It is not recommended to leave objects, such as FlashCopy mappings or host mappings, in an inconsistent state. To ensure full information, only run system backup operations on the system when objects are in a consistent state.

Migration scenarios that require reconfiguration of these objects can be varied; however, two basic migration scenarios are available:

Migration Scenario 1: Child pools exist on the system

In this scenario, child pools exist on the system and a security administrator can create the ownership group and assign it to the existing child pool. All volumes that are associated with the child pool automatically inherit the ownership group of that child pool.

The following graphics show a configuration where child pools currently exist on the system and the process of assigning ownership groups to these child pools. The first graphic represents the current configuration before ownership groups are assigned to existing child pools. The second graphic shows the how some objects inherit the ownership group of the child pools and how others need to be added explicitly.

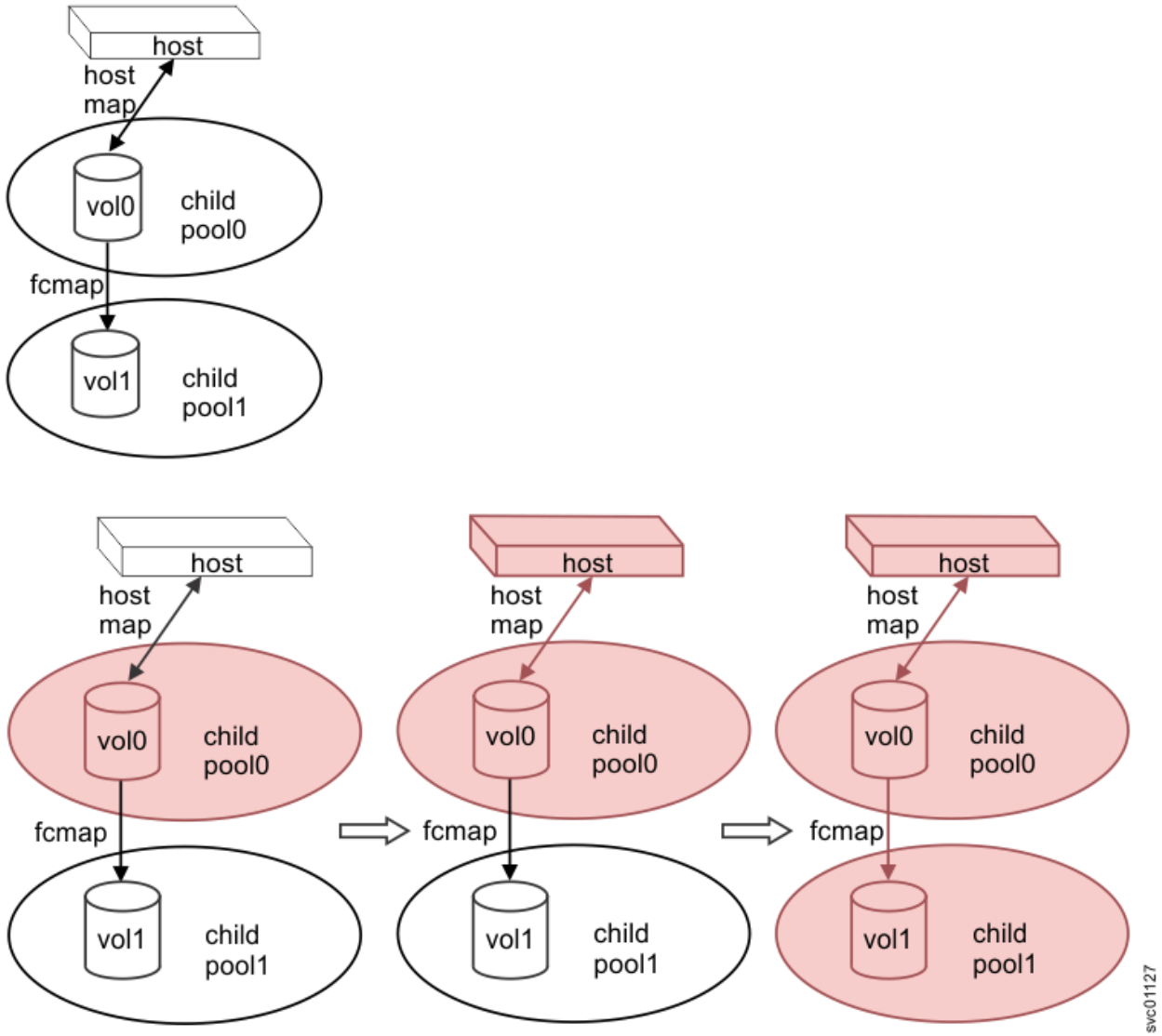


Figure 21. Example: Child pools are configured on the system

Using the management GUI

To use ownership group with existing child pools, complete the following steps:

1. In the management GUI, select **Access** > **Ownership Groups**.
2. On the **Ownership Group** page, select **Create Ownership Group**.
3. On the **Create Ownership Group** page, enter a name for the ownership group and select **Create**.
4. On the **Ownership Group** page, select **Assign Child Pools**.
5. On the **Assign Child Pools** page, select the child pool that you want to add to the ownership group and click **Assign**. In the example, *child pool0* is added to the ownership group. Because the volume inherits the ownership group of *child pool0* automatically.
6. If the system detects dependent resources for volumes that are within that child pool, the management GUI displays the **Additional Resources to Add** page where it lists hosts or host clusters that have mappings to volumes that inherit the ownership group. For example, in the preceding graphic, *child pool0* is added to the ownership group and *vol0* automatically inherits the ownership group of the child pool0. However, the host mapping between *vol1* and *host* does not inherit the ownership group from *vol0* because it also depends on the host. On the **Additional Resources to Add** page, select the host or host clusters to add to the ownership group. Click **Proceed**.

7. On the **Ownership Groups** page, verify that the volumes and hosts are assigned to the new ownership group.
8. In the example, *pool1* also needs to be assigned to the same ownership group as *pool0*. Repeat this process for all child pools that you want ownership groups to manage access.
9. After you have assigned ownership groups to all child pools, create a user group to include users that are the owners of the ownership group. Select **Access** > **Users** and click **Create User Group**.
10. On the **Create User Group** page, enter the following information:

    Name
    Enter the name of the user group.

    Role
    Select the role for all users within the user group. User groups that are assigned in ownership groups cannot use the Security Administrator role.

    Ownership Group
    Select the ownership group that you created earlier and assign it to this user group.
11. On the **Access** > **Users** page, select existing users or create new users to assign to the user group. These users automatically inherit the ownership group that is assigned to the user group. These users can now start creating objects like hosts and volumes within this ownership group, using capacity in the child pools that are assigned to the ownership group.

Using the command-line interface

Migration Scenario 2: Child pools do not exist on the system

In this scenario, child pools do not exist on the system, so they must be created from existing parent pools, and assigned to an ownership group. Volume mirroring can be used to create a copy of the volume in the new child pool. The volume copy inherits the ownership group that was assigned to the child pool.

The following graphics show a configuration where child pools do not exist on the system and the process of assigning ownership groups. The first graphic represents the current configuration before ownership groups are assigned.

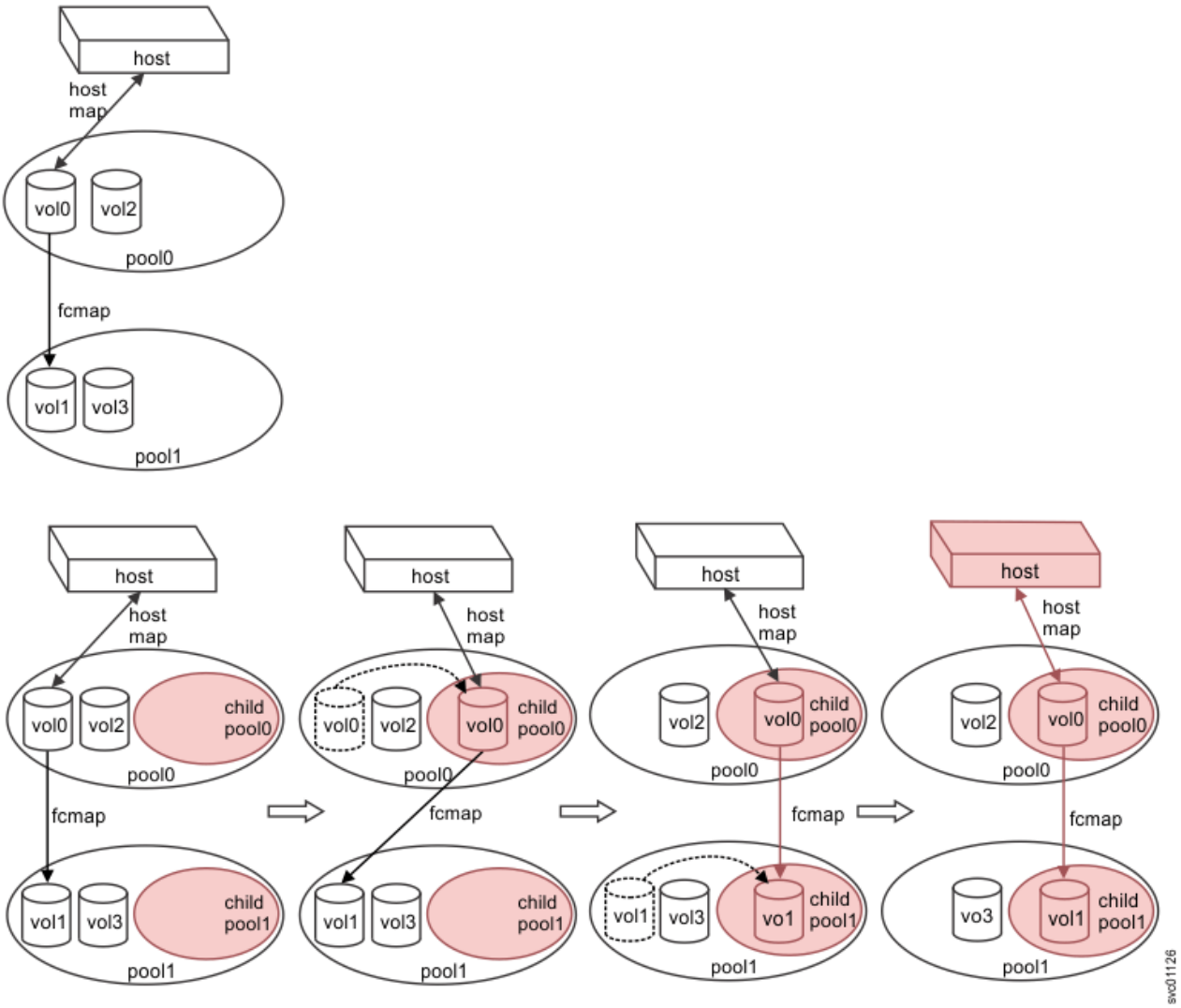


Figure 22. Example: Child pools are not configured on the system

In this example, the first graphic shows a system with two parent pools that contain two volumes each. Two volumes (*vol0* and *vol2*) use a parent pool (*pool0*) and two other volumes (*vol1* and *vol3*) use the other parent pool (*pool1*). The volume (*vol0*) is mapped to the host and is the source volume in a FlashCopy mapping with *vol0* in *pool1*. In this case, you want to limit access to *vol0* and *vol1* and its related resources.

The second graphic shows the required steps to use ownership groups in this environment. First, two child pools are created from the parent pools. After the child pools are created, a mirrored copy of each volume is created in the new child pools. After the volume copies are in the corresponding child pool, the child pools are assigned to an ownership group and the volumes with those child pools inherit that ownership group. The original copy of the pool can be deleted from the parent pool. More resources can be added to the ownership group, such as hosts and host clusters. The last graphic in the sequence shows the new configuration.

Using the management GUI

To migrate ownership groups on a system without child pools, complete the following tasks:

1. **Create child pools from existing parent pools:**
    a. In the management GUI, select **Pools** > **Pools**.

b. On the **Pools** page, right-click the pool that you want to create the child pool from and select **Create Child Pool**.

c. Repeat this step for all the pools that you want to use ownership groups. In the example, a separate child pool is created from both parent pools (*pool0* and *pool1*).

2. **Migrate volumes from parent pool to child pools:**

a. In the management GUI, select **Volumes** > **Volumes by Pools**.

b. Select the parent pool that you created the child pool from, and then right-click the volume and select **Add Volume Copy**.

c. On the **Add Volume Copy** page, select the child pool that you want to copy the volume to and click **Add**.

d. If necessary, delete the original copy from the parent pool.

e. Repeat steps "2.b" on page 226 to "2.c" on page 226 for each volume.

Note: This process uses volume mirroring to create a copy of an existing volume in the child pool. You can also use the **Migrate volume to another pool** action.

3. **Create ownership groups and assign child pools:**

a. In the management GUI, select **Access** > **Ownership Groups**.

b. On the **What is an Ownership Group?** page, enter a name of the ownership group and select **Create Ownership Group**.

c. On the **Create Ownership Group** page, expand **Child Pools** and select **Assign**.

d. On the **Assign Child Pools** page, select the child pools that you want to add to the ownership group and click **Assign**.

e. If the system detects dependent resources for volumes that are within that child pool, the management GUI displays the **Additional Resources to Add** page. The page lists hosts or host clusters that have mappings to volumes that inherit the ownership group. For example, in the preceding graphic, *child pool0* is added to the ownership group and *vol0* automatically inherits the ownership group of *child pool0*. However, the host mapping between *vol0* and *host* does not inherit the ownership group from *vol0* because it also depends on the host. On the **Additional Resources to Add**, select the host or host clusters to add to the ownership group. Click **Proceed**.

f. On the **Ownership Groups** page, verify that child pool is assigned to the new ownership group.

g. Repeat steps "3.c" on page 226 to "3.f" on page 226 for any other child pool.

h. After you assign ownership groups to all child pools, create a user group to include users that are the owners of the ownership group. Select **Access** > **Users** and click **Create User Group**

i. On the **Create User Group** page, enter the following information:

Name
Enter the name of the user group.

Role
Select the role for all users within the user group. User groups that are assigned in ownership groups cannot use the Security Administrator role.

Ownership Group
Select the ownership group that you created earlier and assign it to this user group.

j. On the **Access** > **Users** page, select existing users or create new users to assign to the user group. These users automatically inherit the ownership group that is assigned to the user group. These users can now start creating objects like hosts and volumes within this ownership group, using capacity in the child pools that are assigned to the ownership group.

Deleting an ownership group

You can use the management GUI or the command-line interface (CLI) to delete ownership groups.

An ownership group defines a subset of users and objects within the system.

Using the management GUI

1. Select **Access** > **Ownership Groups**. The **Ownership Group** page opens.
2. Click the menu button next to the ownership group that you want to delete, and then select **Remove Group**. The **Remove Ownership Group** page opens.
3. Select **Remove**.

Using the command-line interface

Use the **rmownershipgroup** command to delete an ownership group.

For example:

```
rmownershipgroup -keepobjects group_id_or_name
```

where *group_id_or_name* specifies ID or name of the ownership group to delete. The **keepobjects** parameter specifies to keep the objects that currently belong to the ownership group and to modify them to have no ownership group. You must specify the **keepobjects** parameter to delete an ownership group that still contains associated objects. You cannot delete an ownership group that still contains user groups, even if you specify the **keepobjects** parameter.

Users

Each user of the management GUI must provide a username and a password to sign on.

Each user also has an associated role, such as monitor or security administrator. These roles are defined at the system level. For example, a user can be the administrator for one system, but the security administrator for another system.

Security Administrators can create role-based user groups where any users that are added to the group adopt the role that is assigned to that group. Roles apply to both local and remote users on the system and are based on the user group to which the user belongs. A local user can belong only to a single group; therefore, the role of a local user is defined by the single group to which that user belongs. Roles are defined at the system level, which means that a user can be an administrator on one system, but a security administrator on another system. Remote users that use either LDAP or single sign-on require a user group on the system that matches a group defined on the remote authentication server. Security Administrators can also select to enable multifactor authentication per user group. With multifactor authentication, any users that are assigned to a user group must present a second factor to access the system. Multifactor authentication requires that an authentication server is configured to verify the identity of all users within the group.

You can assign the following roles to your user groups:

Monitor
: Users have access to all actions that are related to viewing objects and processes on the system. Monitor-role users cannot change the state of the system nor change the resources that the system manages. Monitor-role users can access all information-related GUI functions and commands, back up configuration data, and change their own passwords.

Service
: Users can set the time and date on the system, delete dump files, add and delete nodes, apply service, and shut down the system. Users can also complete the same tasks as users in the monitor role.

Administrator
: Users can manage all functions of the system except those functions that manage users, user groups, authentication, and encryption. Administrator-role users can run the system commands that the security-administrator-role users can run from the CLI, except for commands that deal with users, user groups, authentication, and encryption.

Security Administrator
Users can manage all functions of the system, including managing users, user groups, user authentication, and configuring encryption. Users with the Security Administrator role can run any system commands from the command-line interface (CLI). However, they cannot run the **satask** command from the CLI. Only the **superuser** ID can run **satask** command.

Restricted Security Administrator
Users with the security administrator role is changed to restricted security administrator when two person integrity (TPI) is enabled on the IBM Storage Virtualize. TPI requires two security administrators to work together to complete critical or risky tasks. For example, a restricted security administrator with an approved TPI request can remove Safeguarded snapshots.

User groups

Users with the Security Administrator role can organize users of the system by role through user groups.

The following user groups are configured by default:

SecurityAdmin
Users access all functions on the system, including managing users, user groups, and user authentication. Users can also configure encryption on the system.

Administrator
Users can complete most of the same tasks as users who are in the SecurityAdmin role. However, these users cannot access functions that deal with managing users, user groups, and authentication.

RestrictedAdmin
Users can complete the same tasks and run most of the same commands as administrator-role users. Support personnel can be assigned this role to help resolve errors and fix problems.

Service
Users can set the time and date on the system, delete dump files, add and delete nodes, apply service, and shut down the system. Users can also complete the same tasks as users in the monitor role.

Monitor
Users can view objects and the system configuration settings but cannot configure, modify, or manage the system or its resources.

Locking user accounts

The Security Administrator can manually lock or unlock a user account at any time. By default, the superuser is exempt from the system-wide policy for manual or automatic account locking.

When a system is created, a single local user with Security Administrator privileges, called superuser, is created. The superuser contains maximum privileges to complete system setup and configuration. For new systems, the default superuser password must be changed on first login to the system. Although the superuser cannot be deleted, you may want to lock or disable the superuser to prevent access to the system. In addition, the following system-wide policies can be set:

- Automatic lock after (0-10) consecutive failed login attempts. A value of 0 means that the account does not lock because of failed login attempts.
- Set length of time auto-lock applies for (0-10080) minutes (7 days maximum). A value of 0 means that the account remains locked out indefinitely.
- Superuser account can be locked from general login.

User account can be manually locked using the **chuser** command, or automatically locked depending on the locking policy defined on the system.

Locking the superuser account

For systems with a dedicated technician port, the system supports locking the superuser account for environments with strict security requirements.

The superuser account is the only account on the system that can complete these tasks:

- The superuser account is the only user that exists in service mode and can access the service assistant or run or view service-related command-line operations (or **satask sainfo** commands).
- The superuser account is the only user that can run the system recovery procedure and the backup and restore system configuration procedure.

Other Security Administrator accounts cannot complete these tasks if the superuser account becomes locked. If **Allow locking of the superuser account** is checked on the **Password Policy** page, these critical actions are disabled.

If you plan to allow the superuser account to be locked, ensure that you complete the following tasks on the system:

Enable Support Assistance
Support assistance enables support personnel to access the system either remotely from the support center, or locally from a console. If the superuser account is locked, a support person can log in to service the system when Support Assistance is enabled.

Ensure that password reset for the superuser account is enabled
You must ensure that the superuser password can be rest if you plan on allowing locks on the superuser account. This feature is enabled by default. If this feature is not enabled, unlocking the account by resetting the password on the service assistant GUI or service command-line interface is not possible. Use the command **svctask setpwdreset -enable** to enable this feature.

Ensure that an account with the role of SecurityAdmin exists
Ensure that a separate account with the user role of SecurityAdmin exists, either locally or on a remote LDAP server. This account can be used to unlock the superuser account.

The following table describes different use cases where the superuser account is locked and the possible recovery actions:

Locked superuser use case	Recovery action
Superuser account is locked after specified failed login attempts for a specified lockout time period.	Wait for the specified time for the lockout to expire and reattempt log in.
Superuser account is locked after the specified login attempts but an indefinite lockout was selected, or superuser account is manually disabled because it is not intended for use.	Complete one of these tasks to unlock the superuser account: • Manually unlock the superuser account that uses another account with the SecurityAdmin user role. • Create a remote support assistance request. • Use the technician port to access the service assistant GUI.
Superuser account is locked after two person integrity (TPI) is enabled.	Complete one of these tasks to unlock the superuser account: • Disable TPI. • Request remote support assistance. • Use the technician port to access the service assistant GUI.

Managing users and user groups

You can use the management GUI or the command-line interface to manage user and user groups on the system.

Changing user groups

You can use the command-line interface (CLI) to change user groups. User groups organize users of a system by role.

About this task

Roles apply to both local and remote users on the system and are based on the user group to which the user belongs. A local user can belong only to a single group; therefore, the role of a local user is defined by the single group that the user belongs to. Remote users can belong to one or more groups; therefore, the roles of remote users are assigned according to the groups that the remote user belongs to.

In addition to assigning roles to user groups, you can use user groups to control how user access different management interfaces and functions that are supported on the system. You can create a user group with specific roles and further control how users within that group interact with objects on the system by allowing or restricting the user group to functions or interfaces. With this capability, you can also control access from scripts or automated services, such as IBM Storage Insights that access the system automatically. The following interfaces and functions can be enabled or disabled with user groups:

Table 35. Supported access control for management interfaces and functions with user groups

Management interface or function	**Supported interface for configuration**
Management GUI access	Command-line interface
Command-line interface access	Command-line interface
REST API access	Command-line interface
Multifactor authentication	Management GUI and command-line interface
Ability to require a password and key for access	Command-line interface

Note: If any remote user groups on the system are configured with multifactor authentication, then it is not possible to use the BatchMode=yes setting when logging in using SSH for any remote users on the system.

You can use this feature to enable or disable access to interfaces from automated scripts or services, such as data collection for IBM Storage Insights or other automated processes. You can further secure the system by creating an ID for the automation service with only the role that they require to access the information necessary to the application or service. Generally, to configure a user group and access controls for an automation service, you need to complete the following steps:

1. Create a separate user group for automation IDs.
2. Create the automation users with appropriate role. For example, for data collection for IBM Storage Insights, the Monitor role allows the service to view objects and the system configuration settings. Other automated services or scripts can require different role-based access depending on the operations the service completes on the system
3. Disables interfaces for these automation IDs. Depending on the automated service, you can control which interfaces the automation service can use to complete tasks on the system. For example, with IBM Storage Insights data collection, you can limit these automation IDs to command-line access only, since other interfaces and functions are not necessary.
4. In cases where SSH keys are used to establish a connection between the system and the automated service, ensure that the key itself is secured and use these SSH keys with the automated service only.

To change a user group in the management GUI, select **Access** > **Users**. Select a user group and select **Properties** from the **Actions** menu.

To change a user group using the CLI, use the **chusergrp** CLI command.

Procedure

1. Use the **chusergrp** CLI command to change the attributes of an existing user group.
   For example, enter the following command:

   ```
   chusergrp -role role_name -remote yes | no group_id_or_name
   ```

   where *role_name* specifies the role that is associated with any users that belong to this group and *group_id_or_name* specifies the group to be changed. The **remote** parameter specifies whether the group is visible to the authentication server.
2. Issue the **lsusergrp** CLI command to display the user groups that were created on the system.
   For example, enter the following command:

   ```
   lsusergrp usergrp_id_or_name
   ```

   where *group_id_or_name* specifies the user group to view. If you do not specify a user group ID or name, all user groups on the system are displayed.

Example

If you plan to disable or enable interface access or functions, you can use the either the **mkusergrp** or **chusergrp** command. Both commands provide the following parameters to disable and enable access and functions:

-disablegui
Specifies if the user group has management GUI access disabled or not. Value can be yes or no.

-disablecli
Specifies if the user group has command-line interface access disabled or not. Value can be yes or no.

-disablerest
Specifies if the user group has REST-API access disabled or not. Value can be yes or no.

-passwordkeyrequired
Specifies if the user group requires a password and SSH key. Value can be yes or no.

-multifactor
Specifies if the user group has multifactor authentication enabled or not. Value can be yes or no.

The following examples show the command operations for some common uses cases for these parameters:

Limiting access to automated service
Create or change a user group with the appropriate role and restrictions:

```
mkusergrp -name automatedservice -role Monitor -passwordkeyrequired yes -disablecli no
-disablerest yes
Modifying the authentication setting for this user group will affect logins for all users
in the group.
Are you sure you want to continue? (y/yes to confirm)  yes
```

To change an existing user group, run the following command:

```
chusergrp -name automatedservice -role Monitor -disablegui yes -disablecli no -disablerest
yes
Modifying the authentication setting for this user group will affect logins for all users
in the group.
Are you sure you want to continue? (y/yes to confirm)  yes
```

In both of these examples, a user group was dedicated to an automated service, such as the data collector for IBM Storage Insights, that disabled access to the management GUI and REST APIs, but allows access through the command-line interface. Users in this group can only perform functions that are allowed by the Monitor role, such as viewing and collecting data that is related to operations on the system.

Enabling a password and SSH key for a user group

```
chusergrp -name serviceusergroup -role Service -passwordkeyrequired yes
Modifying the authentication setting for this user group will affect logins for all users
in the group.
Are you sure you want to continue? (y/yes to confirm)  yes
```

Enabling multifactor authentication for a user group

```
chusergrp -name mfagroup -role Administrator --multifactor yes
Modifying the authentication setting for this user group will affect logins for all users
in the group.
Are you sure you want to continue? (y/yes to confirm)  yes
```

Note:

- Before you can enable multifactor authentication on the user group, you must ensure all prerequisites are completed as well as configuring a multifactor authentication service on the system.
- Multifactor authentication is not supported on the Common Information Model (CIM). Therefore, user groups configured with multifactor authentication, or password and SSH key authentication are disabled for Common Information Model (CIM) interface.

Deleting user groups

You can use the command-line interface (CLI) to delete user groups. User groups organize users of a volume storage system by role.

About this task

Roles apply to both local and remote users on the system and are based on the user group to which the user belongs. A local user can belong only to a single group; therefore, the role of a local user is defined by the single group that the user belongs to. Remote users can belong to one or more groups; therefore, the roles of remote users are assigned according to the groups that the remote user belongs to.

To delete a user group in the management GUI, select **Access** > **Users**. Select a user group and select **Delete** from the **Actions** menu.

To delete a user group with the CLI, complete these steps:

Procedure

Issue the **rmusergrp** CLI command to delete a user group:
For example:

```
rmusergrp -force group_id_or_name
```

where *group_id_or_name* specifies the group to delete. The **force** parameter specifies to delete the group even if there are users in the user group.

Important: Using the **force** parameter might result in a loss of access. Use it only under the direction of the support center.

Changing users

You can use the command-line interface (CLI) or the management GUI to change users on the system.

Before you begin

System users must provide either a password, a Secure Shell (SSH) key, or both. Local users are authenticated through the authentication methods that are on the system.

You can create two categories of users that access the system. These user types are based on how they authenticate to the system:

- Some users must provide an SSH password (or if not possible an SSH key).
- If a user needs access to the management GUI, a password is needed for the user.
- If the user requires access to the command-line interface (CLI), a valid password and SSH key can be used.
- Users must be in a user group that is defined on the system.

Remote users can also configure local credentials if they need to access the system when the remote service is down. Remote users have their groups that are defined by the remote authentication service.

To change a user in the management GUI, select **Access** > **Users**. Right-click the user and select **Modify** from the **Actions** menu. For local users, you can change the user name, password, and the SSH Public key file. For remote users, you can change the user name and the SSH Public key file. The password is maintained and authenticated through the remote authentication service.

About this task

To change a user in the CLI, follow these steps:

Procedure

1. Use the **chuser** CLI command to change the attributes of an existing user.
   For example, enter the following command:

   ```
   chuser -usergrp group_id_or_name user_id_or_name
   ```

   where the *group_id_or_name* specifies the new group for the user and *user_id_or_name* specifies the user to be changed.
2. Use the **chcurrentuser** CLI command to change the attributes of the current user.
   For example, enter the following command:

   ```
   chcurrentuser -nokey
   ```

   where the **nokey** parameter specifies that the SSH key of the user is to be deleted.
3. Use the **lscurrentuser** CLI command to display the name and role of the logged-in user.
   For example, enter the following command:

   ```
   lscurrentuser
   ```

   The name and the role of the user are displayed.

Deleting users

You can use the command-line interface (CLI) or the management GUI to delete users on the system.

About this task

To delete a user in the management GUI, select **Access** > **Users**. Select a user and select **Delete** from the **Actions** menu.

To delete a user from the CLI, complete these steps:

Procedure

1. Issue the **rmuser** CLI command to delete a user:
   For example:

   ```
   rmuser user_id_or_name
   ```

   where *user_id_or_name* specifies the user to be removed.
2. Issue the **lsuser** CLI command to display a list of users that were created on the system.

For example:

```
lsuser
```

If the user ID or name is not specified, all users on the system are displayed. Verify that the deleted user was removed from the system.

Locking and unlocking user account automatically

Accounts can be locked automatically by the system after a certain number unsuccessful login attempts. A user with Security Administrator access can unlock the account to provide access again.

The account can be locked for a specific amount of time, or the account can be locked indefinitely. A user with the Security Administrator access can unlock the account to provide access again

Enable automatic account locking

Enable automatic account locking by using management GUI or command-line interface.

Using the management GUI

To enable automatic account locking using the management GUI, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Password Policies**.
2. On **Password Policies** page, select **Password expiration and account lockout**.
3. Select the checkbox **After a number of failed login attempts, automatically lock accounts for an amount of time**.
4. Enter the number of consecutive failed login attempts before the account is locked (You can select 1-10 attempts).
5. Select either Define time period (minutes) or Indefinite lockout. If you choose to define the time period, enter the number of minutes that an account should remain locked for before automatically unlocking (You can select 1-10080 minutes). If you select indefinite lockout, accounts remain locked until a user with the Security Administrator access unlocks the account manually. See “Unlocking user accounts manually” on page 236.
6. Select **Allow locking of the superuser account** to enable automatic account locking for the superuser account, for more information see “Locking and unlocking superuser accounts” on page 236.

   Note: If **Allow locking of the superuser account** is not enabled, then the superuser account will not be automatically locked when the specified number of failed attempts is reached.
7. Click **Save**.

Using the command-line interface

To enable automatic account locking using the command-line interface, enter the following command:

1. Decide how many consecutive failed login attempts (1-10) should lock the account, and how long the account should remain locked for when it becomes locked (0-10080 minutes).

   Note: A time of 0 minutes means that the account will be locked indefinitely until manually unlocked by a user with the Security Administrator access.
2. Specify required values, enter the following command:

   ```
   svctask chsecurity -maxfailedlogins 5 -lockoutperiod 30
   ```
3. To enable the superuser locking, enter the following command:

   ```
   svctask chseurity -superuserlocking enable
   ```

Disable automatic account locking

Disable automatic account locking by using management GUI or command-line interface.

Using the management GUI

To disable automatic account locking using the management GUI, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Password Policies**.
2. On **Password Policies** page, select **Password expiration and account lockout**.
3. Deselect the checkbox **After a number of failed login attempts, automatically lock accounts for an amount of time**.
4. Choose whether to deselect the checkbox **Allow locking of the superuser account**. This setting should not be deselected if there is a requirement to manually lock (or disable) the superuser account.
5. Click **Save**.

Using the command-line interface

To disable automatic account locking using the command-line interface, enter the following command:

1. To disable automatic account locking for all users, enter the following command:

```
svctask chsecurity -maxfailedlogins 0
```

2. To disable the superuser locking, enter the following command:

```
svctask chseurity -superuserlocking disable
```

Locking and unlocking user accounts manually

A user with the Security Administrator role can lock or unlock user accounts. This allows you to prevent the log in of the specific user if there is suspicious activity that is associated with this account.

Note: You can manually lock the user account even if automatic locking is not enabled. You cannot manually lock your own user account.

Locking user accounts manually

You can use the management GUI or the command-line interface to lock user accounts manually.

Before you begin

You must be assigned with the Security Administrator role to lock user accounts manually.

Using the management GUI

To lock the user account, complete the following steps:

1. In the management GUI, click **Access** > **Users by Group**. The **Users by Group** window is displayed, which shows the **All Users** section.
2. From the **All users** section, right-click on the user and click **Lock** to lock the selected user account.

Using the command-line interface

To lock the user account manually, complete these following steps:

1. To view the ID and name list of the user accounts, enter the following command:

```
lsuser
```

2. To lock the user account, enter the following command:

```
svctask chuser -lock <name_or_id>
```

Where <name_or_id> specifies the name or ID of the user account. The user account remains locked until you manually unlock the user account.

Unlocking user accounts manually

You can use the management GUI or the command-line interface to unlock user accounts manually.

Before you begin

You must be assigned with the Security Administrator role to unlock user accounts manually.

Using the management GUI

To unlock the user account, complete the following steps:

1. In the management GUI, click **Access** > **Users by Group**. The **Users by Group** window is displayed, which shows the **All Users**section.
2. From the **All users** section, right-click on the locked user and click **Unlock** to unlock the selected user account.

Using the command-line interface

To unlock the user account manually, complete these following steps:

1. To view the ID and name list of the user accounts, run **lsuser** command.
2. To unlock the user account, enter the following command:

```
svctask chuser -unlock <name_or_id>
```

Where <name_or_id> specifies the name or ID of the locked user account.

Locking and unlocking superuser accounts

A Security Administrator can manually lock or unlock a user account at any time. By default, the superuser is exempt from the system-wide policy for manual or automatic account locking.

For systems with a dedicated technician port, the system supports locking the superuser account for environments with strict security requirements.

When a system is created, a single local user with Security Administrator privileges, called superuser, is created. The superuser contains maximum privileges to complete system setup and configuration. For new systems, the default superuser password must be changed on first login to the system. Although the superuser cannot be deleted, you may want to lock or disable the superuser to prevent access to the system.

The superuser account is the only account on the system that can complete these tasks:

- The superuser account is the only user that exists in service mode. This user can access the service assistant to run service-related command-line operations, such as (**satask** and **sainfo** commands).
- The superuser account is the only user that can run the system recovery procedure and the backup and restore system configuration procedure.

Other Security Administrator accounts cannot complete these service tasks. If the superuser account is locked, then these service actions are unavailable until the superuser account is unlocked.

If you plan to allow the superuser account to be locked, consider completing the following tasks on the system to ensure that superuser access can be restored when needed:

Enable Support Assistance
Support assistance enables support personnel to access the system either remotely from the support center, or locally from a console. If the superuser account is locked, a support person can log in to service the system when Support Assistance is enabled.

Ensure that password reset for the superuser account is enabled
You must ensure that the superuser password can be reset if you plan on allowing locks on the superuser account. This feature is enabled by default. If this feature is not enabled, unlocking the account by resetting the password on the service assistant GUI or service command-line interface is not possible. Use the command **`svctask setpwdreset -enable`** to enable this feature.

Ensure that an account with the role of SecurityAdmin exists
Consider creating a separate account on the system with the user role of SecurityAdmin, either locally or on a remote LDAP server. This account can be used to unlock the superuser account.

Manually locking a superuser account

The system allows to lock or disable the superuser account by using both the management GUI and the command-line interface.

Note: Superuser locking must be enabled before the system-wide automatic locking settings are applied to the superuser account.

Using the management GUI

To lock or disable the superuser account using the management GUI, complete the following steps:

1. Log into the management GUI as a user with the Security Administrator role.
2. Select **Settings** > **Security** > **Password Policies**.
3. Select **Password expiration and account lockout** and then select **Allow locking of the superuser account**.
4. Click **Save** to enable superuser locking.
5. Select **Access** > **Users by Group**.
6. Right-click the superuser account under **All Users**.
7. Click **Lock** to manually lock the superuser account.

Using the command-line interface

To lock or disable the superuser account by using the command-line interface (CLI), complete the following steps:

1. Log in to the CLI as a user with the Security Administrator role.
2. To enable superuser locking, run the following command:

```
svctask chsecurity -superuserlocking enable
```

3. To manually lock the superuser account, run the following command:

```
svctask chuser -lock superuser
```

Unlocking a superuser account

The system allows to unlock or re-enable the superuser account by using both the management GUI and the command-line interface.

If the locked superuser account is required to complete a service or recovery action, the account must be re-enabled. Re-enabling the superuser account is possible under following circumstances:

1. If there is another user with the Security Administrator role, then that user can unlock the superuser account using the management GUI or the CLI.

2. If Remote Support Assistance is enabled, then remote support personnel can connect to the system to perform service actions.
3. If dedicated technician port is available, the system allows to unlock the superuser account using that port. This requires physical access to the system. Connect to the system using the superuser account. The service GUI provides an option to unlock the superuser account, or use the service CLI to run the *satask unlocksuperuser* command.

Using the management GUI

To unlock or re-enable the superuser account using the management GUI, complete the following steps:

1. Log in to the management GUI as a user with the Security Administrator role.
2. Select **Access** > **Users by Group**.
3. Right-click the superuser account under **All Users** and click **Unlock**.
4. Select **Settings** > **Security** > **Password Policies**.
5. Select **Password expiration and account lockout** and then deselect **Allow locking of the superuser account**.

   Note: Deselecting the Allow locking of the superuser account check box prevents both manual and automatic locking of the superuser account. Skip the above step to keep automatic locking of the superuser account enabled if automatic locking is also enabled system-wide.
6. Click **Save**.

Using the command-line interface

To unlock or re-enable the superuser account using the command-line interface (CLI), complete the following steps:

1. Log in to the CLI as a user with the Security Administrator role.
2. To unlock the superuser account, run the following command:

   ```
   svctask chuser -unlock superuser
   ```
3. To disable superuser locking, run the following command

   ```
   svctask chsecurity -superuserlocking disable
   ```

   Note: Disabling superuser locking prevents both manual and automatic locking of the superuser account. Skip the above step to keep automatic locking of the superuser account enabled if automatic locking is also enabled system-wide.

The following table describes different use cases where the superuser account is locked and the possible recovery actions:

Locked superuser use case	Recovery action
Superuser account is locked after specified failed login attempts for a specified lockout time period.	Wait for the specified time for the lockout to expire and reattempt log in.
Superuser account is locked after the specified login attempts but an indefinite lockout was selected, or superuser account is manually disabled because it is not intended for use.	Complete one of these tasks to unlock the superuser account: • Manually unlock the superuser account that uses another account with the SecurityAdmin user role. • Create a remote support assistance request. • Use the technician port to access the service assistant GUI.

Locked superuser use case	Recovery action
Superuser account is locked after two person integrity (TPI) is enabled.	Complete one of these tasks to unlock the superuser account: • Disable TPI. • Request remote support assistance. • Use the technician port to access the service assistant GUI.

Working with local and remote users

You can create either a local or a remote user to access a system.

Before you begin

You can create two categories of users that access the system. These types are based on how the users are authenticated to the system. Local users must provide the *username* and password, and if you require command line access without entering a password, a Secure Shell (SSH) key - or both. Local users are authenticated through the authentication methods that are located on the system.

If the local user needs access to management GUI, a password is needed for the user. Access to the command-line interface (CLI) is also possible with the same password or (alternatively) a valid SSH key can be used. An SSH password is required if a user is working with both interfaces. User groups define roles that authorize the users within that group to a specific set of operations on the system.

Local users must be part of a user group that is defined on the system.

A remote user is authenticated on a remote service that is provided by an LDAP server and does not need local authentication methods. The remote user does not need to be added to the list of users on the system. If the user requires command-line access without entering a password, the remote user can be added to the system and an SSH key can be specified for the user.

Remote users cannot access the system when the remote service is down. A local user account must be used until the remote service is restored. The user groups a remote user is a member of are defined by the remote authentication service. To define a remote user, create a user group on the local computer that is also defined on the remote authentication service.

You can connect to the system by using the same username with which you log in to the system.

About this task

Complete these steps to create either a local or remote user:

Procedure

1. Select **Access** > **Users** .
2. Select the appropriate user group.
3. Click **Create User**.
4. Enter the information on the new user and click **Create**.

Audit log

An audit log documents user actions that are submitted through the different management interfaces that run commands on the system, such as the management GUI.

You can use the audit log to monitor user activity on all supported interfaces.

The audit log entries provide the following information:

- The time and date when the action or command was submitted on the system.

Note: When logs are displayed in the command-line interface, the time stamps for the logs in CLI are the system time. However, when logs are displayed in the management GUI, the time stamps are converted to the local time where the Web browser is running.

- The name of the user who completed the action or command.
- The IP address of the system where the action or command was submitted.
- The name of source and target node on which the command was submitted.
- The parameters that were submitted with the command, except confidential information like clear-text passwords.
- The results of the command or action that completed successfully. Failed commands are not added to the audit log.
- The sequence number and the object identifier that is associated with the command or action.
- The origin of the command or action. Possible values are: GUI, CLI, or REST.
- For service-related actions that are completed by support personnel, the audit log displays the following information:

  Challenge
  Indicates the challenge with which the support user authenticated to the system. Support users connect to the system through a challenge-response authentication. If authentication succeeds, a session is opened and support users can conduct service-related actions on the system. This field identifies the support user at the support center. This field is blank for other users.

Login message

You can create or change a message that displays when users log on to the system.

About this task

When users log on to the system with the management GUI, command-line interface, or service assistant, the message displays before they log on to the system. To create or change the login message, you can use the **chbanner** command or the management GUI. If you are using the command, you must create the message in a supported text editor and use Secure Copy (SCP) to copy the file to the configuration node on the system. In the management GUI, select **Settings** > **GUI Preferences** > **Login** to enable and create a login message.

Procedure

To change the login message from a SAN administrator workstation, complete the following steps.

1. Use a suitable text editor to create a file that contains the text of the message.

   Note: The message cannot exceed 4 Kbytes.
2. Use a Secure Copy client to copy the file to the configuration node of the system to be configured. Specify the management IP address of the system to be configured.
3. Log on to the system to be configured.
4. In the command-line interface, type the following command to set the login message:

   ```
   chbanner -file filepath
   ```

   where *filepath* specifies the file that contains the text of the new message.

Example

```
echo "Access to this system is permitted only for authorized personnel." > loginmessage
 scp loginmessage admin@12.23.34.45:/tmp/loginmessage
   (enter password)
 ssh admin@12.23.34.45
   (enter password)
 chbanner -file /tmp/loginmessage
```

Chapter 10. FlashSystem grid

A FlashSystem grid can be used to scale-out a storage system by adding more independent systems to create a federated cluster.

A FlashSystem grid allows for greater scalability than a traditional scale-out system (which uses I/O groups) as it supports up to eight systems per FlashSystem grid with linear increases in performance, capacity, and number of volumes, hosts, and snapshots. Systems in a FlashSystem grid can be updated independently and systems with different hardware models and generations can be included within a single FlashSystem grid. This capability allows for hardware replacement or tiering within the same FlashSystem grid.

Systems in a FlashSystem grid have the following requirements:

- The system supports a maximum of a single I/O group and the compatibility mode for multiple I/O groups must be disabled.
- The system topology must be 'standard'.
- The system layer must be 'replication'.

FlashSystem grid functionality is not supported on some platforms due to hardware limitation. If applicable, the limitations are documented in a Software feature limits page in the relevant Hardware Guide.

Systems in a FlashSystem grid exchange certificates to enable secure communication between members. System-signed certificates are the default. However, a CA-signed certificate can be used instead, provided it meets the requirements. For more information, see Requesting and installing an externally signed certificate.

A FlashSystem grid can be managed by using the manageflashgrid task.

© Copyright IBM Corp. 2024

Chapter 11. Storage partitions

Storage partitions are used to implement policy-based high availability and disaster recovery solutions. Partitions on different systems can be linked to configure replication for disaster recovery between the systems. Replication policies can be assigned to individual volume groups, enabling different policies to be used for different groups within the same partition.

Partitions contain volumes, volume groups, hosts, and host-to-volume mappings. Within a partition:

- All volumes are in volume groups.
- Mappings can be created between volumes and hosts in the same partition only.
- Volumes in a storage partition cannot be mapped to other IBM Storage Virtualize systems.

Each partition that is associated with a high availability replication policy has two properties - the preferred management system and the active management system.

All configuration actions on a storage partition must be performed on the active management system. The storage partition can be monitored on either system.

The preferred management system is the system that you want to be the active management system under ideal conditions. If the active management system and the preferred management system are not the same, the system automatically switches the active management system back to the preferred management system when it is able. You can change the preferred management system.

You can configure more volumes, volume groups, hosts, and host-to-volume mappings at any time, either by adding to an existing partition or by creating a new one. A partition must include all volumes that are mapped to any hosts included in the partition.

Only SCSI hosts that are connected with Fibre Channel are supported in storage partitions.

Draft partitions

A draft partition can be used to simulate modifications to the configuration without influencing the behavior of the system. A draft partition can be used to construct a partition that meets the requirements of a partition. Some of these requirements are not enforced while constructing a draft partition. Some conditions result in errors that identify the issue that must be resolved to update the draft partition. A draft partition can only be published when these requirements are met. The GUI helps to add volume groups to the draft and indicate which volumes and hosts are added automatically. A draft partition has no effect until it is published.

© Copyright IBM Corp. 2024

Chapter 12. Data migration

This topic describes methods to migrate data.

Data migration applications

Data migration has the following characteristics and applications:

- The system supports migrating existing volumes to a data reduction pool with volume mirroring. For details, see “Moving volumes to a data reduction pool” on page 247.
- For systems with encryption enabled, you can migrate existing volumes from non-encrypted pools to encrypted pools. For details, see “Migrating volumes to an encrypted pool” on page 249.
- Changes the MDisks to which the volume is allocated.
- Does not disrupt access to the volume data.
- Can be used when you are removing MDisks from a clustered system.
- Can move data to an MDisk that has appropriate performance for host requirements.
- Can move data from image mode volumes to MDisks when you are adding a system to an existing SAN infrastructure.
- Can migrate striped and sequential volumes to an image mode volume.

For more information on the command-line interface, see **`Migration commands`**.

Migrating storage partitions between systems

IBM Storage Virtualize supports the non-disruptive migration of storage partitions between systems. This enables a partition to be migrated from one Storage Virtualize system to another without any application downtime.

Migrating storage partitions can be used for the following cases:

- Migrating to a new storage array before decommissioning the original system.
- Load-balancing by migrating storage partitions from overloaded systems to other systems.

Migration enables relocation of storage partitions from a source system to a target system location. As a consequence of which, the following flow of events occurs:

- All the resources associated with the storage partition are created on the migration target system, and the copy process starts from the source system to the target system.
- Host I/O is served from the migration target storage system once the synchronization process is complete.
- At the end of the migration process, the storage partition and all of its resources are removed from the source system.

For more information, see **`chpartition`** command to automate the procedure of migration of storage partitions. This automates intermediate steps, such as setting up Fibre Channel partnerships between the systems or establishing a policy-based high availability replication policy.

Limits and restrictions

- Draft storage partitions and storage partitions configured for policy-based high availability are not supported for storage partition migration.
- Storage partition migration is supported only with Fibre Channel partnerships between systems.
- Storage partition migration is supported only with Fibre Channel hosts.

© Copyright IBM Corp. 2024

- Only one storage partition can be migrated at a time from a system. Any consecutive migrations that are attempted get queued, and are scheduled automatically as per the sequence of invocations when the earlier migrations complete.
- If a system has storage partitions configured for high availability, partition migration can only be used with the HA partner system.

Prerequisites

Before you can use non-disruptive storage partition migration function, ensure that the following prerequisites are met:

1. Migrating a storage partition requires both Fibre Channel and IP connectivity between the source and target storage systems. The requirements are the same as that for configuring high availability (HA). Review the high availability requirements to ensure that the storage partition supports being migrated and the host operating systems support this feature. For details, see “Planning high availability” on page 202.
   a. Make sure that both systems have their certificates added to the truststore of each other, with the REST API usage enabled. For more information, see mktruststore .

      Note: It is required that the systems that are involved in storage partition migrations, the time must be synchronized between the systems. The best way to achieve this is to configure each system to use a Network Time Protocol (NTP) service.
2. Confirm that both systems are members of the same FlashSystem grid, or that neither system is a member of a FlashSystem grid, and that both systems meet the requirements for FlashSystem grid. For more details, see FlashSystem grid.
3. Use the **lspartnershipcandidate** command and make sure that both the source and the target systems are correctly zoned and are visible to each other. For more information, see **lspartnershipcandidate** command. Alternatively, if the target system is already part of the partnership, use the **lspartnership** command to check the partnership status between the source and the specific system.
4. Although the systems are visible to each other, an identified system can have multiple suitable storage pools that are able to host the migrated storage partition. Establish a pool link between systems on the current source and target systems. For more information, see **chmdiskgrp** command.

Procedure

1. Run the **chpartition** command with the `-location` option to migrate the storage partition to a different system.
2. The following example initiates a migration of storage partition to the designated location system:

   ```
   chpartition -location <remote_system> mypartition1
   ```

   Note: If there are two or more systems with the same name, then specify remote system ID instead of remote system name while initiating the migration.
3. To check the migration status, run **lspartition**.

   Note: You can monitor the progress of the migration, including the amount of data remaining to be copied, using the **lsvolumegroupreplication** command.
4. Once the storage partition's data and configuration has been migrated to the target storage system, you must confirm that the affected hosts have discovered paths to the volumes in the target storage system.
5. Another event prompts the storage administrator to remove the copy of the data and configuration of the storage partition from the source storage system to complete the migration.

   Note: The migration can be canceled and the partition rolled back to its original system before fixing the event if the administrator observes any issues with the migrated storage partition.

6. An informational event at the target storage system marks the completion of the storage partition migration.

   Note:

   You can monitor the migration using `migration_status` field that is shown by **lspartition** command indicating that there is no migration activity active or queued for that storage partition.

   A ongoing storage partition migration can be aborted by specifying with a new migration location that uses "-override" option. The migration to the new location gets queued behind existing queued migrations, if any. For more information, see **chpartition**.An active or queued migration operation gets aborted when its source system location is provided as a specified location. This needs an -override option since the abort operations remove the data that is replicated until that moment. For more information, refer to the **chpartition** command.

Rolling back a migration

The migrated storage partition serving I/Os from the target location (after fixing the event for host health checks at the source) can be rolled back to its source location. You can roll back a migration from the target system. For more information, refer to the **chpartition** command.

A Rollback operation may be needed if the migrated storage partition does not perform as expected on the target system post migration. Rolling back operation switches the I/O operations to the storage partition at the source location. This operation does not remove the copy-synced storage partition at the specified target location.

Moving volumes to a data reduction pool

Data reduction is a set of techniques that can be used to reduce the amount of usable capacity that is required to store data. An example of data reduction includes data deduplication. Data reduction can increase storage efficiency and performance and reduce storage costs, especially for flash storage. Data reduction reduces the amount of data that is stored on external storage systems and internal drives by reclaiming previously used capacity that are no longer needed by host systems. To estimate potential capacity savings that data reduction can provide on the system, use the Data Reduction Estimation Tool (DRET). This tool analyzes existing user workloads that are being migrated to a new system. The tool scans target workloads on all attached storage arrays, consolidates these results, and generates an estimate of potential data reduction savings for the entire system.

For more information about DRET, see https://www.ibm.com/support/pages/node/6217841. For more information about Comprestimator, see https://www.ibm.com/support/pages/node/6209688.

Note: The Data Reduction Estimation Tool also provides some analysis of potential compression savings for volumes. However, use the management GUI or the command-line interface to run the integrated Comprestimator Utility to gather data for potential compression savings for volumes in data reduction pools.

The system supports data reduction pools which can use different capacity savings methods simultaneously, increasing the capacity savings across the entire pool. Data reduction pools also support deduplication. When deduplication is specified for a volume, duplicate versions of data are eliminated and not written to storage, thus saving more usable capacity. Some models or software versions require specific hardware or software to use this function.

Support for the host SCSI unmap command is enabled by default.

Using the management GUI

To migrate the volume to a data reduction pool, complete these steps:

1. Create a data reduction pool by completing these steps:
   a. In the management GUI, select **Pools** > **Pools**.

b. On the **Pools** page, click **Create**.

c. On the **Create Pool** page, enter a name of the pool and select **Data Reduction**.

d. Click **Create**.

2. Add storage to the data reduction pool by completing these steps:

a. In the management GUI, select **Pools** > **Pools**.

b. Right-click the data reduction pool that you created and select **Add Storage**.

c. Select from the available storage and reserve capacity to the pool. Click **Assign**.

Note: You cannot select from available storage when you are creating a data reduction child pool. A data reduction child pool has access to the complete available capacity of the parent pool without limit. A data reduction child pool cannot be created if the parent pool does not have available capacity.

3. Create a copy of the volume in the standard pool by completing these steps:

a. In the management GUI, select **Pools** > **Volumes by Pools**.

b. Select the existing pool that contains the volumes you want to migrate to the data reduction pool. Right-click the volume and select **Add Volume Copy**.

c. On the **Add Volume Copy** page, select the data reduction pool to create a copy of the volume. Copy 1 is the original volume in the original pool, while copy 2 is created in the data reduction pool. Ensure that either thin-provisioning or compressed is selected for the capacity savings. For the option that you select, you can also select to use deduplication for the volume that you create. For example, you can create a thin-provisioned volume that also uses deduplication to remove duplicate data.

d. Click **Add**.

e. Ensure that the copies are synchronized before you proceed to the next steps. On the **Volumes** page, ensure that **Yes** is displayed for copy 2 in the **Synchronized** column.

Note: The time synchronization takes to complete depends on the size of the volume and system performance. You can increase the synchronization rate by right-clicking the volume and selecting **Modify Mirror Sync Rate**.

4. To complete the migration, remove the original volume copy by completing these steps:

a. In the management GUI, select **Volumes** > **Volumes**.

b. Right-click the volume copy and select **Delete**.

c. Click **Yes** to confirm the deletion.

Migrating extents using the CLI

To improve performance, you can migrate extents using the command-line interface (CLI).

About this task

The system provides various data migration features. These features can be used to move the placement of data both *within* parent pools and *between* parent pools. These features can be used concurrently with I/O operations. You can use either of these methods to migrate data:

1. Migrating data (extents) from one MDisk to another (within the same parent pool). This method can be used to remove highly used MDisks.
2. Migrating volumes from one parent pool to another. This method can be used to remove highly used parent pools. For example, you can reduce the use of a pool of MDisks. Child pools that receive their capacity from parent pools, cannot have extents that are migrated to them.

Notes:

1. The source MDisk must not currently be the source MDisk for any other migrate extents operation.

2. The destination MDisk must not be the destination MDisk for any other migrate extents operation.

Migration commands fail if the target or source volume is offline, there is no quorum disk defined, or the defined quorum disks are unavailable. Correct the offline or quorum disk condition and reissue the command.

You can determine the use of particular MDisks by gathering input/output (I/O) statistics about nodes, MDisks, and volumes. After you collect this data, you can analyze it to determine which MDisks are used frequently. The procedure then takes you through querying and migrating extents to different locations in the same parent pool. This procedure can only be completed using the command-line interface.

If performance monitoring tools indicate that an MDisk in the pool is being overused, you can migrate data to other MDisks within the same parent pool.

Procedure

1. Determine the number of extents that are in use by each volume for the MDisk by issuing this CLI command:

   ```
   lsmdiskextent mdiskname
   ```

   This command returns the number of extents that each volume is using on the MDisk. Select some of these extents to migrate within the pool.
2. Determine the other MDisks that are in the same volume.
   a) To determine the parent pool that the MDisk belongs to, issue this CLI command:

   ```
   lsmdisk mdiskname | ID
   ```

   b) List the MDisks in the pool by issuing this CLI command:

   ```
   lsmdisk -filtervalue mdisk_grp_name=mdiskgrpname
   ```
3. Select one of these MDisks as the target MDisk for the extents. You can determine how many free extents exist on an MDisk by issuing this CLI command:

   ```
   lsfreeextents mdiskname
   ```

   You can issue the **lsmdiskextent** *newmdiskname* command for each of the target MDisks to ensure that you are not just moving the over-utilization to another MDisk. Check that the volume that owns the set of extents to be moved does not already own a large set of extents on the target MDisk.
4. For each set of extents, issue this CLI command to move them to another MDisk:

   ```
   migrateexts -source mdiskname | ID -exts num_extents
   -target newmdiskname | ID -threads 4 -vdisk vdiskid
   ```

   where *num_extents* is the number of extents on the *vdiskid*. The *newmdiskname* | *ID* value is the name or ID of the MDisk to migrate this set of extents to.

   Note: The number of threads indicates the priority of the migration processing, where **1** is the lowest priority and **4** is the highest priority.
5. Repeat the previous steps for each set of extents that you are moving.
6. You can check the progress of the migration by issuing this CLI command:

   ```
   lsmigrate
   ```

Migrating volumes to an encrypted pool

For systems with encryption enabled, you can migrate existing volumes from non-encrypted pools to encrypted pools. Both the management GUI and the command line interface can be used to migrate volumes to encrypted pools.

During system setup in the management GUI, you can activate and enable encryption licenses. Any pools that are created after encryption is enabled are assigned a key that can be used to encrypt and decrypt

data. However, if encryption was configured after volumes were already assigned to non-encrypted pools, you can migrate those volumes to an encrypted pool by using child pools. When you create a child standard pool after encryption is enabled, an encryption key is created for the child pool even when the parent pool is not encrypted. You can then use volume mirroring to migrate the volumes from the non-encrypted parent pool to the encrypted child pool. You can use either the management GUI or the command line interface to migrate volumes to an encrypted pool.

When you create a child data reduction pool, the pool inherits the encryption status of the parent pool, and both pools use the same encryption key. You cannot use the **-encrypt** parameter when you create a child data reduction pool.

The system supports both internal storage and external storage. Internal storage consists of arrays that attached directly to the system through serial-attached SCSI connections. External storage consists of arrays that are attached to the system over a storage area network (SAN). Depending on the type of underlying storage the volume uses, the migration process is different.

If you are migrating volumes that use SAN-attached MDisks, you must complete the migration steps for "Migrating volumes that use SAN-attached MDisks" on page 250.

Migrating volumes that use SAN-attached MDisks

To migrate volumes that use SAN-attached MDisks and either standard pools or data reduction pools, complete these steps:

1. In the management GUI, select **Pools** > **Pools**.
2. Right-click the non-encrypted parent pool that contains the volumes that you want to migrate and select **Create Child Pool**.
3. On the **Create Child Pool** page, enter the name for the child pool. To create a standard child pool, enter the amount of capacity. Ensure that you select enough capacity to accommodate the migrated volumes. Encryption is selected by default when the system is enabled for encryption. To create a data reduction child pool, continue to the next step. Data reduction child pools have access to the complete free capacity of the parent pool without limit. A data reduction child pool cannot be created if the parent pool does not have free capacity.
4. Click **Create**. After the child pool is created, you can migrate the volumes to the child pool by adding volume copies.
5. In the management GUI, select **Volumes** > **Volumes by Pools**.
6. Select the non-encrypted parent pool to display all the volumes.
7. Right-click the volume and select **Add Volume Copy...**.
8. On the **Add Volume Copy** page, select **Basic** for the type of copy that you are creating. From the list of available pools, select the child pool as the target pool for the copy of the volume.
9. Click **Add**.
10. Repeat these steps to add volume copies to the encrypted child pool for the remaining volumes in the parent pool.
11. After all the copies are synchronized in the encrypted child pool, you can delete the primary copies from the parent pool. The empty parent pool must remain unused to use encrypted volumes in the child pool.

To migrate volumes that use SAN-attached MDisks and either standard pools data reduction pools, complete these steps:

1. In the command line interface, enter the following command to create a standard child pool:

   ```
   mkmdiskgrp -name my_encrypted_child_pool -parentmdiskgrp mypool -encrypt yes
   ```

   where `my_encrypted_child_pool` is the name of the new child pool and `mypool` is the name of the parent pool. Enter the following command to create a data reduction child pool:

   ```
   mkmdiskgrp -name my_encrypted_child_pool -parentmdiskgrp mypool -datareduction yes -noquota
   ```

where `my_encrypted_child_pool` is the name of the new child pool and `mypool` is the name of the parent pool.

2. Issue the CLI command `addvdiskcopy` to create mirrored copies of the volumes that are in the parent pool, in the new child pool. In the following example, `my_encrypted_child_pool` is the name of the new child pool and `volume1` is the name of the volume that is being copied.

```
addvdiskcopy -autodelete -mdiskgrp my_encrypted_child_pool -vdisk volume1
```

   Use the **`-autodelete`** operand to automatically delete the primary copy of the volume after the copy synchronizes.

3. Repeat step 2 until all the volumes from the original parent contain mirrored copies in the new child pool. The empty parent pool must remain unused to use encrypted volumes in the child pool.

Migrating volumes that use internal MDisks

If you have internally-attached enclosures that either do not support encryption or that do not have an encryption license that is activated and enabled and want encrypted volumes, you can still migrate volumes to encrypted pools. Before you can migrate volumes, you must upgrade your software to a version that supports encryption and obtain an encryption license. After you receive authorization documents for the encryption license, you must activate the license and enable the encryption function on the system before you attempt migration of any volumes. In addition, you must ensure that the unencrypted pool contains enough free extents to complete the migration. The migration fails if enough free extents are not available. You have two options to ensure that the pool contains enough extents:

- If you have spare drives, create a new encrypted array with the spare drives.
- If you do not have spare drives, add unassigned external MDisks to the parent pool and create a new encrypted child pool. If you chose this option and intend to keep the external MDisk in the pool after migration, then use the instructions in "Migrating volumes that use SAN-attached MDisks" on page 250.

These steps assume that these tasks were completed.

To migrate volumes that use internal MDisks in the management GUI, complete these steps:

1. In the management GUI, select **Pools** > **MDisks by Pools**.
2. Select the pool and right-click the MDisk that represents the array that you want to delete and select **Remove**.
3. To add the array back to the pool, select **MDisk by Pools**. The new available internal storage is displayed at the bottom of the pane. Select the icon that represents the drive class of the drives that were formerly in the array that was deleted and select **Assign**. The new array is encrypted by default. After all the arrays are encrypted, the pool is considered encrypted and data on volumes in that pool is also encrypted.
4. To verify that the data on these volumes was migrated to encrypted arrays and are now encrypted, select **Volumes** > **Volume by Pools**.
5. Select the pool and verify that the volumes that are listed are encrypted. A key icon is displayed under the **Encryption** column to indicate that the data on the volume is encrypted.

For migrated volumes that use internal MDisk in the command line interface, complete these steps:

1. To remove an array from pool, enter the following command:

```
rmarray -mdisk mdisk3 pool2
```

   In this example, `mdisk3` is the array that is removed from `pool2`.

2. To create a new encrypted array and add the array to the storage pool, enter the following command:

```
mkarray -level raid5 -drive 0:1:2:3 -mdiskgrp pool2
```

   In this example, the array is encrypted by default because encryption is enabled on the I/O group for the array. The data that is stored on the volumes that use that array is now encrypted.

3. To verify that the volumes in the pool are encrypted, enter the following command:

```
lsvdisk
```

The encrypt state is **yes** if all the data on the volume is encrypted.

Migration methods for image mode volumes

Several methods can be used to migrate image mode volumes into managed mode volumes.

About this task

To complete any type of migration on an image mode volume, the image mode volume must first be converted into a managed mode volume. The volume is automatically converted into a managed mode volume whenever any kind of migration activity is attempted. After the image-mode-to-managed-mode migration occurs, the volume becomes a managed mode volume and is treated the same way as any other managed mode volume.

If the image mode disk has a partial last extent, this last extent in the image mode volume must be the first to be migrated. This migration is processed as a special case. After this special migration operation occurs, the volume becomes a managed mode volume and is treated in the same way as any other managed mode volume. If the image mode disk does not have a partial last extent, no special processing is completed. The image mode volume is changed into a managed mode volume and is treated the same way as any other managed mode volume.

An image mode disk can also be migrated to another image mode disk. The image mode disk becomes managed while the migration is ongoing, but returns to image mode when the migration is complete.

You can complete the following types of migrations:

- Migrate extents
- Migrate a volume
- Migrate to image mode

Note: Migration commands fail for the following reasons:

- The target or source volume is offline.
- A quorum disk is not defined.
- Defined quorum disks are unavailable.

Correct the offline or quorum disk condition and reissue the command.

Complete the following steps to migrate volumes:

Procedure

1. Dedicate one storage pool to image mode volumes. Ensure that the pool that you select is not a child pool.
2. Dedicate one storage pool to managed mode volumes.
3. Use the migrate volume function to move the volumes.

Migrating volumes between pools using the CLI

You can migrate volumes between pools using the command-line interface (CLI).

About this task

You can determine the usage of particular MDisks by gathering input/output (I/O) statistics about nodes, MDisks, and volumes. After you collect this data, you can analyze it to determine which volumes or MDisks are hot. You can then migrate volumes from one storage pool to another.

Complete the following step to gather statistics about MDisks and volumes:

1. Use secure copy (**scp** command) to retrieve the dump files for analyzing. For example, issue the following command:

   ```
   scp clusterip:/dumps/iostats/v_*
   ```

   This command copies all the volume statistics files to the AIX host in the current directory.
2. Analyze the memory dumps to determine which volumes are hot. It might be helpful to also determine which MDisks are being used heavily as you can spread the data that they contain more evenly across all the MDisks in the storage pool by migrating the extents.

After you analyze the I/O statistics data, you can determine which volumes are hot. You also need to determine the storage pool that you want to move this volume to. Either create a new storage pool or determine an existing group that is not yet overly used. Check the I/O statistics files that you generated and then ensure that the MDisks or volumes in the target storage pool are used less than the MDisks or volumes in the source storage pool.

You can use data migration or volume mirroring to migrate data between storage pools. Data migration uses the command **migratevdisk**. Volume mirroring uses the commands **addvdiskcopy** and **rmvdiskcopy**.

Migrating data using migratevdisk

You can use the **migratevdisk** command to migrate data between two storage pools. When you issue the **migratevdisk** command, a check is made to ensure that the destination of the migration has enough free extents to satisfy the command. If it does, the command proceeds. The command takes some time to complete.

Notes:

- You cannot use the data migration function to move a volume between storage pools that have different extent sizes.
- Migration commands fail if the target or source volume is offline, there is no quorum disk defined, or the defined quorum disks are unavailable. Correct the offline or quorum disk condition and reissue the command.
- The system supports migrating volumes between child pools within the same parent pool or migrating a volume in a child pool to its parent pool. Migration of volumes fails if source and target child pools have different parent pools. However, you can use **addvdiskcopy** and **rmvdiskcopy** commands to migrate volumes between child pools in different parent pools.

When you use data migration, it is possible for the free destination extents to be consumed by another process; for example, if a new volume is created in the destination parent pool or if more migration commands are started. In this scenario, after all the destination extents are allocated, the migration commands suspend and an error is logged (error ID 020005). To recover from this situation, use either of the following methods:

- Add more MDisks to the target parent pool, which provides more extents in the group and allows the migrations to be restarted. You must mark the error as fixed before you reattempt the migration.
- Migrate one or more volumes that are already created from the parent pool to another group. This action frees up extents in the group and allows the original migrations to be restarted.

Complete the following steps to use the **migratevdisk** command to migrate volumes between storage pools:

1. After you determine the volume that you want to migrate and the new storage pool that you want to migrate it to, issue the following CLI command:

   ```
   migratevdisk -vdisk vdisk_name
   ```

   ```
    -mdiskgrp
    mdisk_group_name -threads 4
   ```

2. You can check the progress of the migration by issuing the following CLI command:

```
lsmigrate
```

Migrating data using volume mirroring

When you use data migration, the volume goes offline if either pool fails. Volume mirroring can be used to minimize the impact to the volume because the volume goes offline only if the source pool fails. You can migrate volumes between child pools or from a child pool to a parent pool using the **addvdiskcopy** and **rmvdiskcopy** commands instead of using the **migratevdisk** command. Complete the following steps to use volume mirroring to migrate volumes between pools:

1. After you determine the volume that you want to migrate and the new pool that you want to migrate it to, enter the following command:

```
addvdiskcopy -mdiskgrp mdisk_group_name -autodelete vdisk_name
```

   where *mdisk_group_name* is the name of the new storage pool and *vdisk_name* is the name of the volume that is being copied. Specify **-autodelete** to automatically delete the original copy of the volume after the copies are synchronized.

2. The copy ID of the new copy is returned. The copies now synchronize such that the data is stored in both storage pools. You can check the progress of the synchronization by issuing the following command:

```
lsvdisksyncprogress
```

Chapter 13. Stretched systems

A stretched system is an extended high availability (HA) method that is supported by SAN Volume Controller systems to enable I/O operations to continue after the loss of half of the system.

A stretched system is also sometimes referred to as a *split system* because one half of the system and I/O group is usually in a geographically distant location from the other, often 10 kilometers or more. A third site is required to host a storage system that provides a quorum disk.

Enhanced stretched systems

The system supports an enhanced stretched system configuration. This configuration can be used regardless of whether the stretched system is configured with or without ISLs between nodes.

Enhanced stretched systems provide the following primary benefits:

- In addition to the automatic failover that occurs when a site fails in a standard stretched system configuration, an enhanced stretched system provides a manual override that can be used to choose which of two sites continues operation.
- Enhanced stretched systems intelligently route I/O traffic between nodes and controllers to reduce the amount of I/O traffic between sites, and to minimize the impact to host application I/O latency.
- Enhanced stretched systems include an implementation of extra policing rules to ensure the correct configuration of a standard stretched system.

Note: In an enhanced stretched system, set the site attribute in the node and controller object.

Stretched system configuration details

You can create an enhanced stretched system configuration where each node on the system is physically on a different site.

When used with mirroring technologies, such as volume mirroring or Copy Services, these configurations can be used to maintain access to data on the system in the event of power failures or site-wide outages.

The enhanced stretched system configuration, with the topology attribute of the system set to stretched, is detailed here. It is possible to non-disruptively move to the current enhanced stretched system configuration by following the final configuration steps that are presented here so that you get better availability and disaster recovery. Contact IBM Remote Technical support center for guidance on changing the topology of an existing system.

In a stretched system configuration, each site is defined as an independent failure domain. If one site experiences a failure, the other site can continue to operate without disruption. You must also configure a third site to host a quorum device that provides an automatic tie-break in the event of a potential link failure between the two main sites. The main site can be in the same room or across rooms in the data center, buildings on the same campus, or buildings in different cities. Different kinds of sites protect against different types of failures.

Sites are within a single location
If each site is a different power phase within a single location or data center, the system can survive the failure of any single power domain. For example, one node can be placed in one rack installation and the other node can be in another rack. Each rack is considered a separate site with its own power phase. In this case, if power was lost to one of the racks, the partner node in the other rack could be configured to process requests and effectively provide availability to data even when the other node is offline due to a power disruption.

Each site is at separate locations
If each site is a different physical location, the system can survive the failure of any single location. These sites can span shorter distances, for example two sites in the same city, or they can be spread

© Copyright IBM Corp. 2024

farther geographically, such as two sites in separate cities. If one site experiences a site-wide disaster, the remaining site can remain available to process requests.

If configured properly, the system continues to operate after the loss of one site. The key prerequisite is that each site contains only one node from each pair of nodes. Simply placing a pair of nodes from the same system in different sites for a stretched system configuration does not provide high availability. You must also configure the appropriate mirroring technology and ensure that all configuration requirements for those technologies are properly configured.

The system supports both Fibre-Channel and Ethernet connections between nodes with a stretched system. Ethernet connections are only supported through RDMA-capable adapters that must be installed and configured on each node within the stretched system. The system supports Fibre Channel and Ethernet (iSCSI) connections to hosts in stretched system environments. However, NVMe-based connections to hosts are not supported in stretched system configurations.

Stretched system

A stretched system is designed to continue operation after the loss of one failure domain.

The stretched system cannot guarantee continuous operation after the failure of two failure domains. If the enhanced stretched system function is configured, you can enable a manual override for this situation.

The two partner systems can not be in the same production site. However, they can be collocated with the storage system that provides the active quorum disk for the stretched system.

Configuration steps

These additional configuration steps can be done by using the command-line interface (CLI) or the management GUI.

- Each node in the system must be assigned to a site. Use the **chnode** CLI command. If additional nodes are cabled to the system, you can specify these nodes as hot-spare nodes. Hot-spare nodes can nondisruptively take over host I/O operations if any node on the site becomes unavailable. For more information, see the topic about adding hot-spare nodes.
- Each back-end storage system must be assigned to a site. Use the **chcontroller** CLI command.
- Each host must be assigned to a site. Use the **chhost** CLI command
- After all nodes, hosts, and storage systems are assigned to a site, the enhanced mode must be enabled by changing the system topology to `stretched`.

Planning for a stretched topology system

You can plan to meet the requirements of a stretched topology system.

Make sure that:

- Each node of an I/O group must be at a different site (location or power phase).
- There must be managed disks at each site.
- There must be managed disks at a third site for quorum disks.
- There must be enough fibre optic connections between the nodes on each site and the managed disks on the third site.
- The site attribute must be set to 1, 2 or 3 for each controller before the system topology can be set to stretched.

Note: Expansion enclosures are not supported with stretched system topology.

Connections:

- Nodes that are in the same I/O group and separated by more than 100 meters (109 yards) must use longwave Fibre Channel connections. A longwave small form-factor pluggable (SFP) transceiver can be

purchased as an optional component, and must be one of the longwave SFP transceivers that are listed at the following website: www.ibm.com/support.

- Avoid using inter-switch links (ISLs) in paths between nodes and external storage systems.

Requirements for stretched systems with Fibre Channel connections

If you are configuring a stretched system that uses Fibre Channel connections, ensure that all SAN and Fibre Channel-specific requirements are met.

Notes:

- Stretched systems are not recommended for models that support nodes with internal flash drives.
- Stretched system Fibre Channel configurations with active/passive controllers such as IBM DS5000, IBM DS4000, and IBM DS3000 systems must be configured with sufficient connections such that all sites have direct access to both external storage systems. For iSCSI configurations with two or more active/passive controllers such as Storwize family systems, the systems must be configured with sufficient connections such that all sites have direct access to both external storage systems. Quorum access for stretched system is possible only through the current owner of the MDisk that is being used as the active quorum disk.

Use the following requirements to configure a stretched system with Fibre Channel connections:

- Directly connect each node to two or more SAN fabrics at the primary and secondary sites (2 - 8 fabrics are supported). In iSCSI connections, connect each node to two or more Ethernet fabrics at the primary and secondary sites. Sites are defined as independent failure domains. A failure domain is a part of the system within a boundary such that any failure (such as a power failure, fire, or flood) within that boundary is contained within the boundary and the failure does not propagate or affect parts outside of that boundary. Failure domains can be in the same room or across rooms in the data center, buildings on the same campus, or buildings in different towns. Different kinds of failure domains protect against different types of faults.
- If a storage system is used at the third site, it must support extended quorum disks. More information is available in the interoperability matrixes that are available at the following website: http://www.ibm.com/support.
- Place independent storage systems at the primary and secondary sites, and use volume mirroring to mirror the host data between storage systems at the two sites. Where possible, set the preferred node of each volume to the node in the same site as the host that the volume is mapped to.
- Connections can vary based on fibre type and small form-factor pluggable (SFP) transceiver (longwave and shortwave).
- Nodes that are in the same I/O group and separated by more than 100 meters (109 yards) must use longwave Fibre Channel or Ethernet connections. A longwave small form-factor pluggable (SFP) transceiver can be purchased as an optional component, and must be one of the longwave SFP transceivers listed at the following website: http://www.ibm.com/support.
- Avoid using inter-switch links (ISLs) in paths between nodes and external storage systems. If this is unavoidable, do not oversubscribe the ISLs because of substantial Fibre Channel traffic across the ISLs. For most configurations, trunking is required. Because ISL problems are difficult to diagnose, switch-port error statistics must be collected and regularly monitored to detect failures.
- Using a single switch at the third site can lead to the creation of a single fabric rather than two independent and redundant fabrics. A single fabric is an unsupported configuration.
- Ethernet port 1 on every node must be connected to the same subnet or subnets. Ethernet port 2 (if used) of every node must be connected to the same subnet (this might be a different subnet from port 1). The same principle applies to other Ethernet ports.
- Some service actions require physical access to all nodes in a system. If nodes in a stretched system are separated by more than 100 meters, service actions might require multiple service personnel. Contact your service representative to inquire about multiple site support.

- Use a third, dedicated site to house a quorum disk or an IP quorum application. For Fibre Channel-connected stretched system, quorum disks or IP quorum applications provide redundancy if communication is lost between the primary and secondary site. In addition both contain configuration metadata that is used to recover the system, if necessary. IP quorum applications are used for when the stretched system connects to iSCSI-attached storage systems. iSCSI storage systems cannot be configured on a third site.

A stretched system locates the active quorum disk or an IP quorum application at a third site. If communication is lost between the primary and secondary sites, the site with access to the active quorum disk continues to process transactions. If communication is lost to the active quorum disk, an alternative quorum disk at another site can become the active quorum disk.

Although a system of nodes can be configured to use up to three quorum disks, only one quorum disk can be elected to resolve a situation where the system is partitioned into two sets of nodes of equal size. The purpose of the other quorum disks is to provide redundancy if a quorum disk fails before the system is partitioned.

This figure illustrates an example stretched system configuration. When used with volume mirroring, this configuration provides a high availability solution that is tolerant of a failure at a single site. If either the primary or secondary site fails, the remaining sites can continue doing I/O operations. In this configuration, the connections between the nodes in the system are greater than 100 meters apart, and therefore must be longwave Fibre Channel connections.

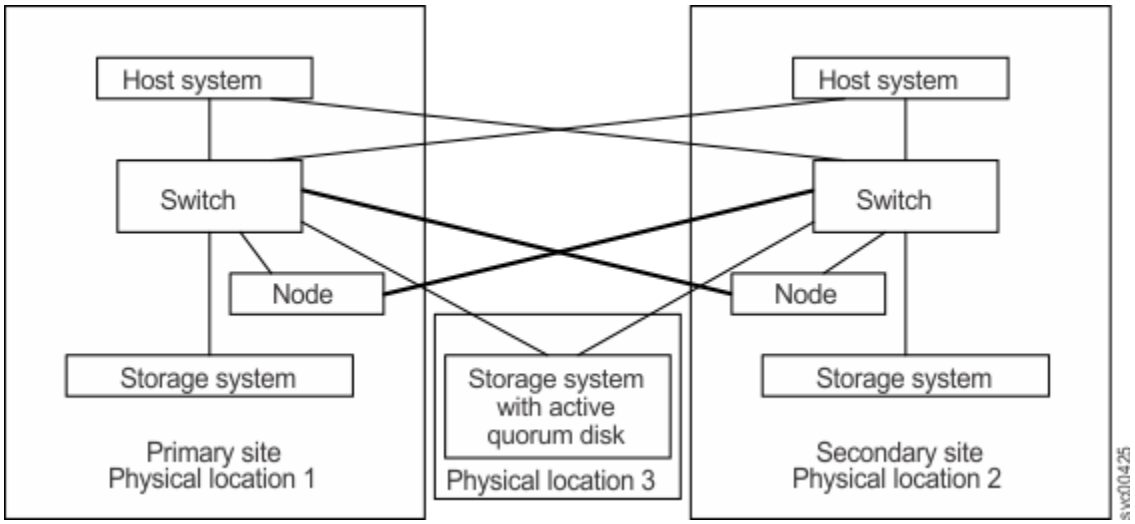


Figure 23. A stretched system with a quorum disk at a third site

In Figure 23 on page 258, the storage system that hosts the third-site quorum disk is attached directly to a switch at both the primary and secondary sites by using longwave Fibre Channel connections. If either the primary site or the secondary site fails, you must ensure that the remaining site retains direct access to the storage system that hosts the quorum disks.

Restriction: Do not connect a storage system in one site directly to a switch fabric in the other site.

An alternative configuration can use an additional Fibre Channel switch at the third site with connections from that switch to the primary site and to the secondary site.

A stretched system configuration is supported only when the storage system that hosts the quorum disks supports extended quorum. Although other types of storage systems can be used to provide quorum disks, access to these quorum disks is always through a single path.

For quorum disk configuration requirements, see the technote *Guidance for Identifying and Changing Managed Disks Assigned as Quorum Disk Candidates*.

Stretched system configuration by using interswitch links

You can use interswitch links (ISLs) in paths between nodes of the same I/O group. If the cable distance between the two production sites exceeds 100 km, potential performance impacts can result.

Using ISLs for node-to-node communication requires configuring two separate SANs, each of them with two separate redundant fabrics:

1. Configure one SAN with two separate fabrics so that it is dedicated for node-to-node communication. This SAN is referred to as a *private* SAN. This private SAN can be used by more than one stretched system if it has enough bandwidth for all these systems. For more information, see "Additional bandwidth requirements" on page 261.
2. Configure one SAN with two separate fabrics so that it is dedicated for host attachment and storage system attachment.

The network hardware that carries the ISLs must maintain the physical separation and independence of the two redundant fabrics. For example, do not connect the two fabrics into a single dark fibre link. If there are two dark fibre links, dedicate one link for each fabric. Do not cross-connect the two fabrics onto each of the two links.

Stretched system configuration rules for stretched system configurations that use ISLs

In a stretched system configuration, a site is defined as an independent failure domain. Different types of sites protect against different types of fault. For example, if configured properly, the system continues to operate after the loss of one failure domain.

However, the system does not guarantee that it can survive the failure of two sites.

- For every storage system, create one zone that contains ports from every node and all storage system ports, unless otherwise stated by the zoning guidelines for that storage system. However, do not connect a storage system in one site directly to a switch fabric in the other site. Instead, connect each storage system only to switched fabrics in the local site. (In stretched system configurations with ISLs in node-to-node paths, these fabrics belong to the public SAN).

  For stretched systems that use the enhanced configuration functions, storage systems that are configured to one of the main sites (1 or 2) need be zoned only to be visible by the nodes in that site. Storage systems in site 3 or storage systems that have no site that is defined must be zoned to all nodes.
- Each node must have direct Fibre Channel connections to at least two SAN fabrics. One is a public fabric and the other is private, in its location. For an example configuration, see Figure 24 on page 260.
- Some service actions require the ability to do actions on the front panel or through the technician port of all nodes in a system within a short-time window. If you use stretched systems, you are required to assist the support engineer and provide communication technology to coordinate these actions between the sites.
- The storage system at the third site must support extended quorum disks. This information is available in the interoperability matrices that are available at the following support website: http://www.ibm.com/support.

Each SAN consists of at least one fabric that spans both production sites. At least one fabric of the public SAN includes also the quorum site. You can use different approaches to configure private and public SANs.

- Use dedicated switches for each SAN.
- Use separate virtual fabrics or virtual SANs for each SAN.

  Note: ISLs must not be shared between private and public virtual fabrics.

To implement private and public SANs with dedicated switches, any combination of supported switches can be used. For the list of supported switches and for supported switch partitioning and virtual fabric options, see the interoperability website: www.ibm.com/support.

Like for every managed disk, all nodes need access to the quorum disk by using the same storage system ports. If a storage system with active/passive controllers (such as IBM DS3000, IBM DS4000, IBM DS5000, or IBM FAStT) is attached to a fabric, the storage system must be connected with both internal controllers to this fabric. This connection is illustrated in Figure 24 on page 260.

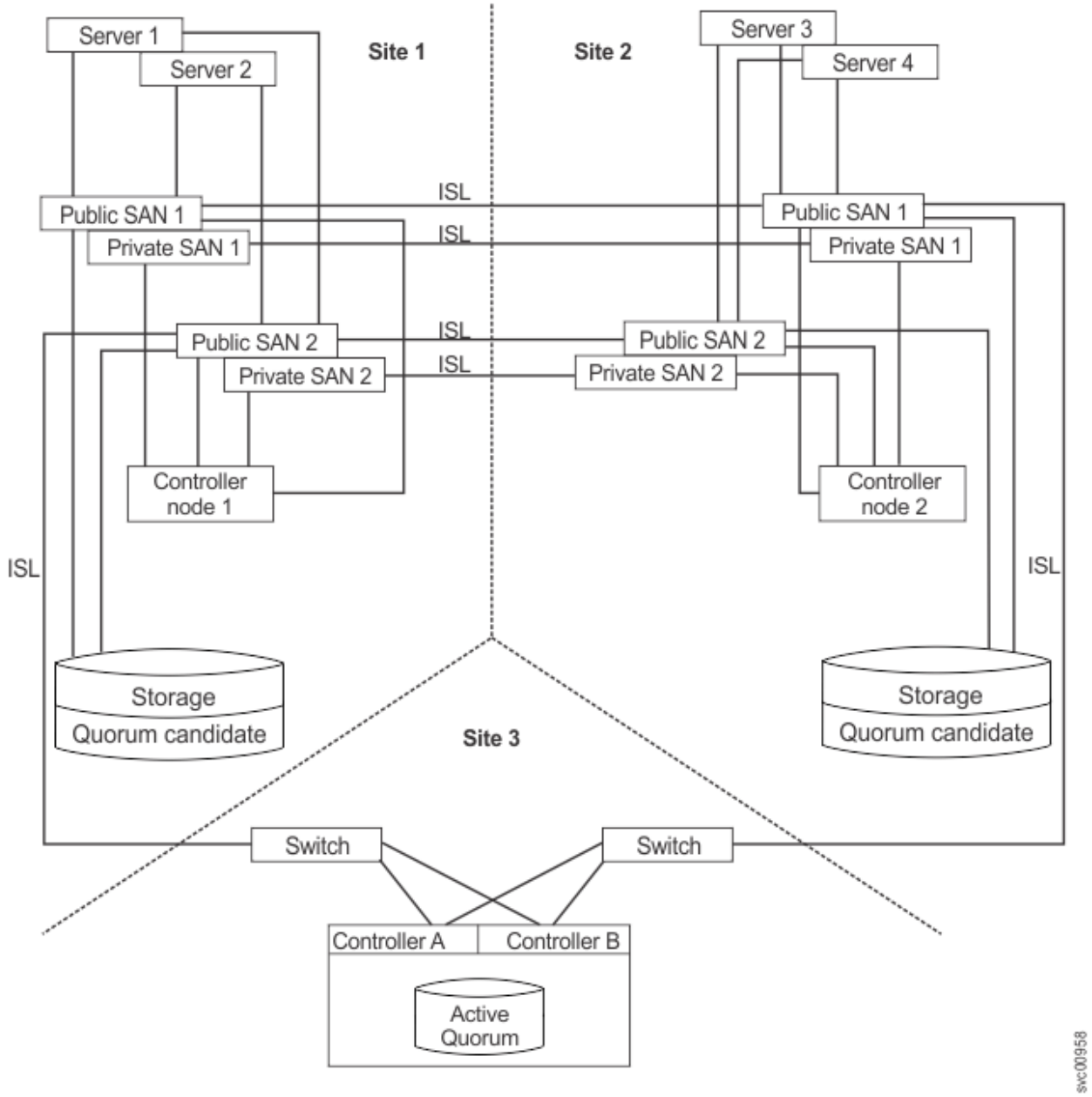


Figure 24. Stretched system nodes with ISLs, DS3000, DS4000, and DS5000 connected to both fabrics

By using FCIP, passive WDM, or active WDM for quorum site connectivity, you can add to the extension. The connections must be reliable. It is strictly required that the links from both production sites to the quorum site are independent and do not share any long-distance equipment. FCIP links are supported also for ISLs between the two production sites in public and private SANs. A private SAN and a public SAN can be routed across the same FCIP link. However, to ensure bandwidth to the private SAN (see also “Additional bandwidth requirements” on page 261), it is typically necessary to configure FCIP tunnels. Similarly, it is permissible to multiplex multiple ISL links across a DWDM link.

Note: It is not required to UPS-protect FCIP routers or active WDM devices that are used only for the node-to-quorum communication.

A stretched system configuration is supported only when the storage system that hosts the quorum disks supports extended quorum. Although the system can use other types of storage systems for providing quorum disks, access to these quorum disks is always through a single path.

For quorum disk configuration requirements, see the *Guidance for Identifying and Changing Managed Disks Assigned as Quorum Disk Candidates* technote at the following website: http://www.ibm.com/support/docview.wss?rs=591&uid=ssg1S1003311.

Additional bandwidth requirements

The connection between each site in the system that is associated with the nodes must ensure a minimum bandwidth of 4 gigabits or two times the peek host-write I/O workload, whichever is higher. The exact usage of this bandwidth depends on whether the site for the system is configured for nodes and storage systems. If the system is not configured as site-aware, half of this bandwidth is required in the public SAN and the other half is required in the private SAN. If the system is configured as site-aware, all of this bandwidth is required in the private SAN. In any case, this bandwidth must be ensured for the traffic between nodes and also between nodes and external storage systems. This bandwidth that is dedicated for system traffic does not include the traffic between nodes and hosts. Hosts that access nodes in the other site require more intersite bandwidth.

Note: To avoid performance bottlenecks, each system node requires two ports that operate at full speed (for example, 2 x 8 gigabits) worth of bandwidth to other nodes within the same system.

For example, a stretched system (no matter how many nodes) with a total of 4 gigabits of stretched system bandwidth on the private SANs is a valid configuration when the peek host-write I/O workload does not exceed 200 megabytes per second.

Operating an eight-node system at its maximum performance requires 64 gigabits of stretched system bandwidth on the private SANs.

Stretched system configuration without using interswitch links

The most simple stretched system configuration is built by attaching each system node directly to the switches in the local and the remote production site. This configuration is done without using interswitch links (ISLs).

Stretched system configuration rules for stretched system configuration without using ISLs

The following rules must be observed when you configure a stretched system without using ISLs:

- For every storage system, create one zone that contains system ports from every node and all storage system ports, unless the zoning guidelines for that storage system state differently. However, do not connect a storage system in one site directly to a switch fabric in the other site. Instead, connect each storage system only to switched fabrics in the local site. .
- Each node must have two direct connections to one or more fabrics at both locations that contain nodes.
- You can include powered components between the system and the switches in a stretched system configuration. For example, if you use Fibre Channel connections, you can use powered dense wavelength division multiplexing (DWDM) Fibre Channel extenders.
- Be prepared to provide and replace longwave SFP transceivers in case they fail.
- Some service actions require the ability to do actions on the front panel or through the technician port of all nodes in a system within a short-time window. If you use stretched systems, you are required to assist the support engineer and provide communication technology to coordinate these actions between the sites.

Configure the stretched system that does not include interswitch links (ISLs) according to the following rules:

- The minimal SAN configuration consists of one switch per production site as two separate fabrics. For highest reliability, use two switches per production site. Single fabric configurations are not supported for stretched I/O group systems.

  Note: Fibre Channel over IP (FCIP) connectivity is not supported between nodes when a stretched system is configured *without* using ISLs. FCIP (or any Fibre Channel extender or SAN router that is supported for intersystem communications) might be used in the connection to the quorum disk at the third site.

  Fibre Channel over IP connectivity is supported (or any Fibre Channel extender or SAN router that is supported for intersystem communications) both between nodes at two different sites and between nodes and the quorum disk at the third site when a stretched system is configured *with* ISLs.
- As with every system, you can use ISLs for host-to-node (with up to three hops) or for node-to-storage (at most one hop) access. If you use Fibre Channel-based connections, configure the SAN zones so that ISLs are not used in paths between nodes. For RDMA-based connections, configure separate physical Ethernet fabrics or use virtual LAN configuration so that ISLs are not used in the paths between nodes.
- In a system with four ports per node, attach half of the ports of each node to the switches in the production site where the node is located.
- In the same system with four ports per node, attach the remaining half of the ports of each node to the switches in the other (remote) production site.
- Connect each storage system at the production sites to switches in the site where the storage system is located.
- Connect the storage system with the active quorum disks to switches in both production sites.
- To avoid fabric topology changes in case of IP errors, it is a good practice to configure Fibre Channel over IP links so that they do not carry ISLs.

  Note: This guideline applies only if Fibre Channel over IP is used to connect to a switch at the third site where the quorum disk is located.

Restriction: Do not connect a storage system in one site directly to a switch fabric in the other site. Connectivity between system ports at one site to storage at the other site that uses ISLs to travel between the locations are not affected by this restriction.

It is strictly required that the links from both production sites to the quorum site are independent and do not share any long-distance equipment.

Configuring an enhanced stretched system

An enhanced stretched system assists with site disaster recovery.

Before you begin

If you are using RDMA-capable Ethernet ports to connect nodes in a stretched system, you must ensure all prerequisites are complete and define all RDMA-capable Ethernet ports before configuring the stretched system. For information, see the topics on Configuration details for using RDMA-capable Ethernet ports to connect nodes and Configuring a system with RDMA-capable Ethernet ports.

About this task

An enhanced stretched topology system can be configured through site allocation and topology change command-line (CLI) commands. The management GUI can also be used to configure a stretched topology with the Modify System Topology wizard. If you are using iSCSI-attached external storage in the stretched system configuration, you also need to add the storage to the system and specify the site for the iSCSI-attached storage system.

You can use either the management GUI or the command-line interface to configure an enhanced stretched system. The management GUI supports a wizard that completes all the steps to create an enhanced stretched system, including site definitions for nodes, hosts, and external storage systems. The CLI process requires that each of these steps are completed separately.

Using the management GUI

1. Select **Monitoring** > **System**.
2. In the **System Actions** menu, select **Modify System Topology**. Use the wizard to complete the configuration.

Using the command line interface

To configure your clustered system site information in the command-line interface, complete these steps.

Procedure

1. Use the following commands **addnode** or **chnode** and **chcontroller** and **chhost** to assign site information to the nodes, hosts and external storage systems in site 1, 2, and 3.

   Note: Site 3 cannot be assigned to any node and contains only external storage systems that are used for quorum.

   For example, enter the following command to change site definitions for a node:

   ```
   chnode -site site_id node node_id
   ```

   where *site_id* is the identifier for the site and *node_id* is the identifier of the node.

   Enter the following command to see the node configuration details for site 1 on node 1:

   ```
   lsnode 1
   ```

   Enter the following command to assign site information to all the external storage systems by using the following command:

   ```
   chcontroller –site site_id controller_id
   ```

   where

   To view the site for the external storage system for and MDisk information, enter the following commands:

   ```
   lscontroller controller_id
   lsmdisk object_id
   ```

   Issue this command to change to site 1 on host 1:

   ```
   chost -site 1 host_name
   ```

   Issue this command to see the node configuration details for site 1 on node 1:

   ```
   lshost host_name
   ```

2. When site awareness is provided to the entities that are wanted in all the three sites, change the system topology to `stretched`.

   In the CLI, enter the following command:

   ```
   chsystem -topology stretched
   ```

3. You can use the **lssystem** command to verify that the topology is `stretched`.

   ```
    lssystem
   ```

   Note: The quorum disk assignment changes automatically after the system topology is `stretched`. The active quorum disk is at site 3. Sites 1 and 2 each have one inactive quorum disk.

Configuring an external iSCSI storage system

About this task

After you configure the stretched system and all nodes and hosts are assigned to sites, you can add any iSCSI-attached storage and assign it a specific site in the stretched system. The **Modify System Topology** wizard automatically detects any Fibre Channel-attached storage but iSCSI storage must be added manually before the site assignments can be configured. You can use either the management GUI or the command-line interface to add supported iSCSI attached storage to the stretched system.

Using the Management GUI

After you configure the stretched topology, add the iSCSI attached storage by completing the following steps:

1. If you are using iSCSI-attached external storage, you must add the storage to the system and specify the site by using the **Add iSCSI External Storage** wizard. To access the wizard, select **Pools** > **External Storage** > **Add iSCSI External Storage**.
2. Select the type of storage system that you are adding. The system supports iSCSI connections to Storwize family products and to Dell Equallogic that are used as external storage.
3. Enter the information for the source and target ports for the iSCSI connection and specify the site for the iSCSI-attached storage.

Using the command-line interface

After you change the system topology to `stretched` and nodes are assigned sites, complete these steps to add iSCSI attached storage systems to the configuration. If you are using the command-line interface to add iSCSI storage, complete these steps:

1. From the nodes of *site 1*, discover the iSCSI target storage system:

```
detectiscsistorageportcandidate –srcportid 1
 -targetip IP1 -site 1
```

2. Check the status of the discovered targets:

```
lsiscsistorageportcandidate id src_port_id target_ipv4 | target_ipv6
 target_iscsiname iogroup_list configured status
 site_id site_name 0 1 IP1 IQN1 1:1:1:1 no Full 1 site1
```

3. From the nodes of *site 1*, add a session to the discovered targets:

```
addiscsitorageport -site 1 0
```

4. From the nodes of *site 1*, view the established sessions with the target:

```
id src_port_id target_ipv4 | target_ipv6 target_iscsiname
 controller_id iogroup_list status site_id site_name
 0 3 IP1 IQN1 5 1:1:1:1 full 1 site1
```

5. The new storage system can be shown as available by entering the **`lscontroller`** command. Assign a site:

```
chcontroller -site 1 controllerID
```

6. From the system nodes of site 2, repeat steps “1” on page 264 through “5” on page 264 for discovering and establishing sessions to the same iSCSI controller or a different iSCSI controller.

Manual quorum override in a stretched system

A manual quorum override is required when you experience a rolling disaster.

In rare situations, the system is subject to what is referred to as a *rolling disaster*. A rolling disaster occurs when an incident has wide scope, and its effects are felt in multiple steps over an extended time period.

The following example scenario describes a rolling disaster and shows how to recover from that rolling disaster.

An example of a rolling disaster occurs when the following situation is true:

1. The link between the two sites fails, at which point one site uses the automatic quorum feature to continue operation.
2. The system site that has control of the quorum device fails (due to a power outage, for example).

This example leaves the second site as the only site that is potentially capable of continuing data I/O. However, it is unable to do so until it gains control of the quorum device. The MDisks in the second site stop. Nodes at the site display the node error 551, indicating that an insufficient number of nodes are available to form a quorum in a stretched system configuration.

In this scenario, you can run the quorum override command to override the automatic quorum device selection and create a new system that contains the nodes in the second site.

Note:

- To ensure that the system is in the correct state before it is used, the quorum override command can be run only with assistance from support.
- If a fabric disruption occurs while the quorum override command is running, it is possible that a subset of the nodes will update their system ID. The updated nodes display the node error 550. The nodes that were not updated display 551 and the nodes are assigned to two different systems. In this situation, you can run the quorum override command again on one of the nodes that reported the error 551. This command updates all the nodes in the two systems with a new cluster (system) ID. You can then recover data.

Enforcing conditions for a quorum

You must run the **chsystem -topology stretched** command as part of the installation process for the system to make the quorum override command available if a rolling disaster occurs. The quorum override command is not available in systems that do not have the topology set to `stretched`. Before you can use the command, the following prerequisites must be met:

- All I/O groups with two nodes are assigned with one node in site 1 and the other in site 2.
- All storage systems with MDisks must have their site that is defined.

When these prerequisites are met and automatic quorum selection is enabled, the system attempts to assign one quorum device within all three sites. If a site does not have an MDisk suitable to be a quorum device, a quorum device is not assigned to it.

Note: After the **chsystem -topology stretched** command is run, you cannot alter the site assignment of any controller except where that controller is a new controller that has only unmanaged MDisks.

It also does not allow site settings for nodes. This enforcement is required to ensure that the system operates correctly to allow the quorum override command to operate correctly.

When you run the **chsystem -topology standard** command, it again is possible to alter the site settings for nodes and controllers. However, this command disables the override quorum feature. Therefore, aim to run **chsystem -topology stretched** when you complete your changes to re-enable this support.

Chapter 14. Legacy copy services

Starting from 8.7.1, policy-based replication and policy-based high availability replace the remote-copy function. Additionally, volume group snapshot supersedes the FlashCopy function.

See the following pages for content on the legacy copy services:

Related concepts

" Replication" on page 201
Policy-based replication can be used to implement high availability and disaster recovery solutions between systems in different locations.

Related information

"Snapshots" on page 102
A volume group snapshot is a consistent point-in-time copy of a volume group or a set of volumes.

FlashCopy function

The FlashCopy function creates a point-in-time copy of data that is stored on a source volume to a target volume. In its basic mode, the FlashCopy function creates copies of content on a source volume to a target volume in a mapping. The function associates a source volume and a target volume in a mapping. If data exists on the target volume, that data is replaced by the copied data.

A FlashCopy mapping defines the relationship that copies data between a source volume and a target volume. A mapping can be created between any two volumes in a system. The volumes do not have to be in the same I/O group or pool. When a FlashCopy operation starts, a checkpoint is made of the source volume. No data is copied at the time a start operation occurs. Instead, the checkpoint creates a bitmap that indicates that no part of the source volume has been copied. Each bit in the bitmap represents one region of the source volume. Each region is called a *grain*. During a read operation to the target volume, the bitmap is used to determine whether the grain was copied. If the grain was copied, the data is read from the target volume. If the grain was not copied, the data is read from the source volume.

After a FlashCopy operation starts, read and write operations to the source volume continue to occur. If new data is written to the source or target volume, the existing data on the source is copied to the target volume before the new data is written to the source or target volume. The bitmap is updated to mark that the grain of the source volume has been copied so that later write operations to the same grain do not recopy the data.

After the copy operation completes, the target volumes contain the contents of the source volumes as they existed at a single point in time, unless target writes are processed. FlashCopy is sometimes described as an instance of a time-zero copy (T 0) or point-in-time copy technology. Although the copy takes some time to complete, the data on the target volume is available immediately. (If needed, data that is still being copied can be accessed from the source.)

Although it is difficult to make a consistent copy of a data set that is constantly updated, point-in-time copy techniques help solve this problem. If a copy of a data set is created that uses a technology that does not provide point-in-time techniques and the data set changes during the copy operation, the resulting copy might contain data that is not consistent. For example, if a reference to an object is copied earlier than the object itself and the object is moved before it is copied, the copy contains the referenced object at its new location, but the copied reference still points to the previous location. You can also assign background copy and cleaning rates to a FlashCopy mapping to control the rate at which updates are propagated to the remote system. FlashCopy mapping copy *rate* values can be from 128 KBps to 2 GBps and can be changed when the FlashCopy mapping is in any state.

The FlashCopy function uses copy-on-write (CoW) to complete write operations to the source volume. The system also supports redirect-on-write (RoW) which is an internal improvement to the existing FlashCopy function. The system automatically activates RoW when the following conditions are met:

- Both source and target volumes are in same data reduction pool with deduplication enabled.

© Copyright IBM Corp. 2024

- Both source and target volumes in the FlashCopy mapping are in the same I/O group.
- Both source and target volumes are not mirrored.

If any of these conditions are not met, the FlashCopy function uses CoW to complete write operations.

More advanced functions allow operations to occur on multiple source and target volumes. Management operations are coordinated to provide a common, single point-in-time for copying target volumes from their respective source volumes. This creates a consistent copy of data that spans multiple volumes. The function also supports multiple target volumes to be copied from each source volume. This can be used to create images from different points in time for each source volume.

FlashCopy mapping guidelines for storage systems

Consider the type of I/O and frequency of update before you create the volumes that you want to use in FlashCopy mappings.

FlashCopy operations complete in direct proportion to the performance of the source and target disks. If you have a fast source disk and slow target disk, the performance of the source disk is reduced because it must wait for the write operation to occur at the target before it can write to the source.

The FlashCopy implementation that is supported on the system copies at least 256 K every time a write is made to the source. This implementation means that any write involves at minimum a read of 256 K from the source, write of the same 256 K at the target, and a write of the original change at the target. Therefore, when an application completes small 4 K writes, this size is converted into 256 K.

Because of this implementation, consider the type of I/O that your application completes during a FlashCopy operation. Ensure that you do not overload the storage. The calculations contain a heavy weighting when the FlashCopy feature is active. The weighting depends on the type of I/O that is completed. Random writes can have a much higher impact than sequential writes. For example, the sequential write would copy the entire 256 K.

You can spread the FlashCopy source volumes and the FlashCopy target volumes between as many storage pools as possible. This configuration limits the potential bottle-necking of a single storage system (assuming that the storage pools contain MDisks from different storage systems). However, this configuration can still result in potential bottlenecks if you want to maintain all your target volumes on a single storage system. You must ensure that you add the appropriate weighting to your calculations.

FlashCopy mapping events

FlashCopy mapping events detail the events that modify the state of a FlashCopy mapping.

Table 36 on page 268 provides a description of each FlashCopy mapping event.

Table 36. FlashCopy mapping events

Event	Description
Create	A new FlashCopy mapping is created between the specified source volume and the specified target volume. The operation fails if any of the following situations is true: • The source volume is already a member of 256 FlashCopy mappings. • The node has insufficient bitmap memory. • The source and target volumes are different sizes.

Table 36. FlashCopy mapping events (continued)	
Event	**Description**
Prepare	The **prepare** command is directed to either a consistency group for FlashCopy mappings that are members of a normal consistency group or to the mapping name for FlashCopy mappings that are stand-alone mappings. The **prepare** command places the FlashCopy mapping into the preparing state.  **Attention:** The **prepare** command can corrupt any data that previously resided on the target volume because cached writes are discarded. Even if the FlashCopy mapping is never started, the data from the target might logically change during the act of preparing to start the FlashCopy mapping.
Flush done	The FlashCopy mapping automatically moves from the preparing state to the prepared state. This is done after all cached data for the source is flushed and all cached data for the target is no longer valid.
Start	When all the FlashCopy mappings in a consistency group are in the prepared state, the FlashCopy mappings can be started. To preserve the cross volume consistency group, the start of all of the FlashCopy mappings in the consistency group must be synchronized correctly regarding I/Os that are directed at the volumes. This alignment is achieved during the **start** command. The following series of events occurs during the **start** command: • New reads and writes to all source volumes in the consistency group are paused in the cache layer until all ongoing reads and writes below the cache layer are completed. • After all FlashCopy mappings in the consistency group are paused, the internal clustered system state is set to allow FlashCopy operations. • After the system state is set for all FlashCopy mappings in the consistency group, read and write operations are unpaused on the source volumes. • The target volumes are brought online. As part of the **start** command, read and write caching is enabled for both the source and target volumes.
Modify	The following FlashCopy mapping properties can be modified: • FlashCopy mapping name • Clean rate • Consistency group • Copy rate (for background copy or stopping copy priority) • Automatic deletion of the mapping when the background copy is complete
Stop	There are two separate mechanisms by which a FlashCopy mapping can be stopped: • You issue a command • An I/O error occurs
Delete	This command requests that the specified FlashCopy mapping is deleted. If the FlashCopy mapping is in the stopped state, the force flag must be used.
Flush failed	If the flush of data from the cache cannot be completed, the FlashCopy mapping enters the stopped state.

Table 36. FlashCopy mapping events (continued)	
Event	**Description**
Copy complete	After all of the source data is copied to the target and there are no dependent mappings, the state is set to copied. If the option to automatically delete the mapping after the background copy completes is specified, the FlashCopy mapping is automatically deleted. If this option is not specified, the FlashCopy mapping is not automatically deleted and can be reactivated by preparing and starting again.
Bitmap Online/ Offline	The node fails.

Grains and the FlashCopy bitmap

When data is copied between volumes, it is copied in units of address space that are known as *grains*.

The grain size can be either 64 KB or 256 KB. The FlashCopy bitmap contains 1 bit for each grain. The bit records whether the associated grain is split by copying the grain from the source to the target. After a FlashCopy mapping is created, the grain size for that FlashCopy mapping cannot be changed. When a FlashCopy mapping is created, if the grain size parameter is not specified and one of the volumes is already part of a FlashCopy mapping, the grain size of that mapping is used. If neither volume in the new mapping is already part of another FlashCopy mapping, and at least one of the volumes in the mapping is a compressed volume, the default grain size is 64 for performance considerations. But other than in this situation, the default grain size is 256.

Write to target volume

A write to the newest target volume must consider the state of the grain for its own mapping and the grain of the next oldest mapping.

- If the grain of the intermediate mapping or the next oldest mapping is not copied, it must be copied before the write is allowed to proceed. This copy is done to preserve the contents of the next oldest mapping. The data that is written to the next oldest mapping can come from a target or source.
- If the grain of the target that is being written is not copied, the grain is copied from the oldest already copied grain in the mappings that are newer than the target (or the source if no targets are already copied). After the copy is complete, the write can be applied to the target.

Read to target volume

If the grain that is being read is split, the read returns data from the target that is being read. If the read is to an uncopied grain on an intermediate target volume, each of the newer mappings is examined to determine whether the grain is split. The read is surfaced from the first split grain that is found. If the newer mappings do not have a split grain, the read is surfaced from the source volume.

Thin-provisioned FlashCopy

You can have a mix of thin-provisioned and standard-provisioned volumes in FlashCopy mappings. One common combination is a standard-provisioned source with a thin-provisioned target, which enables the target to consume a smaller amount of real storage than the source.

For best performance, the grain size of the thin-provisioned volume must match the grain size of the FlashCopy mapping. However, if the grain sizes are different, the mapping still proceeds.

When you create your FlashCopymappings, consider the following information:

- If you are using a standard-provisioned source with a thin-provisioned target, disable background copy and cleaning mode on the FlashCopy map by setting both the background copy rate and cleaning rate to zero. Otherwise, if these features are enabled, all the source is copied onto the target volume. This action causes the thin-provisioned volume to either go offline or to grow as large as the source.

- If you are using only thin-provisioned source, only the space that is used on the source volume is copied to the target volume. For example, if the source volume has a virtual size of 800 GB and a real size of 100 GB of which 50 GB is used, only the used 50 GB are copied.
- A FlashCopy bitmap contains 1 bit for every grain on a volume. For example, if you have a thin-provisioned volume with 1 TB virtual size (100 MB real capacity), you must have a FlashCopy bitmap to cover the 1 TB virtual size even though only 100 MB of real capacity is allocated.

Legacy Safeguarded Copy based on FlashCopy

The Safeguarded Copy function isolates backup copies from production data, so if a cyberattack occurs, you can quickly recover data from Safeguarded copies.

The system supports Safeguarded snapshots, which use the snapshot function to create point-in-time copies of volume groups that are immutable on the system by using an internal scheduler when you do not have external scheduling applications like IBM Copy Services Manager. For more information, see “Safeguarded snapshots” on page 104.

The Safeguarded Copy function supports the following key characteristics that create cyber-resilient copies of your important data.

Separation of duties
Provides more security capabilities to prevent nonprivileged users from compromising production data. Operations related to Safeguarded backups are restricted to only a subset of users with specific roles on the system.

Administrator
Users with the Administrator role can provision and configure Safeguarded copies and related objects, such as volume groups. However, these users cannot remove or change existing Safeguarded snapshots. For auditing, it is recommended that you create a new Administrator user to configure the Safeguarded snapshots or Safeguarded Copy function. Users with this role are limited in how they can manage and interact with Safeguarded Copy operations.

Security Administrator
Users with the Security administrator role can manage users and security on the entire system and can remove and change Safeguarded backups and Safeguarded backup locations.

Superuser
Users with superuser privileges can configure all objects and complete maintenance tasks on the system. These users can remove and change both Safeguarded backups and Safeguarded policies. For more security, this account can be disabled on the system; however, it can be reenabled for remote support assistance or maintenance tasks.

Restricted Security Administrator
Users with the security administrator role is changed to restricted security administrator when two person integrity (TPI) is enabled on the IBM Storage Virtualize. TPI requires two security administrators to work together to complete critical or risky tasks. For example, a restricted security administrator with an elevated role can remove Safeguarded snapshots.

Protected Copies
Safeguarded copies cannot be mapped directly to hosts to prevent any application from changing these copies.

Automation
The system supports Safeguarded snapshots, which use the snapshot function to create point-in-time copies of volume groups that are immutable can be scheduled with an internal scheduler. The system also supports IBM Storage Copy Data Management and IBM Copy Services Manager as external scheduling applications.

The system supports the following external scheduling applications:

IBM Copy Services Manager

The system supports IBM Copy Services Manager as an external scheduling application. IBM Copy Services Manager coordinates and automates Safeguarded Copy function across multiple systems.

IBM Copy Services Manager controls Copy Services in storage environments. Copy Services are features that are used by storage systems, such as IBM Storage Virtualize system, to configure, manage, and monitor data-copy functions. If you do not have an existing IBM Copy Services Manager license, purchase the IBM Copy Manager for IBM Storage Virtualize license, which includes IBM Copy Services Manager version 6.3.4. This license option is available through iERP/AAS, Passport Advantage®, or your IBM Sales team.If you currently have an existing license for IBM Copy Services Manager, download IBM Copy Services Manager version 6.3.4 or later at Latest Downloads for IBM Copy Services Manager.

Note: For Safeguarded snapshot with IBM Copy Services Manager, use IBM Copy Services Manager version 6.3.4 or later.

Safeguarded snapshots with IBM Copy Services Manager

Safeguarded snapshots with IBM Copy Services Manager simplifies the overall configuration and management of snapshots. IBM Copy Services Manager controls both scheduling and creating Safeguarded snapshots which eliminates configuration steps on the IBM Storage Virtualize system. If using the previous version of Safeguarded Copy function that uses a Safeguarded backup policy to manage Safeguarded backups, you must remove the Safeguarded backup policy from the system to use Safeguarded snapshots.

To configure Safeguarded snapshots with IBM Copy Services Manager, you only need to create an administrator user and assign volumes to a volume group. You can also use the create clone or thin-clone feature to create copies an existing Safeguarded snapshot to determine whether the source volumes were compromised. When a clone or thin-clone is created you can examine the data and determine if the data is usable. You can also map the cloned volume group to host applications in IBM Copy Services Manager. With the thin-clone and clone feature you can test and recover data on a volume group basis.

Safeguarded snapshot does not support automatic restore-in-place. However, you can create the clone or thin-clone of Safeguarded snapshot to recover data for analysis and testing.

Safeguarded policy with IBM Copy Services Manager

To use the Safeguarded policy with IBM Copy Services Manager to create the Safeguarded backups, refer documentation version 8.5.0.x (LTS).

IBM Storage Copy Data Management

IBM Storage Copy Data Management provides management of application-consistent Safeguarded copies that use FlashCopy. IBM Storage Copy Data Management is available to order through IBM standard ordering databases. IBM Storage Copy Data Management license is available on iERP/Advanced Administration System (AAS) as part of the IBM Storage Software Suite. The stand-alone version of IBM Storage Copy Data Management is available on IBM Passport Advantage.

You can also purchase IBM Storage Sentinel, a software bundle (PID 5900APZ) license that combines copy data function and ransomware detection software. The IBM Storage Sentinel package combines the ability to create and manage Safeguarded Copy operations with IBM Storage Copy Data Management software and the ability to automatically detect any compromised data in scheduled backups with the built-in ransomware detection. The IBM Storage Copy Data Management license is available through IBM Passport Advantage.

IBM Storage Sentinel includes the following software licenses:

IBM Storage Copy Data Management version 2.2.16 or later

IBM Storage Copy Data Management provides the following actions:

- Creates Safeguarded policies to schedule backups of IBM Storage Virtualize volumes. Safeguarded backup policies are created and managed by IBM Storage Copy Data Management.

- Creates immutable backups of IBM Storage Virtualize volumes.
- Supports easy to use restore operations so compromised data can be tested and restored back to production volumes.

For the Safeguarded Copy function, Safeguarded source volumes have the following restrictions:

- Safeguarded source volumes cannot be in a Safeguarded backup location.
- Safeguarded source volumes cannot be mirrored volumes.
- Safeguarded source volumes cannot be in an ownership group.
- Safeguarded source volumes cannot be used as cloud backups with the transparent cloud tiering function.

As part of the configuration of the Safeguarded Copy function with IBM Storage Virtualize systems, administrators need access to the IBM Storage Copy Data Management interface and must complete basic installation and configuration.

Security scan software
Supports registering dedicated security scan servers to run ransomware detection software on every Safeguarded backup that is generated as part of the IBM Storage Copy Data Management configuration of the Safeguarded Copy function. To use these features for ransomware detection requires more configuration and setup that is outside the scope of IBM Storage Virtualize documentation.

The two software work together to provide detection of ransomware on Safeguarded backups.

Safeguarded backup

A *Safeguarded backup* is a volume in a Safeguarded backup location. As part of the Safeguarded Copy function, you can add volumes to a volume group and assign a Safeguarded policy to that group. The external scheduling application that is configured with the system applies the policy to all the volumes in the group to create Safeguarded backups. Safeguarded backups are created in the same parent pool as the Safeguarded source volumes. A Safeguarded backup is the target of FlashCopy mapping with Safeguarded source volumes as a source.

The *safeguarded_expiration_time* parameter in the **lsvdisk** command stores expiration time of Safeguarded backups. The *safeguarded_expiration_time* parameter notifies Storage Virtualize software to automatically delete the expired Safeguarded backups without giving the system administrator role the capability of deleting these backups directly. The *retentiondays* parameter in the **startfcconsistgrp** command store the retention period scheduled for Safeguarded backups.

To configure a Safeguarded backup volumes, the **mkvdisk** or **mkvolume** command in the system creates a new target volume in a Safeguarded backup location. The command supports only specific parameters for a volume in the Safeguarded backup location. Ensure that the target volume in the Safeguarded backup location must be created before creating the FlashCopy mapping from the Safeguarded source.

The following actions are restricted on the Safeguarded backups:

- A security administrator role only can delete the Safeguarded backups or Storage Virtualize software can automatically delete the Safeguarded backups.
- Safeguarded backup volume can be the target of FlashCopy mapping if the source volume is a Safeguarded source volume.
- Safeguarded backup volume cannot map to a host.
- Safeguarded backup volume mirror copy cannot be created.
- Safeguarded backup volume cannot be removed or changed by using CLI commands.

Safeguarded backup volume characteristics

The following section explains the characteristics of a Safeguarded backup volume in a Safeguarded backup location.

- The Safeguarded source volumes and Safeguarded backup volumes **-iogrp**, **-preferrednode** matches.
- Safeguarded backup volumes and Safeguarded source volumes size is same.
- Safeguarded backup volumes are *auto expand*.
- The Safeguarded backup volume cache mode is **readwrite**.
- In standard pools, the Safeguarded backup volumes are *thin* and *not compressed*.
- In Data Reduction Pools, the Safeguarded backup volumes *compressed* and *deduplicated*.
- A Safeguarded backup volume is of **striped** type.
- The **easytier** attribute is enabled for Safeguarded backup volumes.
- The **name**, **buffer size**, and **grainsize** parameters are user configurable.

To preserve Safeguarded backups from deleting, the `-keeptarget` parameter in `mkfcmap` command specifies that the target volume and source volume availability must be the same. If the Safeguarded backup volume becomes unavailable, then the Safeguarded source volume is also unavailable. The `-keeptarget` parameter supports FlashCopy mappings that contain Safeguarded source volume or the Safeguarded backup volumes as a source volume.

Testing and recovering Safeguarded backups

If the source volumes are compromised during a cyberattack, then the latest Safeguarded backups may not be useful for recovery. Depending on the external scheduling application that you configure with the system, each application has built-in tools and process that can assist an administrator in testing and recovering compromised data. For example, you can configure the system to use IBM Storage Copy Data Management and as the external scheduling application. IBM Storage Copy Data Management provides management of application-consistent Safeguarded copies that use FlashCopy. IBM Storage Copy Data Management is available to order through IBM standard ordering databases. IBM Storage Copy Data Management license is available on iERP/Advanced Administration System (AAS) as part of the IBM Storage Software Suite. The stand-alone version of IBM Storage Copy Data Management is available on IBM Passport Advantage. You can also purchase IBM Storage Sentinel, a software bundle (PID 5900APZ) license that combines copy data function and ransomware detection software. The IBM Storage Sentinel package combines the ability to create and manage Safeguarded Copy operations with IBM Storage Copy Data Management software and the ability to automatically detect any compromised data in scheduled backups with the built-in ransomware detection. The IBM Storage Copy Data Management license is available through IBM Passport Advantage.

IBM Storage Sentinel includes the following software licenses:

IBM Storage Copy Data Management version 2.2.16 or later

IBM Storage Copy Data Management provides the following actions:

- Creates Safeguarded policies to schedule backups of IBM Storage Virtualize volumes. Safeguarded backup policies are created and managed by IBM Storage Copy Data Management.
- Creates immutable backups of IBM Storage Virtualize volumes.
- Supports easy to use restore operations so compromised data can be tested and restored back to production volumes.

For the Safeguarded Copy function, Safeguarded source volumes have the following restrictions:

- Safeguarded source volumes cannot be in a Safeguarded backup location.
- Safeguarded source volumes cannot be mirrored volumes.
- Safeguarded source volumes cannot be in an ownership group.
- Safeguarded source volumes cannot be used as cloud backups with the transparent cloud tiering function.

As part of the configuration of the Safeguarded Copy function with IBM Storage Virtualize systems, administrators need access to the IBM Storage Copy Data Management interface and must complete basic installation and configuration.

Security scan software

Supports registering dedicated security scan servers to run ransomware detection software on every Safeguarded backup that is generated as part of the IBM Storage Copy Data Management configuration of the Safeguarded Copy function. To use these features for ransomware detection requires more configuration and setup that is outside the scope of IBM Storage Virtualize documentation.

The two software work together to provide detection of ransomware on Safeguarded backups.

If you use IBM Copy Services Manager as an external scheduling application with the system, you can determine the uncompromised Safeguarded backups, IBM Copy Services Manager creates a new target volume in the same parent pool by creating another FlashCopy mapping between the Safeguarded backups as source volumes and the new target volume. After the FlashCopy mapping is created, the new target volume is mapped to a host server or a virtual machine. The user can now test the Safeguarded backups by using the external application that is installed on the server system. The new copy volumes that are created for testing and recovery are created according to the provisioning policy associated with the parent pool. IBM Copy Services Manager uses the pool provisioning policy to create recovery volumes. A user can set recovery copies, such as thin provisioned or deduplication by defining a provisioning policy to their parent pool. For more information, see “Provisioning policies” on page 195.

To recover and restore the data, the user would first test the Safeguarded backups. After testing and determining the uncompromised data, user can use the new FlashCopy target volumes as the source volumes. The new source volumes would be added to a volume group with associated Safeguarded backup policy and the original volume group Safeguarded policy would be removed.

The following image shows the testing and recovering a Safeguarded backups scenario:

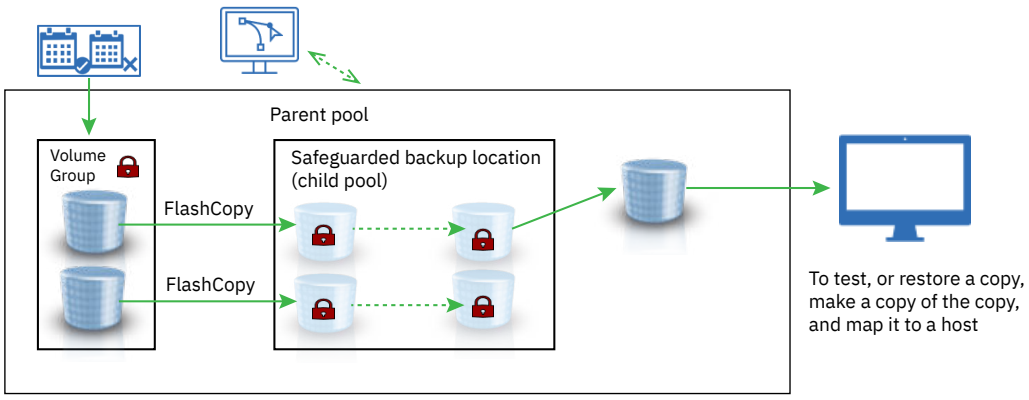


Figure 25. Recovering Safeguarded backups

Suspending Safeguarded Copy operation

You can suspend a Safeguarded copy operation to stop creating copies for a volume group .

If the source volumes are compromised during a cyberattack or during a disaster recovery scenario, the system supports suspending all Safeguarded Copy operations until the compromised data can be determined. During an attack, Safeguarded backups can contain malicious data or software. By suspending Safeguarded Copy functions, you can examine the existing uncompromised Safeguarded backups. The uncompromised Safeguarded backup can be used to restore or recover the backup of the data using your configured external scheduling application. Safeguarded Copy operation is suspended at the system level. When the Safeguarded Copy function is suspended, your configured external scheduling application does not create new backups in the Safeguarded backup location and prevents the expired backups from being removed. The **safeguarded_copy_suspended** parameter in the **lssystem** command displays the suspension status.

In addition to responding to a potential breach, suspending the Safeguarded Copy function can also be used as part of a disaster recovery response. However, suspending Safeguarded Copy function

prevents any new backups being created which results in a lack of new backups that can be used for future recovery. Before suspending Safeguarded Copy function, contact your support representative to assistance. To resume backups, the Safeguarded Copy function must be turned back on.

Configuring Safeguarded Copy function

You can configure Safeguarded Copy function with the management GUI or the command-line interface (CLI).

The system supports Safeguarded snapshots, which use the snapshot function to create point-in-time copies of volume groups that are immutable on the system by using an internal scheduler when you do not have external scheduling applications like IBM Copy Services Manager. For more information, see “Managing snapshots” on page 105 and “Snapshot policies” on page 199.

Configuring Safeguarded snapshots with IBM Copy Services Manager

You can use IBM Copy Services Manager to manage and create Safeguarded snapshots on the system.

Prerequisites

Before you can configure the Safeguarded Copy function on your system, ensure you meet the following prerequisites:

IBM Copy Services Manager Requirements

Ensure that the following requirements are met for IBM Copy Services Manager:

1. If you do not have an existing IBM Copy Services Manager license, purchase the IBM Copy Manager for IBM Storage Virtualize license, which includes IBM Copy Services Manager version 6.3.4. This license option is available through iERP/AAS, Passport Advantage, or your IBM Sales team.
2. If you currently have an existing license for IBM Copy Services Manager, download IBM Copy Services Manager version 6.3.4 or later at Latest Downloads for IBM Copy Services Manager.

   Note: If you are using an existing license, ensure that the licensed capacity is adequate for use of the Safeguarded Copy function. If you need more capacity for Safeguarded Copy function, contact you IBM sales representative to update your licensed capacity for IBM Copy Services Manager.
3. After you download IBM Copy Services Manager, complete the instructions for your installation. IBM Copy Services Manager supports several installation options on different environments. For more information, see Installing IBM Copy Services Manager.
4. During installation, license files can be imported for IBM Copy Services Manager. If the license was not imported during the installation, you need to apply the license to the installation. For more information, see Applying license files after installation.

System Requirements

All systems must be running the 8.5.2 or later release.

For existing systems, ensure that you have completed capacity planning for Safeguarded Copy function.

Creating an Administrator user for IBM Copy Services Manager

Before you can establish the system as a connection endpoint in IBM Copy Services Manager, you need to configure a user with the Administrator role on the IBM Storage Virtualize system. For auditing, it is recommended that you create a new Administrator user to configure the Safeguarded Copy function. Users with this role are limited in how they can manage and interact with Safeguarded Copy operations. The IBM Copy Services Manager uses this role to create Safeguarded snapshots.

Using the management GUI

To create new Administrator user for IBM Copy Services Manager, complete these steps:

1. In the management GUI, select **Access** > **Users by Groups** > **Create User Group**.

2. On the **Create User Group** page, enter a name of the user group, and select **Administrator** for the role.
3. Click **Create**.
4. In the list of user groups, select the user group that you created and select **Create Users**.
5. On the **Create Users** page, enter the name of the user, and select **Local**.
6. Enter and confirm a password.
7. Click **Create**.

After you create the Administrator user, create a connection to the system in IBM Copy Services Manager.

Using the command-line interface (CLI)

To create new Administrator user for IBM Copy Services Manager, complete these steps:

1. To configure a user with an Administrator role, you need to create a user group. Enter the following command.

   ```
   mkusergrp -name group_name -role Administrator
   ```

   where *group_name* specifies the unique user group name and *Administrator* specifies the role for all users that belong to this user group.
2. After creating user group, create a local user to access a system. Enter the following command.

   ```
   mkuser -name user_name -usergrp group_name | group_id -password
   ```

   where *user_name* specifies unique user name. The *group_name* | *group_id* specifies the ID or name of the user group that was created in the previous step. The parameter **-password** specifies the password to be associated with the user.

Creating volume groups and assigning source volumes

A volume group allows for many related volumes to be grouped so that they can be managed together. A volume group also ensures consistency across all volumes in the group when used with snapshot or replication functions.

A Safeguarded volume group describes a set of source volumes that can span different pools and are backed up collectively with the Safeguarded Copy function. Safeguarded snapshots are supported on the system through an internal scheduler that is defined in the snapshot policy or can be configured with an external snapshot scheduling application such as IBM Copy Services Manager.

Using the management GUI

To create a volume group, complete these steps:

1. In the management GUI, select **Volumes** > **Volumes Groups**.
2. Click **Create Volume Group**.
3. On the **Create Volume Group** page, enter the volume group name. If you do not specify a volume group name, the system automatically assigns the name.
4. To assign volumes, select the following options:

   Choose existing volumes
   Select this option if you want to add existing volumes to a volume group by completing these steps:

   a. Select **Choose existing volumes** and click **Next**.
   b. From the list, select the volumes that you want to add to the volume group.
   c. Click **Create Volume Group**.

Using the command-line interface (CLI)

To create and configure a new volume group and assign volumes to that group, complete these steps:

1. To create the new volume group, enter the following command:

   ```
   mkvolumegroup -name volumegroup_name
   ```

   where *volumegroup_name* specifies a volume group name. This command creates the volume group.
2. After creating the volume group, either create new volumes or change existing volumes and assign them to volume group that you created in Step 1.

   If you are assigning a new volume to the volume group, enter the following command:

   ```
   mkvolume -volumegroup <volumegroup_name_or_id> -size <disk_size>
   ```

   where *<volumegroup_name_or_id>* is the name or identifier of the volume group; and *<disk_size>* indicates the capacity that is provisioned for the volume from the parent pool.

   If you are assigning existing volumes to the volume group, enter the following command:

   ```
   chvdisk -volumegroup <volumegroup_name_or_id> <name_or_id>
   ```

   where *<volumegroup_name_or_id>* is the name or identifier of the volume group and *<name_or_id>* of the volume.

Configuring Safeguarded Copy function with IBM Storage Copy Data Management

You can configure Safeguarded Copy function with IBM Storage Copy Data Management.

Prerequisites

Before you can configure the Safeguarded Copy function on your system with IBM Storage Copy Data Management, ensure you meet the following prerequisites:

IBM Storage Copy Data Management Requirements

Purchase a license for IBM Storage Copy Data Management or IBM Storage Sentinel. IBM Storage Copy Data Management provides management of application-consistent Safeguarded copies that use FlashCopy. IBM Storage Copy Data Management is available to order through IBM standard ordering databases. IBM Storage Copy Data Management license is available on iERP/Advanced Administration System (AAS) as part of the IBM Storage Software Suite. The stand-alone version of IBM Storage Copy Data Management is available on IBM Passport Advantage. You can also purchase IBM Storage Sentinel, a software bundle (PID 5900APZ) license that combines copy data function and ransomware detection software. The IBM Storage Sentinel package combines the ability to create and manage Safeguarded Copy operations with IBM Storage Copy Data Management software and the ability to automatically detect any compromised data in scheduled backups with the built-in ransomware detection. The IBM Storage Copy Data Management license is available through IBM Passport Advantage.

IBM Storage Sentinel includes the following software licenses:

IBM Storage Copy Data Management version 2.2.16 or later

IBM Storage Copy Data Management provides the following actions:

- Creates Safeguarded policies to schedule backups of IBM Storage Virtualize volumes. Safeguarded backup policies are created and managed by IBM Storage Copy Data Management.
- Creates immutable backups of IBM Storage Virtualize volumes.
- Supports easy to use restore operations so compromised data can be tested and restored back to production volumes.

For the Safeguarded Copy function, Safeguarded source volumes have the following restrictions:

- Safeguarded source volumes cannot be in a Safeguarded backup location.
- Safeguarded source volumes cannot be mirrored volumes.
- Safeguarded source volumes cannot be in an ownership group.
- Safeguarded source volumes cannot be used as cloud backups with the transparent cloud tiering function.

As part of the configuration of the Safeguarded Copy function with IBM Storage Virtualize systems, administrators need access to the IBM Storage Copy Data Management interface and must complete basic installation and configuration.

Security scan software
Supports registering dedicated security scan servers to run ransomware detection software on every Safeguarded backup that is generated as part of the IBM Storage Copy Data Management configuration of the Safeguarded Copy function. To use these features for ransomware detection requires more configuration and setup that is outside the scope of IBM Storage Virtualize documentation.

The two software work together to provide detection of ransomware on Safeguarded backups.

IBM Storage Virtualize system requirements
All systems must be running the 8.5.1 or later release.

For existing systems, ensure that you complete capacity planning for Safeguarded Copy function. If the Safeguarded Copy function is configured with IBM Copy Services Manager and you want to use IBM Storage Copy Data Management, you unassign all Safeguarded policies that are managed by IBM Copy Services Manager and migrate to IBM Storage Copy Data Management as your scheduling application.

Creating an Administrator user for IBM Storage Copy Data Management

Before you can establish the system as a connection endpoint in IBM Storage Copy Data Management, you need to configure a user with the Administrator role on the IBM Storage Virtualize system. For auditing, it is recommended that you create a new Administrator user to configure the Safeguarded Copy function. Users with this role are limited in how they can manage and interact with Safeguarded Copy operations. The IBM Storage Copy Data Management uses this role to create FlashCopy mappings between the source volumes and the Safeguarded backups on the system.

To create new Administrator user for IBM Storage Copy Data Management, complete these steps:

1. In the management GUI, select **Access** > **Users by Groups** > **Create User Group**.
2. On the **Create User Group** page, enter a name of the user group, and select **Administrator** for the role.
3. Click **Create**.
4. In the list of user groups, select the user group that you created and select **Create Users**.
5. On the **Create Users** page, enter the name of the user, and select **Local**.
6. Enter and confirm a password.
7. Click **Create**.

Using the command-line interface (CLI)

To create new Administrator user for IBM Storage Copy Data Management , complete these steps:

1. To configure a user with an Administrator role, you need to create a user group. Enter the following command.

```
mkusergrp -name group_name -role Administrator
```

where *group_name* specifies the unique user group name and *Administrator* specifies the role for all users that belong to this user group.

2. After creating user group, create a local user to access a system. Enter the following command.

```
mkuser -name user_name -usergrp group_name | group_id -password
```

where *user_name* specifies unique username. The *group_name* | *group_id* specifies the ID or name of the user group that was created in the previous step. The parameter **-password** specifies the password to be associated with the user.

Registering the IBM Storage Virtualize system in IBM Storage Copy Data Management

After you create an Administrator user for IBM Storage Copy Data Management, you must configure and register the system in IBM Storage Copy Data Management interface.

To configure the system in IBM Storage Copy Data Management, complete these steps:

Note: The following instructions are based on another product's interface and can be subject to change. For the most recent version of these steps, see IBM Storage Copy Data Management documentation.

1. In a supported browser, enter the following URL:

   `https://<HOSTNAME>:8443/portal/`

   Where <HOSTNAME> is the IP address of the virtual machine where the IBM Storage Copy Data Management is configured. You now connected to IBM Storage Copy Data Management management GUI.
2. Enter the administrator username and password.
3. After you log in to IBM Storage Copy Data Management GUI, you need to do the following tasks:

   Add the IBM Storage Virtualize system as a site
   Create sites to define a grouping of providers based on their location.

   Registering an Application Server - InterSystems Database
   Establishes a secure communication channel between SCDM and application server.

   Registering an IBM Storage Virtualize provider
   Register the IBM Storage Virtualize system as a storage provider for IBM Storage Copy Data Management.

   Note: As part of registering, you need to add the administrator username that you created in

Specifying IBM Storage Copy Data Management as the external scheduling application

After you configure the system on IBM Storage Copy Data Management, specify IBM Storage Copy Data Management as the external scheduling application.

The management GUI supports two alternatives to specify the external scheduling application. You can set this value for the entire system on the **External Scheduling Applications** page or set the external scheduling application on the **Volume Groups** page.

To specify IBM Storage Copy Data Management as your external scheduling application, complete the following steps:

1. In the management GUI, select **Settings** > **GUI preferences**, and ensure **Display Safeguarded backup policy tile and External Scheduler Application Settings:** is **Visible**.
2. Select **Settings** > **System** > **External Scheduling Applications**.
3. On the **External Scheduling Applications** page, select IBM Storage Copy Data Management.
4. Click **Save**.

To specify IBM Storage Copy Data Management as the external scheduling application on the **Volume Groups** page, complete the following steps:

1. In the management GUI, select **Volumes** > **Volume Groups**.
2. On the **Volume Groups** page, select a volume group and click **Assign Safeguarded Backup Policy**.
3. On the **Assign Safeguarded policy** page, select IBM Storage Copy Data Management.
4. Click **Next** to save changes.

If you are using IBM Copy Services Manager as external scheduling application and you need to migrate to IBM Storage Copy Data Management, see .

Creating Safeguarded backup locations

You can use the GUI and command- line interface for creating Safeguarded backup location.

Using the management GUI

To create a Safeguarded backup location, complete these steps:

1. In the management GUI, select **Pools** > **Pools**.
2. Right-click a parent pool and select **Create Child Pool**.
3. On the **Create Child Pool** page, enter a name of the child pool.
4. If the parent pool is a standard pool, enter the amount of capacity that is dedicated to the child pool. If the parent pool is a data reduction pool, the child pool shares capacity with the parent pool.
5. Select **Safeguard** to indicate that the child pool is used as the Safeguarded backup location for immutable backup copies of source volumes.

   Note: Before a Safeguarded policy can be applied, create one Safeguarded backup location for each parent pool. After you create the corresponding Safeguarded backup locations for each parent pool, create a volume group for the all the affected source volumes.
6. Select **Encrypted** to encrypt the child pool. If the parent pool is encrypted, the child pool adopts encryption from the parent.
7. Click **Create**. Child pools that are used as Safeguarded backup locations are marked with a shield icon on the **Pools** page.

After you create the Safeguarded backup locations, create volume groups.

Using the command-line interface (CLI)

To creates a new child pool to be used as the Safeguarded backup location for parent pool, enter the following command.

```
mkmdiskgrp -size mdiskgrp_size -parentmdiskgrp mdiskgrp_id | mdiskgrp_name -name pool_name
-safeguarded
```

where *mdiskgrp_size* specifies the child pool capacity. The *mdiskgrp_id* | *mdiskgrp_name* specifies the parent pool from which the volume extents of the child pool are allocated when you create a child pool. The **-safeguarded** parameter specifies whether the child pool is used as a Safeguarded backup location.

Creating volume groups and assigning source volumes

Create a volume group and add source volumes.

Using the management GUI

To create a volume group, complete these steps:

1. In the management GUI, select **Volumes** > **Volumes Groups**.
2. Click **Create Volume Group**.
3. On the **Create Volume Group** page, enter the volume group name. If you do not specify a volume group name, the system automatically assigns the name.
4. To assign volumes, select the following options:

   Choose existing volumes
   Select this option if you want to add existing volumes to a volume group by completing these steps:

a. Select **Choose existing volumes** and click **Next**.

b. From the list, select the volumes that you want to add to the volume group.

c. Click **Create Volume Group**.

Using the command-line interface (CLI)

To create and configure a new volume group and assign volumes to that group, complete these steps:

1. To create the new volume group, enter the following command:

   ```
   mkvolumegroup -name volumegroup_name -external
   ```

   where *volumegroup_name* specifies a volume group name. The *external* defines that the policy is scheduled by an external scheduling application. This command creates the volume group and assigns the policy to the volume group.

2. After creating the volume group, either create new volumes or change existing volumes and assign them to volume group that you created in Step 1.

   If you are assigning a new volume to the volume group, enter the following command:

   ```
   mkvolume -pool <pool_name_or_id> -volumegroup <volumegroup_name_or_id> -size <disk_size>
   ```

   where *<pool_name_or_id>* is the name or identifier of the parent pool that contains the Safeguarded backup location; *<volumegroup_name_or_id>* is the name or identifier of the volume group; and *<disk_size>* indicates the capacity that is provisioned for the volume from the parent pool.

   If you are assigning existing volumes to the volume group, enter the following command:

   ```
   chvdisk -volumegroup <volumegroup_name_or_id> <name_or_id>
   ```

   where *<volumegroup_name_or_id>* is the name or identifier of the volume group and *<name_or_id>* of the volume.

Assigning a Safeguarded policy with IBM Storage Copy Data Management

You can assign one external Safeguarded policy to all the volume groups that are controlled by the IBM Storage Copy Data Management interface.

The system supports one external policy. The external policy is not a predefined policy. An external safeguarded policy cannot be converted into a safeguarded policy after it is created. Only a new created policy can be defined as external at time of its creation.After you create a volume group and assign source volumes in the management GUI, you need to create and assign the Safeguarded policy with IBM Storage Copy Data Management. To create and assign a Safeguarded policy to configured volume groups, access IBM Storage Copy Data Management interface, which is a part of IBM Storage Copy Data Management software bundle.

To create and assign a Safeguarded policy to configured volume groups, complete the following tasks in IBM Storage Copy Data Management:

Note: The following instructions are based on another product's interface and can be subject to change. For the most recent version of these steps, see IBM Storage Copy Data Management documentation.

1. In a supported browser, enter the following URL:

   `https://<HOSTNAME>:8443/portal/`

   Where <HOSTNAME> is the IP address of the virtual machine where the IBM Storage Copy Data Management is configured. This connects you to IBM Storage Copy Data Management management GUI.

2. Enter the administrator username and password.

3. After logging in to IBM Storage Copy Data Management GUI, you need to do the following task:

4. **Creating a Safeguarded Copy SLA policy**
   Create a Safeguarded backup policy in IBM Storage Copy Data Management and assign the policy to detected volume groups on IBM Storage Virtualize system.

Using the command-line interface (CLI)

To assign an external safeguard policy to volume groups, enter the following command:

```
mksafeguardedpolicy -name sgpolicy0 -external
```

Where the *external* parameter defines that the schedule is set by an external software.

Managing the Safeguarded Copy function

Safeguarded Copy function supports the ability to create cyber-resilient point-in-time copies of volumes that cannot be changed or deleted through user errors, malicious actions, or ransomware attacks.

Safeguarded snapshots are supported on the system through an internal scheduler that is defined in the snapshot policy. When the policy is assigned to a volume group, you can select Safeguarded option. The policy creates immutable snapshots of all volumes in the volume group. The system supports internal and external snapshot scheduling applications such as IBM Copy Services Manager and IBM Storage Copy Data Management. For Safeguarded snapshots with internal scheduler, refer to “Managing snapshots” on page 105.

After you configure Safeguarded Copy function on your system, ensure to regularly test the configuration to ensure that Safeguarded backups are ready in the event of a cyberattack. In addition to testing and recovering, you can also manage objects that are related to the Safeguarded Copy function on the system. These tasks include adding source volumes to Safeguarded volume groups, managing Safeguarded backups after expiration, and other actions. Some administrative actions are completed on the system, while others are completed on the management interface for the external scheduling application, like IBM Storage Copy Data Management or IBM Copy Services Manager.

IBM Copy Services Manager

The system supports IBM Copy Services Manager as an external scheduling application. IBM Copy Services Manager coordinates and automates Safeguarded Copy function across multiple systems. You can manage Safeguarded Copy function with IBM Copy Services Manager using the management GUI and command- line interface.

Adding new source volumes to a Safeguarded volume group

You can add new source volumes to an existing Safeguarded volume group. New volumes use the schedule in the external scheduling software or by using the internal snapshot scheduler.

For releases earlier than 8.5.2, the Safeguarded policy contains copy schedule information for the systems using FlashCopy technology with IBM Copy Services Manager as an external scheduler. For more information, you can refer to the “Configuring Safeguarded snapshots with IBM Copy Services Manager” on page 276.

Using the management GUI

To add source volumes to a volume group, complete these steps:

1. In the management GUI, select **Volumes** > **Volume Groups**.
2. Select the volume group that you want to add volumes to and select **Group Actions** > **Add Volumes**.
3. On the **Add Volume** page, select the volume and select **Add Volume**.

Using the command-line interface (CLI)

To add a volume to a volume group, enter the following command:

```
chvdisk -volumegroup volumegroup_name_or_id volume_name_or_id
```

where *volumegroup_name_or_id* indicates the name or ID of the volume group and *volume_name_or_id* is the name or ID of the source volume that are being added to the volume group.

Recovering a Safeguarded backup

After you configure Safeguarded Copy function on your system, ensure to regularly test the configuration to ensure that Safeguarded backups can be recovered and restored if necessary.

If a cyberattack occurs, a Safeguarded source volume can be compromised for an indefinite amount of time until the breach is detected. In this situation, the most recent Safeguarded backups are not useful for restoring data on the production volume. Effective testing includes identifying versions of Safeguarded backups that can be used to restore the compromised data. Follow the guidelines and direction of your business continuity plan and your recovery point objectives to determine the frequency of testing your configuration.

For recovering the safeguarded backup, refer to the recovery commands for IBM Copy Services Manager.

Migrating to Safeguarded snapshots

If you are currently using IBM Copy Services Manager as your external scheduling application, you can migrate to Safeguarded snapshots. Previous versions of Safeguarded Copy function with IBM Copy Services Manager used FlashCopy mappings to create backups by using an external Safeguarded backup policy. However, IBM Copy Services Manager support Safeguarded snapshots, which reduces the configuration and management complexity of previous versions.

You must remove the Safeguarded policy assigned to the volumes groups in IBM Storage Virtualize for migrating to Safeguard snapshots.

Before you can migrate to Safeguarded snapshots, complete the following prerequisites:

- Update IBM Storage Virtualize to version 8.5.2 or later.
- Update IBM Copy Services Manager to version 6.3.4 or later.

Using management GUI

Configure IBM Copy Services Manager system

To configure IBM Copy Services Manager system, complete these steps:

1. Create an IBM Copy Services Manager snapshot session, and associate it to the volume group in the IBM Copy Services Manager system. For more information, see Creating a Storage Virtualize Snapshot session using a Volume Group.
2. Set up scheduled task in the IBM Copy Services Manager system that runs **`Create Safeguarded Snapshot`** command on the session that is created in step 1. For more information, see Creating scheduled tasks.
3. Validate the creation of Safeguarded snapshot in IBM Copy Services Manager. You can manage and validate Safeguarded snapshots with IBM Copy Services Manager by creating clone or thin-clone. For more information, see Snapshot for Storage Virtualize commands.
4. Repeat these steps for all the Safeguarded backup volume groups as IBM Copy Services Manager sessions have one to one mapping with IBM Storage Virtualize volume groups.

Remove Safeguarded policy in IBM Storage Virtualize

To remove Safeguarded policy in IBM Storage Virtualize, complete these steps:

1. In the management GUI, select **Volumes** > **Volume Groups**.
2. On the **Volume Groups** page, select a volume group and from the overflow menu, select the **View Details**.
3. Select the **Policies** tab, and click **Manage Safeguarded Backup Policy**.
4. On the **Manage Safeguarded Policy** page, select **Remove policy from the volume group**.

5. Click **Apply changes**.
6. Repeat these steps for all the Safeguarded backup volume groups.
7. Verify that the existing Safeguarded backup policies are removed. On the **Volumes Group** page, select a volume group and select the **Policies** tab. The Safeguarded backup policy is not assigned message is displayed. Repeat this step for all the volume groups.

Migrating Safeguarded Copy with FlashCopy to Safeguarded snapshots with internal scheduler

If you are currently using any external scheduling application, you can migrate to Safeguarded snapshot with internal scheduler.

To migrate from Safeguarded Copy policy to Safeguarded snapshot, complete these steps:

Remove Safeguarded policy in IBM Storage Virtualize

To remove Safeguarded policy in IBM Storage Virtualize, complete these steps:

1. In the management GUI, select **Volumes** > **Volume Groups**.
2. On the **Volume Groups** page, select a volume group and from the overflow menu, select the **View Details**.
3. Select the **Policies** tab, and click **Manage Safeguarded Backup Policy**.
4. On the **Manage Safeguarded Policy** page, select **Remove policy from the volume group**.
5. Click **Apply changes**.
6. Repeat these steps for all the Safeguarded backup volume groups.
7. Verify that the existing Safeguarded backup policies are removed. On the **Volumes Group** page, select a volume group and select the **Policies** tab. The Safeguarded backup policy is not assigned message is displayed. Repeat this step for all the volume groups.

Disable external scheduling application in IBM Storage Virtualize

To disable the external scheduling application, complete these steps:

1. In the management GUI, select **Settings** > **GUI preferences** > **GUI Features**, and ensure **Display Safeguarded backup policy tile and External Scheduler Application Settings:** is **Hidden**.

Assign snapshot policy in IBM Storage Virtualize

To assign a snapshot policy to a volume group by selecting **Volume Groups**, complete these steps:

1. In the management GUI, select **Volumes** > **Volume Groups**.
2. Select the volume group from the list of volume groups. In the overflow menu, select **View Details**.
3. On the **Policies** tab, click **Assign Internal Snapshot Policy**.
4. From the **Assign Internal Snapshot Policy** page, select the snapshot policy from the list of snapshot policies that you want to assign to a volume group. If you select a volume group that is already assigned a snapshot policy, it replaces that policy with this new policy that you selected.
5. If any volumes in the selected volume groups are source volumes in a FlashCopy mapping, the **Ignore existing FlashCopy mappings** is displayed. If you are creating new snapshots or a Safeguarded snapshot, select this option. If this option is not selected, snapshots cannot be created. After you select this option and the snapshot policy is assigned successfully to a volume group, this option no longer displays.

   If volumes in a volume group are target volumes, you cannot create snapshot and **Ignore existing FlashCopy mappings** is not visible. If volumes in a volume group are both source and target volumes, **Ignore existing FlashCopy mappings** is visible, however you cannot create snapshot.
6. Select **Date** and **Choose a time (optional)** for the policy to begin. Based on selected frequency of snapshots, the start time defaults to whenever the policy is assigned to a volume group.
7. Select **Safeguarded** to safeguard the snapshots. If **Safeguarded** is selected, the snapshot policy creates Safeguarded snapshots for all the volumes in the group based on the schedule.

8. Click **Assign Policy** to assign the policy.

Note: If a snapshot policy is assigned to a volume group, the policy's schedule may result in more volume group snapshots being created than can be accommodated within the system's maximum limits. The frequency of snapshots in the snapshot policy should be adjusted to ensure that the limits are not exceeded.

Migrating to IBM Storage Copy Data Management as external scheduling application

If you are currently using IBM Copy Services Manager as your external scheduling application, you can migrate to IBM Storage Copy Data Management.

Using the management GUI

Migration to IBM Storage Copy Data Management from IBM Copy Services Manager requires that all active Safeguarded backup policies are removed from volume groups that use IBM Copy Services Manager as an external scheduling application. After backup policies are removed, remove the URL for the IBM Copy Services Manager interface from the management GUI. Currently, the system only supports one scheduling provider.

To migrate from IBM Copy Services Manager to IBM Storage Copy Data Management, complete these steps:

1. In the management GUI, select **Volumes** > **Volume Groups**.
2. On the **Volume Groups** page, select a volume group, select the **Policies** tab, and click **Manage Safeguarded Backup Policy**.
3. On the **Manage Safeguarded Policy** page, select **Remove policy from the volume group** check box.
4. Click **Apply changes**.
5. Repeat these steps for all the Safeguarded backup volume groups.
6. To verify that the existing Safeguarded backup policies are removed, on the **Volume Groups** page, select a volume group and select the **Policies** tab. Ensure that you see this message: **Safeguarded backup policy is not assigned**. Repeat this step for all the volume groups.
7. In the management GUI, select **Settings** > **GUI preferences**, and ensure **Display Safeguarded backup policy tile and External Scheduler Application Settings:** is **Visible**.
8. Select **Settings** > **System** > **External Scheduling Applications**.
9. On the **External Scheduling Applications** page, select **Remove URL**.
10. Click **Remove** on the verification panel.

Using the command-line interface

Alternatively, you can use following command-line interface steps to migrate to IBM Storage Copy Data Management:

1. Enter the following command:

```
rmsafeguardedpolicy -removefromvolumegroups <safeguarded_policy_id | safeguarded_policy_name>
```

   where *<safeguarded_policy_id | safeguarded_policy_name>* is the ID or name of the Safeguarded policy.

2. Enter the following command:

```
mksafeguardedpolicy -external
```

After removing the URL for IBM Copy Services Manager you can configure IBM Storage Copy Data Management as your external scheduling application.

Changing the system topology

You can use the command-line interface or the management GUI to change the system topology.

About this task

In a standard topology, the system is configured in a single site. IBM FlashSystem supports only the standard topology. IBM SAN Volume Controller supports the stretched topology to provide high availability where nodes in an I/O group are configured in different sites with volumes configured with a mirrored copy in each site.

Note: The recommended solution for high availability is policy-based high availability.

Use the **-topology** parameter in the **chsystem** command to change the topology of the system. In the management GUI, select **Monitoring** > **System**. In the **System Actions** menu, select **Modify System Topology**.

Procedure

1. To change the system topology from a stretched topology to a standard topology, enter the following command:

   ```
   chsystem -topology standard
   ```

2. To verify that the site definitions have been removed from the enclosure, enter this command:

   ```
   lsnode
   ```

   If either the **site_id** or **site_name** parameters are listed, remove the site definitions from the node by using the **chnode** command.

FlashCopy statistics collection

The system collects FlashCopy specific metrics, which allows support to analyze performance issue.

Introduction

FlashCopy statistics measures the I/O amplification and magnification in FlashCopy layer, and it also measures the latency of the FlashCopy I/O operations. The various cache statistics gives information regarding the various FlashCopy I/O stages, though it is not obvious where the amplification or magnification is originating. Therefore, FlashCopy will track and measure the internal operations that are required to process an I/O in FlashCopy.

Statistics types

There are two types of statistics.

- Aggregate statistics: The aggregate duration in microseconds(XXag) and the maximum duration in microseconds (XXmx).
- Count statistics: A count(XXcn) is just a simple count, this doesn't get reset.

The aggregate statistics(XXag) and count statistics(XXcn) are cumulative. This way if there are two different files (from any time), you can calculate the average response time between the two files. The maximum will be reset after each sample period.

Tables

The following tables describe the information that is reported for individual volumes.

The below table describes the statistics collection for Volume aggregate statistics.

Table 37. Statistics collection for volume aggregate	
Statistic name	**Description**
wlag, wlcn, wlmx	Indicates Write Lock Requests, these will be against the client volume. **Note:** This will include cleaning I/O.
rlag, rlcn, rlmx	Indicates Read Lock Requests, these will be against the client volume.
bcag, bccn, bcmx	Indicates Background Copy Operations, these will be against the target, that is, the volume being copied to.
clag, clcn, clmx	Indicates Cleaning Operations, these will be against the target being cleaned, that is, the volume being copied from.
rwag, rwcn, rwmx	Indicates ROW copy Operations, these will be against the client volume.
twag, twcn, twmx	Indicates Host Write I/O Operations at top of FlashCopy, these will be against the client volume.
trag, trcn, trmx	Indicates Host Read Operation at top of FlashCopy, these will be against the client volume.
bwag, bwcn, bwmx	Indicates Write Operations at bottom of host I/O, these will be against the client volume of the host IO.
brag, brcn, brmx	Indicates Read Operations at bottom of FlashCopy for host I/O, these will be against the client volume of the host IO.

The below table describes the statistics collection for Volume count statistics.

Table 38. Statistics collection for volume count	
Statistic name	**Description**
hwsg	Indicates Host Write Split Grain. The number of grains split due to host I/O for this volume, this will be against the client volume.
hwdg	Indicates Host Write Diff Grain. The number of grains marked as different due to host I/O for this volume, this will be against the client volume.
srcp	Indicates Successful ROW copy. The number of successful ROW copy operations, this will be against the client volume.
twbl	Indicates Host Write Blocks at top of FlashCopy, this will be against the client volume.
trbl	Indicates Host Read blocks at top of FlashCopy, this will be against the client volume.
bwbl	Indicates Write Blocks at bottom of FlashCopy for Host I/O, this will be against the client volume for the host I/O.
brbl	Indicates Read Blocks at bottom of FlashCopy for Host I/O, this will be against the client volume for the host I/O.

Example snippets for volume statistics files.

```
id="mtfc_vdisk0"
ro="431608" wo="134744" wou="16792" rb="256827679" wb="65293344"
rl="1384624" wl="5200676" rlw="305664" wlw="313986" xl="1325749"
wxl="612068" rxl="713681" oro="0" owo="0"
orl="0" owl="0" oiowp="0"
uo="0" ub="0" uou="0" ul="0" ulw="0">
<fc wlag="49953663" wlcn="171517" wlmx="20924" rlag="0" rlcn="0" rlmx="0" bcag="0" bccn="0"
bcmx="0" clag="0" clcn="0" clmx="0" rwag="0" rwcn="0" rwmx="0" twag="174414058"
twcn="13531" twmx="0" trag="0" trcn="0" trmx="0" bwag="1644964" bwcn="11657" bwmx="0"
brag="169298247" brcn="11657" brmx="0" hwsg="0" hwdg="78695" srcp="0" twbl="3018747" trbl="0"
```

```
bwbl="5968384" brbl="5968384" />
<ca rh="3509455" d="65314336" ft="128128" wt="0"
fw="65188256" wh="0" v="1920" m="1920"
ri="431608" wi="134789" r="253318224" dav="35731" dcn="36073" sav="2150" scn="30609" teav="0"
tsav="0"  tav="3000"  pp="0"/>
<cl bup="0" bdn="0"/><cpy idx="0">
<ca p="23479360" rh="629009376" ph="23234248" d="67061321"
ft="488576" wt="0" fw="85798816" wh="18677066"
v="7594176" m="553216" pm="1136" ri="1281284"
wi="185325" r="21743776" dav="234688" dcn="51900"
sav="14551" scn="6209" pav="16011" pcn="1365"
teav="0"  tsav="0"  tav="0"  pp="0"/>
```

Converting a volume to a stretched volume

You can use the management GUI or the command-line interface to add copies to volumes to create a stretched volume.

About this task

The **addvolumecopy** command can be used to add a copy to a basic volume, which changes a basic volume into a stretched volume with copies on two separate sites. Before you add a copy to convert a basic volume to a stretched volume, the system must be configured as a stretched system topology. To change a basic volume to a stretched volume in the management GUI, select **Volumes** and right-click the basic volume and select **Add Volume Copies**. On the **Add Volume Copy** page, select the appropriate volume type and specify the pools at each site where the copies are located.

Procedure

To add a copy to basic volume to convert it to a stretched volume, enter the following command, where *site2pool* is the storage pool on the other site.

```
addvolumecopy -pool site2pool volume5
```

Chapter 15. Notifications

The system supports the use of Simple Network Management Protocol (SNMP) traps, syslog messages, and Call Home emails to notify you and IBM Support Center when significant events are detected.

Syslog notifications

The syslog protocol is a standard protocol for forwarding log messages from a sender to a receiver on an IP network. The system can send syslog messages that notify personnel about an event. You can set up syslog event notifications with either the management GUI or the command-line interface (CLI).

The system can transmit syslog messages in either expanded or concise format. Servers configured with facility values of 0 - 3 receive syslog messages in concise format. Servers configured with facility values of 4 - 7 receive syslog messages in fully expanded format. The default value is 0. The facility number that is used in syslog messages also identifies the origin of the message to the receiving server. You can use a syslog manager to view the syslog messages that the system sends. For error, warning, and information notifications the format that messages are sent in depends on the facility setting. Audit (-audit) and authentication (-login) messages are sent in a single format so for these messages there is no distinction between concise and expanded format. The system supports TCP, UDP, and TLS transmission protocols to send the syslog message to the specified syslog servers. You can specify up to a maximum of six syslog servers with either an IP address or a fully qualified domain name and its corresponding port. The default port for the TCP protocol is port 6514, and the default port for UDP transmissions is 514. If you are using a domain name to identify a syslog server, ensure that a DNS server is configured on the system. Domain names cannot exceed 40 characters.

The system supports the following syslog notifications and message types:

Error notifications
Select this option to send error notifications that can indicate a serious problem with the system.

Warning notifications
Select this option to send warning notifications that can indicate a problem or unexpected condition with the system. Always immediately investigate this type of notification to determine the effect that it might have on your operation, and make any necessary corrections.

Information notifications
Select this option to send information messages that indicate an expected operation has completed on the system.

Audit log messages
Select this option to include any CLI or management GUI operations on the specified syslog servers.

Authentication log messages
Select this option to send successful and failed authentication attempts to the specified syslog servers.

Syslog over TLS

The system supports encrypted communication with a syslog server using a TLS connection.

Before you begin
The system and the syslog server use mutual TLS to establish a secure connection. The system must verify the syslog server certificate, and the syslog server must verify the certificate that is presented by the system.

The syslog server certificate must be installed in a trust store on the system that has the syslog tag turned on. If the syslog server's certificate is signed by a Certificate Authority (CA), then the CA certificate must be installed in a trust store with the syslog tag that is turned on. See **Truststore management commands** for more information about creating and managing trust stores.

© Copyright IBM Corp. 2024

The system only supports the use of one CA to sign the server certificates. All server certificates must be signed by the same CA. If using self-signed certificates, then all servers must use the same certificate. The system's certificate must be installed on the syslog server. If the system's certificate is signed by a Certificate Authority (CA), then the CA certificates must be installed on the SYSLOG server too.

Note: The system uses the same certificate for all services that use certificate authentication. If changes are made to the system certificate, then services that use certificate authentication may be interrupted. If any services are interrupted, add the new certificate to the necessary trust stores.

Some syslog servers may have requirements about details that must be included in the system certificate. The cluster's fully qualified domain name, and the cluster IP address, should be included in the subject alternative name fields. See **`System certificates`** for more details about generating a new system certificate.

The system and the syslog server must agree to use a cipher suite that is supported by both parties. See Changing security protocol levels for more information about changing the system's list of supported cipher suites.

Using the management GUI

To configure or work with syslog notification settings in the management GUI, select **Settings** > **Notifications** > **Syslog**.

Creating a Syslog server

To configure Syslog notifications, you must create a Syslog server.

1. Go to **Settings** > **Notifications** > **Syslog** > **Create Syslog Server** to create a Syslog server.
2. In the **Create Syslog Server** window, you can define the following protocols:

   UDP
   Use this protocol to reduce system and network overhead

   TCS
   It is a reliable delivery protocol.

   TLS
   This option uses TLS to send encrypted syslog messages.
3. If you select **TLS**, the **Drag and drop or click or click here to upload the syslog server certificate** option appears.
   a. If the syslog server's certificate is signed by a CA, upload the root CA certificate.
   b. If the server's certificate is self-signed, upload the self-signed certificate.
   c. If you are creating multiple syslog servers that use the same certificate and the certificate has already been uploaded, select the box **I have already uploaded the certificate**.

   Note: If the server certificate is signed by a chain of CAs that includes a root CA and intermediate CAs, then the server should be configured to present its server certificate and any intermediate CA certificates when establishing a connection.

   When you create a Syslog server that uses TLS, the management GUI creates a new trust store that contains the uploaded server certificates. The CLI must be used to modify or remove trust store entries for existing servers.
4. Click **Create.**

Using the Command Line interface

The following commands can be used to configure and manage syslog notifications:

1. See **`mksyslogserver`** command to create a syslog notification.
2. See **`chsyslogserver`** command to modify a syslog notification.

3. See **rmsyslogserver** command to delete a syslog notification.
4. See **lssyslogserver** command to display a concise list of syslog notification.

SNMP notifications

Simple Network Management Protocol (SNMP) is a standard protocol for managing networks and exchanging messages. The system can send SNMP messages that notify personnel about an event. You can use an SNMP manager to view the SNMP messages that the system sends. The system supports both SNMP version 2 and version 3.

About this task

Event notifications are reported to the SNMP destinations of your choice. You can specify SNMP destinations by creating SNMP server objects on the system. You must provide a valid IP address or fully qualified domain name. A DNS server must be created on the system to use a fully qualified domain name. A maximum of six SNMP destinations can be specified.

You can use the Management Information Base (MIB) file for SNMP to configure a network management program to receive SNMP messages that are sent by the system. This file can be used with SNMP messages from all versions of the software. You can use your browser to download the MIB file by clicking **Download MIB** in the management GUI by navigating to **Settings** > **Notifications** > **SNMP**.

For version 2 SNMP servers, the community string is required and the default value is public. SNMP version 3 introduces improved encryption and authentication mechanisms. The system supports two different security models when using version 3. The first security model is the User-based Security Model (USM), which provides authentication and privacy at the message level. The second security model is the Transport Security Model (TSM), which uses Transport Layer Security (TLS) to send messages over a secure connection. For more information about using TLS, see . Table 1 displays the security levels that are supported when using SNMP version 3 with USM.

These different security levels depend on the credentials that you configure to authenticate successfully to the SNMP server.

Table 39. Security levels and required credentials for SNMP version 3 servers using the user-based security model

Security Level	Description	Required security credentials
None	No additional authentication or encryption is used to send SNMP notifications.	• Engine ID • Security Name
Authentication	Notifications are authenticated, but message content is not encrypted.	• Engine ID • Security Name • Authentication Protocol • Authentication Passphrase
Authentication and Privacy	Notifications are authenticated and message content is encrypted.	• Engine ID • Security Name • Authentication Protocol • Authentication Passphrase • Privacy Protocol • Privacy Passphrase

For SNMP Version 3, use a unique engine ID for every system. The system does not automatically generate the engine ID. You can use an existing engine ID from an SNMP server. If you are setting up SNMP for the first time, you must generate a new engine ID.

Note:

- SNMP will not accept any configuration get requests from SNMP server.

Using the management GUI

1. To configure or manage SNMP servers in the management GUI, select **Settings** > **Notifications** > **SNMP**.
2. To configure a new server, select **Add SNMP Server** or right-click an existing instance and select **Modify**.
3. To send a test trap from the system to an SNMP server, right-click an existing instance and select **Test**. The system cannot provide any feedback about whether the test is a success, but you can check on the specified server that the trap has arrived successfully.

Using the Command Line interface

To configure the SNMP server, use the following commands:

- See **mksnmpserver** to add an SNMP Version 2 or Version 3 server.
- See **chsnmpserver** to change the settings of an existing SNMP server.
- See **rmsnmpserver** to remove an existing SNMP server from the system.
- See **lssnmpserver** to display either a concise list or a detailed view of the SNMP servers that are detected by the system.
- See **testsnmpserver** to send a test trap to an existing server.

SNMP over TLS

The system supports SNMP v3 servers that use the Transport Security Model (TSM). Traps are sent to the SNMP server by using Transport Layer Security (TLS) and the Transmission Control Protocol (TCP). The system does not support DTLS (or TLS over UDP).

One benefit of using TLS and TCP is that the TCP protocol handles timeouts and retries to improve delivery of traps. This can be compared to the User-based Security Model (USM) where traps are sent from the system by using UDP, with no confirmation that a trap has reached the SNMP server. Using SNMP over TLS may be preferable if your organization already has an existing public-key infrastructure in place.

Note: IPv6 is not supported with SNMP over TLS. Ensure that the system and SNMP servers are using IPv4 addresses.

Before you begin

The system and the SNMP server use mutual TLS to establish a secure connection. The system must verify the SNMP server certificate, and the SNMP server must verify the certificate that is presented by the system.

The SNMP server certificate must be installed in a trust store on the system, with the snmp tag turned on. If the SNMP server's certificate is signed by a Certificate Authority (CA), then the CA certificate must be installed in a trust store with the snmp tag that is turned on.

Note: The system only supports the use of one CA to sign the server certificates. All server certificates must be signed by the same CA. If using self-signed certificates, then all servers must use the same certificate.

The system's certificate must be installed on the SNMP server. If the system's certificate is signed by a Certificate Authority (CA), then the CA certificates must be installed on the SNMP server too.

Note:

The system uses the same certificate for all services that use certificate authentication. If changes are made to the system certificate, then services that use certificate authentication may be interrupted. If any services are interrupted, add the new certificate to the necessary trust stores.

Some SNMP managers may have requirements about details that must be included in the system certificate. The certificate may need to contain a username in a specific field (such as the Common Name field). The cluster's fully qualified domain name, and the cluster IP address, should be included in the subject alternative name fields. See for more details about generating a new system certificate.

The system and the SNMP server must agree to use a cipher suite that is supported by both parties. See for more information about changing the system's list of supported cipher suites.

Creating a new SNMP server that uses TLS using the management GUI

1. In the management GUI, select **Setting** > **Notifications** > **SNMP**.
2. To configure a new server, select **Add SNMP Server** and select the TLS checkbox. When the TLS checkbox is selected, the **SNMP certificate** box appears.
3. If the SNMP server's certificate is signed by a CA, upload the root CA certificate. If the server's certificate is self-signed, upload the self-signed certificate.

   Note:

   If the server certificate is signed by a chain of CAs that includes a root CA and intermediate CAs, then the server should be configured to present its server certificate and any intermediate CA certificates when establishing a connection.

   When you create an SNMP server that uses TLS, the management GUI creates a new trust store that contains the uploaded server certificates.

Testing the SNMP server and resolving common problems

1. In order to testing the SNMP server and resolving common problems, select **Settings** > **Notifications** > **SNMP**, right-click a server and select **Test** to send a test trap to an SNMP server. When using TLS, this checks that the TLS connection can be established, and sends a trap to the server.
2. The SNMP log on the server should be examined to verify that the test trap has arrived. If the test returns an error, or the test trap has not arrived on the server, check the SNMP log on the SNMP server for the following issues:
   a. If the log shows SSL or TLS errors at the time of the test trap, check that the correct certificates have been installed and trusted on the system and on the SNMP server. Also check that the system and the SNMP server are using compatible cipher suites.
   b. If there are no errors at the time of the test trap, check that the system can ping the SNMP server. See **ping** for information about running the command. If the system can ping the SNMP server, check that the system certificate contains any required usernames in the correct field of the certificate. Check that the user is configured correctly on the SNMP server.

Examples of configuration using net-snmp

The following example demonstrates the configuration settings used to create a connection to SNMP servers using Net-SNMP on Linux. In this example, the system certificate and the SNMP server certificate are both signed by the IBM Example CA. The storage system certificate contains the username flashsystem in the `Common Name` field. See "System Certificates" on page 389 for more details about generating the system certificate.

Configuring the server

A configuration file must be at a location accessible to snmptrapd, and the following example illustrates this configuration file at `/usr/local/share/snmp/snmptrapd.conf` location:

```
# Logging Settings
[snmp] logOption f /var/log/SNMP/snmptraps.log
format 1 "%02.2h:%02.2j TRAP%w.%q from %B\n"
format 2 "%02.2h:%02.2j TRAP%w.%q from %B\n"
authCommunity log public
```

```
# TLS Settings
[snmp] localCert snmpd.crt
[snmp] tlsMinVersion tls1_2
[snmp] tlsMaxVersion tls1_3
[snmp] trustCert ibm-example-ca
certSecName 10 flashsystem.crt --cn flashsystem
authUser log -s tsm flashsystem
```

- snmpd.crt is the signed server certificate and added to `/usr/local/share/snmp/tls/certs/snmpd.crt` location.

  The subject alternative name field should contain the fully qualified domain name and IP address of the SNMP server.

- The private key that matches snmpd.crt is added to `/usr/local/share/snmp/tls/private/snmpd.key` location.

- ibm-example-ca is the name of the root CA used to sign both snmpd.crt and flashsystem.crt. The root certificate has been added to `/usr/local/share/snmp/tls/ca-certs/ibm-example-ca.crt`

  If a chain of CA certificates is used in the signing process, then all of the CA certificates should be added to this directory, and an entry for each CA should be added to the configuration file.

- flashsystem.crt is the signed system certificate that was exported from the **Security** > **System Certificates** panel on the storage system, and added to `/usr/local/share/snmp/tls/certs/flashsystem.crt`. --cn flashsystem indicates that the username flashsystem is located in the Common Name field of the certificate.

Configuring the storage system

1. In this example we are signing the SNMP server's certificate and the storage system's certificate with our organization's CA (The IBM Example CA). If the system's certificate needs to be configured, use the **Security** > **System Certificates** panel in the management GUI to generate a new certificate request.
2. In this example, we have added flashsystem to the Common Name field as username. We have also added the system's fully qualified domain name, and the cluster IP, to the subject alternative name fields. The certificate request is then signed by the IBM Example CA, and the signed certificate is installed in **Security** > **System Certificates**.
3. In the management GUI, navigate to **Settings** > **Notifications** > **SNMP**.
4. Select **Add SNMP Server**.
5. Enter the IP and port details for the server.
6. Select the TLS checkbox.
7. In the SNMP certificate box, upload the root CA certificate used to sign the SNMP server certificate. In this example, we are using ibm-example-ca.crt.

Now that the SNMP server and the storage system have both been configured, we can start snmptrapd on the server and have it listen for TLS connections. We are using the default trap port 10162

```
snmptrapd -c /usr/local/share/snmp/snmptrapd.conf -L o tlstcp:10162
```

We can now send a test trap to the server in **Settings** > **Notifications** > **SNMP** by right-clicking the server and selecting `Test`. In our example, the trap is logged in `/var/log/SNMP/snmptraps.log` on the server.

Chapter 16. Network

The system relies on a combination of network infrastructure and connectivity technologies to facilitate data storage, retrieval, and management.

The connectivity in Storage Systems includes Ethernet Connectivity, Fibre Channel (FC) connectivity, Portsets, and Partnerships. You can connect systems to enable migration, data replication, and high-availability solutions.

Ethernet connectivity

IBM FlashSystem storage systems can use Ethernet networks to communicate with hosts.

Routing considerations

When you configure any of the protocols for event notifications, consider these routing decisions.

Routing considerations for event notification and Network Time Protocol

The system supports the following protocols that make outbound connections:

- Email
- Simple Network Mail Protocol (SNMP)
- Syslog
- Network Time Protocol (NTP).

These protocols operate only on a port that is configured with a management IP address. When it is making outbound connections, the system uses the following routing decisions:

- If the destination IP address is in the same subnet as one of the management IP addresses, the system sends the packet immediately through the Ethernet port with the matching subnet as the destination IP.
- If the destination IP address is not in the same subnet as either of the management IP addresses, the system sends the packet to the default gateway for Ethernet port 1.
- If the destination IP address is not in the same subnet as either of the management IP addresses and Ethernet port 1 is not connected to the Ethernet network, the system sends the packet to the default gateway for Ethernet port 2.

When you configure any of these protocols for event notifications, use these routing decisions to ensure that error notification works correctly if the network fails.

Configuring VLAN

Consider these guidelines and tasks before you configure a virtual local area network (VLAN).

Following are the general guidelines before configuring a VLAN:

- Virtual local area network (VLAN) tagging is supported for Ethernet-based traffic.
- VLAN provides network traffic separation at the layer 2 level for transport.
- VLAN tagging by default is disabled for any IP address of a node port. You can use the command-line interface (CLI) to optionally set a VLAN ID for port IP addresses.
- To prevent connectivity issues when a VLAN ID is configured for the IP addresses that are used for host attachment, appropriate VLAN settings on the Ethernet network and servers must also be properly configured.
- A VLAN ID can be configured for IPv4 addresses and IPv6 addresses of Ethernet ports.
- The same VLAN ID can be configured for all the Ethernet port IP addresses, or a different VLAN ID can be configured on each Ethernet port IP address.

© Copyright IBM Corp. 2024

- To ensure reliable connectivity to storage during a node failure, configure the same VLAN ID on the local and failover port on two nodes of an I/O group.
- If different VLAN IDs are set for local and failover port on two nodes of an I/O group, configure both VLAN IDs on the Ethernet switch ports. By doing this, a failover of IP addresses from a failing node to a surviving node succeeds.
- After VLAN is configured, changes to the VLAN settings disrupt traffic to or from those ports.
- Different VLAN configurations on different ports but with the same subnet on the ports is not supported.

Configuring VLAN for hosts

Virtual local area network (VLAN) tagging is supported for Ethernet-based traffic.

- Setting VLAN tags for host attachment sessions that are already configured with a system that is running software before version 7.4.0 is a disruptive process.
- To configure VLAN tags for existing host attachment sessions:
  - Configure VLAN on host ports that are used for Ethernet connectivity.
  - Configure all intervening switches with appropriate VLAN tags according to port connectivity.
  - Configure system ports with appropriate VLAN tags.
- If VLAN tagging needs to be done without disrupting I/Os from hosts, the host should have multiple ports. Multipath drivers on hosts must be correctly configured with multiple Ethernet sessions to each system port. Then, when you configure VLANs, sessions must be brought down one at a time so that I/Os can continue through alternative paths to the system. Depending on switch connectivity, this process might not be possible.

Configuring VLAN ID for iSCSI

You can use management GUI and the command-line interface (CLI) to configure a virtual local area network (VLAN) ID for a node port IP address in multiple ways.

Using the management GUI

Complete these steps to configure VLAN tags on existing host sessions:

1. On the hosts, configure VLAN on the host ports that are used for Ethernet-based connections. Consult the appropriate host documentation to complete this step.
2. On all intervening switches, configure the switches with the VLAN tags per port connectivity. Consult the appropriate switch documentation to complete this step.
3. Verify that the port on which you are configuring VLAN tags is used for host attachment. In the management GUI, select **Settings** > **Network** > **Ethernet Ports**. Ensure that **Yes** is displayed in the host attach column for the selected port. If **No** is displayed, right-click the port and select **Modify iSCSI hosts** and select **Enabled**.
4. Configure ports on the system with appropriate VLAN tags by completing these steps:
   a. In the management GUI, select **Settings** > **Network** > **Ethernet Ports**.
   b. Right-click the port on which to configure VLAN and select **Manage IP Addresses**.
   c. On **Manage IP Addresses** page, select **Add IP Address**. Enter the following information for the IP address that you are adding to the selected port:

      IP address
      Enter the IP address to associate with the selected port.

      Type
      Select the IP protocol version of the IP address.

      Subnet Mask or Prefix
      Enter the subnet mask for the IPv4 addresses or enter the prefix for IPv6 addresses.

VLAN
Enter the corresponding VLAN tag that this IP address belongs to.

Portset
Select the name or ID of the portset where the portset type is **Host Attachment**.

d. Click **Save**.

e. If you are updating an existing IP address, select the IP address and select **Edit** from the overflow menu. Enter the new VLAN tag in the **VLAN** field. Click **Save**.

5. After you configure VLAN settings for host attachment, ensure that you set the priority flow control (PFC) setting for the host attachment traffic.

   Note: If you are sharing the same port for different types of traffic, then PFC and Enhanced Transmission Selection (ETS) settings must be configured.

Using the command-line interface

1. To configure a VLAN ID along with a port IPv4 and IPv6 address to a specified Ethernet port, use the **-vlan** parameter in the **mkip** command, as in the following command-line interface (CLI) command:

   ```
   mkip -node node_name | node_id -port port_id -ip ipaddr
   -prefix prefix -vlan vlan_id
   ```

   where *node_name* | *node_id* specifies the name or ID of the node that is being configured, *ipaddr* is the IPv4 or IPv6 address for the Ethernet port, *prefix* is subnet prefix for IPv4 or IPv6, *port_id* specifies the Ethernet port ID. The **-vlan** parameter sets the virtual local area network (VLAN) ID to the value specified by *vlan_id* variable.

2. To remove the VLAN ID and the IP address from a node Ethernet port, enter the following command:

   ```
   rmip ip_address_id
   ```

   where *ip_address_id* specifies the name of the IP address. To list the valid values for the IP address ID, enter the **lsip** command.

System IP addresses

The system IP address provides access to the system management interfaces and access to remote services.

The following requirements apply to System IP addresses:

- System IP requirements: The system IP address provides access to the system management interfaces, including the management GUI, CLI, and RestAPI. The system IP address is also used to access remote services like authentication servers, NTP, SNMP, SMTP, and syslog systems, if configured.
  - A system address (IPv4 or IPv6) must be configured on Ethernet port 1 (default) or any other Ethernet port.
  - Optionally, a second system address can be configured on any Ethernet port except the one where the first system address is configured.
  - A maximum of two system addresses, either IPv4, IPv6, or a mix of both, can be configured for system management.
  - To ensure system IP failover operations, the Ethernet ports where the system addresses are configured must be connected to the same subnet. The system IP address can fail over to any node in the system.
  - System addresses can be configured on any ports.
  - IP addresses configured for system management and service access must not be used for Ethernet-based host I/O.

To assign an IP address to any Ethernet interface I/O port, use the management GUI or the **mkip** command. To change the MTU parameter for specific Ethernet port, use the management GUI or **chportethernet** command and specifies the maximum transmission unit (MTU) to improve iSCSI performance.

Note: MTU changes are not supported on 100 Gbps Ethernet adapter ports.

If the node fails, the address becomes unavailable and the host loses communication with the system via that node. To allow hosts to maintain, access to data, the node-port IP addresses for the failed node are transferred to the partner node in the I/O group. The partner node handles requests for both its own node-port IP addresses and also for node-port IP addresses on the failed node. This process is known as node-port IP failover.

Priority flow control

Priority flow control (PFC) is an Ethernet standard that supports the ability to select the priority of different types of traffic within the network.

With PFC, administrators can reduce network congestion by slowing or pausing certain classes of traffic on ports, thus providing better bandwidth for more important traffic. The Storage Virtualize system supports PFC on various supported Ethernet-based protocols on three types of traffic classes: system, host attachment, and storage traffic.

You can configure a priority tag for each of these traffic classes. The priority tag can be any value 0 - 7. You can set identical or different priority tag values to all these traffic classes. You can also set bandwidth limits to ensure QoS for these traffic classes by using the Enhanced Transmission Selection (ETS) setting on the network. When you plan to configure priority flow control (PFC), follow these guidelines and examples.

To use PFC and ETS, ensure that the following tasks are completed:

1. Ensure that ports support 10 GB or higher bandwidth to use PFC settings.
2. Configure VLAN on the system to use PFC capabilities.
3. Ensure that the same VLAN settings are configured on the all entities, including all switches between the communicating end points.
4.
5. To enable priority flow for host attachment traffic on a port, make sure that the host flag is set to *yes* on the configured IP address on that port.
6. To enable priority flow for storage traffic on a port, make sure that storage flag is set to *yes* on the configured IP address on that port.
7. On the switch, enable Data Center Bridging Exchange (DCBx). DCBx enables switch and adapter ports to exchange parameters that describe traffic classes and PFC capabilities. For these steps, check your switch documentation for details.
8. For each supported traffic class, configure and enable the same priority tag on the switch as COS value in step 4 . For example, if you plan to use priority tag 3 for storage traffic, ensure that it is configured on the IBM Storage Virtualize system through **chsystemethernet -storagecos** command and the priority tag 3 is also enabled on the switch.
9. If you are planning on using the same port for different types of traffic, ensure that ETS settings are configured on the network.

Using the management GUI

To set PFC on the system, complete these steps:

1. In the management GUI, select **Settings** > **Network** > **Priority Flow Control**.
2. For each of following classes of service, select the priority setting for that traffic type:

System
Set a value in the range 0 - 7 for the system traffic that includes communication between nodes within the system. The system priority tag is supported on iSCSI connections and on systems that support RDMA over Ethernet connections between nodes. Ensure that you set the same priority tag on the switch to use PFC capabilities.

Host Attachment
Set the priority tag in the range 0 - 7 for system to host traffic. The host attachment priority tag is supported on iSCSI connections and on systems that support NVMe over RDMA connections. Ensure that you set the same priority tag on the switch to use PFC capabilities.

Storage Virtualization
Set the priority tag in the range 0 - 7 for system to external storage traffic. The storage virtualization priority tag is supported on storage traffic over iSCSI connections. Ensure that you set the same priority tag on the switch to use PFC capabilities.

3. Click **Save**.
4. Make sure that IP is configured with VLAN.

Using the command-line interface

To use PFC on the system by using the command-line interface, follow these steps:

1. To use priority flow control for host attachment and storage traffic, use the **chsystemethernet** command to set the priority tag (COS) values before you set any data IPs by using the **mkip** command. All the host attachment and storage connections can then inherit the COS values that are already set.

   Note: You can also set it after you configure data IPs as existing sessions do not refresh the PCP value immediately. Then, when you try to connect again, the new connection takes the newly configured COS values.
2. Enter the **lssystemethernet** command to determine the current settings for host attachment, system, or storage priority tags. The resulting output displays values in the range 0 - 7 such as the following example:

   ```
   host_attach_cos  0
   system_cos       0
   storage_cos      0
   ```
3. To set the host attachment, system, or storage priority tags, enter the **chsystemethernet** command with values in the range 0 - 7 as in the following example:

   ```
   chsystemethernet -hostattachcos 4 -systemcos 5 -storagecos 6
   ```

   Resulting output:

   ```
   host_attach_cos  4
   system_cos       5
   storage_cos      6
   ```

   Where the host attachment priority tag (class of service) is set to 4, the system priority tag is set to 5, and the storage priority tag is set to 6.
4. Make sure that IP is configured with VLAN. If the IP is not configured with VLAN, see “Modifying VLAN for priority flow control” on page 302 to configure the VLAN for host attachment, system, and storage traffic.

Known considerations and limitations for PFC

- PFC is supported on 16 Gbps, 100 Gbps, and 25 Gbps Ethernet adapters.
- PFC is not supported by onboard Ethernet ports.

Considerations for system_cos

- After an upgrade to 8.5.2 or later releases, you must run **satask stopnode -warmstart** orderly on all the nodes in the system, if the value of *system_cos* is set to a nonzero value.
- On version 8.5.2 or later releases, if you change *system_cos* values with the **chsystemethernet** command, you must run **satask stopnode -warmstart** orderly on all the nodes in the system.

Warmstart ensures that the *system_cos* value is reflected correctly for system traffic.

Modifying VLAN for priority flow control

You can change Virtual local area network (VLAN) settings to separate network traffic for Ethernet transport.

The system supports VLAN configuration for system, host attachment, storage virtualization, and IP replication traffic. VLAN is a prerequisite for using priority flow control (PFC). VLAN tags msvc_chnodeipust be assigned to IP versions for which you want to use PFC capabilities and VLAN should be configured end to end including all the switches. PFC creates different classes for all types of traffic that improves network performance during times of congestion. To ensure that PFC operates as expected, the network and its components must be configured correctly. For example, if a single port is used for multiple types of traffic (host attachment, and storage) then the performance of these traffic types can be affected. The system supports setting system-wide PFC-setting on traffic for host attachment traffic and storage virtualization traffic. PFC capabilities are not supported on IP replication traffic.

Configuring VLAN for host attachment

When you configure VLAN tagging for host connections, follow these guidelines:

- To configure VLAN tagging without disrupting I/O from hosts, all hosts need multiple ports and multipath drivers. In addition, each host must be configured with multiple sessions to each node on the system.
- Each session needs to be stopped one at a time to ensure that I/O can continue on alternative paths to the system.

If you are running previous versions of the system, configuring VLAN tags for host sessions is a disruptive process.

Complete these steps to configure VLAN tags on existing host sessions:

1. On the hosts, configure VLAN on the host ports that are used for Ethernet-based connections. Consult the appropriate host documentation to complete this step.
2. On all intervening switches, configure the switches with the VLAN tags per port connectivity. Consult the appropriate switch documentation to complete this step.
3. Verify that the port on which you are configuring VLAN tags is used for host attachment. In the management GUI, select **Settings** > **Network** > **Ethernet Ports**. Ensure that **Yes** is displayed in the host attach column for the selected port. If **No** is displayed, right-click the port and select **Modify iSCSI hosts** and select **Enabled**.
4. Configure ports on the system with appropriate VLAN tags by completing these steps:
   a. In the management GUI, select **Settings** > **Network** > **Ethernet Ports**.
   b. Right-click the port on which to configure VLAN and select **Manage IP Addresses**.
   c. On **Manage IP Addresses** page, select **Add IP Address**. Enter the following information for the IP address that you are adding to the selected port:

      IP address
      Enter the IP address to associate with the selected port.

      Type
      Select the IP protocol version of the IP address.

Subnet Mask or Prefix
Enter the subnet mask for the IPv4 addresses or enter the prefix for IPv6 addresses.

VLAN
Enter the corresponding VLAN tag that this IP address belongs to.

Portset
Select the name or ID of the portset where the portset type is **Host Attachment**.

d. Click **Save**.

e. If you are updating an existing IP address, select the IP address and select **Edit** from the overflow menu. Enter the new VLAN tag in the **VLAN** field. Click **Save**.

5. After you configure VLAN settings for host attachment, ensure that you set the priority flow control (PFC) setting for the host attachment traffic.

Note: If you are sharing the same port for different types of traffic, then PFC and Enhanced Transmission Selection (ETS) settings must be configured.

Configuring VLAN for storage virtualization

VLAN tagging of the IP addresses configured for external storage systems is supported. For the systems and external storage systems to access each other, the same VLAN ID must be configured for ports on nodes on the system and the ports on the external storage system. In addition, VLAN settings must be configured on the Ethernet network to prevent any connection issues. When more than two ports are available, configure different ports for iSCSI hosts and storage systems.

Complete these steps to configure VLAN tags on storage systems:

1. On the storage systems, configure VLAN on the ports that are used for Ethernet connections. Consult the appropriate storage documentation to complete this step.
2. On all intervening switches, configure the switches with the VLAN tags per port connectivity. Consult the appropriate switch documentation to complete this step.
3. Verify that the port on which you are configuring VLAN tags is used for storage system attachment. In the management GUI, select **Settings** > **Network** > **Ethernet Ports**. Ensure that **Enabled** is displayed in the Storage Port column for the selected port. If **Disabled** is displayed, right-click the port and select **Modify Storage Ports** to enable the port for iSCSI connections to storage systems.
4. Configure ports on the system with appropriate VLAN tags by completing these steps:

a. In the management GUI, select **Settings** > **Network** > **Ethernet Ports**.

b. Right-click the port on which to configure VLAN and select **Manage IP Addresses**.

c. On **Manage IP Addresses** page, select **Add IP Address**. Enter the following information for the IP address that you are adding to the selected port:

IP address
Enter the IP address to associate with the selected port.

Type
Select the IP protocol version of the IP address.

Subnet Mask or Prefix
Enter the subnet mask for the IPv4 addresses or enter the prefix for IPv6 addresses.

VLAN
Enter the corresponding VLAN tag that this IP address belongs to.

Portset
Select the name or ID of the portset where the portset type is **Storage**.

Note: Portset3 (ID=3) is created by default and is reserved for storage traffic. Portset3 (ID =3) is the only portset of type Storage.

d. Click **Save**.

e. If you are updating an existing IP address, select the IP address and select **Edit** from the overflow menu. Enter the new VLAN tag in the **VLAN** field. Click **Save**.

5. After you configure VLAN settings for storage virtualization, ensure that you set PFC settings for the system-to-storage traffic.

   Note: If you are sharing the same port for different types of traffic, then PFC and Enhanced Transmission Selection (ETS) settings must be configured.

Planning for a direct-attached iSCSI configuration

A direct-attached configuration dedicates entire port bandwidth to be used for a specific connection. Planning must account for the volume of expected traffic before you decide how many ports are used in a direct-attached configuration. Direct attached connections are supported on the same speed at both ends.

The system offers multiple options in deciding how to create a direct-attached configuration. Bandwidth requirements for a direct-attached configuration make it important to determine the volume of expected traffic in deciding the number of ports. Ensure you have accounted for sufficient bandwidth, in particular node to node as the bandwidth between nodes determines write speed for replicated workloads.

For example, if a system to system direct-attached link is configured between nodes in an I/O group, then the link carries cache replication traffic and it might carry inter-node traffic such as FlashCopy data. Hence, enough system to system bandwidth needs to be available so that:

- Cache replication for host writes can occur without any bandwidth bottleneck
- All possible intra-node traffic can be carried without any bandwidth bottleneck.

The following guidelines are provided for direct-attached iSCSI configurations.

Note: One or more Ethernet host adapters must be installed for the system to support these direct attachment configurations.

1. iSCSI direct attachment to host systems.

   The system allows direct attachment connectivity between its iSCSI ports and the host ports. Host systems can be connected to 10 Gbps, 25 Gbps, or 100 Gbps Ethernet ports on the system. No special configuration is needed for host systems that are using this configuration.

2. iSCSI direct attachment between nodes in a system.

   Direct connection of the system Ethernet ports without using an Ethernet switch is supported. Such direct connections between the system nodes might be useful in small configurations where no switch exists. It can also be used to connect nodes in the same I/O group to provide a dedicated connection for mirroring the fast write cache data. Ensure that sufficient bandwidth is provisioned between nodes to accommodate all the intra-node traffic.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Note:

- AIX host attach is generally not supported with direct-attached iSCSI configuration and requires SCORE approvals for enabling support.
- iSCSI direct-attached over policy-based high availability configuration is generally not supported and requires SCORE approvals.
- For each connection, the host interfaces must be on different subnets. For instance, two interfaces from the same host must be on different subnets.

Fibre channel connectivity

The Fibre Channel protocol (FCP) is used to establish connectivity over FC infrastructure.

Systems are defined by supported configuration and zoning rules. A system requires single-initiator zoning for all large configurations that contain more than 64 host objects. See zoning for hosts and zoning for external storage system.

Fabric Device Management Interface

The Fabric Device Management Interface (FDMI) enables any storage endpoint to register itself to the Fibre Channel (FC) fabric and query the HBA and port details of the entire fabric.

The FDMI functionality is divided into these two categories:

FDMI registration
The IBM Storage Virtualize system registers with the FC fabric and provides the system information to the FDMI database, such as node names, port speed, and software version.

FDMI discovery
The IBM Storage Virtualize system queries the FC fabric and collects the information about all the other devices on the fabric. Information is also collected about the devices that are not zoned to connect to the IBM Storage Virtualize system.

To view the FDMI information, run the **`lsfabricport`** command. FDMI discovery runs automatically every 6 hours and collects information about the fabric.

The same details are captured as a part of Call Home inventory data.

Fibre Channel port numbers and worldwide port names

Fibre Channel (FC) ports are identified by their physical port number and by a worldwide port name (WWPN).

The physical port numbers identify FC adapters and cable connections when you run service tasks. Worldwide port names (WWPNs), which uniquely identify the devices on the SAN, are used for tasks such as FC switch configuration. The WWPNs are derived from the worldwide node name of the node in which the ports are installed.

The vital product data (VPD) for each controller enclosure contains a unique WWNN value for each node canister. Each WWNN, in turn, defines the WWPN that each node canister allocates to the ports on an FC host interface adapter. All systems can support an optional quad-port 16 Gbps Fibre Channel (FC) host interface adapter.

Table 40 on page 305 shows the WWPN values for each port when a quad-port FC host interface adapter is installed in a control enclosure.

Table 40. Examples of WWPN values for the system

Host interface adapter	Port number	Canister 1 WWPN	Canister 2 WWPN
Quad-port 16 Gbps	1	5005 0768 0D00 0000	5005 0768 0D00 0001
	2	5005 0768 0D04 0000	5005 0768 0D04 0001
	3	5005 0768 0D08 0000	5005 0768 0D08 0001
	4	5005 0768 0D0C 0000	5005 0768 0D0C 0001

N-Port ID Virtualization

The system supports N_Port ID virtualization (NPIV) technology for Fibre Channel (FC) connections. NPIV is an industry standard technology that provides the capability to share one physical FC adapter as multiple virtual ports. One physical port on an FC adapter can use several world wide port names (WWPNs) and world wide node names (WWNNs) to register with the SAN fabric.

NPIV is a method for virtualizing a physical Fibre Channel port that is used for host I/O. When NPIV is enabled, ports do not come up until they are ready to service I/O, which improves host behavior around node unpends. In addition, path failures due to an offline node are masked from host multipathing.

NPIV functionality allows different WWPNs to be used for host I/O and clustering on the same physical Fibre Channel port. When NPIV is enabled, host I/O WWPNs do not come online until they are ready to service I/O, which can improve host behavior when a node is coming online. In addition, path failures due to an offline node are masked from hosts and their multipathing drivers do not need to do any path recovery because the I/O group partner can take the WWPNs on the offline node.

By default, NPIV is enabled on the system. However, you must accurately zone your system for NPIV configuration. For more information, see "Zoning requirements for N_Port ID virtualization" on page 306.

Zoning requirements for N_Port ID virtualization

All of the N_Port IDs for the two nodes in an I/O group must be associated with the same fabric.

The port ID of a virtual port remains constant regardless of the node that it is active on. Ports of the two nodes in an I/O group must be attached to the same fabric, while ports of different I/O groups can be on different fabrics. Figure 26 on page 306 shows a configuration that is not valid when using NPIV.

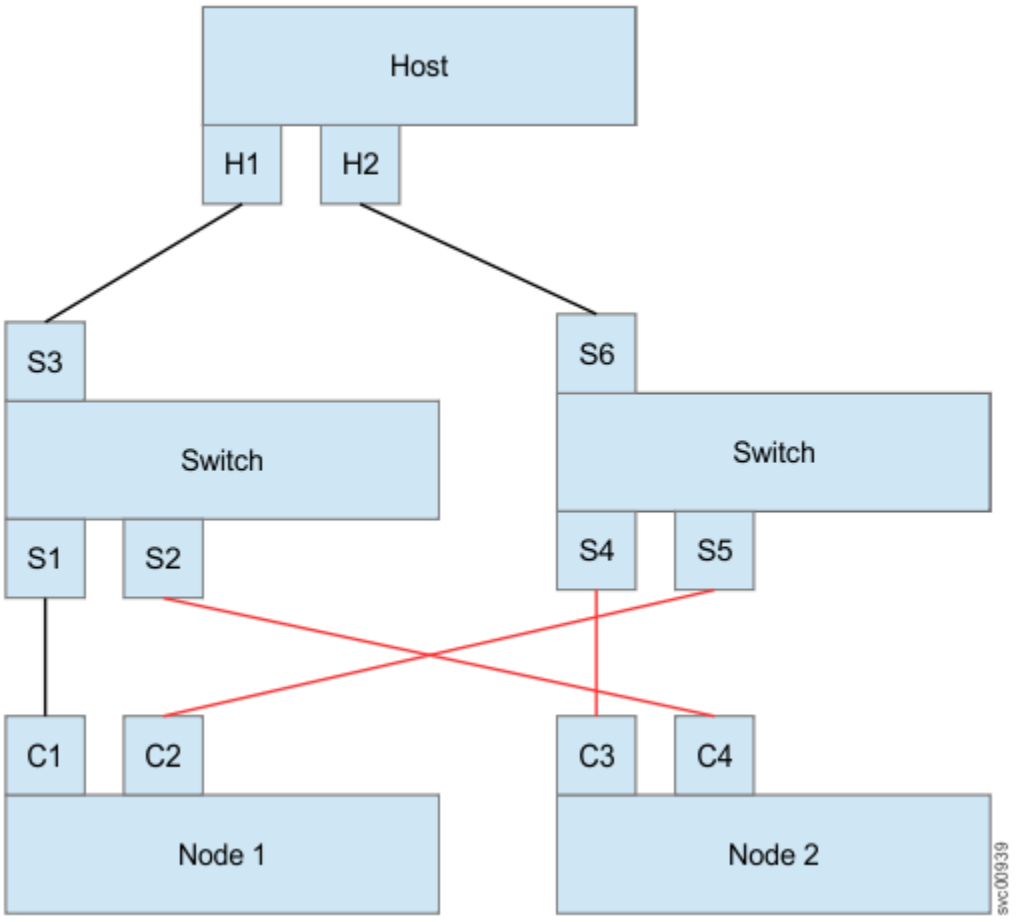


Figure 26. A configuration that is not valid for use with NPIV

The connections to switch ports S2, S4, and S5 must be exchanged, leading to the following correct configuration that is shown in Figure 27 on page 307.

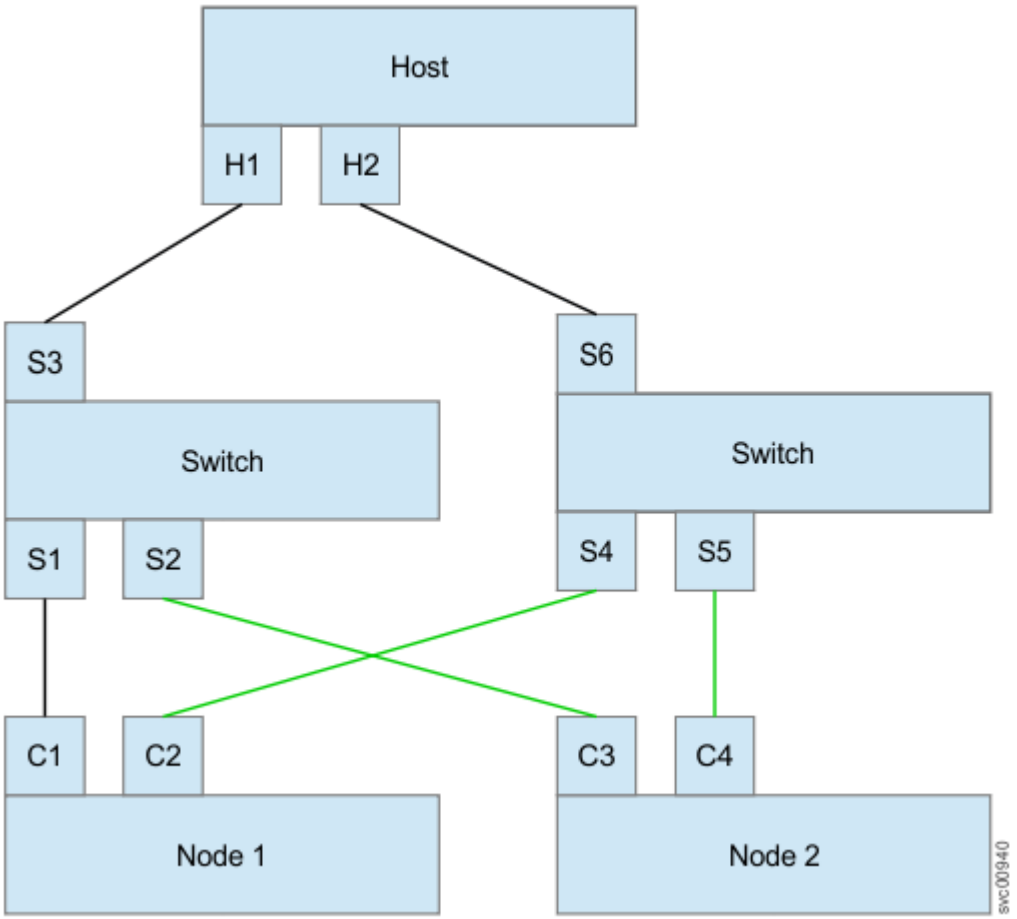


Figure 27. A configuration that is valid for use with NPIV and (WWPN-based) zoning

The same situation can occur when using a single switch fabric with hard (port based) zoning, as shown in Figure 28 on page 308.

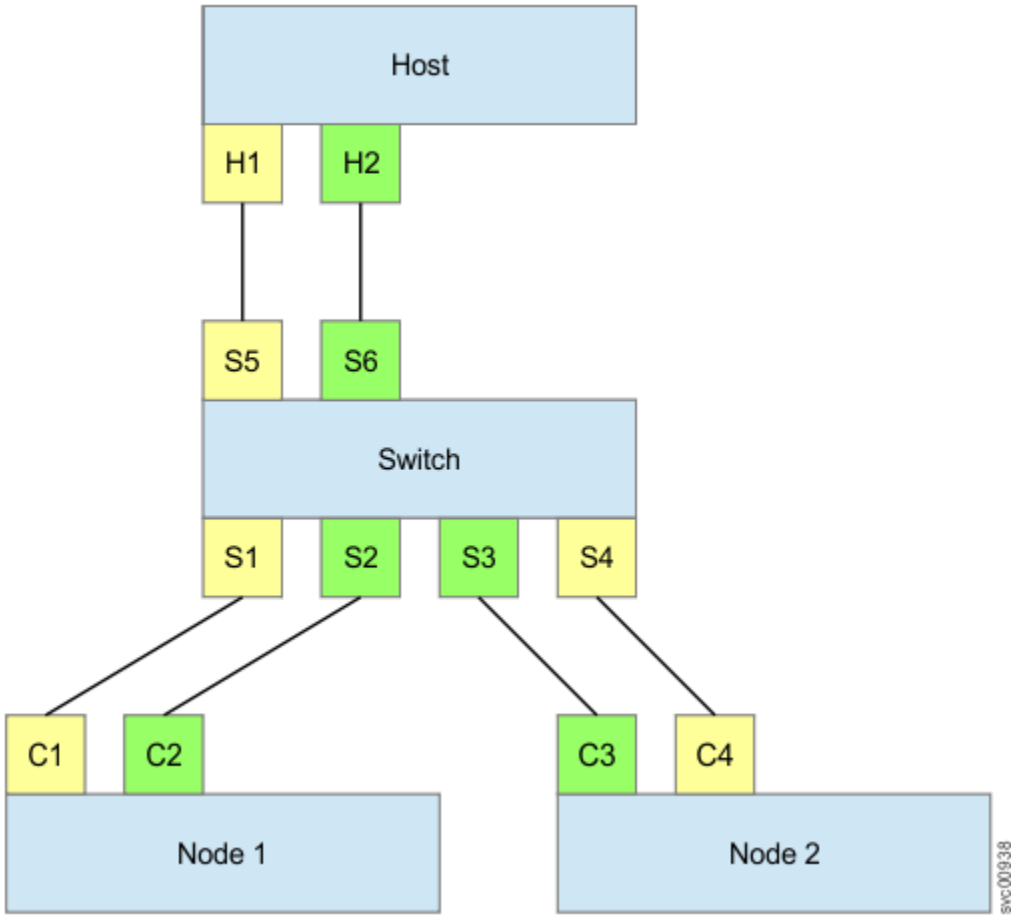


Figure 28. A single switch configuration that is not valid for use with NPIV and hard (port-based) zoning

Two nodes are connected to one host by using one switch. However, port-based zoning is used to ensure that each port of the host can detect only one port of each node. S1, S4, and S5 are in one zone; and S2, S3, and S6 are in the other zone.

During an IP failover, NPIV ports move from C1 to C3 and C2 to C4 (or the other way around). The configuration is not valid with NPIV as the visibility of ports move between H1 and H2. To use port-based zoning in this situation, two zones are required: One zone contains S1, S3, S5, C1 and C3. Another zone contains S2, S4, S6, C2 and C4. This configuration is shown in Figure 29 on page 309.

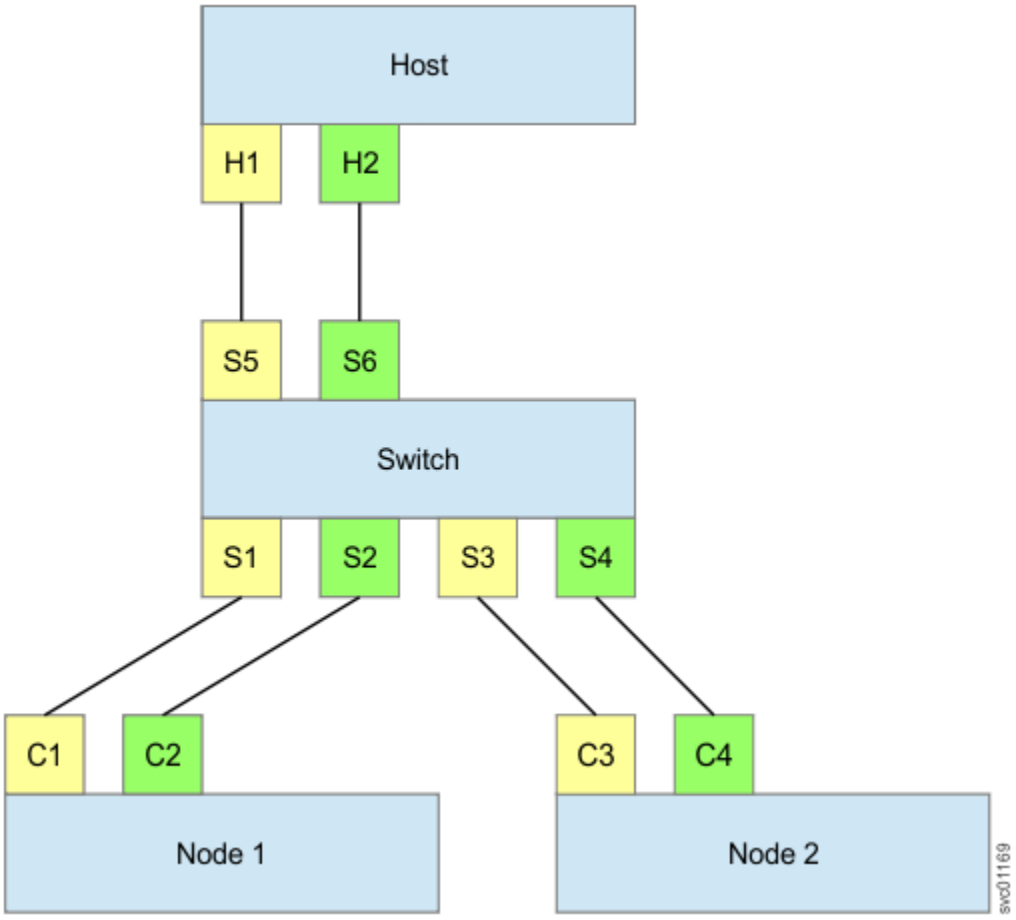


Figure 29. A single switch configuration that is valid for use with NPIV and hard (port-based) zoning

This situation does not arise when using a single switch fabric with soft (WWPN based) zoning. However, care needs to be taken in a multiswitch fabric where the node-to-host route passes through different switches for the two nodes in an I/O group.

If fabric changes are required, make them before you switch each I/O group to transitional state. I/O groups are switched to transitional state to help move them to NPIV. In transitional state, an I/O group contains both physical and virtual Fibre Channel ports that are currently being used. After you switch to transitional state, an attempt is made by the system software to warn about mismatched fabric issues. It does not warn about issues that are caused by port-based zoning.

Failure to adhere to the fabric zoning rules might cause undesirable behavior on the host during NPIV port failure. Examples include a 2-minute pause in I/O or reduced path redundancy.

Planning for connectivity

Plan to provide the network infrastructure and the storage network infrastructure that your system requires.

Before you begin

Event notifications and call home notifications from the storage enclosures are done by email. This process requires the IP address of an email server that can access the internet.

Suitability for telecommunication environment: This product is not intended to connect directly or indirectly by any means whatsoever to interfaces of public telecommunications networks.

The SAN must contain only supported switches, Fibre Channel extenders, and SAN routers. See the following website for specific firmware levels and the latest supported hardware: www.ibm.com/support.

Planning for a direct-attached configuration

A direct-attached configuration dedicates entire port bandwidth to be used for a specific connection. Planning must account for the volume of expected traffic before you decide how many ports are used in a direct-attached configuration. Direct attached connections are supported on the same speed at both ends.

The system offers multiple options in deciding how to create a direct-attached configuration. Bandwidth requirements for a direct-attached configuration make it important to determine the volume of expected traffic in deciding the number of ports. Ensure you have accounted for sufficient bandwidth, in particular node to node as the bandwidth between nodes determines write speed for replicated workloads.

For example, if a system to system direct-attached link is configured between nodes in an I/O group, then the link carries cache replication traffic and it might carry inter-node traffic such as FlashCopy data. Hence, enough system to system bandwidth needs to be available so that:

- Cache replication for host writes can occur without any bandwidth bottleneck
- All possible intra-node traffic can be carried without any bandwidth bottleneck.

The following guidelines are provided for direct-attached configurations.

Note: One or more Fibre Channel host adapters must be installed for the system to support these direct attachment configurations.

1. Fibre Channel direct attachment to host systems.

   The system allows direct attachment connectivity between its Fibre Channel ports and host ports. Host systems can be connected to 16 Gbps to 64 Gbps Fibre Channel ports on the system. No special configuration is needed for host systems that are using this configuration.
2. Fibre Channel direct attachment between nodes in a system.

   Direct connection of the system Fibre Channel ports without using a Fibre Channel switch is supported. Such direct connections between the system nodes might be useful in small configurations where no Fibre Channel switch exists. It can also be used to connect nodes in the same I/O group to provide a dedicated connection for mirroring the fast write cache data. Ensure that sufficient bandwidth is provisioned between nodes to accommodate all the intra-node traffic.

FlashSystem supports 16 Gb Fibre Channel ports.

Note: Direct attached connections are supported on the same speed at both ends.

For more information about host compatibility, see IBM System Storage Interoperation Center (SSIC).

Planning for more than four fabric ports per node canister

You can use more than four fabric ports per node canister to improve peak load I/O performance, but careful planning is needed.

A *fabric port* is a Fibre Channel port. If you use more than four fabric ports per node, you must either use the **`localfcportmask`** and **`partnerfcportmask`** commands or be careful with your fabric zoning configuration.

Careful zoning improves resilience and prevents lease expires due to port errors, low fabric buffer credits, shared adapter resources, or push-back from remote links. Table 41 on page 310 displays a port allocation scheme.

Table 41. Port allocation scheme

Adapter / Port	4 ports	8 ports	12 ports	16 ports	SAN Fabric
Adapter 1 Port 1	Host and Storage	Host and Storage	Host and Storage	Host and Storage	A

Table 41. Port allocation scheme (continued)

Adapter / Port	4 ports	8 ports	12 ports	16 ports	SAN Fabric
Adapter 1 Port 2	Host and Storage	Host and Storage	Host and Storage	Host and Storage	B
Adapter 1 Port 3	Intracluster and Replication	Intracluster	Intracluster	Intracluster	A
Adapter 1 Port 4	Intracluster and Replication	Intracluster	Intracluster	Intracluster	B
Adapter 2 Port 1	-	Host and Storage	Host and Storage	Host and Storage	A
Adapter 2 Port 2	-	Host and Storage	Host and Storage	Host and Storage	B
Adapter 2 Port 3	-	Intracluster or Replication	Replication or Host and Storage	Replication or Host and Storage	A
Adapter 2 Port 4	-	Intracluster or Replication	Replication or Host and Storage	Replication or Host and Storage	B
Adapter 3 Port 1	-	-	Host and Storage	Host and Storage	A
Adapter 3 Port 2	-	-	Host and Storage	Host and Storage	B
Adapter 3 Port 3	-	-	Intracluster	Intracluster	A
Adapter 3 Port 4	-	-	Intracluster	Intracluster	B
Adapter 4 Port 1	-	-	-	Host and Storage	A
Adapter 4 Port 2	-	-	-	Host and Storage	B
Adapter 4 Port 3	-	-	-	Replication or Host and Storage	A
Adapter 4 Port 4	-	-	-	Replication or Host and Storage	B
localfcportmask	1100	11001100 or 00001100	110000001100	0000110000001100	-
remotefcportmask	1100	00000000 or 11000000	000011000000	1100000011000000	-

- Host refers to host objects defined in the system.
- Replication refers to nodes that are part of a different cluster.
- Storage refers to controller objects defined in the system if external storage is being used.
- Intracluster refers to nodes within the same cluster.

- The word "and" indicates that both types are used.
- The word "or" indicates that one of the options must be selected. If using replication, preference should be given to replication.

If a node canister receives more than 16 logins from another node, then it causes node error 860.

You can ensure that no more than 16 logins are received by following these guidelines:

- Zone the SAN fabric to reduce the number of paths between the nodes.
- Apply a local Fibre Channel port mask (if the nodes are within the same system) or partner Fibre Channel port mask (if the nodes are in different systems) to reduce the number of ports that are being used for node-to-node communication.
- Provide a combination of the two (zoning and port masks).

To avoid receiving node error 860 and to maximize performance on your system, follow these guidelines:

- For redundancy, use a minimum of two ports. If ports are being dedicated for different types of traffic, assign a minimum of two ports for each type of traffic.
- Within a system, up to 50% of the overall data that is transferred across the SAN is transmitted between the nodes. However, for read-intensive workloads, the figure is much less. Therefore, if ports are being dedicated for different types of traffic, assign between 1/4 and 1/2 of the overall ports for node-to-node communication within a system.
- For replication between systems, the connection between the systems is usually the bottleneck. Other than for redundancy, there is no point in having more SAN connections than there is bandwidth. For example, if two sites are connected with a 10 Gbps link, two 8 Gbps ports are sufficient. Systems usually are configured with two ports per node for replication traffic primarily for redundancy. For systems with larger numbers of nodes, it might be appropriate to have connections only from a subset of the nodes for replication to a remote system. The system automatically forwards replication traffic between local nodes so that all nodes can still participate in replication.

In the examples that follow, consider this configuration that is shown in Figure 30 on page 313:

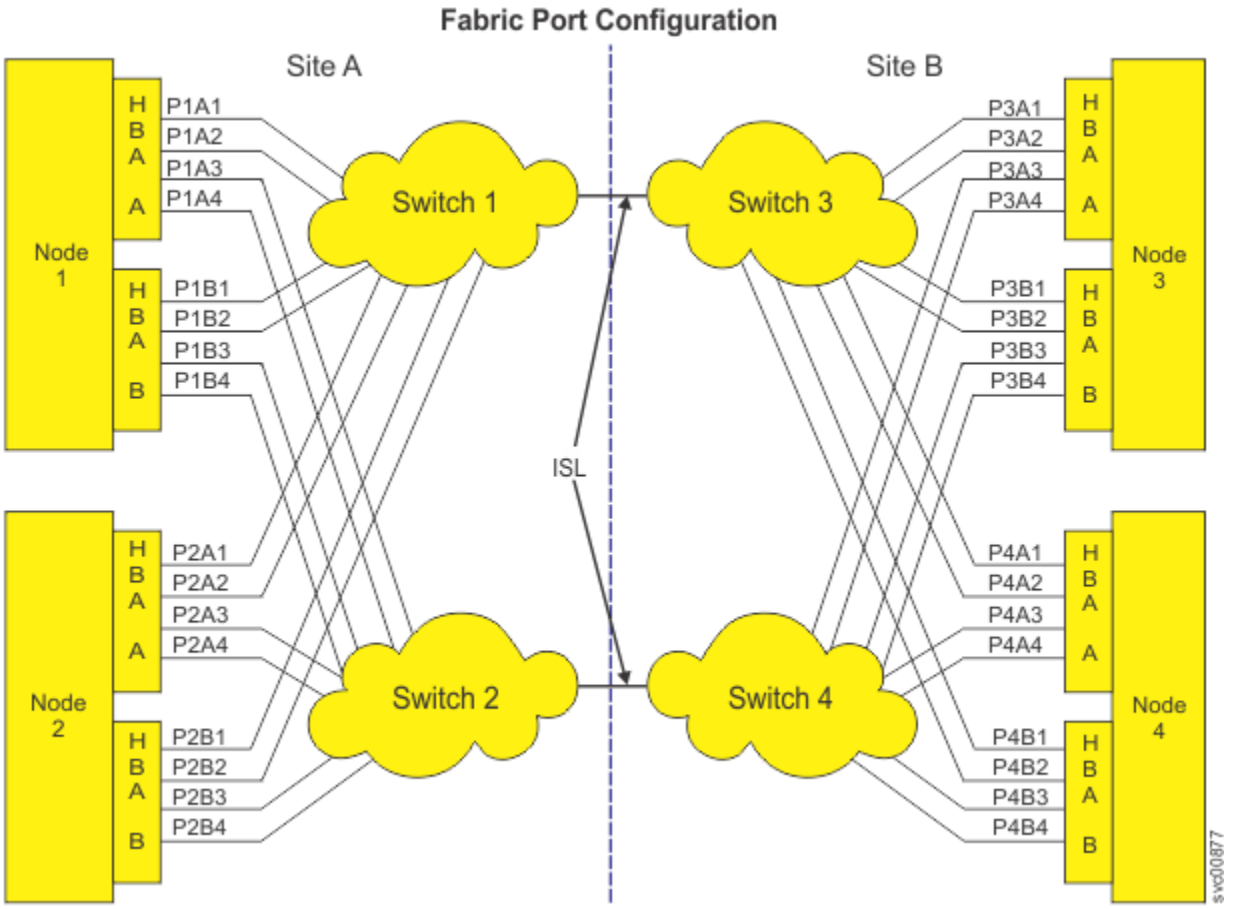


Figure 30. Fabric port configuration

- Four nodes, where nodes 1 and 2 are in a system on site A. Nodes 3 and 4 are in a system on site B.
- Each site has 2 switches (switch 1 and 2 on site A; switch 3 and 4 on site B).
- Each node has 2 adapters (A and B).
- Each adapter has 4 ports where the port is named P *nodeid adapterid portnumber*. For example, port 3 on adapter B of node 4 is named P4B3.

Example 1: Zoning only

Each node in the system is configured in the following way:

- Four ports are used for host, controller, and local node connections.
- Four ports are used for replication.

The following zones are required for two sites, A and B:

Zone sets for site A:

- Zone for storage controllers and local node connections (Fibre Channel adapter HBA)
  - Zone 1 - [Node ports P1A1, P1A2, P2A1, P2A2 plus all storage controller ports in site A on switch 1]
  - Zone 2 - [Node ports P1A3, P1A4, P2A3, P2A4 plus all storage controller ports in site A on switch 2]
- Zones for host attachment (Fibre Channel adapter HBA A)
  - The same ports on the first adapter are zoned to be visible to hosts. It is recommended that each host is in a separate zone, and a host with two ports that are connected to two fabrics have two zones, one per fabric. The actual number of zones might be two multiplied by the number of hosts. A host with four ports that are connected to two fabrics should have four zones. In this case, the actual

number of zones might be four multiplied by the number of hosts. In the four-port host case, the zones for one host might be:

- Zone 3 - [host port, P1A1 and P2A1 for multipath fail over] on switch 1
- Zone 4 - [host port, P1A2 and P2A2 for multipath fail over] on switch 1
- Zone 5 - [host port, P1A3 and P2A3 for multipath fail over] on switch 2
- Zone 6 - [host port, P1A4 and P2A4 for multipath fail over] on switch 2

If the host only had two ports, it might have Zone 3 and Zone 5, or Zone 4 and Zone 6.

- Zones for replication (Fibre Channel adapter HBA B only):

To ensure replication use the second Fibre Channel adapter exclusively, it is required that each local or remote port pair has its own zone:

Zone 7 - [P1B1,P3B1]
Zone 8 - [P1B1,P3B2]
Zone 9 - [P1B1,P4B1]
Zone 10 - [P1B1,P4B2]
Zone 11 - [P1B2,P3B1]
Zone 12 - [P1B2,P3B2]
Zone 13 - [P1B2,P4B1]
Zone 14 - [P1B2,P4B2]
Zone 15 - [P2B1,P3B1]
Zone 16 - [P2B1,P3B2]
Zone 17 - [P2B1,P4B1]
Zone 18 - [P2B1,P4B2]
Zone 18 - [P2B1,P4B2]
Zone 19 - [P2B1,P4B2]
Zone 20 - [P2B2,P3B2]
Zone 21 - [P2B2,P4B1]
Zone 22 - [P2B2,P4B2]

Zone sets for site B:

This zone set is essentially a mirror image of the zoning for site A.

- Zone for storage controllers and local node connections (Fibre Channel adapter HBA A):
  - Zone 23 - [Node ports P3A1, P3A2, P4A1, P4A2 plus all storage controller ports in site B on switch 3]
  - Zone 24 - [Node ports P3A3, P3A4, P4A3, P4A4 plus all storage controller ports in site B on switch 4]
- Zones for host attachment (Fibre Channel adapter HBA A)
  - The same ports on the first adapter are zoned to be visible to hosts. Only zones for one host are listed. It is recommended that each host is in a separate zone. A host with two ports that are connected to two fabrics has two zones, one per fabric. The actual number of zones might be two multiplied by the number of hosts. A host with four ports that are connected to two fabrics should have four zones. In this case, the actual number of zones might be four multiplied by the number of hosts. In the four-port host case, the zones for one host might be:

    - Zone 25 - [host ports, P3A1 and P4A1 for multipath fail over] on switch 3
    - Zone 26 - [host ports, P3A2 and P4A2 for multipath fail over] on switch 3
    - Zone 27 - [host ports, P3A3 and P4A3 for multipath fail over] on switch 4
    - Zone 28 - [host ports, P3A4 and P4A4 for multipath fail over] on switch 4

    If the host had only two ports, it might have Zone 25 and Zone 27, or Zone 26 and Zone 28.
- Zones for replication (Fibre Channel adapter HBA B only):

To ensure replication use the second Fibre Channel adapter exclusively, it is required that each local or remote port pair has its own zone:

Zone 29 - [P1B3,P3B3]
Zone 30 - [P1B3,P3B4]
Zone 31 - [P1B3,P4B3]
Zone 32 - [P1B3,P4B4]
Zone 33 - [P1B4,P3B3]
Zone 34 - [P1B4,P3B4]
Zone 35 - [P1B4,P4B3]
Zone 36 - [P1B4,P4B4]
Zone 37 - [P2B3,P3B3]
Zone 38 - [P2B3,P3B4]
Zone 39 - [P2B3,P4B3]
Zone 40 - [P2B3,P4B4]
Zone 41 - [P2B4,P3B3]
Zone 42 - [P2B4,P3B4]
Zone 43 - [P2B4,P4B3]
Zone 44 - [P2B4,P4B4]

Example 2: Port masking only

Assuming the same configuration as in example 1, you can achieve the same result by using port masking rather than zoning. The positions in the mask represent the Fibre Channel I/O port IDs with ID 1 in the rightmost position. These IDs are shown by the **lsportfc** command. In this example ports A1, A2, A3, A4, B1, B2, B3, and B4 correspond to FC I/O port IDs 1, 2, 3, 4, 5, 6, 7 and 8.

For remote partnership communication, apply a mask that limits communication to ports B1, B2, B3, and B4 by applying a port mask of 11110000 to both systems. Use the command `chsystem -partnerfcportmask 11110000`.

For local node-to-node communication, apply a mask that limits communication to ports A1, A2, A3, and A4 by applying a port mask of 00001111 to both systems. Use the command `chsystem -localfcportmask 00001111`.

It is then possible to have a single zone that contains all the node ports, as the port masking ensures that the number of logins is no more than 16.

With this setup, zoning for host connectivity and storage connectivity does not need to consider the rules about more than 16 logins.

Zoning examples

The following zoning examples describe ways for zoning a switch. In the examples, a list of port names that are inside brackets ([]) represent a single zone whose zone members are the list of ports shown.

Example 1

Consider the SAN environment in the following example:

- Two nodes (nodes A and B).
- Nodes A and B each have four ports.
  - Node A has ports A0, A1, A2, and A3.
  - Node B has ports B0, B1, B2, and B3.
- Two hosts that are called P and Q.
- Each of the two hosts has two ports, as described in the following table.

Table 42. Two hosts and their ports	
P	**Q**
P0	Q0
P1	Q1

- Two switches that are called X and Y.
- Two storage systems I and J.
- Each of the two storage systems has ports as described in the following table.

Table 43. Two storage systems and their ports	
I	**J**
I0	J0
I1	J1
I2	
I3	

The following tasks comprise an example configuration:

1. Attach half the host and node ports 1 (A0, A1, B0, B1, P0, Q0) to switch X.
2. Attach half the host and node ports 3 (A2, A3, B2, B3, P1, Q1) to switch Y.
3. Attach half the storage system ports (I0, I1, J0) to switch X.
4. Attach half the storage system ports (I2, I3, J1) to switch Y.
5. Create one zone per host port (one port per node) on switch X:

   [A0, B0, P0]
   [A1, B1, Q0]

Figure 31 on page 317 illustrates this example.

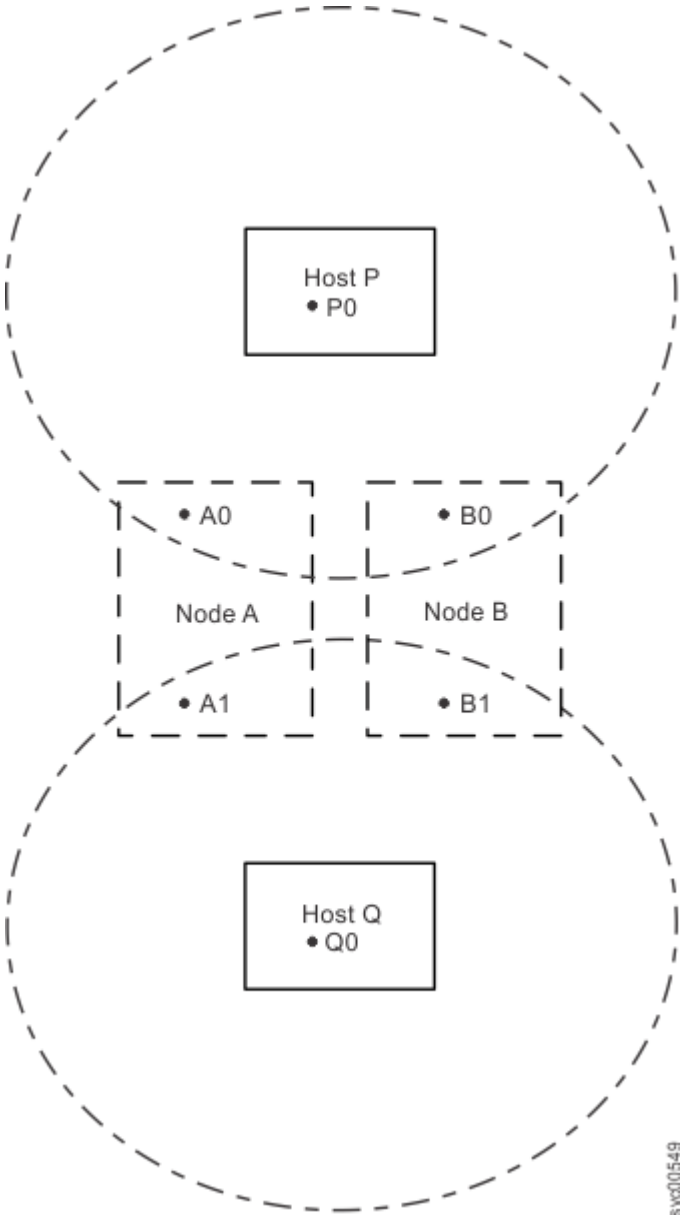


Figure 31. An example of a host zone

6. Create one storage zone per storage system on switch X:

   [A0, A1, B0, B1, I0, I1]
   [A0, A1, B0, B1, J0]

   Figure 32 on page 318 illustrates this example.

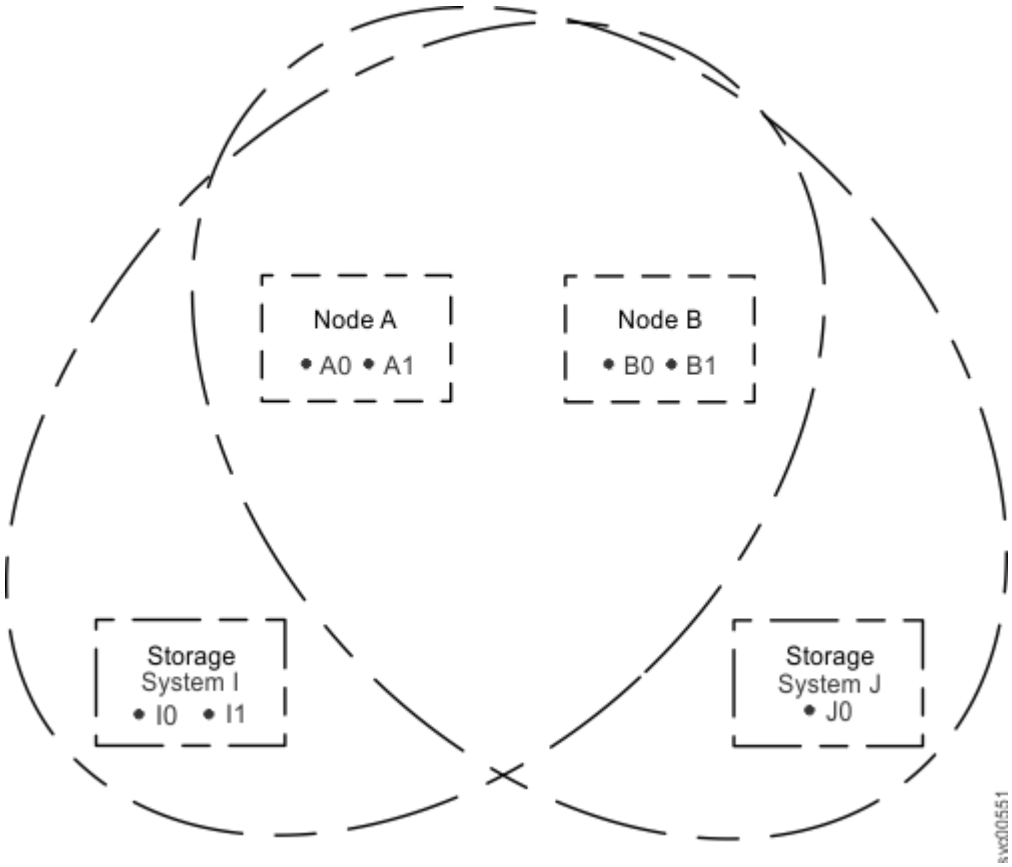


Figure 32. An example of a storage system zone

7. Create one internode zone on switch X:

   [A0, A1, B0, B1]

   Figure 33 on page 318 illustrates this example.

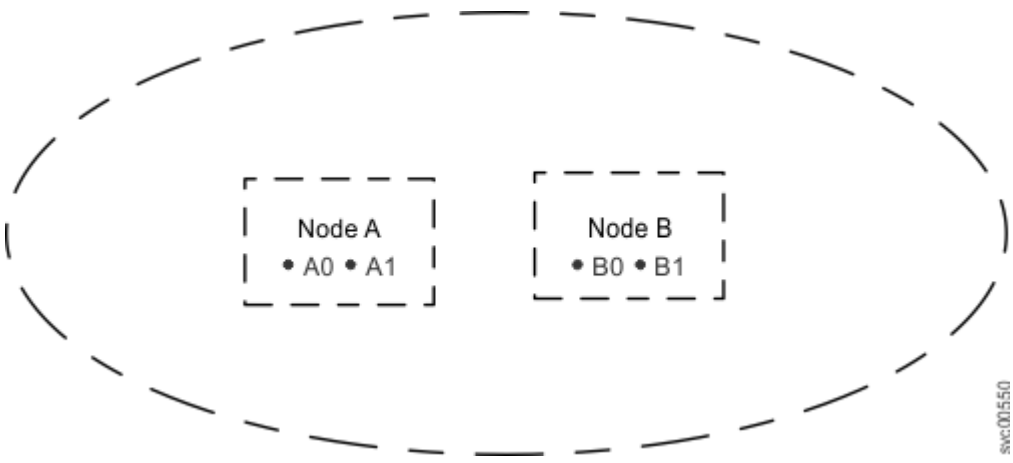


Figure 33. An example of a system zone

8. Follow the same steps "5" on page 316 through "7" on page 318 to create the following list of zones for switch Y:

   One zone per host port:

   [A2, B2, P1]
   [A3, B3, Q1]

   Storage zone:

[A2, A3, B2, B3, I2, I3]
[A2, A3, B2, B3, J1]

One internode zone:

[A2, A3, B2, B3]

Example 2

The following example describes a SAN environment that is like the previous example except for the addition of four hosts that have two ports each.

- Two nodes that are called A and B.
- Nodes A and B have four ports each.
  - Node A has ports A0, A1, A2, and A3.
  - Node B has ports B0, B1, B2, and B3.
- Six hosts that are called P, Q, R, S, T, and U.
- Four hosts have four ports each and the other two hosts have two ports each as described in the following table.

Table 44. Six hosts and their ports

P	**Q**	**R**	**S**	**T**	**U**
P0	Q0	R0	S0	T0	U0
P1	Q1	R1	S1	T1	U1
P2	Q2	R2	S2		
P3	Q3	R3	S3		

- Two switches that are called X and Y.
- Three storage systems
- Each storage system has ports as described in the following table.

Table 45. Three storage systems and their ports

I	**J**	**K**
I0	J0	K0
I1	J1	K1
I2		K2
I3		K3
		K4
		K5
		K6
		K7

The following tasks comprise a different example configuration:

1. Attach half of the host and node ports 1 (A0, A1, B0, B1, P0, P1, Q0, Q1, R0, R1, S0, S1, T0, U0) to switch X.
2. Attach half of the host and node ports 1 (A2, A3, B2, B3, P2, P3, Q2, Q3, R2, R3, S2, S3, T1, U1) to switch Y.
3. Attach half of the storage system ports (I0, I1, J0, K0, K1, K2, K3) to switch X.

4. Attach half of the storage system ports (I2, I3, J1, K4, K5, K6, K7) to switch Y.
5. Create one zone per host port (one port per node) on switch X:

   [A0, B0, P0]
   [A1, B1, P1]
   [A0, B0, Q0]
   [A1, B1, Q1]
   [A0, B0, R0]
   [A1, B1, R1]
   [A0, B0, S0]
   [A1, B1, S1]
   [A0, B0, T0]
   [A1, B1, U0]

   Attention: Hosts T and U (T0 and U0) and (T1 and U1) are zoned to different ports so that each port is zoned to the same number of host ports.
6. Create one storage zone per storage system on switch X:

   [A0, A1, B0, B1, I0, I1]
   [A0, A1, B0, B1, J0]
   [A0, A1, B0, B1, K0, K1, K2, K3]
7. Create one internode zone on switch X:

   [A0, A1, B0, B1]
8. Follow the same steps “5” on page 320 through “7” on page 320 to create the following list of zones for switch Y:

   One zone per host port:

   [A2, B2, P2]
   [A3, B3, P3]
   [A2, B2, Q2]
   [A3, B3, Q3]
   [A2, B2, R2]
   [A3, B3, R3]
   [A2, B2, S2]
   [A3, B3, S3]
   [A2, B2, T1]
   [A3, B3, U1]

   Storage zone:

   [A2, A3, B2, B3, I2, I3]
   [A2, A3, B2, B3, J1]
   [A2, A3, B2, B3, K4, K5, K6, K7]

   One internode zone:

   [A2, A3, B2, B3]

Fibre Channel port masking

With Fibre Channel port masking, you control the use of Fibre Channel ports. You can control whether the ports are used to communicate to other nodes within the same local system, and if they are used to communicate to nodes in partnered systems. Fibre Channel port masking does not affect host or storage traffic. It gets applied only to node-to-node communications within a system and replication between systems.

Usefulness of Fibre Channel port masking

- The setup of Fibre Channel port masks is useful when you have more than four Fibre Channel ports on any node in the system, as it saves setting up many SAN zones.
- Fibre Channel I/O ports are logical ports, which can exist on Fibre Channel platform ports.
- The system has two Fibre Channel port masks. The local port mask controls connectivity to other nodes in the same system, and the partner port mask control connectivity to nodes in remote, partnered systems. By default, all ports are enabled for both local and partner connectivity.
- The port masks apply to all nodes on a system; a different port mask cannot be set on nodes in the same system. You do not have to have the same port mask on partnered systems.

What is Fibre Channel port masking

Port numbers refer to the Fibre Channel I/O port IDs that are shown by the **lsportfc** command.

A port mask is a string of zeros and ones. The last digit in the string represents port one. The previous digits represent ports two, three, and so on. If the digit for a port is "1", the port is enabled and the system attempts to send and receive traffic on that port. If it is "0", the system does not send or receive traffic on the port. If there are not sufficient digits in the string to specifically set a port number, the port is disabled for traffic.

For example, if the local port mask is set to "00001100" on a node with eight Fibre Channel ports, ports 3 and 4 are able to connect with other nodes in the system. The other Fibre Channel ports are not able to be used for node-to-node communication.

Note: The **lsfabric** CLI command shows all of the paths that are possible on the system (as defined by zoning) independent of their usage. Therefore, the command output includes paths that will not be used because of port masking.

The port masks are set by using the **chsystem** CLI command.

Port masking example

The following table shows recommended port masking for systems with 4, 8, and 12 Fibre Channel ports. The number of 0s and 1s in the port mask examples match the number of ports.

Table 46. Port masking configuration

Card / Port	4 ports	8 ports	12 ports
Card 1 Port 1	Host/Storage/Inter-node	Host/Storage	Host/Storage
Card 1 Port 2	Host/Storage/Inter-node	Host/Storage	Host/Storage
Card 1 Port 3	Host/Storage/ Replication*	Inter-node	Inter-node
Card 1 Port 4	Host/Storage/ Replication*	Inter-node	Inter-node
Card 2 Port 1		Host/Storage	Host/Storage
Card 2 Port 2		Host/Storage	Host/Storage
Card 2 Port 3		Host/Storage/ Replication*	Host/Storage/ Replication*
Card 2 Port 4		Host/Storage/ Replication*	Host/Storage/ Replication*
Card 3 Port 1			Host/Storage

Table 46. Port masking configuration (continued)			
Card / Port	**4 ports**	**8 ports**	**12 ports**
Card 3 Port 2			Host/Storage
Card 3 Port 3			Host/Storage
Card 3 Port 4			Host/Storage
localfcportmask	0011	00001100	000000001100
partnerfcportmask	1100	11000000	000011000000
* Use for host/storage when no replication is in place.			

Fibre Channel SAN configuration details

Apply the following configuration details for Fibre Channel, Fibre Channel switches and Fibre Channel Over Ethernet (FC/FCoE Gateway, FCF) switches to ensure that you have a valid configuration.

Configuring your SAN with at least two independent switches, or networks of switches, ensures a redundant fabric with no single point of failure. If one of the two SAN fabrics fails, the configuration is in a degraded mode, but is still valid. Maintain separate fabrics for FCoE and FC. If you attempt to combine these fabrics, you might risk adding more paths to the volumes. Supported configurations allow a maximum of eight paths. A SAN with only one fabric is a valid configuration but risks loss of access to data if the fabric fails. SANs with only one fabric are exposed to a single point of failure.

For Fibre Channel connections, the nodes must be connected to either SAN switches or directly connected to a host port. The system requires that a minimum of two Fibre Channel ports from each node to be connected to the SAN. This configuration provides connections to each of the counterpart SANS that are in the redundant fabric. When iSCSI hosts are attached to nodes, Ethernet switches must be used.

All back-end storage systems must always be connected to SAN switches only. Multiple connections are permitted from redundant storage systems to improve data bandwidth performance. A connection between each redundant storage system and each counterpart SAN is not required. For example, in an IBM DS4000 configuration that contains two redundant storage systems, only two storage system minihubs are usually used. Storage system A is connected to counterpart SAN A, and storage system B is connected to counterpart SAN B. Any configuration that uses a direct physical connection between the node and the storage system is not supported.

When you attach a node to a SAN fabric that contains core directors and edge switches, connect the node ports to the core directors. Then, connect the host ports to the edge switches. In this type of fabric, the next priority for connection to the core directors is the storage systems, leaving the host ports connected to the edge switches.

A SAN must follow all switch manufacturer configuration rules, which might place restrictions on the configuration. Any configuration that does not follow switch manufacturer configuration rules is not supported.

Fibre Channel switches and interswitch links

The system supports distance-extender technology, including dense wavelength division multiplexing (DWDM) and Fibre Channel over IP (FCIP) extenders, to increase the overall distance between local and remote systems. If this extender technology involves a protocol conversion, the local and remote fabrics are regarded as independent fabrics, limited to three ISL hops each.

With ISLs between nodes in the same system, the inter-switch links (ISLs) are considered a single point of failure. Figure 34 on page 323 illustrates this example.

- If Link 1 or Link 2 fails, the system communication does not fail.
- If Link 3 or Link 4 fails, the system communication does not fail.

- If ISL 1 or ISL 2 fails, the communication between Node A and Node B fails for some amount of time. The node is not recognized, even though the nodes still have a connection.

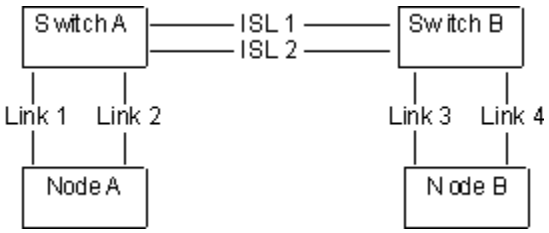


Figure 34. Fabric with ISL between nodes in a system

To ensure that a Fibre Channel link failure does not cause nodes to fail when ISLs are between nodes, you must use a redundant configuration. This configuration is illustrated in Figure 35 on page 323. With a redundant configuration, if any one of the links fails, communication on the system does not fail.

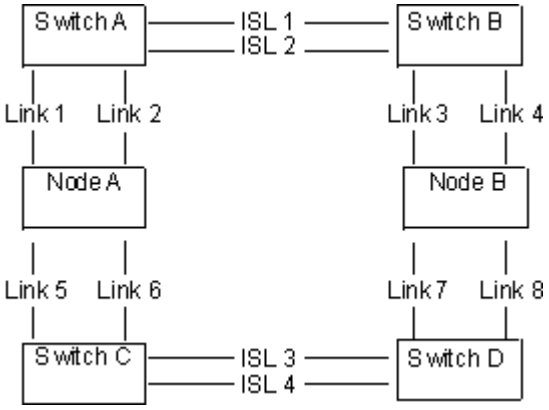


Figure 35. Fabric with ISL in a redundant configuration

Fibre Channel over Ethernet servers and system connections to the existing Fibre Channel SAN

FCoE servers and systems can be connected in several different ways. The following examples show the various supported configurations.

Figure 36 on page 324 shows a system that is connected to a Fibre Channel forwarder switch along with any FCoE hosts and FCoE storage systems. The connections are 10 Gbps Ethernet. The Fibre Channel forwarder is linked to the existing Fibre Channel SAN by using Fibre Channel ISLs. Any Fibre Channel hosts or storage systems remain on the existing Fibre Channel SAN. The connection to the system can be via the SAN (if the system is connected via Fibre Channel) or via the Fibre Channel forwarder switch to the FCoE ports on the system.

1.

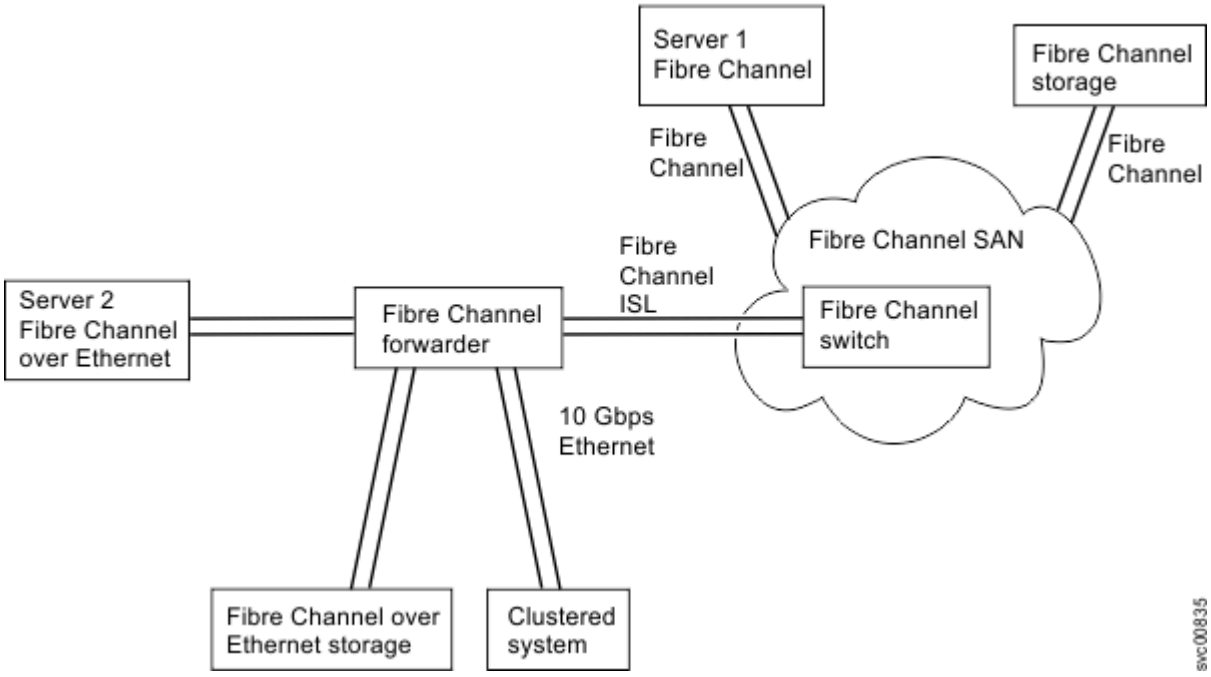


Figure 36. Fibre Channel forwarder linked to existing Fibre Channel SAN

The second example, Figure 37 on page 324, is almost the same as the first example but without an existing Fibre Channel SAN. It shows a system that is connected to a Fibre Channel forwarder switch along with any FCoE hosts and FCoE storage systems. The connections are 10 Gbps Ethernet.

2.

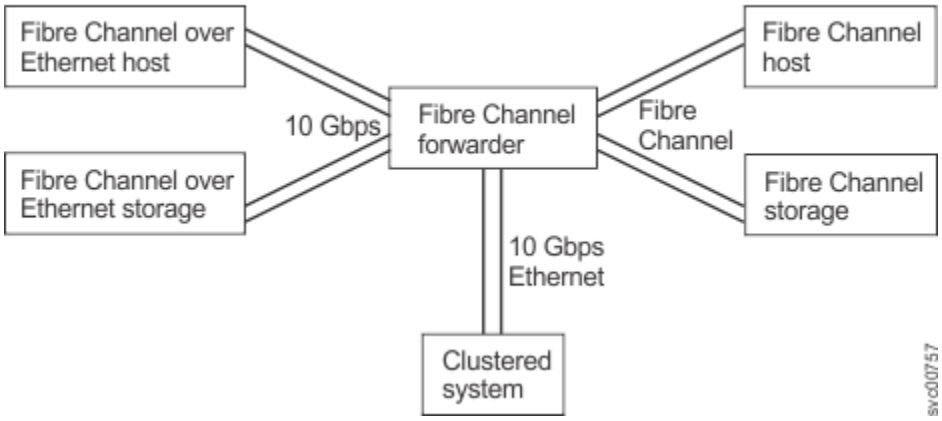


Figure 37. Fibre Channel forwarder linked to hosts and storage systems without an existing Fibre Channel SAN

In the third example, Figure 38 on page 325, a Fibre Channel host connects into the Fibre Channel ports on the Fibre Channel forwarder. The system is connected to a Fibre Channel forwarder switch along with any FCoE storage systems. The connections are 10 Gbps Ethernet. The Fibre Channel forwarder is linked to the existing Fibre Channel SAN by using Fibre Channel ISLs. Any Fibre Channel hosts or storage systems remain on the existing Fibre Channel SAN. The FCoE host connects to a 10 Gbps Ethernet switch (transit switch) that is connected to the Fibre Channel forwarder.

3.

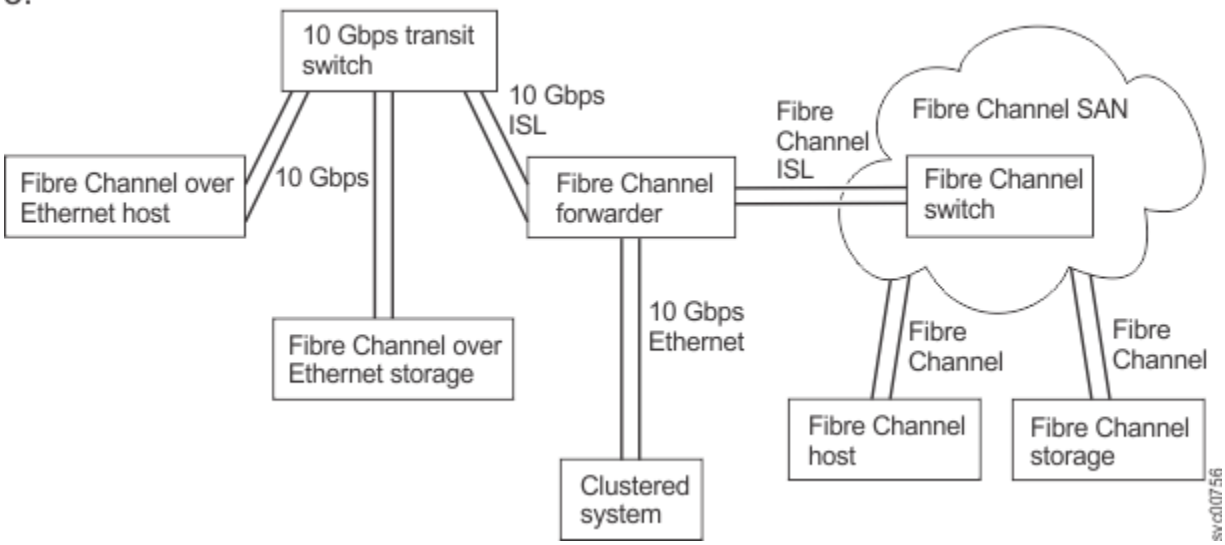


Figure 38. Fibre Channel host connects into the Fibre Channel ports on the Fibre Channel forwarder

The fourth example, Figure 39 on page 325, is about the same as the previous example but without an existing Fibre Channel SAN. The Fibre Channel hosts connect to Fibre Channel ports on the Fibre Channel forwarder.

4.

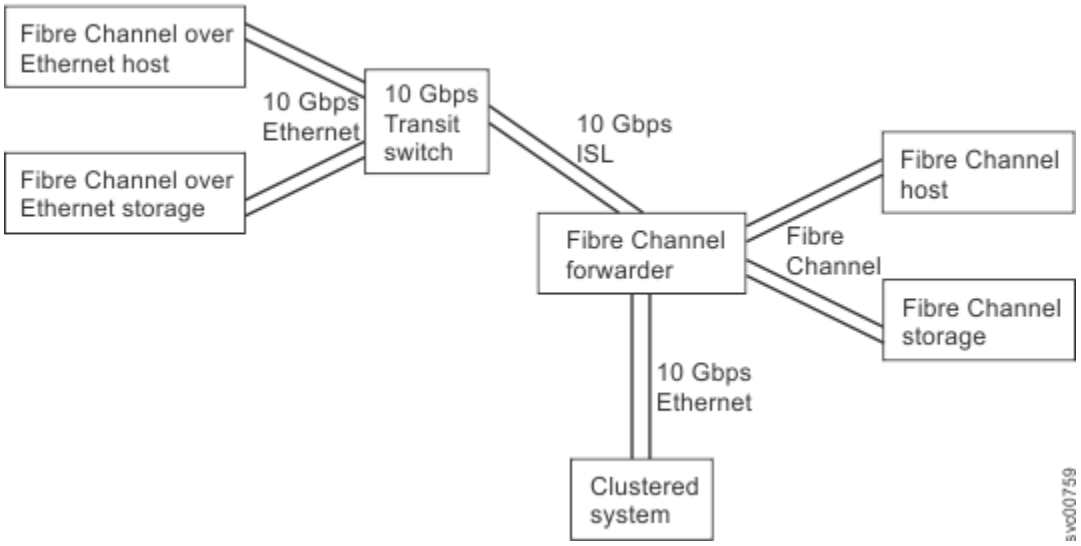


Figure 39. Fibre Channel host connects into the Fibre Channel ports on the Fibre Channel forwarder without an existing Fibre Channel SAN.

ISL oversubscription

Complete a thorough SAN design analysis to avoid ISL congestion. Do not configure the SAN to use system to system traffic or a system to storage system traffic across ISLs that are oversubscribed. For host to system traffic, do not use an ISL oversubscription ratio that is greater than 7 to 1. Congestion on the ISLs can result in severe performance degradation and I/O errors on the host.

Example SAN configurations

Examples of typical ways to configure your system and an associated Fibre Channel network are provided to guide you in setting up your SAN configuration.

SAN fabric function is available only if a Fibre Channel option is installed. Otherwise, this information does not apply.

Note: The illustrations in these examples do not show expansion enclosures as they have no relationship to SAN connectivity.

Figure 40 on page 326 illustrates a small SAN configuration. Two Fibre Channel switches are used to provide redundancy. The control enclosures and each host system are connected to both Fibre Channel switches. If you are using external storage systems with the system, connect as shown in this illustration.

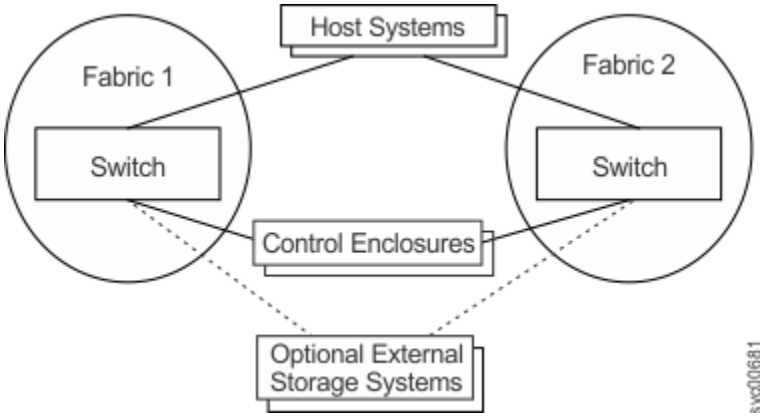


Figure 40. Simple SAN configuration

Figure 41 on page 326 illustrates a medium-sized configuration with external storage systems. The Fibre Channel SAN fabric consists of switches that are interconnected with interswitch links (ISLs). For redundancy, connect each control enclosure and external storage system to two fabrics. The example fabric attaches the control enclosures and the storage systems to the core switch. No ISL hops exist between node canister ports or between the control enclosures and the external storage systems. The example fabric attaches the control enclosure and the storage systems to the core switch. No ISL hops exist between node canister ports or between the control enclosure and the external storage systems.

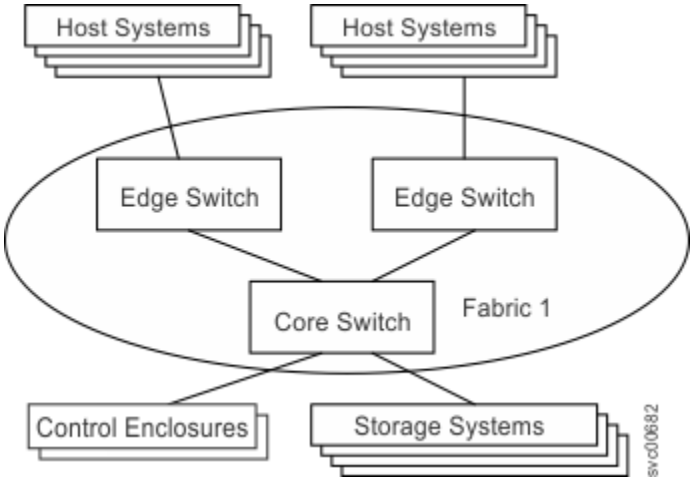


Figure 41. SAN configuration with a medium-sized fabric

Figure 42 on page 327 illustrates a large fabric that consists of two core Fibre Channel switches and edge switches that are interconnected with ISLs. For redundancy, use two fabrics with each host system, node, and storage system that is being connected. Both fabrics attach the nodes to both core fabrics and distribute the storage systems between the two core switches. This configuration ensures that no ISL hops exist between nodes or between nodes and the storage systems.

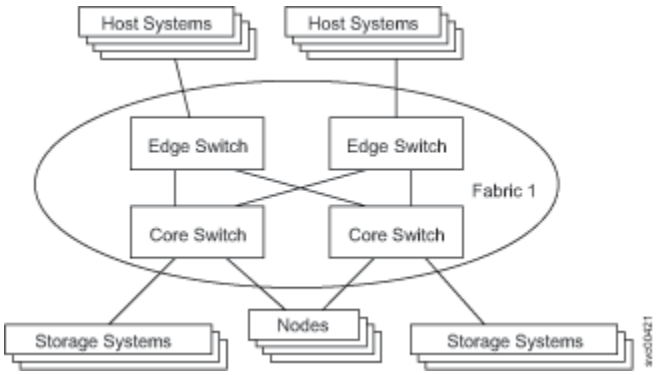


Figure 42. SAN configuration with a large fabric

Figure 43 on page 327 illustrates a fabric where the host systems are at two different sites. A long-wave optical link is used to interconnect switches at the different sites. For redundancy, use two fabrics and at least two separate long-distance links. If many host systems are at the remote site, use ISL trunking to increase the available bandwidth between the two sites.

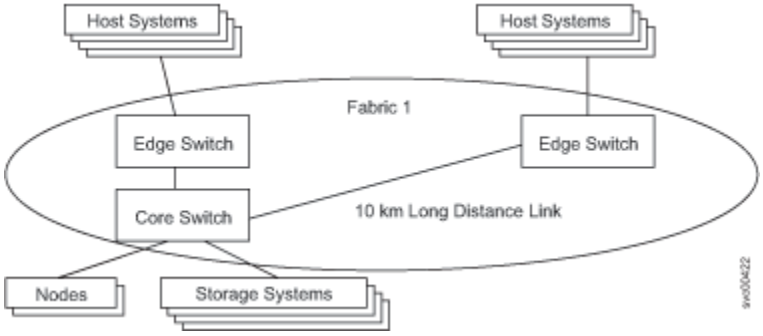


Figure 43. SAN configuration across two sites

System configuration by using SAN fabrics with long-distance fiber connections

Each system that uses SAN fabric switches can use shortwave or longwave optical fiber connections to connect to application hosts, storage systems, or remote systems.

The maximum distance between the system and host or the system and the storage system depends on the Fibre Channel speed and cable type. For all speeds and cable types, the maximum distance is 300 m for shortwave optical connections and 10 km for longwave optical connections.

When you use longwave optical fiber connections, follow these guidelines:

- For disaster recovery, each system must be regarded as a single entity, which includes the storage system that provides the quorum disks for the system. Therefore, the system and quorum disks must be co-located.
- The maximum distance between nodes within a system is 100 meters. There can be a large fiber-cable distance between the same nodes in the system. However, the nodes must be physically co-located for effective service and maintenance.

  For example, the nodes can be 300 meters away for 2 Gbs connections. For 4 Gbs connections, nodes can be 150 meters away from the SAN fabric. In both configurations, the maximum cable distance between the nodes is 600 meters. However, the nodes must be physically within 100 meters of each other.
- All nodes in a system must be on the same IP subnet so that the nodes can assume the same system or service IP address.

Determining and changing the Fibre Channel network speed

The speed of the Fibre Channel (FC) ports is auto-negotiated with the FC switch or host to which the ports are connected.

Determining the FC port speed

The ports operate at the highest speed at which a reliable connection can be established. Different ports and nodes in a system can operate at different speeds.

You can use several methods to determine the port speed of a FC port on a node.

Determining the FC port speed with the management GUI

If the node is in a system that is operational, you can determine the FC port speed with the management GUI.

To view the FC port speed in the management GUI, complete these steps:

1. Select **Monitoring** > **System**.
2. On the **System -- Overview** page, select the directional arrow next to the node that contains the FC port.
3. On the **Node Details** page, under the rear view of the node, hover over the FC port to display details about the port including the port speed. You can also select **Fibre Channel Port** to display port information for all configured ports in the **Component Details** tile.

Determining the FC port speed with the service assistant GUI

You can determine the FC port speed with the service assistant GUI.

To view the FC port speed in the service assistant GUI, complete these steps:

1. Point your web browser to `serviceaddress/service` for the node that you want to work on.
2. Log on to the service assistant using the superuser password.
3. Select the **Port** tab under **Node Details** to view the speed of the FC port.

Determining the FC port speed from the CLI

If the node is in a system that is operational, you can determine the FC port speed using the CLI commands.

To view the FC port speed of a system, issue the following command:

```
svcinfo lsportfc 0
```

The following output is an example of what you might see. The FC port speed is shown in bold.

```
id 0
fc_io_port_id 1
port_id 1
type fc
port_speed 16Gb
node_id 1
node_name node1
WWPN 500507680C212C72
```

HTTP proxy server

As part of Call Home with cloud services configuration, you can choose to use your existing HTTP proxy server within a network to manage connections between the system and the support center. A HTTP or

internal proxy server manages connections between your internal network and any entity outside of your network.

Call Home with cloud services or remote support assistance requires connections to the support center for quick and efficient problem resolution for your system. Instead of opening ports for these external connections through your firewall, you can specify an internal proxy server for better security. In addition to these features, an internal proxy server can be defined to connect to cloud-based services, such as an authentication service provider. To use an internal proxy server, a DNS server is also required to resolve hostnames to IP addresses. To create a DNS server, select **Settings** > **Network** > **DNS**.

To define an internal proxy server, use the **Settings** > **Network** > **Internal Proxy Server** panel in the management GUI.

To define an internal proxy server using the command-line interface, see **mkdnsserver** and **mkproky** commands.

Portsets

Portsets are groupings of logical addresses that are associated with the specific traffic types.

A portset restricts a host to access only a specific set of IP addresses of a node. A host can access only those IP addresses that are configured on a portset and is mapped to that host. A portset object is a system-wide object and might contains IP addresses from every I/O group. To access multiple nodes in a system, a portset must be configured with the IP address of the nodes that the host wants to access. A portset can be of **Host Attach**, **Replication**, and **Storage** type. The default portset is the **Host Attach** type. A portset of a specific type can be used only for that function, for example, a host attach type portset cannot be used for replication partnership.

Each portset is identified by a unique name. The portset 0 is an Ethernet default porset, and portset 64 is a default Fibre Channel portset configured when the system is created or updated, while portset 1 and portset 2 are replication porset and portset 3 is storage portset. A portset can be created and managed by using the command-line interface and GUI.

In the management GUI, go to **Settings** > **Network** > **Portsets** to configure Portsets.

Note: A new FC-NVMe host can be mapped to a user-defined Fibre Channel portset only. Pre-existing FC-NVMe host objects which are part of the default Fibre Channel portset must be moved from the default portset to user-defined Fibre Channel portset. For specific information about FC-NVMe enhanced scaling and system requirements, see “NVMe over Fibre Channel host attachments” on page 182.

In the command-line interface, use the following commands to configure and manage portsets:

- **lsportset**
- **mkportset**
- **rmportset**
- **chportset**

Ethernet Portsets

Planning for multiple IP address involves creating a portset, assigning the IP address to portset, and mapping portset to host object or IP partnership. *Portsets* are groupings of logical addresses that are associated with the specific traffic types.

Each physical Ethernet Port can have multiple IP addresses with each IP on a unique portset. However, for each port the IP address can be shared between multiple unique portsets for different functions. Each port can bind to only a single IP address per portset for specific Ethernet functions like host attachment (iSCSI), backend storage connectivity (iSCSI only), and IP replication. Usually, host attach IP addresses can be separated by VLANs or subnets or a combination of both for multi-tenant scenarios.

The limits for configuring IP addresses for host attachment, iSCSI storage virtualization, and IP replication are described in Configuration limits. If you are using host bus adapters, the following limits apply to multiple IP addresses and VLANs:

Table 47. Limits for Host Bus Adapters	
Type of HBA	**Limit**
25Gbps Ethernet host interface adapter	• Maximum of 32 IP addresses per port[1]
100 Gbps Ethernet host interface adapter	For iSCSI only connections: • Maximum of 64 VLANs per ports • Maximum of 64 IP addresses per ports
[1] If a VLAN is not configured on these ports, the limit is still 32 IP addresses per port. You cannot add any more IP addresses or VLANs after this limit is reached.	

Requirements for Ethernet portsets

In general, Ethernet portsets have the following requirements:

Ethernet Portsets

- Each IP address in a portset must be configured on a separate Ethernet port.
- Same ports can share IP addresses across different portsets that allow the same IP address to be used for host, storage, and replication traffic. All shared IP addresses must use the same port and have the same VLAN, gateway, and prefix. When IP addresses are shared among multiple portsets, the system creates a logical copy of the IP address and its attributes, rather than a new IP address.
- Portsets that are owned by different ownership groups can share an IP address.
- The **Port Type** field value defines whether the portset is Ethernet.
- Each port can be configured with only one unique routable IP address (gateway specified). The routable IP can be shared among multiple portsets.
- Portset 0 is a default portset that is automatically configured when the system is updated or created. Portset 0 is a host portset by default and cannot be deleted even if it is empty. Portset 0 serves as the default portset for any IP addresses and host objects that are configured without a portset specified. Portset 0 allows administrators to continue with an original configuration that does not require multi-tenancy. After an update, all configured Ethernet-based host objects are automatically mapped to portset 0.

Requirements for host portsets

In addition to portset 0, you can create more portsets for host traffic. These requirements are specific to host portsets:

- Portsets can have a maximum of 4 IP addresses per node.
- A single portset can either contain IPv4 or IPv6 or a mix of IPv4 and IPv6 addresses.
- For a host to log in to nodes on the system, the host must be mapped to a portset that contains at least one IP address from any of nodes on that system.

For more information on host portsets, see Host portsets.

Requirements for replication portsets

You can create replication portsets for IP partnership. These requirements are specific to replication portsets:

- Replication portset can have a maximum of 1 IP address per node.
- All IP addresses in replication portsets must be either IPv4 or IPv6 addresses. You cannot mix IP protocol versions on replication portsets.

- Each IP partnership can be mapped to two portsets, one for each link between systems. For a partnership with a single link, a single portset can be defined in the **portset1** field on the **Create Partnership** page in GUI. For a partnership with dual links, a second portset must be defined in the **portset2** field in GUI.
- Portsets replace the requirement for creating replication groups for IP partnerships. During updates of the software, any IP addresses that are assigned to groups with an existing IP partnership are automatically moved to a corresponding portset. For example, if group 1 is defined on the system before the update then IP addresses from that replication group are mapped to portset 1 after the update. Similarly, the IP address in group 2 is mapped to portset 2.

For more information about IP partnership requirements, see "IP partnership requirements" on page 338.

Requirements for storage portsets

The system supports a maximum of one portset of storage type, that is, portset 3. These requirements are specific to storage portsets:

- The maximum number of IP addresses for a storage portset is equal to the number of Ethernet ports on the node.
- A single portset can either contain IPv4 or IPv6 IP or a mix of IPv4 and IPv6 addresses.
- The storage type portset must be already created.

Requirements of ownership groups for portsets

You can assign portsets to ownership groups to further restrict portset to a specific set of users:

- Restricted users can create a portset and IP addresses and assign the hosts to portset.
- Restricted users cannot view portsets that are assigned to a different ownership group.
- Restricted users cannot own replication or storage type portset.

Creating a portset

You can use the management GUI or the command-line interface (CLI) to create a portset. The system supports both Fibre Channel and Ethernet portsets.

Creating portsets by using the management GUI

To create an Ethernet portset in the management GUI, complete the following steps:

1. Select **Settings** > **Network** > **Portsets**.
2. Select **Create Portset**.
3. On the **Create Portset** page, enter a name of the portset, and select one of these options for the portset type:
   - Host attachment:

     Indicates that the IP addresses that are added to the portset are used for host attachment only.

     For the **Port Type**, select **Ethernet** or **Fibre Channel**.

     or

     Select one of these options for the portset type:

     – Replication:

       Indicates that the IP addresses that are added to the portset are used for IP partnerships only.

     – High-speed replication:

       Indicates that the IP addresses that are added to the portset are used for IP (short distances by using RDMA).

4. For the **Port Type**, select **Ethernet**.
5. Select the ownership group for the portset. An ownership group defines a subset of users and objects within the system. When you define an ownership group for portsets, you can limit, and restrict users to view and manage only specific portsets.
6. Click **Create**.

After you create the portset, you can assign IP addresses to that portset.

1. Select **Settings** > **Network** > **Ethernet Ports**.
2. Right-click the port and select **Manage IP addresses**.
3. On the **Manage IP Addresses** page, select **Add IP Address**. Enter the following information for the IP address that you are adding to the selected port:
   - IP address

     Enter the IP address to associate with the selected port.
   - Type

     Select the IP protocol version of the IP address.
   - Subnet Mask or Prefix

     Enter the subnet mask for the IPv4 addresses or enter the prefix for IPv6 addresses.
   - VLAN

     Enter the corresponding VLAN tag that this IP address belongs to.
   - Portset

     Select the name or ID of the portset and ensure that the portset type matches the traffic type that is assigned to the port.
4. Click **Back** to return to the **Ethernet Ports** page. Verify that the port displays the **Configured** state. Select another port and add more IP addresses to corresponding portsets.
5. Right-click the port and select either **Modify Replication**, **Modify iSCSI hosts**, or **Modify Storage Ports**. The traffic type for the port must match the traffic for the portset that you created.

To create a Fibre Channel portset, complete the following steps:

1. Select **Settings** > **Network** > **Portsets**.
2. Select **Create Portset**.
3. On the **Create Portset** page, enter a name of the portset, and select the **Host attachment** option for the portset type. The **Host attachment** indicates that the assigned Fibre Channel I/O ports can be used only for host attachment.
4. Select **Port Type** as **Fibre Channel**.
5. Click **Create**.

After you create the Fibre Channel portset, you can assign the Fibre Channel ports to that portset.

1. Select **Settings** > **Network** > **Fibre Channel Ports**.
2. Right-click the port and select **Assign Portset**.
3. On the **Assign Portset** page, select **Portsets** that you want to map with the Fibre Channel port. Multiple selections are allowed.
4. Select **Assign Portset** to save your changes.

After you create Fibre Channel portsets and Fibre Channel ports, you can assign hosts and host clusters to the portset for those traffic types.

Creating portsets by using the command-line interface

To configure Ethernet portsets and assign IP addresses, complete these steps:

To create a Fibre Channel portset, complete the following steps in the command-line interface:

1. If you want to change any Fibre Channel IO Ports that were added automatically to default Fibre Channel portset (**portset64**), remove these Fibre Channel IO Ports from the default portset first. To remove Fibre Channel IO Ports from the default portset, enter the following command:

   ```
   rmfcportsetmember -portset 64 -fcioportid 2
   ```

2. To create a Fibre Channel portset, enter the following command:

   ```
   mkportset <name> -porttype fc
   ```

   Where **<name>** is the name of the portset.

3. To associate a Fibre Channel IO port ID with portset, enter the following command:

   ```
   addfcportsetmember -portset <portset_name | portset_id> -fcioportid <fc_io_port_id>
   ```

   where *<portset_name | portset_id>* is the name or ID of the portset and *<fc_io_port_id>* is the ID for the Fibre Channel IO port that you want to assign the portset to.

4. To verify that the mapping between the Fibre Channel IO port ID and the associated portset, enter the following command:

   ```
   lsfcportsetmember
   ```

   In the results that display, verify that the portset is assigned to the correct Fibre Channel IO port ID. In the following example, the Fibre Channel IO port ID 1 was assigned to portset4:

   ```
   id   fc_io_port_id portset_id portset_name owner_id owner_name
    4    1             4          portset4     0        ownershipgroup0
   ```

Resolving a problem with PFC settings

This procedure provides a solution for Internet Small Computer Systems Interface (iSCSI) with PFC settings.

About this task

The portsets are associated with the different functions like host attach, storage, and replication. For each portset **lossless** attribute of **lsportset** command indicates that all the IP addresses for the portset are `lossless`. The *ON* value in the `lossless` attribute of **lsportset**command indicates that PFC is enabled for that portset. Even if a single IP address is not configured as `lossless`, the portset is not a `lossless` portset. Use **lsip** to view **lossless** parameter per IP address.

Procedure

Troubleshoot problems with PFC settings.

In the **lsip** command output, the `lossless` parameter indicates whether PFC is enabled for the IP address or not. The *ON* value indicates PFC is enabled on the IP address. However, several off states indicate the problem that might prevent PFC to be enabled on the IP address:

a. If the *off-not_supported* is displayed in the `lossless`, it indicates that IP address is associated with portset of type replication. Verify the followings:

   - Check the portset for the IP address by using **lsip** command.
   - Change the portset to host attach if the IP address needs to be PFC enabled.

b. If the *off-vlan_not_configured* is displayed in the `lossless`, it indicates that IP address is associated with non-VLAN interface. Verify the followings:

   - Check the VLAN information for IP using **lsip** command.

- Add the IP address with VLAN if same needs to be PFC enabled.

c. If the *off-eth_port_not_enabled_for_pfc* is displayed in the `lossless`, it indicates that IP address is associated with the port where PFC is not configured correctly.

If **`lsip`** command output displays *eth_port_not_enabled_for_pfc* as `lossless` attribute value, then verify the followings:

a. Whether DCBx is enabled for that port:

- **`-dcbx_state`** attribute of **`lsportethernet`** command displays the Data Center Bridging Capability Exchange protocol (DCBx) state of the port. You must enable it for PFC to work. DCBx state does not support for networks with the speed lesser than 10G.

b. For the host attach type portsets and the associated IP addresses, check `pfc_enabled_host_attach` field of **`lsportethernet`** command. The field displays whether a PFC tag for host attach is enabled *on* to the connected switch port or not. It displays the followings:

- `no-link_inactive` if the link is inactive.
- `no-priority_tag_not_enabled_on_switch` if the host attach *cos* value is set by using **`chsystemethernet`** command is not enabled on switch.
- *Yes* if PFC settings are done correctly on that port.

c. For the storage type portsets and the associated IP addresses, check `pfc_enabled_storage` field of **`lsportethernet`** command that displays whether a PFC tag for storage is enabled on the connected switch port or not. The field displays the followings:

- `no-link_inactive` if the link is inactive.
- `no-priority_tag_not_enabled_on_switch` if the storage *cos* value that is set using **`chsystemethernet`** command is not enabled on switch.
- *Yes* if PFC settings are done correctly on that port.

.

Partnerships

Partnerships are used to connect systems together to enable migration, data replication, and high-availability solutions.

A system can have partnerships with up to three remote systems. The connectivity for each partnership can be either Fibre Channel or IP.

A partnership configuration requires actions on both systems involved. This ensures that there is authority to access each system and share data between them.

You can create a partnership in the following ways:

- By using Fibre Channel connectivity.
- By using “Partnerships using IP Connectivity” on page 337.

Policy-based Replication

To use partnerships for policy-based replication (asynchronous or high-availability), IP connectivity is required between management IP addresses of partnered systems.

Replication management requires access to the REST API on the remote system. Replication setup using the management GUI requires access to the management GUI of the remote system. Ensure that firewalls between the systems allow inbound traffic to port 7443 and 443 on the system management IP address.

Note: The communication on port 7443 and port 443 between the two partnered systems must occur on the default management IP address of each system.

The management traffic uses authentication certificates to prevent unauthorized access and ensure secure communications between the systems. Therefore, ensure that valid authentication certificates are installed on both the systems. For more information, see “Configuring system certificates” on page 391.

Replication between IBM Storage Virtualize systems

Systems that run IBM Storage Virtualize software are in one of two layers: the replication layer or the storage layer.

- A SAN Volume Controller system is always in the replication layer.
- A FlashSystem is in the replication layer by default, but the system can be configured to be in the storage layer instead.

To create a partnership between systems, both systems must be in the same layer. For more information, including how to change the layer, see .

Partnership states

The state of the partnership helps determine whether the partnership operates as expected. A partnership can have the following states:

Configured
Both the local and remote systems have a partnership that is defined and are running as expected.

Partial Local

For the partnership to be fully configured, you must create a partnership from the remote system to the local system.

Local Stopped
Indicates that the partnership is defined on both the systems, but the partnership is stopped on the local system.

Remote Stopped
Indicates that the partnership is the defined on both the systems, but the partnership is stopped on remote system.

Partial Local Stopped
Indicates that only the local system has the partnership that is defined and the partnership is stopped on the local system.

Local Excluded
Indicates that both the local system and the remote system have the partnership that is defined, but the local system is excluding the link to the remote system. This state usually occurs when the link between the two systems is compromised by too many errors or slow response times of the partnership.

Remote Excluded
Indicates that both the local system and the remote system are defined in a partnership, but the remote system is excluding the link to the local system. This state usually occurs when the link between the two systems is compromised by too many errors or slow response times of the partnership.

Exceeded
Indicates that the partnership is unavailable because the network of systems exceeds the number of systems that are allowed in partnerships. To resolve this error, reduce the number of systems that are in partnerships in this network.

Not Present
Indicates that the remote system is not visible. This state can be caused by a problem with the connectivity between the local and remote system or if the remote system is unavailable.

Partnerships using Fibre Channel connectivity

Partnerships can be established over Fibre Channel infrastructure.

Environment

See the following pages to understand the Fibre Channel network and port configuration needed to support Fibre Channel partnerships.

-
-
-

Creating a partnership using Fibre Channel connectivity by using the management GUI

1. Connect to the GUI on either system. Select **Settings** > **System** > **Partnerships** and select **Create Partnership**.

Creating a partnership using Fibre Channel connectivity by using the command-line interface

To create a Fibre Channel partnership between the two systems, use the following command on both systems:

```
mkfcpartnership -linkbandwidthmbits <link_bandwidth_in_mbps> -backgroundcopyrate <percentage>
<remote_system_id | remote_system_name>
```

where **mkfcpartnership** command defines a new partnership created over a Fibre Channel connection. For more information, see **mkfcpartnership command**.

Zoning for partnerships

Partnerships that are configured by using Fibre-Channel connectivity require fabric zoning.

Redundant Fabrics shows an example of a configuration that uses dual redundant fabrics that can be configured for Fibre Channel connections. Part of each fabric is at the local system and the remote system. There is no direct connection between the two fabrics.

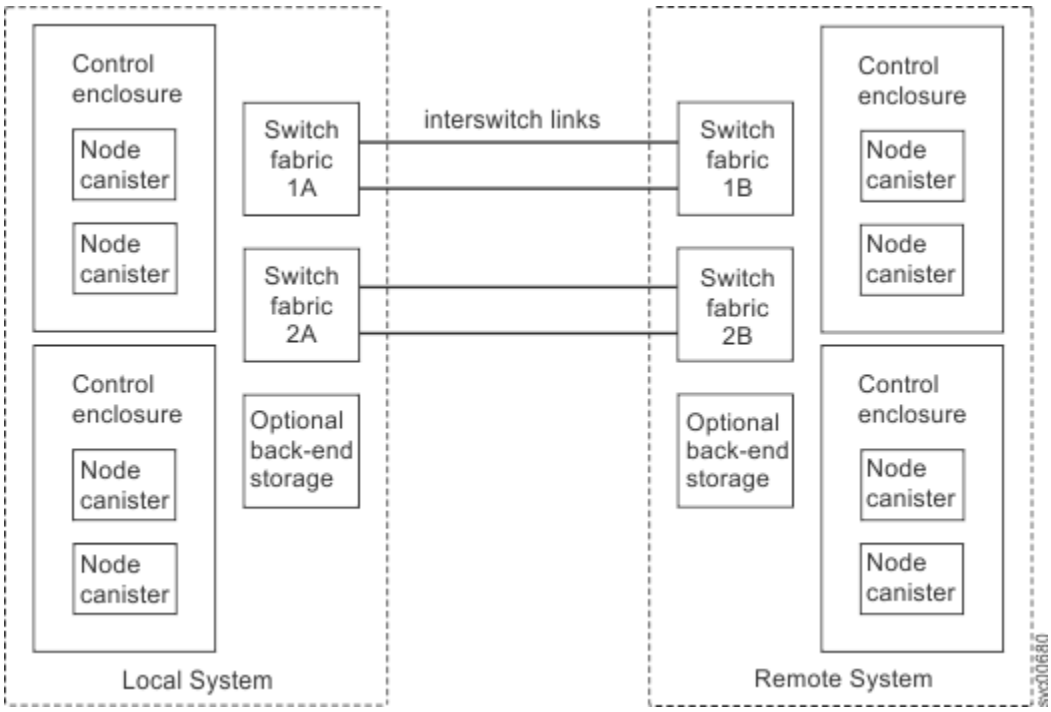


Figure 44. Redundant fabrics

ISL vs FCIP configurations

Where the Fibre Channel fabric spans two physical sites, the inter-site connectivity can be achieved either by using ISLs to connect the sites, or by using Fibre Channel over IP (FCIP) routers to connect the sites by using an IP link.

Where FCIP is used, the configuration must provide guaranteed bandwidth for private (local node to node) traffic either by using separate dedicated links or by implementing QoS.

Public or Private ports

Best practice is to ensure that cluster ports performing node to node traffic are separated from those performing inter-system replication, or host I/O. This ensures that if there is an issue with a host, or with inter-system replication, any Fibre Channel credit loss does not impact local system's node to node communication. 2-4 ports per node have visibility of nodes in the I/O group on the partnered system. Usually this is 1 port per redundant fabric.

Policy-based replication (Asynchronous or HA) zoning

Zone two ports from each node or canister in the first cluster I/O group with two ports from each canister in the second cluster I/O group. If dual-redundant fabrics are available, zone one port from each node across each fabric to provide the greatest fault tolerance. No other Fibre Channel ports on any node should have remote zones.

Host zoning

For policy-based High Availability, hosts need to be zoned so they can see the host ports from both systems.

Partnerships using IP Connectivity

Partnerships can be established over Ethernet links that use the IPv4 and IPv6 addresses associated with Ethernet ports. These IP partnerships can be connections through Ethernet switches, or direct

connections between local and partner systems. Partnerships can be created using an IPv4 address, IPv6 address, or the domain name of the remote system.

Partnership over native IP brings with it changes to some fundamental assumptions in that no name server is in an Ethernet fabric to detect and discover replication ports. There are specific configuration settings such that the local and remote systems can discover each other's replication capable IP addresses. TCP/IP sessions can then be set up across those IP addresses to establish a partnership.

Partnerships over IP Connectivity are provided as follows:

- "Short-distance partnerships using RDMA" on page 345.
- Long-distance partnerships using TCP.
- "Long-distance Ethernet partnerships using TCP" on page 342

IP partnership requirements

When you create and manage IP partnerships, consider the following requirements and constraints:

- A system can be part of up to three IP partnerships.
- You can create IP partnerships between systems that are in the same layer. In other words, both systems in the partnership must be at the storage layer or both systems must be at the replication layer.
- A secured IP partnership uses certificate-based authentication to create a secure communication channel between the partnered systems. You must ensure that the following requirements are met so that replicated data is encrypted and communication between partnered systems is secure. For more information, see "Planning secured IP partnerships" on page 341.
- You cannot use link-local addressing.
- If you use IPv4 addressing, the management IP addresses on both systems must be IPv4-compliant, and these addresses must have connectivity with each other.
- If you use IPv6 addressing, the management IP addresses on both systems must be IPv6-compliant, and these addresses must have connectivity with each other.
- You must configure all links between the production and recovery site with either IPv4 or IPv6 addresses.
- A system can have simultaneous partnerships over Fibre Channel and IP but with separate systems.
- Data compression is supported for IPv4 or IPv6 partnerships. To enable data compression, both systems in an IP partnership must be running a software level that supports IP partnership compression.
- To fully enable compression in an IP partnership, each system must enable compression. When compression is enabled on the local system, by using the **mkippartnership** or **chpartnership** commands, data sent to the remote system is compressed. To send compressed data to the local system, the remote system in the IP partnership must also enable compression.
- The use of WAN optimization devices such as Riverbed are not supported in native Ethernet IP partnership configurations.
- Bandwidth limiting on IP partnerships between both sites is supported.
- Review the appropriate hardware guide to understand which ports can be used for IP partnerships.
- Using an Ethernet switch to convert a 25 Gbps to a 1 Gbps IP partnership, or a 10 Gbps to a 1 Gbps partnership is not supported.
- The IP infrastructure on both partnership sites must match.
- When an IP partnership is configured between two systems with one system on an older build and system management IP is configured on the port other than the default port in a newer build. The **lssystemip** command on the older build shows stale port_id value as 1 and 2 for remote system. Until the other system is upgraded with latest build, it continues to show stale port_ids. To get actual port_id, **lsip** can be used in the remote system.

Port configuration requirements

- Specific ports are used by systems for IP partnerships (long-distance using TCP and short-distance using RDMA) communications. See the table TCP and UDP ports that are supported in section *TCP and UDP ports* for more information.
- The maximum supported round-trip time between systems is 80 milliseconds (ms).
- *Portsets* are groupings of logical addresses that are associated with the specific traffic types.
- Each IP partnership can be mapped to two portsets, one for each link between systems. For a partnership with a single link, a single portset can be defined with the **-link1** attribute in the **mkippartnership** command for partnerships with a single intersite link. For a partnership with dual links, a second portset must be defined with the **-link2** attribute to specify the second portset for a dual link configuration. You can also use the management GUI to specify portset for each intersite link.
- If your system supports Ethernet ports of several speeds, **only one** of those speeds can be used to configure all of the replication links between the local and remote site.

  Note: You can use 25 Gbps Ethernet ports to establish an IP partnership between systems. However, 25 Gbps Ethernet ports do not provide an additional performance advantage over IP partnerships that are established using 10 Gbps Ethernet ports.
- iSCSI hosts can access volumes over IP ports that are participating in an IP partnership; however, this access might result in an impact on performance.
- VLAN tagging of the IP addresses configured for replication is supported.
- IP partnerships do not support Network Address Translation (NAT) traversal. Most NAT implementations change the IP address and ports of the IP packets when the packets traverse the NAT or a network firewall. IP replication does not work with transport protocols that change the IP header source and destination IP addresses and their associated port numbers.
- If you configure two intersite WAN links, you must assign each WAN link to separate portset, one for each link.
- If you have one intersite link, you must configure one replication port group for that link.
- No more than two intersite links are supported.
- Only one port in a node can be configured in an IP partnership.
- If you connect systems by directly attaching them without switches, you must have only two direct-attach links. Both direct-attach links must be on the same I/O group. You should use two port groups, where a port group contains only the two ports that are directly linked.

If you choose to use compression and IP partnership in the same system, certain hardware updates or configuration choices might help increase IP partnership performance:

- Use a different port for IP partnership traffic, which is not used for iSCSI host, iSCSI virtualization, iSCSI backend, and RDMA clustering. Also, use a different VLAN ID for iSCSI host I/O and IP partnership traffic.

Requirements for short-distance partnerships using RDMA

When you create and manage short-distance partnerships using RDMA, consider the following requirements and constraints:

- Up to two short-distance partnership using RDMA can be configured per system.
- You can create short-distance partnership using RDMA between systems that are in the replication layer. In other words, systems in the short-distance partnership using RDMA must be at the storage layer or both systems must be at the replication layer.
- The system supports a maximum of 6 highspeedreplication portsets.
- You cannot use link-local addressing.
- Configure all links between the production and recovery site with either IPv4 or IPv6 addresses.

- A system can have simultaneous partnerships over native-IP and short-distance partnership using RDMA but with separate remote systems.
- Systems that are configured in active short-distance partnership using RDMA must not be zoned with each other for Fibre Channel.
- VLAN tagging of the IP addresses configured for highspeedreplication portsets is supported.
- Short-distance partnership using RDMA do not support Network Address Translation (NAT) traversal. Most NAT implementations change the IP address and ports of the IP packets when the packets traverse the NAT or a network firewall. IP replication does not work with transport protocols that change the IP header source and destination IP addresses and their associated port numbers.
- If you have one intersite link, you must assign one highspeedreplication portset for that link that is, all ports that belong to the portset must be connected by using that link.
- If you configure two inter-site links, you can configure two highspeedreplication portset and assign one portset per link.
- No more than two intersite links are supported.
- If you have one highspeedreplication portset, then configure maximum of two ports from each node in one I/O group in that highspeedreplication portset.
- If you have two highspeedreplication portsets and one I/O group, then on each system, configure maximum of two ports from one node in the first highspeedreplication portset. Then, configure maximum of two ports from the other node in the second highspeedreplication portset.
- If you connect systems by directly attaching them without switches, you must have only two direct-attach links. Both direct-attach links must be on the same I/O group. Use two port groups, where a port group contains only the two ports that are directly linked.

Related tasks

“Planning secured IP partnerships” on page 341
A secured IP partnership uses certificate-based authentication to create a secure communication channel between the partnered systems. You must ensure that the following requirements are met so that replicated data is encrypted and communication between partnered systems is secure.

Configuring VLAN for IP partnerships

To configure VLAN when you use IP (Internet Protocol) partnerships, consider the following requirements and procedures.

Before you begin

- VLAN tagging is supported for IP partnership traffic between systems.
- VLAN provides network traffic separation at the layer 2 level for Ethernet transport.
- VLAN tagging by default is disabled for any IP address of a node port. You can use the management GUI or the command-line interface (CLI) to optionally set the VLAN ID for port IPs on systems in the IP partnership.
- When a VLAN ID is configured for the port IP addresses that are mapped to the portsets, appropriate VLAN settings on the Ethernet network must also be properly configured to prevent connectivity issues.
- Setting VLAN tags for a port is disruptive. Therefore, VLAN tagging requires that you stop the partnership first before you configure VLAN tags. Then, restart again when the configuration is complete.

About this task

Follow this procedure to configure VLAN tags for existing IP partnership setups:

Procedure

1. Stop the partnership between the local and remote system.

2. Configure VLAN on node ports in the portsets on the local system.
3. Configure all intervening switches with appropriate VLAN tags.
4. Configure VLAN on node ports in the portsets on the remote system.
5. Check to see whether connectivity between the local and remote sites are restored.
6. Restart the partnership.

Planning secured IP partnerships

A secured IP partnership uses certificate-based authentication to create a secure communication channel between the partnered systems. You must ensure that the following requirements are met so that replicated data is encrypted and communication between partnered systems is secure.

Before you begin

To enable a secured IP partnerships, install an encryption license on both of the partnered systems. See "Licensing encryption" on page 454 for more information, or contact support for assistance.

You must plan what type of system certificate the partnered systems will use. The system certificate is used for all functions on the system that use certificate authentication. A system certificate can be an internally signed certificate or an externally signed certificate. An internally signed certificate is a certificate that has been issued by the system's root certificate authority. If the system is currently using a self-signed certificate, the self-signed certificate can be used until it expires. After it expires, you must use either an internally signed certificate or an externally signed certificate. Externally signed certificates are issued and signed by a third-party CA. A signing request must be generated on the system and presented to the third-party CA, which signs the request and returns a signed certificate that can be installed. For more information about managing system certificates, see "Configuring system certificates" on page 391.

Mutual authentication

A secured IP partnership between the partnered systems requires the systems to be mutually authenticated. Mutual authentication requires that the system certificate or certificate authority certificate for each local system is installed in a truststore on the partner system. The local system presents its certificate and chain of signing CA certificates over the network when the secured partnership is established. The partner system authenticates the certificate presented by the local system by using the certificates and authorities that are installed in its truststore.

Setting up mutual authentication

To set up mutual authentication, you must install certificates on the partnered systems based on the type of certificate you choose. The system supports both internally-signed certificates and externally signed certificates.

Internally signed certificates

- Create an internally signed certificate with either the management GUI or use the **svctask chsystemcert -mksystemsigned** command. The certificate is installed on the system automatically. To configure an internally-signed certificate, see "Updating an internally signed certificate" on page 391.
- Export the system's root certificate to the partnered system's truststore.

Externally signed certificates

- – Generate a certificate signing request (CSR) with the management GUI or use the **svctask chsystemcert -mkrequest** command.
  – Send the CSR that is generated to your external CA.
  – When the third-party CA returns the signed certificate, install the certificate and the intermediate CA certificates on the system.
  – Install the third-party CA's root certificate in the partnered system's truststore. To configure an externally-signed certificate, see "Requesting and installing an externally signed certificate" on page 394.

The certificates are added to a system-wide truststore on the partnered system and managed with a set of CLI commands. For more information, see **Truststore management commands**.

When using an externally signed certificate, the certificate can be signed by the third-party root CA or by an intermediate CA which itself is signed by the third-party root CA. If an intermediate CA is used then there is a certificate chain, and the complete chain of trust must be established in order for one system to authenticate the other. The complete chain of trust can be established in the following ways:

- Only the signed system certificate is installed on the local system. The local system presents only the endpoint certificate over the network and the intermediate CA certificates and root CA certificate must be installed in the truststore on the partner system.
- The signed system certificate along with the intermediate CA certificates (the root CA certificate can optionally be included but is not required) are installed on the local system. The local system presents the endpoint certificate and the intermediate CA certificates over the network, and only the root CA certificate must be installed in the truststore on the partner system.

It is recommended that the full certificate chain is installed on the local system. If the full chain is not installed then errors in other functions that use certificate authentication such as the management GUI can occur.

Best practices for secured IP partnerships:

- In a secured IP partnership, the partnered systems mutually authenticate with each other using certificates. These certificates have a validity period for which they are valid, and expire after this time. The certificate for each system can be configured separately, and may have a different validity period to the partner system's certificate. Therefore, it is possible for one system's certificate to expire while the partner system's certificate is still valid. In this scenario, the certificate authentication for the IP partnership will fail. It is the best practice to:
  - Ensure the certificates of each system in the partnership are configured with similar validity periods to avoid disruption in the IP partnership.
  - Synchronize the date and time of each system.
- User gets better encryption efficiency and better replication throughput in secured IP partnerships with higher MTU size. However, larger MTU sizes might not be possible outside controlled IP networks. Users need to plan this carefully with network administrators.
- If MTU sizes configured at endpoints of a secured or unsecured IP partnership do not match, the partnership may remain in *not_present* state. Hence it is recommended to configure same MTU size at all endpoints such as IBM Storwize devices and network devices that are involved for establishing an IP partnership.

Related concepts

"Intersite link planning" on page 344
Specific intersite link requirements must be met when you are planning to use IP partnership for replication.

Related reference

"IP partnership requirements" on page 338
When you create and manage IP partnerships, consider the following requirements and constraints:

Long-distance Ethernet partnerships using TCP

Two types of IP partnerships can be configured between systems:

1. IP (long distances by using TCP)
2. IP (short distances by using RDMA)

Secured IP partnerships secure the data as it travels through an untrusted network between production and recovery systems. Although securing IP partnerships is optional, it supports authentication of the production and recovery systems, and verifies the confidentiality and integrity of the replicated data.

Secured IP partnerships minimize the risk of hackers manipulating or intercepting data in untrusted networks. Secured IP partnerships use Internet Protocol Security (IPsec) suite of protocols that covers important security aspects such as:

- Enhanced mutual authentication
- Stronger encryption algorithms
- Encryption key management mechanisms

In secured IP partnerships, the partner systems authenticate to each other, negotiate the security parameters, exchange encryption keys, and establish secured network tunnels through which encrypted data travels. Partner systems are authenticated by certificates issued by internal root certificate authorities (CA) or trusted third-party's root CA or intermediate CA. Secured IP partnerships are created when the necessary certificates and authorities are installed on the partner systems.

To enable secured partnerships, you must purchase an encryption license and activate the license on both partnered systems. For more information, see "Licensing encryption" on page 454.

Portsets replace the requirement for creating groups for IP partnerships. Dedicated portsets can be created for replication traffic. The dedicated portsets provide group of IP addresses for IP Partnerships. Each node can have one IP address that is assigned to a portset for traffic. If the local system in the IP partnership contains four nodes, a portset can be created that defines four IP addresses, one per each node. Similarly, the remote system with four nodes, a portset on that system can also have four IP addresses to handle traffic exclusively. Before you can configure a new IP partnership, you need to define a portset and assign IP addresses to nodes.

You can configure portsets so that each IP partnership can be mapped to two portsets, one for each WAN link between systems. For network configurations that have a single link between systems in an IP partnership, a single portset can be defined in the **Portset Link 1** field on the **Create Partnership** page from GUI. For a partnership with dual links, a second portset must be defined in the **Portset Link 2** field.

Supported IP partnership configurations

The following general configurations are supported.

Configuration 1: In this configuration, only a single WAN intersite link is available. Therefore, only one portset is configured on each node.

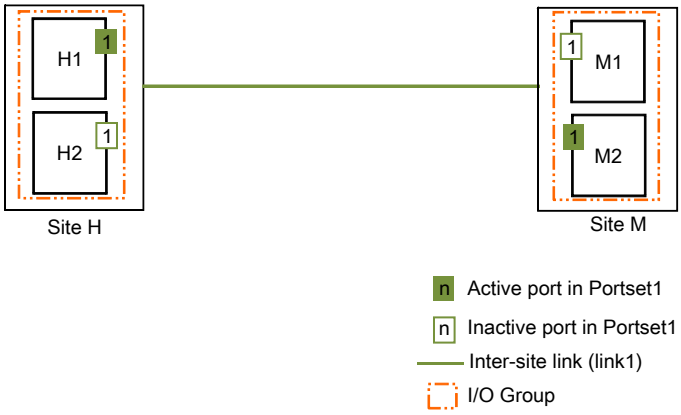


Figure 45. One intersite link, one I/O group per system

Only one port from either of the nodes in each system actively participates in the IP partnership. The other port acts as the failover port. If a critical failure is observed on node H1 in Site H, the IP partnership will fail over to node H2 and continue. Replication might stop momentarily during the failover.

Note: The system supports two IP partnerships that use the same physical link between systems however the single link can have lower throughput than a single IP partnership.

Configuration 2: In this configuration, two intersite links are available; therefore, two portsets are configured. One port from each node in each system actively participates in the IP partnership. If a

critical failure is observed on node H1 in Site H, the IP partnership continues over the other port on node H2.

No failure occurs in this scenario; however, the effective bandwidth is reduced to half; only one of the two links is available to facilitate IP partnership traffic. When the failure is corrected, ports will fail back and the IP partnership continues to operate over both links.

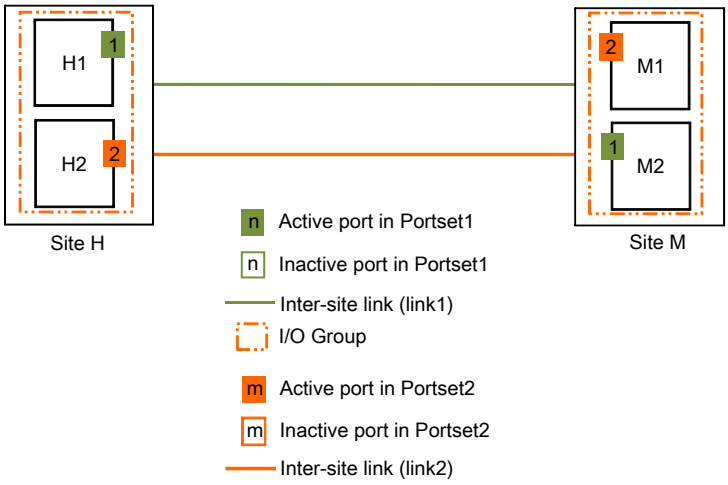


Figure 46. Two intersite links, one I/O group per system

Intersite link planning

Specific intersite link requirements must be met when you are planning to use IP partnership for replication.

If you use IP partnership, you must meet the following requirements:

- IP partnerships can also support data compression. Both systems in the IP partnership must be running a software level that supports IP partnership data compression. Compression must also be functional on each system. For more information, see “Long-distance Ethernet partnerships using TCP” on page 342.
- The amount of intersite heartbeat traffic is 1 megabit per second (Mbps) per link.
- The minimum bandwidth requirement for the intersite link is 10 Mbps.

IBM Storage Virtualize maintains near line-speed performance by masking the latency of the link up to a point. Even as the line latency increases, the performance of the technology enables the throughput to continue to exceed that of a plain link. The bandwidth limit defined on the partnership is the uncompressed data rate, which is measured as the remote node send data rates.

1 Gbps ports maintain line-rate until approximately 20 ms of RTT and 10 Gbps ports start to reduce the throughput after the RTT exceeds 5 ms.

IP partnership performance best practices

A number of factors affect the performance of an IP partnership. Some of these factors are latency, link speed, number of intersite links, host I/O, MDisk latency, and hardware.

The following actions might help improve this performance. See “IP partnership requirements” on page 338 for additional considerations.

To avoid losing synchronization during node failover, use policy-based replication with change volumes. These modes of replication provide consistency protection and restarts replication automatically.

Short-distance partnerships using RDMA

Short-distance partnerships using RDMA allows data replication with high bandwidth over a short distance. Therefore, enterprises can use their existing Ethernet infrastructure to fully adopt an end-to-end Ethernet-based solution.

Data replication is important for the integrity, availability, and disaster recovery of critical business information. High-performance computing environments and ultra-fast network infrastructures need replication solutions that match technological advancements. Short-distance partnership using RDMA offers several advantages:

Increased Throughput
Short-distance partnership using RDMA is suited for throughput-hungry applications, offering compatibility with link bandwidths that surpass 10 Gbps. Experience data replication at unprecedented speeds.

Optimized for Short Distances
For enterprises that require disaster recovery (DR) solutions within a 100-kilometer range, short-distance partnership using RDMA provides dependable links, minimizing frame loss and maximizing data integrity.

Low Round-Trip Time
Short-distance partnership using RDMA is ideal in scenarios up to 1 ms Round-Trip Time (RTT). Replicate data seamlessly across racks, within data centers, or throughout city-based campuses, while the highest standards of performance are maintained.

Comprehensive Ethernet-Based Solution
Suitable for enterprises who want Ethernet-only environment such as Ethernet data centers deployments for Business Continuity and Resiliency.

Note: SVC nodes would still require FC clustering to form a cluster.

Total Cost of Ownership (TCO)
Short-distance partnership using RDMA is cheaper than traditional FC partnerships because expensive FCIP routers are not required for this solution.

Some of the major use cases that this solution enables include:

- *3-Site replication.* All three sites can operate only in the Ethernet environment.
- *Setting up DR Sites in sync and async environments.* This solution enables enterprises to set up multiple DR sites, where replication partnerships over IP networks maintain data volume copies, both in a synchronous and asynchronous environment.

Planning for short-distance partnerships using RDMA

Planning for short-distance partnership using RDMA involves defining the partnerships between two systems and associating portsets.

Note: Short-distance partnership using RDMA is only supported on RDMA ports. In addition, only iWARP ports are allowed by default. To configure short-distance partnership by using RoCE adapters, contact IBM Support.

For short-distance partnerships that uses RDMA partnerships, you must specify highspeedreplication portsets. The following are the requirements for highspeedreplication portset:

- Node IP addresses in Service GUI must not be preconfigured in the system.
- Both link1 and link2 portsets must be of type *highspeedreplication* portsets.
- All ports in a *highspeedreplication* portset must be of an identical type; that is either both must be iWARP or RoCE.
- All the IP addresses within a *highspeedreplication* type portset must be of the identical type that is, either IPv4 or IPv6.

The following are the system limits for short-distance partnerships by using RDMA:

Table 48. System limits	
Limit	**Description**
Max IP addresses per portset per node	2
Max *highspeedreplication* portset per system	6
Max portset per link	1
Max links per partnership	2

Note: When a node is added to a system configured for a short-distance partnership using RDMA, any IP addresses configured on the node are removed.

Creating IP partnership

After creating portset and assigning IP addresses to the portset, you have to create IP partnership for replication.

Using the management GUI

To create IP partnership using the management GUI, complete these steps:

1. Select **Settings** > **System** > **Partnerships** and select **Create Partnership**.
2. On the **Create Partnership** page, select either IP (long distances using TCP) or IP (short distances using RDMA).
3. Enter the following information to configure the partnership:

   Partner IP/FQDN
   Select the IP address or the fully qualified domain name (FQDN) of the partner system. If you specify a fully qualified domain name, a DNS server must be configured on your system. To configure a DNS server for the system, select **Settings** > **Network** > **DNS**.

   Secured IP partnerships

   Enter the following information to obtain the certificate:

   a. Select **Secured IP partnerships** to secure the connection between systems.
   b. Click **Test Connection** to retrieve and validate the authorized-signed certificate of the remote system.
   c. Select **View certificate** to display the certificate from the remote system. Verify that the certificate is signed by a trusted certificate authority, and that the subject and issuer names are the expected values.
   d. If the retrieved certificate is signed by an unrecognized authority, select **Upload File** to upload the root certificate or certificate chain from the certificate authority that signed the partner system's certificate.

   Secured IP partnerships support Both internally signed certificates and externally signed certificates.

   Link Bandwidth
   Enter the amount of bandwidth used for replication between systems in the partnership.

   Partner System's CHAP Secret (Optional):
   Specify the CHAP secret for the partner system if you plan to use Challenge Handshake Authentication Protocol (CHAP) to authenticate connections between the systems in the partnership.

Compression enabled
Select this option if data on the local system is compressed before it is sent to the partner system. To fully enable compression in an IP partnership, each system must support compression.

Portset Link 1
In single link configurations between partnered systems, select the portset that provides IP addresses for replication traffic.

Portset Link 2 (optional)
In dual link configurations between partnered systems, select the portset that provides IP addresses for replication traffic. Dual links provide redundancy for connections between systems.

Using the CLI

For creating secured IP partnerships, the authorities and certificate installation is a prerequisite.

You can export the certificate from the remote system and import it on the local system in order to allow configuration changes to be coordinated across systems without manual authentication. Use the **chsystemcert** command to create and manage certificates on the system.

If the certificate is signed by the system's root certificate authority, or it is a self-signed certificate, then use the following command to export the entire certificate chain on the remote system. If the certificate is signed by a trusted third-party certificate authority, then the root certificate must be retrieved from the third-party certificate authority.

```
chsystemcert –export
```

The certificate is exported to the `/dumps/certificate.pem` directory on the configuration node.

After you complete all these steps, run **lspartnership** command on each system in the configuration and verify the following items:

- Portsets are configured for all links between the systems.
- All systems are in the Fully Configured state.

For more information, see **lspartnership** command.

Creating portset

You can create portset for IP partnership by using both the management GUI and the CLI.

Using the management GUI

To create portset using the management GUI, complete these steps:

1. Select **Settings** > **Network** > **Portsets**.
2. Select **Create Portset**.
3. On the **Create Portset** page, enter a name of the portset, and select **Host Attachment**, **Replication**, or **High speed replication** for the portset type.
4. **(Optional)** Select the ownership group for the portset. An ownership group defines a subset of users and objects within the system. When you define an ownership group for portsets, you can limit, and restrict users to view and manage only specific portsets.
5. Click **Create**.

Using the CLI

To create portset using the CLI, enter the following command:

```
mkportset -name portset_name -type portset_type -ownershipgroup owner_name
```

where *portset_name* is the name of the portset and *portset_type* is **replication**. The value *owner_name* indicates the name of the ownership group to which the portset belongs. It is an optional value.

Assigning IP addresses to the portset

After creation of the portset for the traffic type, you must assign the IP addresses to portset. This is required to assign the portset to the new IP partnership.

Using the management GUI

To assign IP addresses to the portset using the management GUI, complete these steps:

1. Select **Settings** > **Network** > **Ethernet Ports**.
2. Right-click the port and select **Manage IP addresses**.
3. On the **Manage IP Addresses** page, select **Add IP Address**. Enter the following information for the IP address that you are adding to the selected port:
   - IP address

     Enter the IP address to associate with the selected port.
   - Type

     Select the IP protocol version of the IP address.
   - Subnet Mask or Prefix

     Enter the subnet mask for the IPv4 addresses or enter the prefix for IPv6 addresses.
   - VLAN

     Enter the corresponding VLAN tag that this IP address belongs to.
   - Portset

     Select the name or ID of the portset and ensure that the portset type matches the traffic type that is assigned to the port.
4. Click **Back** to return to the **Ethernet Ports** page. Verify that the port displays the **Configured** state. Select another port and add more IP addresses to corresponding portsets.
5. Right-click the port and select either **Modify Replication**, **Modify iSCSI hosts**, or **Modify Storage Ports**. The traffic type for the port must match the traffic for the portset that you created.

Using the CLI

To assign IP addresses to the portset using the CLI, enter the following command:

```
mkip -node <id | name> -port <id> -portset <id | name> -ip <ip address> -prefix <subnet
prefix> -gw <gateway address> -vlan <vlan_id>
```

where *id* | *name* is the name or ID of the node; *id* is the port identifier; and *id* | *name* indicates the name or ID of the portset that was created earlier. Enter either a valid IP address for the **-ip** and include the values for the address. This address that is assigned to the portset and more addresses can be added to portset with the **mkip** command.

Configuring VLAN for IP partnerships

To configure VLAN when you use IP (Internet Protocol) partnerships, consider the following requirements and procedures.

- VLAN tagging is supported for IP partnership traffic between systems.
- VLAN provides network traffic separation at the layer 2 level for Ethernet transport.
- VLAN tagging by default is disabled for any IP address of a node port. You can use the management GUI or the command-line interface (CLI) to optionally set the VLAN ID for port IPs on systems in the IP partnership.

- When a VLAN ID is configured for the port IP addresses that are mapped to the portsets, appropriate VLAN settings on the Ethernet network must also be properly configured to prevent connectivity issues.
- Setting VLAN tags for a port is disruptive. Therefore, VLAN tagging requires that you stop the partnership first before you configure VLAN tags. Then, restart again when the configuration is complete.

Follow this procedure to configure VLAN tags for existing IP partnership setups:

1. Stop the partnership between the local and remote system.
2. Configure VLAN on node ports in the portsets on the local system.
3. Configure all intervening switches with appropriate VLAN tags.
4. Configure VLAN on node ports in the portsets on the remote system.
5. Check to see whether connectivity between the local and remote sites are restored.
6. Restart the partnership.

Migrating partnerships

It is possible to migrate existing partnerships from IPv4 to IPv6 deployments, and from one Ethernet speed link to a different Ethernet speed link. It is not possible to convert a partnership between Fibre Channel, IP or RDMA.

Migrate an existing partnership from IPv4 to IPv6 deployments

To migrate from IPv4 to IPv6, the following requirements must be met:

- System IPs have IPv6 addresses configured.
- Data path IPs (IPs configured by using **mkip**) have IPv6 addresses.

  Note: You can assign IPv6 addresses to ports while IP partnerships are active. However, you cannot add them to portsets.

Complete the following procedure:

1. Stop the IP partnership on both systems.
   a. On system C1: #svctask **chpartnership -stop** <**systemid C2**>
      - The command **lspartnership** reports as `fully_configured_stopped` on system C1.
      - The command **lspartnership** reports as `fully_configured_remote_stopped` on system C2.
   b. On system C2: #svctask **chpartnership -stop** <**systemid C1**>
      - The command **lspartnership** reports as `fully_configured_stopped` on system C1.
      - The command **lspartnership** also reports as `fully_configured_stopped` on system C2.
2. Add the IPv6 IP addresses, which are configured on data path IP Ports (by using **mkip**), in respective portsets.
   - On system C1: **mkip -node** <**id | name**> **-port** <**id**> **-portset** <**id | name**> **-ip** <**ipv6 address**> **-prefix** <**subnet prefix**> **-gw** <**ipv6 gateway address**> **-vlan** <**vlan_id**>
   - On system C2: **mkip -node** <**id | name**> **-port** <**id**> **-portset** <**id | name**> **-ip** <**ipv6 address**>**-prefix** <**subnet prefix**> **-gw** <**ipv6 gateway address**> **-vlan** <**vlan_id**>
3. Modify the IP partnerships to do discovery over IPv6 addresses.
   - On system C1: #svctask **chpartnership -type ipv6 -clusterip** <**ipv6_ipaddr_of_cluster_c2**> <**systemid C2**>
   - On system C2: #svctask **chpartnership -type ipv6 -clusterip** <**ipv6_ipaddr_of_cluster_c1**> <**systemid C1**>
   - On system C1: #svctask **chpartnership -start** <**systemid C2**>

- On system C2: #svctask **chpartnership -start** <**systemid C1**>

  Note: Since new data path over IPv6 addresses are now available, partnership will first change to `not_present`, and then to `fully_configured`. If it remains in `not_present`, monitor the node error/DMP if it is getting triggered and check the appropriate DMP in the troubleshooting section.

This same procedure can be applied to migrate from IPv6 to IPv4 by applying suitable substitutes for IPv4 instead of IPv6.

Migrate an existing partnership to new IP addresses

When you change one or both system and data path IP addresses of the system, replication needs to be temporarily stopped during the procedure. However, host I/O can continue. After the procedure is complete, the v need to resynchronize only the host write operations that are submitted while the partnership was stopped.

Complete the following procedure:

1. Stop the IP partnership on both systems.
   a. On system C1: #svctask **chpartnership -stop** <**systemid C2**>
      - The command **lspartnership** reports as `fully_configured_stopped` on system C1.
      - The command **lspartnership** reports as `fully_configured_remote_stopped` on system C2.
   b. On system C2: #svctask **chpartnership -stop** <**systemid C1**>
      - The command **lspartnership** reports as `fully_configured_stopped` on system C1.
      - The command **lspartnership** also reports as `fully_configured_stopped` on system C2.
2. Reconfigure the IP addresses on one or both systems by using the following commands:

   chsystemip -clusterip <**new ip**> **-port** <**ethernet port number**>

   mkip -node <**id | name**> **-port** <**id**> **-portset** <**id | name**> **-ip** <**ipv4 | ipv6 address**> **-prefix** <**subnet prefix**> **-gw** <**ipv4 | ipv6 gateway address**> **-vlan** <**vlan_id**>
3. For each reconfigured system IP, reconfigure the partnership on the remote system:

   If system IP changed on system C1, then on system C2: #svctask **chpartnership -clusterip** <**system IP of system C1**> <**systemid C1**>

   If system IP changed on system C2, then on system C1: #svctask **chpartnership -clusterip** <**system IP of system C2**> <**systemid C2**>
4. Restart the partnerships:

   On system C1: #svctask **chpartnership -start** <**systemid C2**>

   On system C2: #svctask **chpartnership -start** <**systemid C1**>
5. Restart any stopped relationships.

Migrate an existing partnership to different links

Before you attempt to migrate existing deployments from to a different link, refer to the limitations and considerations requirements in the IP partnership configuration topic.

Note: You cannot mix link speeds. For example, if there are two links, both links should either be 10 Gbps links or both should be 1 Gbps links.

This procedure applies to two systems that are in IP partnership with each other over two 1 Gbps links.

Complete the following procedure:

1. Stop the IP partnership on both systems.
   a. On system C1: #svctask **chpartnership -stop** <**systemid C2**>

- The command **lspartnership** reports as `fully_configured_stopped` on system C1.
- The command **lspartnership** reports as `fully_configured_remote_stopped` on system C2.

b. On system C2: #svctask **chpartnership -stop** <**systemid C1**>

- The command **lspartnership** reports as `fully_configured_stopped` on system C1.
- The command **lspartnership** also reports as `fully_configured_stopped` on system C2.

2. Add the IP addresses, which are configured on data path IP Ports (by using **mkip**), on 10 Gbps links in respective portsets.
   - On system C1: #svctask **mkip -node** <**id | name**> **-port** <**id**> **-portset** <**id | name**> **-ip** <**ipv4 | ipv6 address**>**-prefix** <**subnet prefix**> **-gw** <**ipv4 | ipv6 gateway address**> **-vlan** <**vlan_id**>
   - Remove the existing the 1 Gbps ports from the portset by using the following command:

     #svctask **rmip ip_ address_id**
   - On system C2: #svctask **mkip -node** <**id | name**> **-port** <**id**> **-portset** <**id | name**> **-ip** <**ipv4 | ipv6 address**>**-prefix** <**subnet prefix**> **-gw** <**ipv4 | ipv6 gateway address**> **-vlan** <**vlan_id**>
   - Remove the existing the 1 Gbps ports from the portset by using the following command:

     #svctask **rmip ip_ address_id**
3. Start the IP partnerships.
   - On system C1: #svctask **chpartnership -start** <**systemid C2**>
   - On system C2: #svctask **chpartnership -start** <**systemid C1**>

     Note: Since new data path over IPv6 addresses are now available, partnership will first change to `not_present`, and then to `fully_configured`. If it remains in `not_present`, monitor the node error/DMP if it is getting triggered and check the appropriate DMP in the troubleshooting section.

This procedure completes the migration from 1 Gbps links to 10 Gbps links. This same procedure can be applied to migrate from 10 Gbps to 1 Gbps or 25 Gbps by applying suitable substitutes for 1 Gbps or 25 Gbps ports instead of 10 Gbps ports.

Intersystem link for host traffic

If you use the intersystem link for host traffic, ensure that you have sufficient bandwidth to support all sources of load.

Scenario: The hosts in a local clustered system can read and write to the volumes in a remote system

In this scenario, the hosts in the local system also exchange heartbeats with the hosts that are in the remote system. Because the intersystem link is being used for multiple purposes, you must have sufficient bandwidth to support the following sources of load:

- Local host to remote volume I/O traffic or remote host to local volume I/O traffic.
- Local-host-to-remote-host heartbeat traffic.

Long-distance requirements for partnerships

The links between systems in a partnership that are used for replication must meet specific configuration, latency, and distance requirements.

- The following table lists the maximum round-trip latency for each type of partnership.

Table 49. Maximum supported round-trip latency between sites	
Partnership	
FC	**IP**
250 ms	80 ms

- Replication requires a specific amount of bandwidth for intersystem heartbeat traffic. When the system uses a Fibre Channel partnership, the amount of traffic depends on the number of nodes that are in both the local system and the remote system. Table 50 on page 352 provides a guideline for the intersystem heartbeat traffic between the systems. These numbers represent the total traffic between two systems when no I/O operations run on the copied volumes. Half of the data is sent by the local system and half of the data is sent by the remote system. Therefore, traffic is evenly divided between all of the available intersystem links. If you have two redundant links, half of the traffic is sent over each link.

Table 50. Intersystem heartbeat traffic in Mbps				
System 1	**System 2**			
	2 nodes	**4 nodes**	**6 nodes**	**8 nodes**
2 nodes	5	6	6	6
4 nodes	6	10	11	12
6 nodes	6	11	16	17
8 nodes	6	12	17	21

- If the link between two sites is configured with redundancy so that it can tolerate single failures, the link must be sized so that the bandwidth and latency statements are correct during single failure conditions.

Chapter 17. Cyber resilience

Ransomware threat detection

Ransomware threat detection is a process that identifies and responds to security threats before they can damage data or systems.

Ransomware threat detection uses information theory concepts such as Shannon Entropy and other statistics to characterize the current workload. This type of analysis requires byte-by-byte processing and is computationally expensive.

FlashCore Module (FCM) drives provide computational storage offload to compute and summarize these statistics, which the FlashSystem aggregates and analyzes to notify Storage Insights of a potential threat. With drives that run FCM 4.1 firmware or later, statistics about data are processed within the FCM drives, and presented to the FlashSystem on a per volume basis. This is collected on a regular basis, and aggregated across all participating drives, resulting in a set of statistics per volume. These statistics are then processed on the system to determine its threat potential.

© Copyright IBM Corp. 2024

Chapter 18. Security

IBM Storage Virtualize based storage systems are secure storage platforms that implement various security-related features for both system-level security and data-level security.

Security overview

IBM Storage Virtualize based storage systems are secure storage platforms that implement various security-related features for both system-level security and data-level security.

The security features are broadly categorized as *System security* and *Data security*. These security features protect and prevent unauthorized access and use of the system, its resources, and the data that is stored on the system.

IBM Storage Virtualize provides the following security features on the storage systems.

System Security

System security describes the controls that protect the system and its resources from both internal and external disruption. In general, these features prevent unauthorized access to system resources and provide event notifications and alerts that warn security administrators of any unauthorized access attempts. IBM Storage Virtualize systems support the following system security features:

User authentication

The system supports both local users, and remote users who are authenticated to the system through a remote authentication service. You can create local users who can access the system. These user types are defined based on the administrative privileges that they have on the system. Local users must provide either a password, a Secure Shell (SSH) key, or both. Local users are authenticated through the authentication methods that are configured on the system. If the local user needs access to the management GUI, a password is needed for the user. If the user requires access to the command-line interface (CLI) through SSH, either a password or a valid SSH key file is necessary. Local user passwords are securely stored by using the PBKDF2 hashing algorithm. Local users must be part of a user group that is defined on the system. User groups define roles that authorize the users within that group to a specific set of operations on the system. For more information, see “User authentication” on page 421.

Remote authentication

Remote authentication allows users to authenticate to the system using credentials that are stored on an external authentication service. When you configure remote authentication, you do not need to configure users on the system or assign more passwords. Instead, you can use your existing passwords and user groups that are defined on the remote service to simplify user management and access to enforce password policies more efficiently, and to separate user management from storage management. For more information, see “Remote Authentication” on page 422.

Role-based access control

Each user of the management GUI must provide a username and a password to sign on. Each user also has an associated role, such as monitor or security administrator. These roles are defined at the system level. For example, a user can be the administrator for one system, but the security administrator for another system. For more information, see “Users” on page 227.

Default user

When a system is created, a single local user with Security Administrator privileges, called superuser, is created. The superuser contains maximum privileges to complete system setup and configuration. For new systems, the default superuser password must be changed on first login to the system. Although the superuser cannot be deleted, you may want to lock or disable the superuser to prevent access to the system. For more information, see “Locking user accounts” on page 228.

© Copyright IBM Corp. 2024

Object-based access control
An ownership group defines a subset of users and objects within the system. For more information on configuring an ownership group, see “Configuring ownership groups” on page 220.

Login interfaces

The system provides several management interfaces that allow user to authenticate and manage the system and its objects. These interfaces include the management GUI, command-line interface, service assistant interface and commands, and REST APIs. All interfaces use in-flight encryption for data to secure the login and all subsequent connections with system. You can create a user group with specific roles to allow or restrict the user group to interfaces. With this capability, you can also control access from scripts or automated services, as well as individual users. For more information, see “Changing user groups ” on page 230.

Password Policy
With password policy support, system administrators can set security requirements that are related to password creation and expiration, timeout for inactivity, and actions after failed logon attempts. For more information, see “Password policy” on page 481.

Account locking

The Security Administrator can manually lock or unlock a user account at any time. By default, the superuser is exempt from the system-wide policy for manual or automatic account locking. For more information, see “Locking user accounts” on page 228.

Session timeouts

You can configure session timeout for both the management GUI and the CLI. Session timeouts automatically signs a user out of a session if it has been idle for a specified amount of time. The timeout values can be 5 - 240 minutes. For more information, see the **chsecurity** command.

Login banner

You can create or change a message that displays when users log on to the system. When users log on to the system with the management GUI, command-line interface, or service assistant, the message displays before they log on to the system. For more information, see “Login message” on page 240.

Multifactor authentication
Multifactor authentication requires users to provide multiple pieces of information when they log in to the system to prove their identity. Multifactor authentication uses any combination of two or more methods, called *factors*, to authenticate users to your resources and protect those resources from unauthorized access. For more information, see “Multifactor authentication” on page 399.

Single sign-on
Single Sign-on (SSO) authentication requires users to register their credentials only once when the user signs on to the application for the first time. The user information is stored at the Identity Provider (IdP) that manages the user credentials and determines whether the user is required to authenticate again or not. For more information, see “Single Sign-on” on page 425.

Secured IP partnership

Secured IP partnerships secure the data as it travels through an untrusted network between production and recovery systems. Secured IP partnerships minimize the risk of hackers manipulating or intercepting data in untrusted networks as replicated data travels between partnered systems. For more information, see “Partnerships using IP Connectivity” on page 337.

Auditing and reporting

The system includes an internal tamper-proof audit log that traces all successful commands and identifies the user who issued the commands. This information includes the user details, the IP address where they connected to the system, and timestamps. For more information, see **Audt log commands**.

Secure and trusted boot

Secure boot encrypts the file systems and relies on a hardware root of trust that extends all the way through to the operating system and **initrd** to unlock the file system. The system supports hardware

root of trust and secure boot operations, which protects against unauthorized physical access to the hardware and prevents malicious software from running on the system.

Secure sockets and Secure Shell settings

Secure sockets (SSL/TLS) and Secure Shell (SSH) are used to establish secure connection to management interfaces, such as the management GUI, CLI, and REST APIs. These security protocols are also used to authenticate to remote servers, such as email, LDAP, and key management servers. The system supports different levels of these security protocols that define the cipher suites and key exchange algorithms that can be used. For more information, see "Security protocol levels" on page 359.

SSL/TLS certificates

During system initialization and set up an initial internally signed certificate is created to secure connections between nodes on the system and a supported web browser for management GUI access. After system setup, you can either create internally signed certificate from the native certificate authority or generate a certificate signing request to an external, third-party CA. For more information, refer to "Configuring system certificates" on page 391.

Disabling USB ports

This feature can be combined with encryption key management on local USB flash drives to keep the ports disabled during day-to-day use and only reenabled when a key must be provided. For more information, see the **satask chnodeusb** command.

Network Time Protocol

To mitigate against attacks that are based on the system time/date being out of sync with other services, the system supports a configurable Network Time Protocol (NTP) server to control the system time. An NTP server can be configured only by a user with the Security Administrator role. For more information, see the **chsystem** command.

Preventing access as root

The management software runs as an internal Linux user without root privileges, preventing unauthorized users from accessing the system. Although external users cannot access this software, an attacker might use a weakness to compromise the system. Even in this circumstance, they do not have root privileges.

Disabling service assistant password reset

The superuser is the only user account that is permitted to run service assistant commands on the system. You can use the command-line interface (CLI) to view and change the status of the password reset feature for the system. .

Data security

Data security protects the data that is stored on the system against theft, loss, or attack.

Encrypting data at rest

All NVMe drives that are supported by the system, including IBM FlashCore® Modules (FCMs) and a range of other third-party drives, are self-encrypting drives (SEDs) that encrypt data within the electrical circuit of each individual drive.

Secure data deletion

The system provides methods to securely erase data from a drive or from a boot drive when a node or node canister is decommissioned. For more information, refer to "Secure data deletion" on page 479.

Volume protection

Volume protection prevents active volumes or host mappings from being deleted inadvertently if the system detects recent I/O activity. For more information, see "Volume protection " on page 84.

Safeguarded Copy function

Safeguarded Copy function supports the ability to create cyber-resilient point-in-time copies of volumes that cannot be changed or deleted through user errors, malicious actions, or ransomware

attacks. Safeguarded snapshots are supported on the system through an internal scheduler that is defined in the snapshot policy. When the policy is assigned to a volume group, you can select Safeguarded option. The policy creates immutable snapshots of all volumes in the volume group. The system supports internal and external snapshot scheduling applications such as IBM Copy Services Manager and IBM Storage Copy Data Management.

For more information, refer to the "Configuring Safeguarded Copy function" on page 276.

Logical port isolation

The system provides means to define logical port sets. These port sets can be used to further restrict the login traffic from a host, or set of hosts, to isolate traffic to specific SAN paths. *Portsets* are groupings of logical addresses that are associated with the specific traffic types. For more information, refer to "Portsets" on page 329.

Encrypting data in flight

If you use secured IP partnerships to secure connections between partnered systems, you also require an encryption license. If you have not purchased a license, contact a customer representative to purchase an encryption license. For more information, refer to "Configuring encryption" on page 453.

Managing security

The system supports several security-related features that can help protect system data and resources from unauthorized access.

User security and authentication

The system supports both local users, and remote users who are authenticated to the system through a remote authentication service. You can create local users who can access the system. These user types are defined based on the administrative privileges that they have on the system. Local users must provide either a password, a Secure Shell (SSH) key, or both. Local users are authenticated through the authentication methods that are configured on the system. If the local user needs access to the management GUI, a password is needed for the user. If the user requires access to the command-line interface (CLI) through SSH, either a password or a valid SSH key file is necessary. Local user passwords are securely stored by using the PBKDF2 hashing algorithm. Local users must be part of a user group that is defined on the system. User groups define roles that authorize the users within that group to a specific set of operations on the system.

A remote user is authenticated on a remote LDAP server. A remote user does not need to be added to the list of users on the system, although they can be added to configure optional SSH keys. For remote users, an equivalent user group must be created on the system with the same name and role as the group on the remote LDAP server. Remote users cannot access the system when the remote LDAP server is down. In that case, a local user account must be used until the LDAP service is restored. Remote users have their groups that are defined by the remote authentication server.

Encryption is a technology that uses cryptography to help ensure confidentiality of sensitive information. Encryption uses keys to encode information so that it cannot be understood by unauthorized parties. Depending on your model, the system supports both encryption of data-at-rest and encryption of data-in-flight.

SSL/TLS security controls

The system supports a choice of security levels (2 to 7) to enforce a minimum level of SSL (Secure Sockets Layer) or TLS (Transport Layer Security) that can be used to access the system. The security levels 5, 6, and 7 support TLS 1.3, which provides more secure cipher suites along with a faster handshake process. Only clients that support the minimum SSL or TLS level that is enforced by the system are able to establish secure connections.

Security protocol levels

Security administrators can change the security protocol level for either SSL or SSH protocols. When you change the security level for either of these security protocols, you can control which encryption algorithms, ciphers, and version of the protocol are permitted on the system.

Depending on your security requirements for your organization or geography, you can change the level for both SSL and SSH protocols.

The system supports OpenSSL and Java SSL ciphers to provide strong encryption for secure connections using the SSL or TLS protocols. On a new system, the default SSL protocol level is 5, and the default SSH protocol level is 3. If you want to allow the use of more cipher suites for compatibility with some applications, you can select a lower value. Selecting a higher value further restricts the list of supported cipher suites.

By default, the system uses the suggested SSL protocol and SSH protocol levels. If the suggested SSL and SSH protocol levels change on a future system upgrade, the system applies the new levels automatically. However, if you manually select a new level, then the system no longer uses the suggested levels and does not modify the level on future system upgrades. To use automatic suggestions, reset the SSL and SSH protocol levels using the management GUI or CLI.

Note:

The suggested SSL and SSH protocol levels might be increased in future code upgrades as security requirements change. To automatically update the protocol level to the new suggested level whenever you upgrade the system, select **Automatic** for the protocol level.

For servers or services that do not support TLS 1.3, do not set the security level to a level that supports only TLS 1.3, such as levels 6 and 7. Currently, the following servers or services do not support TLS 1.3:

- KeySecure key servers
- Duo Security for Multifactor Authentication on the Command Line Interface (CLI)
- Transparent Cloud Tiering (TCT)

If you use KeySecure key servers and want to use TLS 1.3 for secure communication, migrate to CipherTrust Manager key servers.

Note: For services or features involving two systems (system A and system B) that communicate with each other (for example, policy-based replication or secured IP partnerships), using mutual TLS authentication, you can configure different security levels on each system. Do not configure system A to only support TLS 1.2 (SSL protocol levels 2-4) and system B to support only TLS 1.3 (levels 6-7), or vice versa.

The following table describes each security level, minimum version of SSL or TLS allowed and the supported ciphers for each level:

Table 51. Supported SSL/TLS security levels

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
1	Sets the system to disallow SSL version 3.0.	TLS 1.1	• SSL_ECDHE_ECDSA_WITH_AES_256_CBC_SHA384 • SSL_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384 • SSL_ECDHE_RSA_WITH_AES_256_CBC_SHA384 • SSL_ECDHE_RSA_WITH_AES_256_GCM_SHA384 • SSL_RSA_WITH_AES_256_CBC_SHA256 • SSL_RSA_WITH_AES_256_GCM_SHA384 • SSL_ECDH_ECDSA_WITH_AES_256_CBC_SHA384 • SSL_ECDH_RSA_WITH_AES_256_GCM_SHA384 • SSL_DHE_RSA_WITH_AES_256_CBC_SHA256 • SSL_DHE_RSA_WITH_AES_256_CBC_SHA256 • SSL_DHE_RSA_WITH_AES_256_GCM_SHA384 • SSL_DHE_DSS_WITH_AES_256_CBC_SHA256 • SSL_DHE_DSS_WITH_AES_256_GCM_SHA384	• ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-ECDSA-AES256-GCM-SHA384 • ECDHE-RSA-AES256-SHA384 • ECDHE-ECDSA-AES256-SHA384 • ECDHE-RSA-AES256-SHA • ECDHE-ECDSA-AES256-SHA • DHE-DSS-AES256-GCM-SHA384 • DHE-RSA-AES256-GCM-SHA384 • DHE-RSA-AES256-SHA256 • DHE-DSS-AES256-SHA256 • DHE-RSA-AES256-SHA • DHE-DSS-AES256-SHA • DHE-RSA-CAMELLIA256-SHA • DHE-DSS-CAMELLIA256-SHA • ECDHE-ECDSA-DES-CBC3-SHA • DHE-RSA-SEED-SHA • DHE-DSS-SEED-SHA • DHE-RSA-CAMELLIA128-SHA • ECDH-RSA-AES256-GCM-SHA384 E • ECDH-ECDSA-AES256-GCM-SHA384 • ECDH-RSA-AES256-SHA384

Table 51. Supported SSL/TLS security levels (continued)

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
	Sets the system to disallow SSL version 3.0.	TLS 1.1	• SSL_RSA_WITH_AES_256_CBC_SHA • SSL_ECDH_ECDSA_WITH_AES_256_CBC_SHA • SSL_ECDH_RSA_WITH_AES_256_CBC_SHA • SSL_DHE_RSA_WITH_AES_256_CBC_SHA • SSL_DHE_DSS_WITH_AES_256_CBC_SHA • SSL_ECDHE_ECDSA_WITH_AES_128_CBC_SHA256 • SSL_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256 • SSL_ECDHE_RSA_WITH_AES_128_CBC_SHA256 • SSL_ECDHE_RSA_WITH_AES_128_GCM_SHA256 • SSL_RSA_WITH_AES_128_CBC_SHA256 • SSL_RSA_WITH_AES_128_GCM_SHA256 • SSL_ECDH_ECDSA_WITH_AES_128_CBC_SHA256 • SSL_ECDH_ECDSA_WITH_AES_128_GCM_SHA256 • SSL_ECDH_RSA_WITH_AES_128_CBC_SHA256 • SSL_ECDH_RSA_WITH_AES_128_GCM_SHA256 • SSL_DHE_RSA_WITH_AES_128_CBC_SHA256 • SSL_DHE_RSA_WITH_AES_128_GCM_SHA256 • SSL_DHE_DSS_WITH_AES_128_CBC_SHA256 • SSL_DHE_DSS_WITH_AES_128_GCM_SHA256 • SSL_ECDHE_ECDSA_WITH_AES_128_CBC_SHA • SSL_ECDHE_RSA_WITH_AES_128_CBC_SHA	• ECDH-ECDSA-AES256-SHA384 • ECDH-RSA-AES256-SHA • ECDH-ECDSA-AES256-SHA • AES256-GCM-SHA384 • AES256-SHA256 • AES256-SHA • CAMELLIA256-SHA • PSK-AES256-CBC-SHA • ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES128-GCM-SHA256 • ECDHE-RSA-AES128-SHA256 • ECDHE-ECDSA-AES128-SHA256 • ECDHE-RSA-AES128-SHA • ECDHE-ECDSA-AES128-SHA • DHE-DSS-AES128-GCM-SHA256 • DHE-RSA-AES128-GCM-SHA256 • DHE-RSA-AES128-SHA256 • DHE-DSS-AES128-SHA256 • DHE-RSA-AES128-SHA • DHE-DSS-AES128-SHA • ECDHE-RSA-DES-CBC3-SHA

Table 51. Supported SSL/TLS security levels (continued)

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
	Sets the system to disallow SSL version 3.0.	TLS 1.1	• SSL_ECDHE_ECDSA_WITH_AES_256_CBC_SHA • SSL_ECDHE_RSA_WITH_AES_256_CBC_SHA • SSL_RSA_WITH_AES_128_CBC_SHA • SSL_ECDH_ECDSA_WITH_AES_128_CBC_SHA • SSL_ECDH_RSA_WITH_AES_128_CBC_SHA • SSL_DHE_RSA_WITH_AES_128_CBC_SHA • SSL_DHE_DSS_WITH_AES_128_CBC_SHA • SSL_ECDHE_ECDSA_WITH_3DES_EDE_CBC_SHA • SSL_ECDHE_RSA_WITH_3DES_EDE_CBC_SHA • SSL_RSA_FIPS_WITH_3DES_EDE_CBC_SHA • SSL_RSA_WITH_3DES_EDE_CBC_SHA • SSL_ECDH_ECDSA_WITH_3DES_EDE_CBC_SHA • SSL_ECDH_RSA_WITH_3DES_EDE_CBC_SHA • SSL_DHE_DSS_WITH_3DES_EDE_CBC_SHA • SSL_DHE_RSA_WITH_3DES_EDE_CBC_SHA	• DHE-DSS-CAMELLIA128-SHA • EDH-RSA-DES-CBC3-SHA • EDH-DSS-DES-CBC3-SHA • ECDH-RSA-AES128-GCM-SHA256 • ECDH-ECDSA-AES128-GCM-SHA256 • ECDH-RSA-AES128-SHA256 • ECDH-ECDSA-AES128-SHA256 • ECDH-RSA-AES128-SHA • ECDH-ECDSA-AES128-SHA • ECDH-RSA-DES-CBC3-SHA • ECDH-ECDSA-DES-CBC3-SHA • AES128-GCM-SHA256 • AES128-SHA • SEED-SHA • CAMELLIA128-SHA • DES-CBC3-SHA • PSK-AES128-CBC-SHA • PSK-3DES-EDE-CBC-SHA • KRB5-DES-CBC3-SHA

Table 51. Supported SSL/TLS security levels (continued)

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
2	Sets the system to disallow SSL version 3.0, TLS version 1.0, and TLS version 1.1.	TLS 1.2	• SSL_ECDHE_ECDSA_WITH_AES_256_CBC_SHA384 • SSL_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384 • SSL_ECDHE_RSA_WITH_AES_256_CBC_SHA384 • SSL_ECDHE_RSA_WITH_AES_256_GCM_SHA384 • SSL_RSA_WITH_AES_256_CBC_SHA256 • SSL_RSA_WITH_AES_256_GCM_SHA384 • SSL_ECDH_ECDSA_WITH_AES_256_CBC_SHA384 • SSL_ECDH_RSA_WITH_AES_256_GCM_SHA384 • SSL_DHE_RSA_WITH_AES_256_CBC_SHA256 • SSL_DHE_RSA_WITH_AES_256_CBC_SHA256 • SSL_DHE_RSA_WITH_AES_256_GCM_SHA384 • SSL_DHE_DSS_WITH_AES_256_CBC_SHA256 • SSL_DHE_DSS_WITH_AES_256_GCM_SHA384 • SSL_ECDHE_ECDSA_WITH_AES_256_CBC_SHA • SSL_ECDHE_RSA_WITH_AES_256_CBC_SHA • SSL_RSA_WITH_AES_256_CBC_SHA	• ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-ECDSA-AES256-GCM-SHA384 • ECDHE-RSA-AES256-SHA384 • ECDHE-ECDSA-AES256-SHA384 • DHE-DSS-AES256-GCM-SHA384 • DHE-RSA-AES256-GCM-SHA384 • DHE-RSA-AES256-SHA256 • ECDH-RSA-AES256-GCM-SHA384 E • ECDH-ECDSA-AES256-GCM-SHA384 • ECDH-RSA-AES256-SHA384 • ECDH-ECDSA-AES256-SHA384 • AES256-GCM-SHA384 • AES256-SHA256 • AES256-SHA • ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES128-GCM-SHA256

Table 51. Supported SSL/TLS security levels (continued)

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
	Sets the system to disallow SSL version 3.0, TLS version 1.0, and TLS version 1.1.	TLS 1.2	• SSL_ECDH_ECDSA_WITH_AES_256_CBC_SHA • SSL_ECDH_RSA_WITH_AES_256_CBC_SHA • SSL_DHE_RSA_WITH_AES_256_CBC_SHA • SSL_DHE_DSS_WITH_AES_256_CBC_SHA • SSL_ECDHE_ECDSA_WITH_AES_128_CBC_SHA256 • SSL_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256 • SSL_ECDHE_RSA_WITH_AES_128_CBC_SHA256 • SSL_ECDHE_RSA_WITH_AES_128_GCM_SHA256 • SSL_RSA_WITH_AES_128_CBC_SHA256 • SSL_RSA_WITH_AES_128_GCM_SHA256 • SSL_ECDH_ECDSA_WITH_AES_128_CBC_SHA256	• ECDHE-RSA-AES128-SHA256 • ECDHE-ECDSA-AES128-SHA256 • DHE-DSS-AES128-GCM-SHA256 • DHE-RSA-AES128-GCM-SHA256 • DHE-RSA-AES128-SHA256 • DHE-DSS-AES128-SHA256 • ECDH-RSA-AES128-GCM-SHA256 • ECDH-ECDSA-AES128-GCM-SHA256 • ECDH-RSA-AES128-SHA256 • ECDH-ECDSA-AES128-SHA256 • AES128-GCM-SHA256 • AES128-SHA256 • AES128-SHA • DES-CBC3-SHA • ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES128-GCM-SHA256

Table 51. Supported SSL/TLS security levels (continued)

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
	Sets the system to disallow SSL version 3.0, TLS version 1.0, and TLS version 1.1.	TLS 1.2	• SSL_ECDH_ECDSA_WITH_AES_128_GCM_SHA256 • SSL_ECDH_RSA_WITH_AES_128_CBC_SHA256 • SSL_ECDH_RSA_WITH_AES_128_GCM_SHA256 • SSL_DHE_RSA_WITH_AES_128_CBC_SHA256 • SSL_DHE_RSA_WITH_AES_128_GCM_SHA256 • SSL_DHE_DSS_WITH_AES_128_CBC_SHA256 • SSL_DHE_DSS_WITH_AES_128_GCM_SHA256 • SSL_ECDHE_ECDSA_WITH_AES_128_CBC_SHA • SSL_ECDHE_RSA_WITH_AES_128_CBC_SHA • SSL_RSA_WITH_AES_128_CBC_SHA • SSL_ECDH_ECDSA_WITH_AES_128_CBC_SHA • SSL_ECDH_RSA_WITH_AES_128_CBC_SHA • SSL_DHE_RSA_WITH_AES_128_CBC_SHA • SSL_DHE_DSS_WITH_AES_128_CBC_SHA	• ECDHE-ECDSA-AES128-GCM-SHA256 • ECDHE-RSA-AES128-SHA256 • ECDHE-ECDSA-AES128-SHA256 • DHE-DSS-AES128-GCM-SHA256 • DHE-RSA-AES128-GCM-SHA256 • DHE-RSA-AES128-SHA256 • DHE-DSS-AES128-SHA256 • ECDH-RSA-AES128-GCM-SHA256 • ECDH-ECDSA-AES128-GCM-SHA256 • ECDH-RSA-AES128-SHA256 • ECDH-ECDSA-AES128-SHA256 • AES128-GCM-SHA256 • AES128-SHA256 • AES128-SHA • DES-CBC3-SHA

Table 51. Supported SSL/TLS security levels (continued)

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
3	Sets the system to disallow SSL version 3.0, TLS version 1.0, and TLS version 1.1 and to allow cipher suites that are exclusive to TLS version 1.2.	TLS 1.2	• SSL_ECDHE_ECDSA_WITH_AES_256_CBC_SHA384 • SSL_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384 • SSL_ECDHE_RSA_WITH_AES_256_CBC_SHA384 • SSL_ECDHE_RSA_WITH_AES_256_GCM_SHA384 • SSL_RSA_WITH_AES_256_CBC_SHA256 • SSL_RSA_WITH_AES_256_GCM_SHA384 • SSL_ECDH_ECDSA_WITH_AES_256_CBC_SHA384 • SSL_ECDH_RSA_WITH_AES_256_GCM_SHA384 • SSL_DHE_RSA_WITH_AES_256_CBC_SHA256 • SSL_DHE_RSA_WITH_AES_256_CBC_SHA256 • SSL_DHE_RSA_WITH_AES_256_GCM_SHA384 • SSL_DHE_DSS_WITH_AES_256_CBC_SHA256 • SSL_DHE_DSS_WITH_AES_256_GCM_SHA384 • SSL_ECDHE_ECDSA_WITH_AES_256_CBC_SHA • SSL_ECDHE_RSA_WITH_AES_256_CBC_SHA • SSL_RSA_WITH_AES_256_CBC_SHA	• ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-ECDSA-AES256-GCM-SHA384 • ECDHE-RSA-AES256-SHA384 • ECDHE-ECDSA-AES256-SHA384 • DHE-DSS-AES256-GCM-SHA384 • DHE-RSA-AES256-GCM-SHA384 • DHE-RSA-AES256-SHA256 • ECDH-RSA-AES256-GCM-SHA384 E • ECDH-ECDSA-AES256-GCM-SHA384 • ECDH-RSA-AES256-SHA384 • ECDH-ECDSA-AES256-SHA384 • AES256-GCM-SHA384 • AES256-SHA256

Table 51. Supported SSL/TLS security levels (continued)

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
	Sets the system to disallow SSL version 3.0, TLS version 1.0, and TLS version 1.1 and to allow cipher suites that are exclusive to TLS version 1.2.	TLS 1.2	• SSL_ECDH_ECDSA_WITH_AES_256_CBC_SHA • SSL_ECDH_RSA_WITH_AES_256_CBC_SHA • SSL_DHE_RSA_WITH_AES_256_CBC_SHA • SSL_DHE_DSS_WITH_AES_256_CBC_SHA • SSL_ECDHE_ECDSA_WITH_AES_128_CBC_SHA256 • SSL_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256 • SSL_ECDHE_RSA_WITH_AES_128_CBC_SHA256 • SSL_ECDHE_RSA_WITH_AES_128_GCM_SHA256 • SSL_RSA_WITH_AES_128_CBC_SHA256 • SSL_RSA_WITH_AES_128_GCM_SHA256 • SSL_ECDH_ECDSA_WITH_AES_128_CBC_SHA256 • SSL_ECDH_ECDSA_WITH_AES_128_GCM_SHA256 • SSL_ECDH_RSA_WITH_AES_128_CBC_SHA256 • SSL_ECDH_RSA_WITH_AES_128_GCM_SHA256 • SSL_DHE_RSA_WITH_AES_128_CBC_SHA256 • SSL_DHE_RSA_WITH_AES_128_GCM_SHA256 • SSL_DHE_DSS_WITH_AES_128_CBC_SHA256 • SSL_DHE_DSS_WITH_AES_128_GCM_SHA256	• ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES128-GCM-SHA256 • ECDHE-RSA-AES128-SHA256 • ECDHE-ECDSA-AES128-SHA256 • DHE-DSS-AES128-GCM-SHA256 • DHE-RSA-AES128-GCM-SHA256 • DHE-RSA-AES128-SHA256 • DHE-DSS-AES128-SHA256 • ECDH-RSA-AES128-GCM-SHA256 • ECDH-ECDSA-AES128-GCM-SHA256 • ECDH-RSA-AES128-SHA256 • ECDH-ECDSA-AES128-SHA256 • AES128-GCM-SHA256 • AES128-SHA256

Table 51. Supported SSL/TLS security levels (continued)

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
4	Sets the system to disallow SSL version 3.0, TLS version 1.0, and TLS version 1.1, and to allow cipher suites that are exclusive to TLS version 1.2. Sets the system to disallow RSA ciphers and static key exchange ciphers.	TLS 1.2	• SSL_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384 • SSL_DHE_DSS_WITH_AES_256_GCM_SHA384 • SSL_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256 • SSL_DHE_DSS_WITH_AES_128_GCM_SHA256	• ECDHE-ECDSA-AES256-GCM-SHA384 • DHE-DSS-AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-GCM-SHA256 • DHE-DSS-AES128-GCM-SHA256
5	Sets the system to disallow SSL version 3.0, TLS version 1.0, and TLS version 1.1 and to allow cipher suites that are exclusive to TLS version 1.2 and 1.3.	TLS 1.2, TLS 1.3	**For TLS 1.3** • TLS_AES_256_GCM_SHA384 • TLS_CHACHA20_POLY1305_SHA256 • TLS_AES_128_GCM_SHA256 • TLS_AES_128_CCM_8_SHA256 • TLS_AES_128_CCM_SHA256	**For TLS 1.3** • AES-256-GCM-SHA384 • CHACHA20-POLY1305-SHA256 • AES-128-GCM-SHA256 • AES-128-CCM-8-SHA256 • AES-128-CCM-SHA256

Table 51. Supported SSL/TLS security levels (continued)

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
	Sets the system to disallow SSL version 3.0, TLS version 1.0, and TLS version 1.1 and to allow cipher suites that are exclusive to TLS version 1.2 and 1.3.	TLS 1.2, TLS 1.3	**For TLS 1.2** The security level 5 supports all the Java SSL ciphers supported at the security level 3. A few additional TLS 1.2 specific Java SSL ciphers supported at level 5 are: • TLS_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384 • TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256 • TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384 • TLS_RSA_WITH_AES_256_GCM_SHA384 • TLS_ECDH_ECDSA_WITH_AES_256_GCM_SHA384 • TLS_ECDH_RSA_WITH_AES_256_GCM_SHA384 • TLS_DHE_DSS_WITH_AES_256_GCM_SHA384 • TLS_DHE_RSA_WITH_AES_256_GCM_SHA384	**For TLS 1.2** The security level 5 supports all the OpenSSL ciphers supported at the security level 3. A few additional TLS 1.2 specific OpenSSL ciphers supported at level 5 are: • ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-ECDSA-AES256-GCM-SHA384 • ECDHE-RSA-AES256-SHA384 • ECDHE-ECDSA-AES256-SHA384 • DHE-RSA-AES256-GCM-SHA384 • DHE-RSA-AES256-SHA256 • AES256-GCM-SHA384 • AES256-SHA256 • ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES128-GCM-SHA256

Table 51. Supported SSL/TLS security levels (continued)

Security level	Description	Minimum security allowed	Supported Java SSL ciphers	Supported OpenSSL ciphers
	Sets the system to disallow SSL version 3.0, TLS version 1.0, and TLS version 1.1 and to allow cipher suites that are exclusive to TLS version 1.2 and 1.3.	TLS 1.2, TLS 1.3	• TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256 • TLS_RSA_WITH_AES_128_GCM_SHA256 • TLS_ECDH_ECDSA_WITH_AES_128_GCM_SHA256 • TLS_ECDH_RSA_WITH_AES_128_GCM_SHA256 • TLS_DHE_RSA_WITH_AES_128_GCM_SHA256 • TLS_DHE_DSS_WITH_AES_128_GCM_SHA256	• ECDHE-RSA-AES128-SHA256 • ECDHE-ECDSA-AES128-SHA256 • DHE-RSA-AES128-GCM-SHA256 • DHE-RSA-AES128-SHA256 • AES128-GCM-SHA256 • AES128-SHA256
6	Sets the system to disallow SSL version 3.0, TLS version 1.0, TLS version 1.1, and TLS version 1.2 and to allow cipher suites that are exclusive to TLS version 1.3.	TLS 1.3	• TLS_AES_256_GCM_SHA384 • TLS_CHACHA20_POLY1305_SHA256 • TLS_AES_128_GCM_SHA256 • TLS_AES_128_CCM_8_SHA256 • TLS_AES_128_CCM_SHA256	• AES-256-GCM-SHA384 • CHACHA20-POLY1305-SHA256 • AES-128-GCM-SHA256 • AES-128-CCM-8-SHA256 • AES-128-CCM-SHA256
7	Sets the system to disallow SSL version 3.0, TLS version 1.0, TLS version 1.1, and TLS version 1.2 and to allow the TLS 1.3 cipher suites that are FIPS mode compliant.	TLS 1.3	TLS_AES_256_GCM_SHA384	AES-256-GCM-SHA384

The following table describes the SSH security levels supported by the system:

Table 52. SSH algorithms supported at each security level

Security level	Key Exchange	Cipher Suite	MAC Algorithm	Host Key Algorithms
1	• curve25519-sha256 • curve25519-sha256@libssh.org • ecdh-sha2-nistp256 • ecdh-sha2-nistp384 • ecdh-sha2-nistp521 • diffie-hellman-group-exchange-sha256 • diffie-hellman-group16-sha512 • diffie-hellman-group18-sha512 • diffie-hellman-group14-sha256 • diffie-hellman-group14-sha1 • diffie-hellman-group1-sha1 • diffie-hellman-group-exchange-sha1	aes256-ctr aes192-ctr aes128-ctr chacha20-poly1305@openssh.com aes256-gcm@openssh.com aes128-gcm@openssh.com aes256-cbc aes192-cbc aes128-cbc	hmac-sha2-256 hmac-sha2-512 hmac-sha2-256-etm@openssh.com hmac-sha2-512-etm@openssh.com hmac-sha1	rsa-sha2-256 rsa-sha2-512 ecdsa-sha2-nistp521 ecdsa-sha2-nistp521-cert-v01@openssh.com ssh-rsa ssh-rsa-cert-v01@openssh.com

Table 52. SSH algorithms supported at each security level (continued)

Security level	Key Exchange	Cipher Suite	MAC Algorithm	Host Key Algorithms
2	• curve25519-sha256 • curve25519-sha256@libssh.org • ecdh-sha2-nistp256 • ecdh-sha2-nistp384 • ecdh-sha2-nistp521 • diffie-hellman-group-exchange-sha256 • diffie-hellman-group16-sha512 • diffie-hellman-group18-sha512 • diffie-hellman-group14-sha256 • diffie-hellman-group14-sha1	aes256-ctr aes192-ctr aes128-ctr chacha20-poly1305@openssh.com aes256-gcm@openssh.com aes128-gcm@openssh.com	hmac-sha2-256 hmac-sha2-512 hmac-sha2-256-etm@openssh.com hmac-sha2-512-etm@openssh.com hmac-sha1	rsa-sha2-256 rsa-sha2-512 ecdsa-sha2-nistp521 ecdsa-sha2-nistp521-cert-v01@openssh.com ssh-rsa ssh-rsa-cert-v01@openssh.com
3	• curve25519-sha256 • curve25519-sha256@libssh.org • ecdh-sha2-nistp256 • ecdh-sha2-nistp384 • ecdh-sha2-nistp521 • diffie-hellman-group-exchange-sha256 • diffie-hellman-group16-sha512 • diffie-hellman-group18-sha512 • diffie-hellman-group14-sha256	aes256-ctr aes192-ctr aes128-ctr chacha20-poly1305@openssh.com aes256-gcm@openssh.com aes128-gcm@openssh.com	hmac-sha2-256 hmac-sha2-512 hmac-sha2-256-etm@openssh.com hmac-sha2-512-etm@openssh.com	rsa-sha2-256 rsa-sha2-512 ecdsa-sha2-nistp521 ecdsa-sha2-nistp521-cert-v01@openssh.com ssh-rsa ssh-rsa-cert-v01@openssh.com

Table 52. SSH algorithms supported at each security level (continued)				
Security level	**Key Exchange**	**Cipher Suite**	**MAC Algorithm**	**Host Key Algorithms**
4	• curve25519-sha256 • curve25519-sha256@libssh.org • ecdh-sha2-nistp256 • ecdh-sha2-nistp384 • ecdh-sha2-nistp521	aes256-ctr aes192-ctr aes128-ctr aes256-gcm@openssh.com aes128-gcm@openssh.com	hmac-sha2-256 hmac-sha2-512 hmac-sha2-256-etm@openssh.com hmac-sha2-512-etm@openssh.com	rsa-sha2-256 rsa-sha2-512 ecdsa-sha2-nistp521 ecdsa-sha2-nistp521-cert-v01@openssh.com

Restriction: The 3-site-orchestrator does not support SSH protocol level 4.

When you change the SSL security protocol level, you must restart any service using SSL/TLS. All current session are ended to ensure no sessions are open using the old security level. It can take a few minutes for the service to be available.

Using the management GUI

You can use the management GUI to update protocol levels for SSL and SSH connections:

SSL/TLS security protocol level

By default, the SSL protocol level is set to 5, and the SSH protocol level is set to 3 . To change the SSL or SSH security protocol levels, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Security protocol levels**.
2. You can update any of the following details:

SSL protocol level

Note: Changing the **SSL protocol level** causes the GUI to restart.

SSL ensures that the data is securely transferred. By default, security level 5 is set to allow both TLS 1.2 and TLS 1.3. You can select the required **SSL protocol level** from the following options:

- Automatic - Use suggested level of 5.
- 2 - TLS 1.2, allow TLS 1.0, 1.1, and 1.2 ciphers.
- 3 - TLS 1.2, allow TLS 1.2 ciphers.
- 4 - TLS 1.2, allow TLS 1.2 ciphers but disallow RSA and static key exchange ciphers.
- 5 - TLS 1.2 and TLS 1.3, disallow static key exchange ciphers.
- 6 - TLS 1.3, allow only TLS 1.3 ciphers.
- 7 - TLS 1.3, allow only ciphers that support FIPS.

SSH protocol level

Select the SSH protocol level that is used for connections to the command-line interface. Each level supports different algorithms for key exchange. The range is 1 - 4, where 3 is the default value. Select the required **SSH protocol level** from the following options:

- Automatic - Use suggested level of 3.
- 1 - Allow block ciphers.

- 2 - Disallow block ciphers.
- 3 - Disallow SHA1.
- 4 - Disallow Diffie-Hellman.

3. Click **Save**.

Note: The suggested SSL and SSH protocol levels might be increased in future code upgrades as security requirements change. To automatically update the protocol level to the new suggested level whenever you upgrade the system, select **Automatic** for the protocol level.

Using the command-line interface (CLI)

The `chsecurity` command allows you to set the ciphers and protocols that are allowed by secure interfaces to reduce the vulnerability to attack. However, changing the security level might break the connection to external systems such as web browsers and anything that is connected through CIM such as VMWare provisioning utilities or IBM Spectrum Control software.

1. To display your current system SSL, TLS, and SSH security settings, enter the following command:

```
lssecurity
```

   The results show the current setting as shown in the following example:

```
sslprotocol 5
sshprotocol 3
gui_timeout_mins 30
cli_timeout_mins 15
restapi_timeout_mins 60
min_password_length 8
password_special_chars 0
password_upper_case 0
password_lower_case 0
password_digits 0
check_password_history no
max_password_history 6
min_password_age_days 1
password_expiry_days 0
expiry_warning_days 14
superuser_locking disabled
max_failed_login_attempts 0
lockout_period_mins 10
superuser_multi_factor no
ssh_grace_time_seconds 60
ssh_max_tries 6
superuser_password_sshkey_required no
superuser_gui_disabled no
superuser_rest_disabled no
superuser_cim_disabled yes
two_person_integrity_enabled no
two_person_integrity_superuser_locked no
ssl_protocols_enabled TLSv1.2:TLSv1.3
ssl_protocol_suggested yes
ssh_protocol_suggested yes
```

2. To change SSL/TLS settings, enter **`chsecurity -sslprotocol security_level`**, where **`security_level`** is 2, 3, 4, 5, 6, or 7.

   Note: You might lose the connection to the management GUI when the security level is changed. If you lose the connection, use the CLI to decrease the security level to a lower setting.

3. To change SSH settings, enter **`chsecurity -sshprotocol security_level`**, where **`security_level`** is 1, 2, 3, or 4.

Security levels and supported security ciphers

You can use connections based on the Transport Layer Security (TLS), which is the successor of the Secure Sockets Layer (SSL) protocol, to ensure safer communications.

Version

This information about security settings applies to the current release only.

SSL certificates

The system uses a certificate to authenticate SSL connections. For more information about managing certificates, see “Creating and managing certificate authority store by using CLI” on page 398.

TLS or SSL connections and security levels

Note: The terms TLS and SSL are often used interchangeably in the industry.

The system uses TLS or SSL connections to control access to interfaces such as the management GUI, the service assistant GUI, the key server, and RESTful API. TLS or SSL connections use security ciphers to help control access.

You can use security ciphers that are supported by different levels of TLS or SSL. Each level supports ciphers that provide differing strengths of encryption. You can set the security level to 5, 6, or 7 to be compliant with the NIST 800-52 standard. Security level 7 allows only the cipher suite TLS_AES_256_GCM_SHA384, which the NIST recommends for Federal Information Processing Standards (FIPS) mode.

SSL protocol levels 2 to 4 do not support the TLS 1.3 protocol and the cipher suites that are approved by the NIST 800-52 standard. Currently, the TLS or SSL security level 7 is the maximum level that is supported and TLS or SSL security level 2 is the lowest security level supported.

SSL protocols and ciphers supported at each security level

Table 53 on page 375 displays the protocols that are supported at security level 7.

Table 53. Protocols supported at level 7

Protocol level	**Is it supported?**
TLS 1.3	Yes
TLS 1.2	No
TLS 1.1	No
TLS 1.0	No
SSL 3 and earlier	No

Table 54 on page 375 displays Java SSL ciphers that are supported at security level 7.

Table 54. Java SSL ciphers supported at security level 7

Java SSL ciphers
TLS_AES_256_GCM_SHA384

Table 55 on page 376 displays OpenSSL security ciphers that are supported by security level 7.

Table 55. OpenSSL ciphers supported at level 7 (`chsecurity -sslprotocol 7`)		
Cipher	**Bulk encryption algorithm**	**Hashing algorithm**
AES-256-GCM-SHA384	AES-256-GCM	SHA384

Table 56 on page 376 displays the protocols that are supported at security level 6.

Table 56. Protocols supported at level 6	
Protocol level	**Is it supported?**
TLS 1.3	Yes
TLS 1.2	No
TLS 1.1	No
TLS 1.0	No
SSL 3 and earlier	No

Table 57 on page 376 displays Java SSL ciphers that are supported at security level 6.

Table 57. Java SSL ciphers supported at security level 6
Java SSL ciphers
TLS_AES_256_GCM_SHA384
TLS_CHACHA20_POLY1305_SHA256
TLS_AES_128_GCM_SHA256
TLS_AES_128_CCM_8_SHA256
TLS_AES_128_CCM_SHA256

Table 58 on page 376 displays OpenSSL security ciphers that are supported by security level 6.

Table 58. OpenSSL ciphers supported at level 6 (`chsecurity -sslprotocol 6`)		
Cipher	**Bulk encryption algorithm**	**Hashing algorithm**
AES-256-GCM-SHA384	AES-256-GCM	SHA384
CHACHA20-POLY1305-SHA256	CHACHA20-POLY1305	SHA256
AES-128-GCM-SHA256	AES-128-GCM	SHA256
AES-128-CCM-8-SHA256	AES-128-CCM-8	SHA256
AES-128-CCM-SHA256	AES-128-CCM	SHA256

Table 59 on page 376 displays the protocols that are supported at security level 5.

Table 59. Protocols supported at level 5	
Protocol level	**Is it supported?**
TLS 1.3	Yes
TLS 1.2	Yes
TLS 1.1	No
TLS 1.0	No

Table 59. Protocols supported at level 5 (continued)	
Protocol level	**Is it supported?**
SSL 3 and earlier	No

Table 60 on page 377 displays Java SSL ciphers that are supported at security level 5.

Table 60. Java SSL ciphers supported at security level 5
Java SSL ciphers
For TLS 1.3
TLS_AES_256_GCM_SHA384
TLS_CHACHA20_POLY1305_SHA256
TLS_AES_128_GCM_SHA256
TLS_AES_128_CCM_8_SHA256
TLS_AES_128_CCM_SHA256
For TLS 1.2 The security level 5 supports all the Java SSL ciphers that are supported at the security level 3. A few more TLS 1.2 specific Java SSL ciphers that are supported at level 5 are:
TLS_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384
TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256
TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384
TLS_RSA_WITH_AES_256_GCM_SHA384
TLS_ECDH_ECDSA_WITH_AES_256_GCM_SHA384
TLS_ECDH_RSA_WITH_AES_256_GCM_SHA384
TLS_DHE_DSS_WITH_AES_256_GCM_SHA384
TLS_DHE_RSA_WITH_AES_256_GCM_SHA384
TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256
TLS_RSA_WITH_AES_128_GCM_SHA256
TLS_ECDH_ECDSA_WITH_AES_128_GCM_SHA256
TLS_ECDH_RSA_WITH_AES_128_GCM_SHA256
TLS_DHE_RSA_WITH_AES_128_GCM_SHA256
TLS_DHE_DSS_WITH_AES_128_GCM_SHA256

Table 61 on page 377 and Table 62 on page 378 displays OpenSSL security ciphers that are supported by security level 5.

Table 61. OpenSSL ciphers supported at level 5 for TLS 1.3 (`chsecurity -sslprotocol 5`)		
Cipher	**Bulk encryption algorithm**	**Hashing algorithm**
AES-256-GCM-SHA384	AES-256-GCM	SHA384
CHACHA20-POLY1305-SHA256	CHACHA20-POLY1305	SHA256

Table 61. OpenSSL ciphers supported at level 5 for TLS 1.3 (`chsecurity -sslprotocol 5`) (continued)		
Cipher	**Bulk encryption algorithm**	**Hashing algorithm**
AES-128-GCM-SHA256	AES-128-GCM	SHA256
AES-128-CCM-8-SHA256	AES-128-CCM-8	SHA256
AES-128-CCM-SHA256	AES-128-CCM	SHA256

Table 62. OpenSSL ciphers supported at level 5 for TLS 1.2 (`chsecurity -sslprotocol 5`)				
Cipher	**Kx**	**Au**	**Enc**	**Mac**
ECDHE-RSA-AES256-GCM-SHA384	ECDH	RSA	AESGCM(256)	AEAD
ECDHE-ECDSA-AES256-GCM-SHA384	ECDH	ECDSA	AESGCM(256)	AEAD
ECDHE-RSA-AES256-SHA384	ECDH	RSA	AES(256)	SHA384
ECDHE-ECDSA-AES256-SHA384	ECDH	ECDSA	AES(256)	SHA384
DHE-RSA-AES256-GCM-SHA384	DH	RSA	AESGCM(256)	AEAD
DHE-RSA-AES256-SHA256	DH	RSA	AES(256)	SHA256
AES256-GCM-SHA384	RSA	RSA	AESGCM(256)	AEAD
AES256-SHA256	RSA	RSA	AES(256)	SHA256
ECDHE-RSA-AES128-GCM-SHA256	ECDH	RSA	AESGCM(128)	AEAD
ECDHE-ECDSA-AES128-GCM-SHA256	ECDH	ECDSA	AESGCM(128)	AEAD
ECDHE-RSA-AES128-SHA256	ECDH	RSA	AES(128)	SHA256
ECDHE-ECDSA-AES128-SHA256	ECDH	ECDSA	AES(128)	SHA256
DHE-RSA-AES128-GCM-SHA256	DH	RSA	AESGCM(128)	AEAD
DHE-RSA-AES128-SHA256	DH	RSA	AES(128)	SHA256
AES128-GCM-SHA256	RSA	RSA	AESGCM(128)	AEAD
AES128-SHA256	RSA	RSA	AES(128)	SHA256

Table 63 on page 379 displays the protocols that are supported at security level 4.

Table 63. Protocols supported at level 4	
Protocol level	**Is it supported?**
TLS 1.2	Yes
TLS 1.1	No
TLS 1.0	No
SSL 3 and earlier	No

Table 64 on page 379 displays Java SSL ciphers that are supported at security level 4.

Table 64. Java SSL ciphers supported at security level 4
Java SSL ciphers
SSL_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384
SSL_DHE_DSS_WITH_AES_256_GCM_SHA384
SSL_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256
SSL_DHE_DSS_WITH_AES_128_GCM_SHA256

Table 65 on page 379 displays OpenSSL security ciphers that are supported by security level 4.

Table 65. OpenSSL ciphers supported at level 4 (`chsecurity -sslprotocol 4`)				
Cipher	**Kx**	**Au**	**Enc**	**Mac**
ECDHE-ECDSA-AES256-GCM-SHA384	ECDH	ECDSA	AESGCM(256)	AEAD
DHE-DSS-AES256-GCM-SHA384	DH	DSS	AESGCM(256)	AEAD
ECDHE-ECDSA-AES128-GCM-SHA256	ECDH	ECDSA	AESGCM(128)	AEAD
DHE-DSS-AES128-GCM-SHA256	DH	DSS	AESGCM(128)	AEAD

Table 66 on page 379 displays the protocols that are supported at security level 3.

Table 66. Protocols supported at level 3	
Protocol level	**Is it supported?**
TLS 1.2	Yes
TLS 1.1	No
TLS 1.0	No
SSL 3 and earlier	No

Table 67 on page 379 displays Java SSL ciphers that are supported at security level 3.

Table 67. Java SSL ciphers supported at security level 3
Java SSL ciphers
SSL_ECDHE_ECDSA_WITH_AES_256_CBC_SHA384

Table 67. Java SSL ciphers supported at security level 3 (continued)
Java SSL ciphers
SSL_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384
SSL_ECDHE_RSA_WITH_AES_256_CBC_SHA384
SSL_ECDHE_RSA_WITH_AES_256_GCM_SHA384
SSL_RSA_WITH_AES_256_CBC_SHA256
SSL_RSA_WITH_AES_256_GCM_SHA384
SSL_ECDH_ECDSA_WITH_AES_256_CBC_SHA384
SSL_ECDH_RSA_WITH_AES_256_GCM_SHA384
SSL_DHE_RSA_WITH_AES_256_CBC_SHA256
SSL_DHE_RSA_WITH_AES_256_CBC_SHA256
SSL_DHE_RSA_WITH_AES_256_GCM_SHA384
SSL_DHE_DSS_WITH_AES_256_CBC_SHA256
SSL_DHE_DSS_WITH_AES_256_GCM_SHA384
SSL_ECDHE_ECDSA_WITH_AES_256_CBC_SHA
SSL_ECDHE_RSA_WITH_AES_256_CBC_SHA
SSL_RSA_WITH_AES_256_CBC_SHA
SSL_ECDH_ECDSA_WITH_AES_256_CBC_SHA
SSL_ECDH_RSA_WITH_AES_256_CBC_SHA
SSL_DHE_RSA_WITH_AES_256_CBC_SHA
SSL_DHE_DSS_WITH_AES_256_CBC_SHA
SSL_ECDHE_ECDSA_WITH_AES_128_CBC_SHA256
SSL_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256
SSL_ECDHE_RSA_WITH_AES_128_CBC_SHA256
SSL_ECDHE_RSA_WITH_AES_128_GCM_SHA256
SSL_RSA_WITH_AES_128_CBC_SHA256
SSL_RSA_WITH_AES_128_GCM_SHA256
SSL_ECDH_ECDSA_WITH_AES_128_CBC_SHA256
SSL_ECDH_ECDSA_WITH_AES_128_GCM_SHA256
SSL_ECDH_RSA_WITH_AES_128_CBC_SHA256
SSL_ECDH_RSA_WITH_AES_128_GCM_SHA256
SSL_DHE_RSA_WITH_AES_128_CBC_SHA256
SSL_DHE_RSA_WITH_AES_128_GCM_SHA256
SSL_DHE_DSS_WITH_AES_128_CBC_SHA256
SSL_DHE_DSS_WITH_AES_128_GCM_SHA256

Table 68 on page 381 displays OpenSSL security ciphers that are supported by security level 3.

Table 68. OpenSSL ciphers supported at level 3 (`chsecurity -sslprotocol 3`)

Cipher	Kx	Au	Enc	Mac
ECDHE-RSA-AES256-GCM-SHA384	ECDH	RSA	AESGCM(256)	AEAD
ECDHE-ECDSA-AES256-GCM-SHA384	ECDH	ECDSA	AESGCM(256)	AEAD
ECDHE-RSA-AES256-SHA384	ECDH	RSA	AES(256)	SHA384
ECDHE-ECDSA-AES256-SHA384	ECDH	ECDSA	AES(256)	SHA384
DHE-DSS-AES256-GCM-SHA384	DH	DSS	AESGCM(256)	AEAD
DHE-RSA-AES256-GCM-SHA384	DH	RSA	AESGCM(256)	AEAD
DHE-RSA-AES256-SHA256	DH	RSA	AES(256)	SHA256
ECDH-RSA-AES256-GCM-SHA384 E	ECDH/RSA	ECDH	AESGCM(256)	AEAD
ECDH-ECDSA-AES256-GCM-SHA384	ECDH/ECDSA	ECDH	AESGCM(256)	AEAD
ECDH-RSA-AES256-SHA384	ECDH/RSA	ECDH	AES(256)	SHA384
ECDH-ECDSA-AES256-SHA384	ECDH/ECDSA	ECDH	AES(256)	SHA384
AES256-GCM-SHA384	RSA	RSA	AESGCM(256)	AEAD
AES256-SHA256	RSA	RSA	AES(256)	SHA256
ECDHE-RSA-AES128-GCM-SHA256	ECDH	RSA	AESGCM(128)	AEAD
ECDHE-ECDSA-AES128-GCM-SHA256	ECDH	ECDSA	AESGCM(128)	AEAD
ECDHE-RSA-AES128-SHA256	ECDH	RSA	AES(128)	SHA256
ECDHE-ECDSA-AES128-SHA256	ECDH	ECDSA	AES(128)	SHA256
DHE-DSS-AES128-GCM-SHA256	DH	DSS	AESGCM(128)	AEAD
DHE-RSA-AES128-GCM-SHA256	DH	RSA	AESGCM(128)	AEAD

Table 68. OpenSSL ciphers supported at level 3 (`chsecurity -sslprotocol 3`) (continued)				
Cipher	**Kx**	**Au**	**Enc**	**Mac**
DHE-RSA-AES128-SHA256	DH	RSA	AES(128)	SHA256
DHE-DSS-AES128-SHA256	DH	DSS	AES(128)	SHA256
ECDH-RSA-AES128-GCM-SHA256	ECDH/RSA	ECDH	AESGCM(128)	AEAD
ECDH-ECDSA-AES128-GCM-SHA256	ECDH/ECDSA	ECDH	AESGCM(128)	AEAD
ECDH-RSA-AES128-SHA256	ECDH/RSA	ECDH	AES(128)	SHA256
ECDH-ECDSA-AES128-SHA256	ECDH/ECDSA	ECDH	AES(128)	SHA256
AES128-GCM-SHA256	RSA	RSA	AESGCM(128)	AEAD
AES128-SHA256	RSA	RSA	AES(128)	SHA256

Table 69 on page 382 displays the protocols that are supported at security level 2.

Table 69. Protocols supported at level 2	
Protocol level	**Is it supported?**
TLS 1.2	Yes
TLS 1.1	No
TLS 1.0	No
SSL 3 and earlier	No

Table 70 on page 382 displays Java SSL ciphers that are supported at security level 2.

Table 70. Java SSL ciphers supported at level 2
SSL_ECDHE_ECDSA_WITH_AES_256_CBC_SHA384
SSL_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384
SSL_ECDHE_RSA_WITH_AES_256_CBC_SHA384
SSL_ECDHE_RSA_WITH_AES_256_GCM_SHA384
SSL_RSA_WITH_AES_256_CBC_SHA256
SSL_RSA_WITH_AES_256_GCM_SHA384
SSL_ECDH_ECDSA_WITH_AES_256_CBC_SHA384
SSL_ECDH_RSA_WITH_AES_256_GCM_SHA384
SSL_DHE_RSA_WITH_AES_256_CBC_SHA256
SSL_DHE_RSA_WITH_AES_256_CBC_SHA256
SSL_DHE_RSA_WITH_AES_256_GCM_SHA384

Table 70. Java SSL ciphers supported at level 2 (continued)
SSL_DHE_DSS_WITH_AES_256_CBC_SHA256
SSL_DHE_DSS_WITH_AES_256_GCM_SHA384
SSL_ECDHE_ECDSA_WITH_AES_256_CBC_SHA
SSL_ECDHE_RSA_WITH_AES_256_CBC_SHA
SSL_RSA_WITH_AES_256_CBC_SHA
SSL_ECDH_ECDSA_WITH_AES_256_CBC_SHA
SSL_ECDH_RSA_WITH_AES_256_CBC_SHA
SSL_DHE_RSA_WITH_AES_256_CBC_SHA
SSL_DHE_DSS_WITH_AES_256_CBC_SHA
SSL_ECDHE_ECDSA_WITH_AES_128_CBC_SHA256
SSL_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256
SSL_ECDHE_RSA_WITH_AES_128_CBC_SHA256
SSL_ECDHE_RSA_WITH_AES_128_GCM_SHA256
SSL_RSA_WITH_AES_128_CBC_SHA256
SSL_RSA_WITH_AES_128_GCM_SHA256
SSL_ECDH_ECDSA_WITH_AES_128_CBC_SHA256
SSL_ECDH_ECDSA_WITH_AES_128_GCM_SHA256
SSL_ECDH_RSA_WITH_AES_128_CBC_SHA256
SSL_ECDH_RSA_WITH_AES_128_GCM_SHA256
SSL_DHE_RSA_WITH_AES_128_CBC_SHA256
SSL_DHE_RSA_WITH_AES_128_GCM_SHA256
SSL_DHE_DSS_WITH_AES_128_CBC_SHA256
SSL_DHE_DSS_WITH_AES_128_GCM_SHA256
SSL_ECDHE_ECDSA_WITH_AES_128_CBC_SHA
SSL_ECDHE_RSA_WITH_AES_128_CBC_SHA
SSL_RSA_WITH_AES_128_CBC_SHA
SSL_ECDH_ECDSA_WITH_AES_128_CBC_SHA
SSL_ECDH_RSA_WITH_AES_128_CBC_SHA
SSL_DHE_RSA_WITH_AES_128_CBC_SHA
SSL_DHE_DSS_WITH_AES_128_CBC_SHA

Table 71 on page 384 displays OpenSSL security ciphers that are supported by security level 2.

Table 71. OpenSSL ciphers supported at level 2 (`chsecurity -sslprotocol 2`)				
Cipher	**Kx**	**Au**	**Enc**	**Mac**
ECDHE-RSA-AES256-GCM-SHA384	ECDH	RSA	AESGCM(256)	AEAD
ECDHE-ECDSA-AES256-GCM-SHA384	ECDH	ECDSA	AESGCM(256)	AEAD
ECDHE-RSA-AES256-SHA384	ECDH	RSA	AES(256)	SHA384
ECDHE-ECDSA-AES256-SHA384	ECDH	ECDSA	AES(256)	SHA384
DHE-DSS-AES256-GCM-SHA384	DH	DSS	AESGCM(256)	AEAD
DHE-RSA-AES256-GCM-SHA384	DH	RSA	AESGCM(256)	AEAD
DHE-RSA-AES256-SHA256	DH	RSA	AES(256)	SHA256
ECDH-RSA-AES256-GCM-SHA384 E	ECDH/RSA	ECDH	AESGCM(256)	AEAD
ECDH-ECDSA-AES256-GCM-SHA384	ECDH/ECDSA	ECDH	AESGCM(256)	AEAD
ECDH-RSA-AES256-SHA384	ECDH/RSA	ECDH	AES(256)	SHA384
ECDH-ECDSA-AES256-SHA384	ECDH/ECDSA	ECDH	AES(256)	SHA384
AES256-GCM-SHA384	RSA	RSA	AESGCM(256)	AEAD
AES256-SHA256	RSA	RSA	AES(256)	SHA256
AES256-SHA	RSA	RSA	AES(256)	SHA1
ECDHE-RSA-AES128-GCM-SHA256	ECDH	RSA	AESGCM(128)	AEAD
ECDHE-ECDSA-AES128-GCM-SHA256	ECDH	ECDSA	AESGCM(128)	AEAD
ECDHE-RSA-AES128-SHA256	ECDH	RSA	AES(128)	SHA256
ECDHE-ECDSA-AES128-SHA256	ECDH	ECDSA	AES(128)	SHA256
DHE-DSS-AES128-GCM-SHA256	DH	DSS	AESGCM(128)	AEAD
DHE-RSA-AES128-GCM-SHA256	DH	RSA	AESGCM(128)	AEAD

Table 71. OpenSSL ciphers supported at level 2 (`chsecurity -sslprotocol 2`) (continued)				
Cipher	**Kx**	**Au**	**Enc**	**Mac**
DHE-RSA-AES128-SHA256	DH	RSA	AES(128)	SHA256
DHE-DSS-AES128-SHA256	DH	DSS	AES(128)	SHA256
ECDH-RSA-AES128-GCM-SHA256	ECDH/RSA	ECDH	AESGCM(128)	AEAD
ECDH-ECDSA-AES128-GCM-SHA256	ECDH/ECDSA	ECDH	AESGCM(128)	AEAD
ECDH-RSA-AES128-SHA256	ECDH/RSA	ECDH	AES(128)	SHA256
ECDH-ECDSA-AES128-SHA256	ECDH/ECDSA	ECDH	AES(128)	SHA256
AES128-GCM-SHA256	RSA	RSA	AESGCM(128)	AEAD
AES128-SHA256	RSA	RSA	AES(128)	SHA256
AES128-SHA	RSA	RSA	AES(128)	SHA1
DES-CBC3-SHA	RSA	RSA	3DES(168)	SHA1

TCP and UDP ports

You can use firewall protections that restrict Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) ports. You can use external network communications to connect to these ports. Table 72 on page 385 lists all supported ports and describes how they can be used.

Table 72. TCP and UDP ports that are supported				
Service	**Traffic direction**	**Protocol**	**Port**	**Service type**
Email (SMTP) notification and inventory reports	Outbound	TCP	25	Optional
SNMP event notification	Outbound	UDP	162	Optional
Syslog event notification	Outbound	TCP UDP	6514 (TCP) 514 (UDP)	Optional
IPv4 DHCP (Node service address)	Outbound	UDP	68	Optional
IPv6 DHCP (Node service address)	Outbound	UDP	547	Optional
Network time server (NTP)	Outbound	UDP	123	Optional
SSH for command-line interface (CLI) access	Inbound	TCP	22	Mandatory
HTTP to HTTPS redirect for GUI access	Inbound	TCP	80	Optional
HTTPS redirect for GUI access	Inbound	TCP	443	Mandatory

Table 72. TCP and UDP ports that are supported (continued)

Service	Traffic direction	Protocol	Port	Service type
HTTP to HTTPS redirect for GUI access	Inbound	TCP	8080	Optional
HTTPS for GUI access	Inbound	TCP	8443	Mandatory
Remote user authentication service - HTTP	Outbound	TCP	16310	Optional
Remote user authentication service - HTTPS	Outbound	TCP	16311	Optional
Remote user authentication service - Lightweight Directory Access Protocol (LDAP)	Outbound	TCP	389	Optional
iSCSI	Inbound	TCP	3260	Optional
iSCSI iSNS	Outbound	TCP	3260	Optional
IP Partnership management IP communication	Inbound	TCP	3260	Optional
IP Partnership management IP communication	Outbound	TCP	3260	Optional
Long-distance partnerships by using TCP data path connections	Inbound	TCP	3265	Optional
Long-distance partnerships by using TCP data path connections	Outbound	TCP	3265	Optional
Ethernet Clustering data path connections	Inbound	TCP	21455	Optional
Ethernet Clustering data path connections	Outbound	TCP	21456	Optional
Short-distance partnerships by using RDMA data path connections	Inbound	TCP	3265	Optional
Short-distance partnerships by using RDMA data path connections	Outbound	TCP	3265	Optional
VASA Provider	Inbound	TCP	8440	Optional
RESTful API (HTTPS)	Inbound	TCP	7443	Optional

Note: The management GUI is accessed by using an HTTPS connection. For convenience, port 80 is left open but redirects all requests to use an HTTPS connection. The web server for the management GUI runs as a non-privileged process for more security, and requires these settings:

- Port 80 to be redirected to port 8080.
- Port 443 to be redirected to port 8443.

Table 73. SSH algorithms supported at each security level

Security level	Key Exchange	Cipher Suite	MAC Algorithm	Host Key Algorithms
1	• curve25519-sha256 • curve25519-sha256@libssh.org • ecdh-sha2-nistp256 • ecdh-sha2-nistp384 • ecdh-sha2-nistp521 • diffie-hellman-group-exchange-sha256 • diffie-hellman-group16-sha512 • diffie-hellman-group18-sha512 • diffie-hellman-group14-sha256 • diffie-hellman-group14-sha1 • diffie-hellman-group1-sha1 • diffie-hellman-group-exchange-sha1	aes256-ctr aes192-ctr aes128-ctr chacha20-poly1305@openssh.com aes256-gcm@openssh.com aes128-gcm@openssh.com aes256-cbc aes192-cbc aes128-cbc	hmac-sha2-256 hmac-sha2-512 hmac-sha2-256-etm@openssh.com hmac-sha2-512-etm@openssh.com hmac-sha1	rsa-sha2-256 rsa-sha2-512 ecdsa-sha2-nistp521 ecdsa-sha2-nistp521-cert-v01@openssh.com ssh-rsa ssh-rsa-cert-v01@openssh.com

Table 73. SSH algorithms supported at each security level (continued)

Security level	Key Exchange	Cipher Suite	MAC Algorithm	Host Key Algorithms
2	• curve25519-sha256 • curve25519-sha256@libssh.org • ecdh-sha2-nistp256 • ecdh-sha2-nistp384 • ecdh-sha2-nistp521 • diffie-hellman-group-exchange-sha256 • diffie-hellman-group16-sha512 • diffie-hellman-group18-sha512 • diffie-hellman-group14-sha256 • diffie-hellman-group14-sha1	aes256-ctr aes192-ctr aes128-ctr chacha20-poly1305@openssh.com aes256-gcm@openssh.com aes128-gcm@openssh.com	hmac-sha2-256 hmac-sha2-512 hmac-sha2-256-etm@openssh.com hmac-sha2-512-etm@openssh.com hmac-sha1	rsa-sha2-256 rsa-sha2-512 ecdsa-sha2-nistp521 ecdsa-sha2-nistp521-cert-v01@openssh.com ssh-rsa ssh-rsa-cert-v01@openssh.com
3	• curve25519-sha256 • curve25519-sha256@libssh.org • ecdh-sha2-nistp256 • ecdh-sha2-nistp384 • ecdh-sha2-nistp521 • diffie-hellman-group-exchange-sha256 • diffie-hellman-group16-sha512 • diffie-hellman-group18-sha512 • diffie-hellman-group14-sha256	aes256-ctr aes192-ctr aes128-ctr chacha20-poly1305@openssh.com aes256-gcm@openssh.com aes128-gcm@openssh.com	hmac-sha2-256 hmac-sha2-512 hmac-sha2-256-etm@openssh.com hmac-sha2-512-etm@openssh.com	rsa-sha2-256 rsa-sha2-512 ecdsa-sha2-nistp521 ecdsa-sha2-nistp521-cert-v01@openssh.com ssh-rsa ssh-rsa-cert-v01@openssh.com

Table 73. SSH algorithms supported at each security level (continued)

Security level	Key Exchange	Cipher Suite	MAC Algorithm	Host Key Algorithms
4	• curve25519-sha256 • curve25519-sha256@libssh.org • ecdh-sha2-nistp256 • ecdh-sha2-nistp384 • ecdh-sha2-nistp521	aes256-ctr aes192-ctr aes128-ctr aes256-gcm@openssh.com aes128-gcm@openssh.com	hmac-sha2-256 hmac-sha2-512 hmac-sha2-256-etm@openssh.com hmac-sha2-512-etm@openssh.com	rsa-sha2-256 rsa-sha2-512 ecdsa-sha2-nistp521 ecdsa-sha2-nistp521-cert-v01@openssh.com

Restriction: The 3-site-orchestrator does not support SSH protocol level 4.

Interoperability

At SSL security level 4, Google Chrome Version 63.0.3239.132 and higher and Mozilla Firefox Version 52.7.2 and later are known to work with the management GUI. IBM SDK, Java Technology Edition, Version 8 update 1.8.0_161 and later is known to work with the IP quorum application.

SSH rules

Use the **SSH Rules** page to customize settings related to secure shell (SSH) access. SSH is used to authenticate users to the command-line interface.

To update the SSH rules settings, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **SSH Rules**.
2. You can update any of the following details:

   SSH login grace period (seconds)
   Indicates the amount of time in seconds to log in before SSH times out. The range is 15 - 1800.

   Maximum login attempts (SSH)
   Indicates the total number of login attempts allowed per single SSH connection. The range is 1 - 10.
3. Click **Save**.

System Certificates

You can manage the system certificate to secure connections between systems or applications across networks using management GUI or command line interface (CLI).

About this task

Before you create system certificates, determine how certificates are signed based on your current security requirements. The system supports the following ways to sign system certificates:

Self-signed certificate
In 8.5.3.0, self-signed certificates are no longer created by default. After an update to 8.5.3.0 or later, the current self-signed certificate is usable until it expires. Before the self-signed certificate expires, you must update the system certificate to either an internally signed certificate or an externally signed certificate.

Internally signed certificate

The system has a root certificate authority (CA) that can be used to create internally signed system certificates. In 8.5.3.0 or later, system setup creates a certificate that is signed by the root CA to secure connections between the management GUI and the browser. The root certificate can be exported from the system and added to truststores on other systems, browsers, or devices to establish trust. Internally signed certificates can be renewed automatically before they expire. Automatic renewal simplifies the certificate renewal process and prevents security warnings from expired certificates. Automatic renewal is only supported by using an internally signed certificate.

Externally signed certificate

Externally signed certificates are issued and signed by a trusted third-party provider of certificates, called an external certificate authority (CA). This CA can be a public CA or your own organization's CA. Most web browsers trust well-known public CAs and include the root certificate for these CAs in the device or application. Externally signed certificates cannot be renewed automatically because they must be issued by the external CA. Externally signed certificates must be manually updated before they expire by creating a new certificate signing request (CSR) on the system and supplying it to the CA. The CA signs the request and issues a certificate that must be installed on the system. The system raises a warning in the event log 30 days before the certificate expires.

Consider the following additional information about certificates:

Root certificate

From version 8.5.3.0, IBM Storage Virtualize has a root certificate authority that can be used to generate system certificates. The root CA is generated during the first boot on 8.5.3.0 and is shared between all nodes in the clustered system. The root CA cannot be modified by the user.

The root certificate can be exported from the system and added to truststores for web browsers, Thales CipherTrust Manager key servers, and other devices and applications that support chain of trust checking. If chain of trust checking is supported, then the signed system certificate can be renewed and does not need to be exported.

The root certificate can be identified by its subject and issuer distinguished name:

- O represents International Business Machines Corporation.
- OU represents IBM Storage Virtualize Root CA.
- CN represents serial number, for example, 1234567. The common name (CN) field is the enclosure serial number of the configuration node that generated the certificate.

The root certificate is valid for 20 years and uses a 4096-bit RSA key pair.

After the first system upgrade to 8.5.3.0 (or later), the system continues to use its current certificate (either self-signed or externally signed) until a new certificate is generated.

System certificate

The system certificate is signed by the system's root CA, a trusted third-party CA, or is self-signed. This certificate is presented to other devices such as web browsers, key servers, or partner systems in IP partnerships, to authenticate the system and create a secure connection.

The system certificate supports the following key types:

- RSA 4096
- RSA 2048
- ECDSA 521
- ECDSA 384

SSL protocol level 4 restricts the use of RSA key exchange cipher suites. Therefore, you cannot enable SSL protocol level 4 when using a certificate with an RSA key. Likewise, you cannot generate a certificate with an RSA key if SSL protocol level 4 is set.

Note:

- Multifactor Authentication with IBM Security Verify uses the system certificate to sign JSON web tokens (JWTs). The system certificate must be exported and added as a signer certificate in the IBM

Security Verify interface. If the system certificate expires or is renewed, access to the management GUI will be unavailable until the new system certificate is installed in IBM Security Verify. A user must use the CLI to log in to the system to export the new system certificate and install it in IBM Security Verify.

- IBM Security Guardium® Key Lifecycle Manager key servers do not currently support chain of trust checking with IBM Storage Virtualize. The system certificate must be installed on the IBM Security Guardium Key Lifecycle Manager key servers in order to establish a connection. If the system certificate is expired or renewed, then the new system certificate must be installed on the key servers in order for IBM Storage Virtualize to establish a connection with the key servers. Access to encrypted storage is not disrupted while the connection to the key servers is unavailable, unless all nodes in the clustered system are powered off and on at the same time.

During system setup, an initial certificate is created to use for secure connections between web browsers. This certificate is signed by the system's root CA. A new certificate should be generated which includes the relevant DNS or IP entries for the system in the Subject Alternate Name field. The **System Certificates** page in the management GUI suggests the DNS names if a DNS server is added to the system. If a DNS server is not added, then the management GUI suggests the IP addresses.

Configuring system certificates

Secure communications between systems requires that certificates are configured on the system. To ensure that the certificate and public ID of the system are valid and secure, each system has a digital Secure Sockets Layer (SSL) certificate.

About this task

Before you create a request for a certificate, ensure that your current browser does not have restrictions on the type of keys that are used for certificates. Some browsers limit the use of specific types of keys for security and compatibility issues. If your system supports key server-based encryption, the system also uses certificates to secure communications between key servers that are used to distribute and manage encryption keys to the system. If a certificate is changed, the certificate must also be updated on all configured key servers, or access to encrypted data can be lost.

If the system certificate is signed by an intermediate certificate authority (CA), then the full chain of certificates must be installed. To install the signed certificate and certificate authority (CA) certificates, create a single file that contains the full chain of certificates. The file should include the signed certificate and the intermediate CA certificates. The root CA certificate can be included but is optional.

To manage the certificate that is installed on a system, use either the **`svctask chsystemcert`** command-line interface (CLI) command or click **Settings** > **Security** > **System Certificates** in the management GUI to do the following tasks:

- Upload signed certificate
- Update certificate
- Cancel outstanding signing request
- Export system certificate
- Export root certificate

For specific directions on completing these tasks, see “System Certificates” on page 389.

Updating an internally signed certificate

You can create an internally signed certificate for the system by using the management GUI or command-line interface (CLI).

Note: Before updating the certificates, keep the following factors into consideration:

- Updating the certificate logs you out of the current management GUI session, requiring a fresh login. For features that support chain of trust checking, the secure connection is not interrupted when updating the internally signed certificate.

- IBM Security Guardium Key Lifecycle Manager key servers do not currently support chain of trust checking with IBM Storage Virtualize. The new internally signed certificate must be exported to the key servers to re-establish a secure connection.
- If you are using IBM Security Verify multifactor authentication, users who are required to use multifactor authentication cannot log in to the management GUI after the certificate is renewed. The new internally signed certificate must be exported using the CLI and added as a new signer certificate to IBM Security Verify for successful authentication.

Using the management GUI

To configure an internally-signed certificate, complete these steps:

1. If the root certificate has not already been exported, export the root certificate to other systems, web browsers, and devices that require secure communications with the system. For more information, see .
2. In the management GUI, select **Settings** > **Security** > **System Certificates**.
3. On the System Certificate page, the current certificate details are displayed.

   Note: The system does not support the creation of new self-signed certificates. If you currently use a self-signed certificate you can continue to use it until it expires, or generate an internally signed certificate or an externally signed certificate.

   Automatic renewal of the system certificate

   The system certificate can be renewed automatically if it is signed by the system root CA. Turn automatic renewal on by going to **Settings** > **Security** > **System Certificates** and setting automatic renewal to **On**. The default validity period of the system certificate is one year. If automatic renewal is **On**, the system certificate is renewed thirty days before the expiry date. If the validity period of the system certificate is fewer than thirty days, then the certificate will be renewed every eight hours.

   The renewed certificate contains all of the same field values, key type and validity period details as the previous certificate.
4. From the icon menu, select **Update Certificate**.
5. Select **Internally Signed Certificate** for the certificate type.
6. If you are already using certificates, the **Certificate Details** are automatically populated. You can update any of the following details:

   Key type
   Select the cryptographic key type that is used to generate the certificate.

   Validity days
   Enter the number of days the certificate is valid for. The maximum number of days that are allowed is 9000.

   Country
   Enter the two-letter country code or location, for example, 01 for US.

   State
   Enter the name of the state where the system requesting the certificate is located.

   City
   Enter the name of the city where the system is located.

   Organization name
   Enter the name of the organization.

   Organizational unit
   Enter the name of organizational unit.

   Common name
   Enter the common name for the certificate.

Subject alternative name
Subject alternative name is the hostname of the system.

Web browsers, and other features that use certificate authentication, require a **Subject Alternative Name**, which is an extension to the Internet standard for public key certificates. The **Subject Alternative Name** extension is used to match the domain name and site certificate and can be an email address, an IP address, a URI, or a DNS name. A certificate can contain a collection of these values so that the certificate can be used on multiple sites.

The **Subject Alternative Name** field can include the management IP addresses for the cluster or DNS names, the service IP addresses for each node in the cluster or DNS names, and any IP addresses configured for IP replication.

For example, if the system has a management DNS name of `cluster.company.com`, and service DNS names of `node1.company.com` and `node2.company.com`, enter these values in the **Subject Alternate Name** field. For multiple values, list each value on a separate line within the box of the **Subject Alternate Name** field:

```
DNS:cluster.company.com
DNS:node1.company.com
DNS:node2.company.com
IP:196.192.0.20
```

Email Address
Enter the email address.

7. Click **Update**. The certificate is updated in the main panel.
8. If using IBM Security Guardium Key Lifecycle Manager which does not currently support chain of trust checking, export the new system certificate and install it on the key servers.

   If using multifactor authentication with IBM Security Verify which uses the system certificate as a signer certificate, export the new system certificate and install it as a signer certificate in IBM Security Verify. The management GUI is unavailable until the new certificate is added, so the CLI must be used to export the new certificate.

Using the command-line interface (CLI)

Use the following steps to generate an internally signed certificate in the command-line interface.

To generate an internally signed certificate that uses RSA 2048 key type and expires in one year, enter the following command:

```
chsystemcert -mksystemsigned -commonname virtualize -country GB -locality Manchester -org
IBM -orgunit Systems -email certificates@support.ibm.com -keytype rsa2048 -validity 365
-subjectalternativename "DNS:test.ibm.com"
```

After the internally signed certificate is created, it is automatically installed on the system.

Note: You must use the -subjectalternativename parameter to include the management hostname or IP address.

For example, to add a DNS name to the Subject Alternative Name extension, include the following parameter in the **chsystemcert** CLI command: **-subjectalternativename "DNS:dns.mysystem.com"** For multiple values, use a recommended delimiter to separate each entry for the **-subjectalternativename** parameter.
Delimiters can be mixed:

Table 74. Recommended delimiters		
Delimiter Name	**Symbol**	**Example**
Space	(space)	**-subjectalternativename "DNS:dns.myco.com IP:1.2.3.20 URI:http:\ \www.myco.com email:support@myco.com"**
Comma	(,)	**-subjectalternativename "DNS:dns.myco.com,IP:1.2. 3.20,URI:http:\ \www.myco.com,email:suppo rt@myco.com"**
Semi-colon	(;)	**-subjectalternativename "DNS:dns.myco.com;IP:1.2. 3.20;URI:http:\ \www.myco.com;email:suppo rt@myco.com"**
Newline (for Linux or UNIX operating systems)	(\n)	**-subjectalternativename "DNS:dns.myco.com\nIP:1.2 .3.20\nURI:http:\ \www.myco.com\nemail:supp ort@myco.com"**
Tab (for Linux or UNIX operating systems)	(\t)	**-subjectalternativename "DNS:dns.myco.com\tIP:1.2 .3.20\tURI:http:\ \www.myco.com\temail:supp ort@myco.com"**
Carriage return (for Windows operating systems)	(\r)	**-subjectalternativename "DNS:dns.myco.com\rIP:1.2 .3.20\rURI:http:\ \www.myco.com\remail:supp ort@myco.com"**
Carriage return with newline (for Windows operating systems)	(\r\n)	**-subjectalternativename "DNS:dns.myco.com\r\nIP:1 .2.3.20\r\nURI:http:\ \www.myco.com\r\nemail:su pport@myco.com"**

For more information about supported delimiters, see the **chsystemcert** CLI command.

Requesting and installing an externally signed certificate

If your current externally signed certificate expires or is about to expire, you can request a new signed certificate from a third-party certificate authority. The externally signed certificate can be installed by using the management GUI or command line interface (CLI).

Note: Before requesting and installing the certificates, take the following factors into consideration:

- The certificate signing request must be generated on the system. IBM Storage Virtualize does not support certificate signing requests that are generated outside of the system.
- The signed certificate must contain the following X509v3 Key Usage extensions: Digital Signature. It must also contain the following X509v3 Extended Key Usage Extensions: TLS Web Server

Authentication, TLS Web Client Authentication, Any Extended Key Usage. Ensure that the Certificate Authority (CA) used to sign the certificate includes these extensions.

- Updating the certificate logs you out of the current management GUI session, requiring a fresh login. For features that allow chain of trust checking, the secure connection is not interrupted when updating the externally signed certificate.
- By default, IBM Security Guardium Key Lifecycle Manager key servers do not currently enable chain of trust checking with IBM Storage Virtualize. From version 4.2 onwards, this setting can be enabled. Consider enabling this setting if signed certificate chains are used in order to avoid the export of refreshed certificates to the key server.
- If you are using multifactor authentication with IBM Security Verify, the management GUI are unavailable when you update the certificate. The new externally signed certificate must be exported using the CLI and added as a new signer certificate to IBM Security Verify for successful authentication.

In the management GUI, select **Settings** > **Security** > **System Certificates** and select update certificate from the icon menu. In the Update Certificate page, select externally signed certificate and complete the form to create a request for a signed certificate for your system. The resulting certificate signing request (CSR) should be exported to the third-party certificate authority so that it can be signed. After you receive the externally signed certificate from the third-party certificate authority, use the management GUI or command line interface (CLI) to upload and install the signed certificate on the system.

Using the management GUI

To configure an externally signed certificate, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **System Certificates**.
2. From the icon menu, select **Update Certificate**.
3. In the **Update Certificate** page, select **Certificate type** as **Externally Signed Certificate**.
4. If you are already using certificates, the **Certificate Details** are automatically populated. You can update any of the following details:

   Key type
   Select the cryptographic key type that is used to generate the certificate.

   State
   Enter the name of the state where the system requesting the certificate is located.

   City
   Enter the name of the city where the system is located.

   Organization name
   Enter the name of the organization.

   Organizational unit
   Enter the name of organizational unit.

   Common name
   Enter the common name for the certificate.

   Email Address
   Enter the email address.

   Subject alternative name
   Subject alternative name is the hostname of the system.

   Web browsers, and other features that use certificate authentication, require a **Subject Alternative Name**, which is an extension to the Internet standard for public key certificates. The **Subject Alternative Name** extension is used to match the domain name and site certificate and can be an email address, an IP address, a URI, or a DNS name. A certificate can contain a collection of these values so that the certificate can be used on multiple sites.

   The **Subject Alternative Name** field can include the management IP addresses for the cluster or DNS names, the service IP addresses for each node in the cluster or DNS names, and any IP addresses configured for IP replication.

For example, if the system has a management DNS name of `cluster.company.com`, and service DNS names of `node1.company.com` and `node2.company.com`, enter these values in the **Subject Alternate Name** field. For multiple values, list each value on a separate line within the box of the **Subject Alternate Name** field:

```
DNS:cluster.company.com
DNS:node1.company.com
DNS:node2.company.com
IP:196.192.0.20
```

5. Click **Generate Request**. This action presents a dialogue window to download the file. Select a location on your local machine to save the file.
6. If the signing request is generated, you have the option to cancel the outstanding signing request in case of an error. To cancel any outstanding signing request, complete these steps:
   a. Select **Settings** > **Security** > **System Certificates**.
   b. From the icon menu, select **Cancel Outstanding Signing Request**.
   c. Click **Yes**.
7. Share the generated CSR file to the trusted third-party CA. If the CA is a public CA, then it can take some time for the CA to verify your identity before issuing the signed certificate. When it is ready, download the signed certificate file from the CA. You can also download any intermediate CA certificates that were used to sign the request. The files must all be in PEM format.
8. If intermediate CAs are used to sign the certificate request, create a single file that contains the contents of the signed certificate and the contents of each intermediate CA certificate, concatenated together. The root CA certificate is not mandatory, but can optionally be included.
9. Select **Upload Signed Certificate**.
10. On the **Upload Signed Certificate** page, click **Add file** to select the new signed certificate that you downloaded to your device or laptop.
11. Click **Upload**. This action uploads the certificate onto the system. For more information, see .
12. If you are using IBM Security Guardium Key Lifecycle Manager key server, which does not currently support chain of trust checking, export the new system certificate and install it on the key server.
13. If you are using multifactor authentication with IBM Security Verify, which uses the system certificate as a signer certificate, export the new system certificate and install it as a signer certificate in IBM Security Verify. The management GUI is unavailable until the new certificate is added as a signer certificate to IBM Security Verify.

Using the command-line interface (CLI)

In the command-line interface, enter the following command to create a new certificate request and upload the new certificate:

1. Enter the following command to create a new certificate request:

```
chsystemcert -mkrequest -keytype ecdsa521 -country GB -state Hampshire -locality
Hursley -org MYCO -orgunit Storage -commonname svcsystem1.myco.com -email admin@myco.com
-subjectalternativename "DNS:test.ibm.com"
```

The certificate request is automatically written to `/dumps/certificate.csr`.

For example, to add a DNS name to the Subject Alternative Name extension, include the following parameter in the **chsystemcert** CLI command: **-subjectalternativename "DNS:dns.mysystem.com"** For multiple values, use a recommended delimiter to separate each entry for the **-subjectalternativename** parameter.
Delimiters can be mixed:

Table 75. Recommended delimiters

Delimiter Name	Symbol	Example
Space	(space)	**-subjectalternativename "DNS:dns.myco.com IP:1.2.3.20 URI:http:\ \www.myco.com email:support@myco.com"**
Comma	(,)	**-subjectalternativename "DNS:dns.myco.com,IP:1.2 .3.20,URI:http:\ \www.myco.com,email:supp ort@myco.com"**
Semi-colon	(;)	**-subjectalternativename "DNS:dns.myco.com;IP:1.2 .3.20;URI:http:\ \www.myco.com;email:supp ort@myco.com"**
Newline (for Linux or UNIX operating systems)	(\n)	**-subjectalternativename "DNS:dns.myco.com\nIP:1. 2.3.20\nURI:http:\ \www.myco.com\nemail:sup port@myco.com"**
Tab (for Linux or UNIX operating systems)	(\t)	**-subjectalternativename "DNS:dns.myco.com\tIP:1. 2.3.20\tURI:http:\ \www.myco.com\temail:sup port@myco.com"**
Carriage return (for Windows operating systems)	(\r)	**-subjectalternativename "DNS:dns.myco.com\rIP:1. 2.3.20\rURI:http:\ \www.myco.com\remail:sup port@myco.com"**
Carriage return with newline (for Windows operating systems)	(\r\n)	**-subjectalternativename "DNS:dns.myco.com\r\nIP: 1.2.3.20\r\nURI:http:\ \www.myco.com\r\nemail:s upport@myco.com"**

For more information about supported delimiters, see the **chsystemcert** CLI command.

2. Use secure copy (scp) to copy the file /dumps/certificate.csr from the system to your local machine. Share the generated CSR file to the trusted third-party CA. If the CA is a public CA, then it may take some time for the CA to verify your identity before issuing the signed certificate. When it is ready, download the signed certificate file from the CA. You should also download any intermediate CA certificates that were used to sign the request. The files must all be in PEM format.
3. If the intermediate CA's are used to sign the certificate request, create a single certificate chain file that contains the contents of the signed certificate and the contents of each intermediate CA certificate, concatenated together. The root CA certificate is not mandatory, but can optionally be included.
4. Use secure copy (scp) to copy the certificate back onto the system in the file /dumps/certificate.pem, where certificate.pem is the name of the certificate.
5. After you copy the signed CA certificate to the system, enter the following command:

```
chsystemcert -install -file /dumps/certificate.pem
```

where **/dumps/certificate.pem** is the absolute path name of the signed certificate or certificate chain file.

Export certificates

You can export the current root certificate and system certificate by using the management GUI or command line interface (CLI).

You must add the system's root certificates and system certificates to the trust stores of the other systems, web browsers, key servers or any other devices to authenticate the identity and secure the communication channel. If the browser, device, or application supports chain of trust checking, then only the root certificate is needed.

Using the management GUI

To export the root certificate, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **System Certificates**.
2. From the icon menu, select **Export Root Certificate**.
3. In the **Export Root Certificate** page, you can either download the certificate file by clicking **Export Certificate** or copy the contents of the file to clipboard by clicking **Copy to clipboard**.

To export the system certificate, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **System Certificates**.
2. From the icon menu, select **Export System Certificate**.
3. In the **Export System Certificate** page, you can either download the certificate file by clicking **Export Certificate** or copy the contents of the file to clipboard by clicking **Copy to clipboard**.

Using the command line interface (CLI)

In the command-line interface, enter the following command to export the certificate:

1. Enter the following command to export the certificates to file `/dumps/certificate.pem`

```
chsystemcert -export
```

Note: The **chsystemcert -export** command exports the full chain of trust. Alternatively, to export the system's root certificate before the system initialization, use the **satask exportrootcertificate** command. You must run this command on the node that is used to do the system initialization. You can also use the **sainfo lsrootcertificate** command to view the details of the system's root certificate.

2. After the certificate is exported, use secure copy (scp) to copy the certificate from the system.

Creating and managing certificate authority store by using CLI

You can create and manage certificate authority stores by using the command-line interface (CLI). The certificate authority stores contain certificates from external devices that the system uses to establish trust with those devices.

Using the command-line interface (CLI)

To create and manage certificate store, complete the following steps:

1. To create the repository to store a collection of certificates, enter the following command:

```
mktruststore -file file_path -name truststore_name -ipsec on|off
```

where *file_path* specifies the location of the file to import the certificates, *truststore_name* specifies the name of the new certificate store, and *on*|*off* specifies the certificates in the store are used for the secured IP partnership service.

2. To manage the certificate store tags (set or unset), and permits appending more certificates, enter the following command:

```
chtruststore -file file_path -name truststore_name -ipsec on|off <object_id | object_name>
```

where *file_path* specifies the location of the file to import the certificates to append those already in the store, *truststore_name* specifies the name of the new certificate store, *on*|*off* specifies the certificates in the store are used for the secured IP partnership service, and *<object_id | object_name>* specifies the ID or name of the truststore to manage.

3. To display the current certificate stores and to check the space left, enter the following command:

```
lstruststore -bytes <object_id | object_name>
```

where the **-bytes** parameter displays the space in bytes for the specified truststore and *<object_id | object_name>* specifies ID or name of the truststore to display.

Removing certificate authority store by using CLI

You can remove certificate authority store by using command-line interface (CLI). The certificate authority stores contain certificates from external devices that the system uses to establish trust with those devices.

Using the command-line interface (CLI)

To remove the store from the memory, enter the following command:

```
rmtruststore <object_id | object_name>
```

where *<object_id | object_name>* specifies the ID or name of the truststore to remove.

The authentication certificates are kept in a system-wide truststore and managed with a set of CLI commands. For more information, see **Truststore management commands**.

Multifactor authentication

Multifactor authentication requires users to provide multiple pieces of information when they log in to the system to prove their identity. Multifactor authentication uses any combination of two or more methods, called *factors*, to authenticate users to your resources and protect those resources from unauthorized access.

One of the key concepts of multifactor authentication is each factor comes from a different category. These categories include the following:

Something a user knows
Users authenticate with information that only each individual user knows, such as a password or PIN.

Something a user has
The users prove their identity with information that is given to the user by a trusted authentication service, such as one-time passcodes that are generated by an application or mobile device.

Something a user is
Users prove their identity with biometrics, such as fingerprint or retinal scans.

With the adoption of cloud-based services, multifactor authentication increases the control over user access and security settings. First-factor authentication methods alone, such as username and password combinations, do not provide the level of protection and security that is required in cloud and hybrid-cloud environments. With multifactor authentication support, security administrators can reinforce account protection, create granular access for users and user groups, and monitor access more efficiently at a system level.

Multifactor authentication with IBM Security Verify

With IBM Security Verify, security administrators can configure the system as an application that requires two factors for users and user groups to access the system with either the management GUI or CLI.

Multifactor authentication can be used to protect both local users, including superuser, and remote users.

Remote users are users who are defined on a remote LDAP server. For remote users that authenticate with LDAP servers, install and configure IBM Security Verify Bridge for Directory Sync on your LDAP server, such as Windows Active Directory. IBM Security Verify Bridge for Directory Sync duplicates any users and groups that are defined on the source LDAP server into the Cloud Directory in IBM Security Verify. Any subsequent changes that are made to the source LDAP server are copied automatically to the Cloud Directory in IBM Security Verify. For more information, see IBM Security Verify Bridge for Directory Sync in the IBM Security Verify documentation.

IBM Security Verify configures the management GUI and the command-line interface as separate API clients that require separate credentials. For GUI-based logins, the system communicates with IBM Security Verify through the OpenID Connect (OIDC) protocol.

Multifactor authentication with Duo Security

The system integrates with IBM Security Verify or Duo Security , which are cloud-based identity and access management (IAM) service providers. These services provide different factors to validate and verify users who access the system.

You can manage every aspect of your Duo two-factor authentication system from the Duo Admin Panel including creating and managing applications, enrolling and activating users, issuing and managing SMS passcodes and bypass codes, managing mobile devices, fine-tuning the user experience of your Duo installation.

Related information

“Configuring multifactor authentication with IBM Security Verify” on page 401
The system integrates with IBM Security Verify to provide multifactor authentication for system users.

“Configuring user groups for multifactor authentication with IBM Security Verify” on page 405
After you configure multifactor authentication on the system, you must enable multifactor authentication for user groups and add users to those groups in IBM Security Verify.

IBM Security Verify documentation

“Configuring multifactor authentication with Duo Security” on page 415
The system integrates with Duo Security to provide multifactor authentication for system users.

“Configuring user groups for multifactor authentication with Duo Security” on page 417
After you configure multifactor authentication on the system, you must enable multifactor authentication for user groups and add users to those groups in Duo Security.

Duo Security documentation

Configuring multifactor authentication

Multifactor authentication requires users to provide multiple pieces of information when they log in to the system to prove their identity. Multifactor authentication uses any combination of two or more methods, called *factors*, to authenticate users to your resources and protect those resources from unauthorized access.

With the adoption of cloud-based services, multifactor authentication increases the control over user access and security settings. First-factor authentication methods alone, such as username and password combinations, do not provide the level of protection and security that is required in cloud and hybrid-cloud environments. With multifactor authentication support, security administrators can reinforce account protection, create granular access for users and user groups, and monitor access more efficiently at a system level.

The system integrates with IBM Security Verify or Duo Security , which are cloud-based identity and access management (IAM) service providers. These services provide different factors to validate and verify users who access the system.

Configuring multifactor authentication with IBM Security Verify

The system integrates with IBM Security Verify to provide multifactor authentication for system users.

With IBM Security Verify, security administrators can configure the system as an application that requires two factors for users and user groups to access the system with either the management GUI or CLI.

Multifactor authentication can be used to protect both local users, including superuser, and remote users.

Remote users are users who are defined on a remote LDAP server. For remote users that authenticate with LDAP servers, install and configure IBM Security Verify Bridge for Directory Sync on your LDAP server, such as Windows Active Directory. IBM Security Verify Bridge for Directory Sync duplicates any users and groups that are defined on the source LDAP server into the Cloud Directory in IBM Security Verify. Any subsequent changes that are made to the source LDAP server are copied automatically to the Cloud Directory in IBM Security Verify. For more information, see IBM Security Verify Bridge for Directory Sync in the IBM Security Verify documentation.

IBM Security Verify configures the management GUI and the command-line interface as separate API clients that require separate credentials. For GUI-based logins, the system communicates with IBM Security Verify through the OpenID Connect (OIDC) protocol.

Important:

When the system certificate added as the signer certificate expires, you cannot login to the management GUI if you are using multifactor authentication. To export the new system certificate and install it in IBM Security Verify, login using the CLI.

Prerequisites

Ensure that the following prerequisite tasks are completed on the system before you configure multifactor authentication IBM Security Verify:

1. Ensure the system is updated to 8.5.0 or later release.
2. Configure a DNS server. To create a DNS server, select **Settings** > **Network** > **DNS**.
3. Configure an HTTP proxy server or configure your firewall to access IBM Security Verify. To create an HTTP proxy server, **Settings** > **Network** > **Internal Proxy Server**. If your system does not directly connect to the Internet, you can create a firewall exception to allow your system access to IBM Security Verify.

   Important: If the proxy server is not configured correctly, the system cannot communicate with IBM Security Verify and the login to the system fails.
4. For the management GUI and the command-line interface, ensure that the inactivity logout is equal to or greater than the time it takes for a user to receive a one-time passcode (OTP) from the authentication service. The default value for the inactivity timeout is 30 minutes for the management GUI and 15 minutes for the CLI. To set the inactivity timeout for both interfaces in the management GUI, select **Settings** > **Security** > **Inactivity Logout**.
5. Ensure that the SSH grace time for the command-line interface is equal to or greater than the time it takes for a user to receive a one-time passcode (OTP) from the authentication service. The default value for the SSH grace time period is 60 seconds. To set SSH grace time on the system in the command-line interface, use the **`chsecurity -sshgracetime`** command.

The following prerequisite steps on IBM Security Verify must be completed before you can configure multifactor authentication on the system:

1. Create a subscription for IBM Security Verify. You need an IBMid to create a subscription. A 90-day free trial subscription is also available. For more information, see Cloud identity and access management (IAM) solutions. During subscription creation, you specify a tenant that is used to create a URL to access the IBM Security Verify dashboard.

2. Access the IBM Security Verify administrator dashboard by entering the following URL in a web browser:

   ```
   https://tenant.verify.ibm.com/ui/admin
   ```

   Where `tenant` is the name of the tenant that you specified when you created your subscription. Usually this tenant name is associated with your company or organization.
3. In the IBM Security Verify interface, select **Applications** > **Applications** > **Add application**.
4. Select **IBM Storage Virtualize** > **Add Application**.

   Note: Each system must be added as a separate application.

   The following table shows the required fields and actions for the **General** tab in the IBM Security Verify interface.

Table 76. General tab

Field	**Action**
Name	Enter a name to identify the system on IBM Security Verify. If you are adding multiple systems, enter a unique name.
Description	Enter a brief description of the system.
Company name	Name of organization or company.

The following table shows the required fields and actions for the **Sign-on** tab in the IBM Security Verify interface. The Sign-on tab is used to add the management GUI as an API-based client.

Table 77. Sign-on tab

Field	**Action**	**Details**
Application URL	Enter the URL for your system.	Enter the URL that is used to access the management GUI.
Grant type	Select **Authorization code** and **JWT bearer**.	Two grant types are required for setting up MFA for the system. **Authorization code** indicates that the client can request access to protected resources on behalf of users.
Client ID	This value is automatically generated when the system is saved as an application.	This value must be entered to the Multifactor authentication page in the management GUI under **OpenID Credentials**.
Client secret	This value is automatically generated when the system is saved as an application.	This value must be entered to the Multifactor authentication page in the management GUI under **OpenID Credentials**.
User consent	Select **Do not ask for consent**.	

Table 77. Sign-on tab (continued)

Field	Action	Details
Redirect URIs	Enter the locations where the authorization server sends users after they are successfully authorized and granted an authorization code or access token.	Multiple redirect URIs can be specified for both the management GUI and the service assistant GUI. For management GUI access, the redirect URI is comprised of the management IP address or hostname followed by `/mfa`. For the service assistant interface, the redirect URI is comprised of the hostname or IP address for the system followed by `service/mfa`. For example: https://hostname.com/mfa https://hostname/service/mfa
JWT bearer user identification	Select **Username**.	Indicates that the username field in the JWT bearer is used to find users in the Cloud Directory and determines what second factors IBM Security Verify presents to users when they log into the system.
JWT bearer default identity source	Ensure **Cloud Directory** is selected.	Indicates that the IBM Security Verify Cloud Directory is used to look up the second factor for the username. After you configure multifactor authentication on the system, users and user groups must be added to the Cloud Directory.
Generate refresh token	Ensure that this option is unchecked.	
Send all known user attributes in the ID token	Ensure that this option is checked.	
Access policies	Complete these steps: a. Deselect **Use default policy**. b. Click the **Edit** icon. c. Select **Always require 2FA in all devices**. d. Click **OK**.	This action creates an access policy which controls the authentication steps for system access. Access policies can specify different authentication requirements based on properties of the user or connection. In this case, all users must complete a second factor authentication every time they access the system from all devices.
Restrict custom scopes	Ensure this option is unchecked.	

The following table shows the required fields and actions for the **API access** tab in the IBM Security Verify. The **API access** tab is used to add the command-line interface as an API-based client and create credentials for multifactor authentication access for CLI users. To add the command-line as a separate API client, click **Add API client**. On the **Add API client** page, enter the following information:

Table 78. Add API Client action

Field	Action	Details
Name	Enter a name to identify the command line interface as the API client.	
Select the APIs to which you want to grant access	Ensure that all APIs are selected by moving the toggle to display **On**	

5. Click **Save**. After the system is saved as a new application, the **Custom Application** reloads with the **Entitlements** tab selected.
6. On the **Entitlements** tab, select **Automatic access for all users and groups**.
7. Click **Save**.
8. Select **Applications** and select the application name that represents the system.
9. On the **Sign-on** tab, copy the **Client ID** and the **Client secret**. These values must be specified as the OpenID credentials on the Multifactor authentication page in the management GUI.
10. On the **API access** tab, click the edit icon and copy the **Client ID** and the **Client secret**. These values must be specified as the API Client credentials on the Multifactor authentication page in the management GUI.

Using the management GUI

To configure multifactor authentication on IBM Security Verify, complete these steps:

1. Select **Settings** > **Security** > **Multifactor Authentication**.
2. Enter the host name and port of the authentication server. For IBM Security Verify, enter the following:

   ```
   tenant.verify.ibm.com
   ```

   Where `tenant` is the name that is associated with your subscription. Port 443 is the default for the authentication server.
3. For the **OpenID Credentials**, add the **Client ID** and **Client Secret** that you copied on from the Sign-on tab in the IBM Security Verify interface.
4. For the **API Client Credentials**, add the **Client ID** and **Client Secret** that you copied on from the API access tab in the IBM Security Verify interface.
5. On the **Multifactor Authentication** page, click **Export Certificate** to export the system certificate to your device. Copy the system ID alias that displays. This value must be used as the friendly name of the certificate in IBM Security Verify.
6. Access the IBM Security Verify administrator dashboard by entering the following URL in a web browser:

   ```
   https://tenant.verify.ibm.com/ui/admin
   ```

   Where `tenant` is the name of the tenant that you specified when you created your subscription. Usually this tenant name is associated with your company or organization.
7. Select **Security** > **Certificates**.
8. Under **Signer certificates**, select **Add signer certificate**.

9. On the **Add signer certificate** page, select **Add file** and navigate to where you exported the certificate on your device.
10. In the **Friendly name** field, copy the system ID alias that displays on the **Multifactor Authentication** page in the management GUI.
11. Click **OK**.
12. Return to the **Multifactor Authentication** page in the management GUI, and click **Save**. On the confirmation page, click **Confirm** to enable multifactor authentication for the system.

Multifactor authentication is enabled for the system. You can configure user groups to use multifactor authentication. Click **Navigate** to launch the **User Groups** page.

Using the CLI

Before you can enable multifactor authentication on the system, ensure that the system certificate is exported and added as a signer certificate. If the certificate is not added as a signer certificate, users with multifactor authentication enabled cannot sign in to the management GUI. To export and add the system certificate, complete the following steps:

1. To view the system ID alias, enter the following command:

   ```
   lssystem | grep id_alias
   ```

   Note: The system ID alias must be entered in the **Friendly name** field on the **Add Signer Certificate** page in the IBM Security Verify interface.
2. Enter the following command to export the system certificate:

   ```
   chsystemcert -export
   ```

   This command exports the system certificate. Download the resulting file `/dumps/certificate.pem` to your machine and upload to IBM Security Verify. Ensure to add the system ID alias to the **Friendly name** field. For more information, see steps 7 through 11 in the management GUI section.
3. To enable multifactor authentication with IBM Security Verify, enter the following command:

   ```
   chauthmultifactorverify -hostname tenant.verify.ibm.com -openidclientid yyyyyyyy-yyyy-
   yyyy-yyyy-yyyyyyyyyyyy -openidclientsecret yyyyyyyyyy -cliclientid xxxxxxx-xxxx-xxxx-xxxx-
   xxxxxxxxxxxx -cliclientsecret xxxxxxxxxx -enable
   ```

   In the example, `tenant` is the tenant name that is associated with your subscription. The values for the **`-openidclientid`** and the **`-openidclientsecret`** are the Open ID Client and Open ID Secret that are automatically generated when you created your system as a custom application in IBM Security Verify on the Sign-on tab in the IBM Security Verify interface. The values for the **`-cliclientid`** and the **`-cliclientsecret`** are the API Client ID and API Client Secret that are automatically generated when you created your system as a custom application in IBM Security Verify on the API access tab in the IBM Security Verify interface.

Related information

“Multifactor authentication” on page 399
Multifactor authentication requires users to provide multiple pieces of information when they log in to the system to prove their identity. Multifactor authentication uses any combination of two or more methods, called *factors*, to authenticate users to your resources and protect those resources from unauthorized access.

Configuring user groups for multifactor authentication with IBM Security Verify

After you configure multifactor authentication on the system, you must enable multifactor authentication for user groups and add users to those groups in IBM Security Verify.

As part of multifactor authentication configuration, you must enable the function per user group. The system supports enabling multifactor authentication for local and remote user groups.

The security administrator must define all local users manually in the Cloud Directory of IBM Security Verify. When users log into the system with multifactor authentication, their username is used to look up the required second factor in the Cloud Directory in IBM Security Verify. Users can set up multiple second factors to avoid getting locked out of the system.

Note: When you configure multifactor authentication for the first time, follow these guidelines to avoid getting locked out of the system unintentionally:

- Ensure at least one user with Security Administrator role does not have multifactor authentication enabled. If the security administrator gets locked out of the system because of errors in the multifactor authentication set up, the additional user can still access the system.
- Enable multifactor authentication on a user group without logged in users or leave an SSH session active to avoid locking out users on the system.

For remote users that authenticate with LDAP servers, install and configure IBM Security Verify Bridge for Directory Sync on your LDAP server, such as Windows Active Directory. IBM Security Verify Bridge for Directory Sync duplicates any users and groups that are defined on the source LDAP server into the Cloud Directory in IBM Security Verify. Any subsequent changes that are made to the source LDAP server are copied automatically to the Cloud Directory in IBM Security Verify. For local users, each user must be added manually to the Cloud Directory in IBM Security Verify.

Using the management GUI

To enable multifactor authentication on user groups, complete these steps:

For existing user groups:

1. In the management GUI, select **Access** > **Users by Group**.
2. Select the user group from the left navigation and select **User Group Actions** > **Properties**.
3. On the **User Group Properties** page, select **On** under **Multifactor Authentication** to enable second-factor authentication for all users with the user group. These users authenticate with the first factors that are stored on the local system and then are required to provide a second factor to access the system through a supported authentication service.
4. Click **OK**.

For new user groups for local users:

1. In the management GUI, select **Access** > **Users by Group** > **Create User Group**.
2. On the **Create User Group** page, enter the following information:

   Group Name
   Enter a name of the user group.

   Ownership Group
   If ownership groups are configured on your system, you can select an ownership group for the user group.

   Multifactor Authentication
   Select **On** to enable second-factor authentication for local users on the system. These users authenticate with the first factors that are stored on the local system and then are required to provide a second factor to access the system through a supported authentication service.

   Role
   Select a role that for the user group.
3. Click **Create**.
4. Select **Access** > **Users by Group** > **Create User**.
5. On the **Create User** page, enter the following information:

   Name
   Enter a user name for the user. This user name must match the user name that is added to the Cloud Directory on IBM Security Verify.

Authentication mode
Select **Local**.

User Group
Select the name of the user group that the local user belongs to.

Password
Enter a password that is used as the first factor for management GUI access.

SSH key
For CLI users, include a public SSH key that is used as the first factor for CLI access.

6. Click **Create**. Repeat these steps for all local users.

For new user groups for remote users:

1. Select **Access** > **Users** > **Create User Groups**.
2. On the **Create User Group** page, enter the following information:

Group Name
Enter the name of the group that is on the remote LDAP server. The name of the group on the system must match.

Ownership Group
If ownership groups are configured on your system, you can select an ownership group for the user group.

Remote Authentication
Select **LDAP**.

Multifactor authentication
Select **On** to enable second-factor authentication for remote users on the system. These users authenticate with the first factors that are stored on the remote LDAP server and then are required to provide a second factor to access the system through a supported authentication service.

3. Click **Create**.
4. Remote users are defined on the remote authentication service. Only remote users who require access to the command-line interface using a Secure Shell (SSH) key, need to be created. To create remote users, who require access to the command-line interface, select **Access** > **Users by Groups** > **Create User**.
5. On the **Create User** page, enter the following information:

Name
Enter a name for the user.

Authentication mode
Select **Remote**.

SSH key
For remote users who require to access the system, select the public SSH key which stored locally on the system.

6. Click **Create**. Repeat these steps for all remote users.

Using the CLI

To enable multifactor authentication on user groups, enter the following commands:

For existing user groups

```
chusergrp -multifactor yes <id or name of group>
```

For new user groups for local users

```
mkusergrp -name name -role role -multifactor yes
```

For new user groups for remote users

```
mkusergrp -name name -role role -multifactor yes -remote
```

Adding local users to IBM Security Verify
After user groups are enabled for multifactor authentication on the system, add local users to the Cloud Directory in IBM Security Verify.

To add local users to the Cloud Directory in IBM Security Verify, complete these steps:

1. Access the IBM Security Verify administrator dashboard by entering the following URL in a web browser:

   ```
   https://tenant.verify.ibm.com/ui/admin
   ```

   Where `tenant` is the name of the tenant that you specified when you created your subscription. Usually this tenant name is associated with your company or organization.
2. Select **Directory** > **Users & groups**.
3. On the **Users & groups** page, select **Add user**.
4. On the **Add user** page, enter the following information:

   Identity provider
   Select **Cloud Directory**.

   Basic user profile
   Enter the given name, middle name, and surname for the local user.

   Username
   Enter the same username that you defined for the user in the management GUI. Usernames must match otherwise authentication fails.

   User information
   Ensure to include an email address for the user. Each user receives instructions on login to IBM Security Verify which provides step-by-step instructions on selecting second-factors for the user's profile.
5. Click **Save**.
6. Repeat these steps for all local users that are defined on the system.

Depending on your security requirements, you can either require only a certain type of second factor or let users select from the supported list of available factors when they log in for the first time. If you want to control which factors and their settings users can specify, select **Security** from the IBM Security Verify administrator dashboard.

After you have added all users to IBM Security Verify, each user receives an email notification with instructions on setting up the second factors for their account. When these users log in to the management interfaces that are configured to use multifactor authentication, they are directed to IBM Security Verify to present the second factor.

Adding remote users to IBM Security Verify
After user groups are enabled for multifactor authentication on the system, add remote users to the Cloud Directory in IBM Security Verify.

For remote users that authenticate with LDAP servers, install and configure IBM Security Verify Bridge for Directory Sync on your LDAP server, such as Windows Active Directory. IBM Security Verify Bridge for Directory Sync duplicates any users and groups that are defined on the source LDAP server into the Cloud Directory in IBM Security Verify. Any subsequent changes that are made to the source LDAP server are copied automatically to the Cloud Directory in IBM Security Verify.

To add remote users to IBM Security Verify, complete these steps:

1. Download and install IBM Security Verify Bridge for Directory Sync on your supported LDAP server. For more information, see IBM Security Verify Bridge for Directory Sync in the IBM Security Verify documentation.

2. After the installation completes, you need to update the sample JSON file and add properties that define connection settings and the LDAP server. On your supported LDAP server, open the `/DirectorySync/` directory, create a copy of the file `IcbLdapSync.json.ad-sample` and rename it `IcbLdapSync.json`.
3. Open the file in a text editor.
4. Under the `ibm-auth-api` JSON object, complete the following steps:

   `client-id`
   Enter the client ID that was automatically generated when you created the system as an application in IBM Security Verify.

   `obf-client-secret`

   a. Change the name of this JSON object to: **`client-secret`**.

   b. Enter the client secret that was automatically generated when you created the system as an application in IBM Security Verify.

   `host`
   Enter the hostname for the IBM Security Verify tenant.

   `port` (optional)
   Enter the port for your IBM Security Verify tenant.

   `proxy` (optional)
   Enter the hostname for the proxy server.
5. Under the `cloud-bridge` JSON object, add the following information:

   `ldap-poll-time`
   Enter a value in seconds to determine how frequently the LDAP directory and the Cloud Directory are synchronized. The default setting is 4 seconds. However, the example JSON file uses the value of 60 seconds. You can change this value based on your needs.
6. Under the `ldap` JSON object, add the following information:

   `host`
   Enter the hostname of the LDAP server.

   `port`
   Enter the port number for the LDAP server.

   `user`
   Enter the Administrator account distinguished name (DN) format for the LDAP server. Typically the Administrator account has these permissions, for example: `"user":"CN=Administrator,CN=Users,DC=mydomain,DC=com"`

   `obf-password`

   a. Change the name of this JSON object to: **`password`**.

   b. Enter the password for the Administrator for the LDAP server.
7. Under the `user` JSON object, add the following information:

   `realm`
   Change the realm to `"cloudIdentityRealm"`.

   `userCategory`
   Change `userCategory` to `"regular"`.

   `Notification`
   Ensure that you include the following values:

```
"notifyType": "EMAIL",
         "notifyPassword": true,
         "notifyManager": false
```

These values specify that an email notification is sent when they are added to the Cloud Directory inIBM Security Verify. In the email notification, they are instructed to access IBM Security Verify and select second factors to authenticate to the management interfaces.

8. Under the `ldap-to-scim` JSON object, remove or comment out the following string:

```
"append":"@cloudIdentityRealm".
```

9. Save the JSON file. The updated JSON file looks like the following example:

```
{
    "ibm-auth-api":{
        "timeout":100,
        "client-id":"xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
        "client-secret":"xxxxxxxxx",
        "protocol":"https",
        "host":"tenant.verify.ibm.com",
        "port":xxx,
        "proxy":"x.xx.xx.x:xxxx",
        "max-handles":16
    },
    "cloud-bridge":{
        "max-ops": 512,
        /* "trace-file":"c:/tmp/cloudbridge.log", */
        "enable-op-log":true,
        "op-log-rollover": 2097152,
        "do-not-sync-delete":  false,
        "ldap-search-filter":"(&(|(objectClass=user)(objectClass=group))(!
(isCriticalSystemObject=*)))",
        "ldap-is-deleted-attr":"isDeleted",
        "ldap-poll-time":60,
        /* "log-stats-interval": 30, */
        /* "ldap-base-dn": "DC=mycompany,DC=com", */
        "ldap":{
            "host":"ldapserver.example.com",
            "port":xxx,
            "user":"CN=Administrator,CN=Users,DC=mycompany,DC=com",
            "password":"password",
            "use-tls":false,
            "start-tls":false
        },
        "user":{
            "ldap-object-classes": [ "user" ],
            "scim-external-id-attr":"externalId",
            "scim-outline":{
                "urn:ietf:params:scim:schemas:extension:ibm:2.0:User":{
                    "realm":"cloudIdentityRealm",
                    "userCategory":"regular"
                },
                "urn:ietf:params:scim:schemas:extension:ibm:2.0:Notification": {
                    "notifyType": "EMAIL",
                    "notifyPassword": true,
                    "notifyManager": false
                },
                "schemas":[
                    "urn:ietf:params:scim:schemas:core:2.0:User",
                    "urn:ietf:params:scim:schemas:extension:ibm:2.0:User",
                    "urn:ietf:params:scim:schemas:extension:enterprise:2.0:User",
                    "urn:ietf:params:scim:schemas:extension:ibm:2.0:Notification"
                ]
            },
            "ldap-to-scim":[
                {
                    "ldap":"sAMAccountName",
                    "tweaks":{
                        /* "append":"@cloudIdentityRealm" */
                    },
                    "new-attr":{
                        "scim":{"userName":"{{value}}"}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"userName","value":"{{value}}"},
                            "remove":{"op":"remove","path":"userName"},
                            "replace":{"op":"replace","path":"userName","value":"{{value}}"}
                        }
                    }
                },
/* Example of using a custom attribute: */
```

```
/*
                {
                    "ldap":"carLicense",
                    "is-multi-value":true,
                    "new-attr": {
                        "scim":{
                            "urn:ietf:params:scim:schemas:extension:ibm:2.0:User":
{"customAttributes":[{"values":["{{value}}"],"name":"myattrscim"}]}
            }
                    },
                    "mod-attr":{
                        "scim":{
                            "add":
{"op":"add","path":"urn:ietf:params:scim:schemas:extension:ibm:2.0:User:customAttributes",
                                "value":[{"name":"myattrscim","values":["{{value}}"]}]},
                            "remove-all":
{"op":"remove","path":"urn:ietf:params:scim:schemas:extension:ibm:2.0:User:customAttributes[
name eq \"myattrscim\"]"},
                            "remove":
{"op":"remove","path":"urn:ietf:params:scim:schemas:extension:ibm:2.0:User:customAttributes[
name eq \"myattrscim\"].values",
                                "value":["{{value}}"]},
                            "replace":
{"op":"add","path":"urn:ietf:params:scim:schemas:extension:ibm:2.0:User:customAttributes",
                                "value":[{"name":"myattrscim","values":["{{value}}"]}]}
                        }
                    }
                },
*/
                {
                    "ldap":"cn",
                    "new-attr":{
                        "scim":{"displayName":"{{value}}"}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"displayName","value":"{{value}}"},
                            "remove":{"op":"remove","path":"displayName"},
                            "replace":
{"op":"replace","path":"displayName","value":"{{value}}"}
                        }
                    }
                },
                {
                    "ldap":"telephoneNumber",
                    "new-attr":{
                        "scim":{"phoneNumbers":[{"type":"work","value":"{{value}}"}]}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"phoneNumbers","value":
[{"type":"work","value":"{{value}}"}]},
                            "remove":{"op":"remove","path":"phoneNumbers[type eq
\"work\"]"},
                            "replace":{"op":"add","path":"phoneNumbers","value":
[{"type":"work","value":"{{value}}"}]}
                        }
                    }
                },
                {
                    "ldap":"mobile",
                    "new-attr":{
                        "scim":{"phoneNumbers":[{"type":"mobile","value":"{{value}}"}]}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"phoneNumbers","value":
[{"type":"mobile","value":"{{value}}"}]},
                            "remove":{"op":"remove","path":"phoneNumbers[type eq
\"mobile\"]"},
                            "replace":{"op":"add","path":"phoneNumbers","value":
[{"type":"mobile","value":"{{value}}"}]}
                        }
                    }
                },
                {
                    "ldap":"homePhone",
                    "new-attr":{
                        "scim":{"phoneNumbers":[{"type":"home","value":"{{value}}"}]}
                    },
                    "mod-attr":{
                        "scim":{
```

```
                            "add":{"op":"add","path":"phoneNumbers","value":
[{"type":"home","value":"{{value}}"}]},
                            "remove":{"op":"remove","path":"phoneNumbers[type eq
\"home\"]"},
                            "replace":{"op":"add","path":"phoneNumbers","value":
[{"type":"home","value":"{{value}}"}]}
                        }
                    }
                },
                {
                    "ldap":"ipPhone",
                    "new-attr":{
                        "scim":{"phoneNumbers":[{"type":"pager","value":"{{value}}"}]}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"phoneNumbers","value":
[{"type":"pager","value":"{{value}}"}]},
                            "remove":{"op":"remove","path":"phoneNumbers[type eq
\"pager\"]"},
                            "replace":{"op":"add","path":"phoneNumbers","value":
[{"type":"pager","value":"{{value}}"}]}
                        }
                    }
                },
                {
                    "ldap":"facsimileTelephoneNumber",
                    "new-attr":{
                        "scim":{"phoneNumbers":[{"type":"fax","value":"{{value}}"}]}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"phoneNumbers","value":
[{"type":"fax","value":"{{value}}"}]},
                            "remove":{"op":"remove","path":"phoneNumbers[type eq \"fax\"]"},
                            "replace":{"op":"add","path":"phoneNumbers","value":
[{"type":"fax","value":"{{value}}"}]}
                        }
                    }
                },
                {
                    "ldap":"givenName",
                    "new-attr":{
                        "scim":{"name":{"givenName":"{{value}}"}}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"name.givenName","value":"{{value}}"},
                            "remove":{"op":"remove","path":"name.givenName"},
                            "replace":
{"op":"replace","path":"name.givenName","value":"{{value}}"}
                        }
                    }
                },
                {
                    "ldap":"sn",
                    "new-attr":{
                        "scim":{"name":{"familyName":"{{value}}"}}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"name.familyName","value":"{{value}}"},
                            "remove":{"op":"remove","path":"name.familyName"},
                            "replace":
{"op":"replace","path":"name.familyName","value":"{{value}}"}
                        }
                    }
                },
                {
                    "ldap":"displayName",
                    "new-attr":{
                        "scim":{"name":{"formatted":"{{value}}"}}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"name.formatted","value":"{{value}}"},
                            "remove":{"op":"remove","path":"name.formatted"},
                            "replace":
{"op":"replace","path":"name.formatted","value":"{{value}}"}
                        }
                    }
                },
```

```
                {
                    "ldap":"streetAddress",
                    "new-attr":{
                        "scim":{"addresses":[{"type":"work","streetAddress":"{{value}}"}]}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"addresses","value":
[{"type":"work","streetAddress":"{{value}}"}]},
                            "remove":{"op":"remove","path":"addresses[type eq \"work\"]"},
                            "replace":{"op":"add","path":"addresses","value":
[{"type":"work","streetAddress":"{{value}}"}]}
                        }
                    }
                },
                {
                    "ldap":"manager",
                    "is-dn":true,
                    "new-attr":{
                        "scim":
{"urn:ietf:params:scim:schemas:extension:enterprise:2.0:User":{"manager":
{"value":"{{value}}"}}}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":
{"op":"add","path":"urn:ietf:params:scim:schemas:extension:enterprise:2.0:User:manager.value
","value":"{{value}}"},
                            "remove":
{"op":"remove","path":"urn:ietf:params:scim:schemas:extension:enterprise:2.0:User:manager.va
lue"},
                            "replace":
{"op":"replace","path":"urn:ietf:params:scim:schemas:extension:enterprise:2.0:User:manager.v
alue","value":"{{value}}"}
                        }
                    }
                },
                {
                    "ldap":"mail",
                    "new-attr":{
                        "scim":{"emails":[{"type":"work","value":"{{value}}"}]}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"emails","value":
[{"type":"work","value":"{{value}}"}]},
                            "remove":{"op":"remove","path":"emails[type eq \"work\"]"},
                            "replace":{"op":"add","path":"emails","value":
[{"type":"work","value":"{{value}}"}]}
                        }
                    }
                },
                {
                    "ldap":"department",
                    "new-attr":{
                        "scim":
{"urn:ietf:params:scim:schemas:extension:enterprise:2.0:User":{"department":"{{value}}"}}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":
{"op":"add","path":"urn:ietf:params:scim:schemas:extension:enterprise:2.0:User:department","
value":"{{value}}"},
                            "remove":
{"op":"remove","path":"urn:ietf:params:scim:schemas:extension:enterprise:2.0:User:department
"},
                            "replace":
{"op":"replace","path":"urn:ietf:params:scim:schemas:extension:enterprise:2.0:User:departmen
t","value":"{{value}}"}
                        }
                    }
                },
                {
                    "ldap":"objectGUID",
                    "is-binary":true,
                    "new-attr":{
                        "scim":{"externalId":"{{value}}"}
                    }
                },
                {
                    "ldap":"userAccountControl",
                    "new-attr":{
```

```
                        "cond-scim":{
                            "if":{"&":["{{value}}", "2"]},
                            "then":{"active":false},
                            "else":{"active":true}
                        }
                    },
                    "mod-attr":{
                        "cond-scim":{
                            "if":{"&":["{{value}}","2"]},
                            "then":{
                                "add":{"op":"add","path":"active","value":false},
                                "remove":{"op":"remove","path":"active"},
                                "replace":{"op":"replace","path":"active","value":false}
                            },
                            "else":{
                                "add":{"op":"add","path":"active","value":true},
                                "remove":{"op":"remove","path":"active"},
                                "replace":{"op":"replace","path":"active","value":true}
                            }
                        }
                    }
                }
            ]
        },
        "group":{
            "ldap-object-classes": [ "group" ],
            "scim-external-id-attr":"externalId",
            "scim-outline":{
                "schemas":[
                    "urn:ietf:params:scim:schemas:core:2.0:Group",
                    "urn:ietf:params:scim:schemas:extension:ibm:2.0:Group"
                ]
            },
            "ldap-to-scim":[
                {
                    "ldap":"sAMAccountName",
                    "new-attr":{
                        "scim":{"displayName":"{{value}}"}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":{"op":"add","path":"displayName","value":"{{value}}"},
                            "remove":{"op":"remove","path":"displayName"},
                            "replace":
{"op":"replace","path":"displayName","value":"{{value}}"}
                        }
                    }
                },
                {
                    "ldap":"objectGUID",
                    "is-binary":true,
                    "new-attr":{
                        "scim":{"externalId":"{{value}}"}
                    }
                },
                {
                    "ldap":"description",
                    "new-attr":{
                        "scim":{"urn:ietf:params:scim:schemas:extension:ibm:2.0:Group":
{"description":"{{value}}"}}
                    },
                    "mod-attr":{
                        "scim":{
                            "add":
{"op":"add","path":"urn:ietf:params:scim:schemas:extension:ibm:2.0:Group:description","value
":"{{value}}"},
                            "remove":
{"op":"remove","path":"urn:ietf:params:scim:schemas:extension:ibm:2.0:Group:description"},
                            "replace":
{"op":"replace","path":"urn:ietf:params:scim:schemas:extension:ibm:2.0:Group:description","v
alue":"{{value}}"}
                        }
                    }
                },
                {
                    "ldap":"member",
                    "is-dn":true,
                    "is-multi-value":true,
                    "new-attr":{
                        "scim":{"members":[{"type":"{{type}}","value":"{{value}}"}]}
                    },
                    "mod-attr":{
```

```
                        "scim":{
                            "add":{"op":"add","path":"members","value":
[{"type":"{{type}}","value":"{{value}}"}]},
                            "remove":{"op":"remove","path":"members[value eq
{{jq_value}}]"},
                            "remove-all":{"op":"remove","path":"members"},
                            "replace":{"op":"replace","path":"members","value":
[{"type":"{{type}}","value":"{{value}}"}]}
                        }
                    }
                }
            ]
        }
    }
}
```

10. To synchronize the LDAP server and the Cloud Directory for the first time, run `IcbLdapSync.exe`.

Depending on your security requirements, you can either require only a certain type of second factor or let users select from the supported list of available factors when they log in for the first time. If you want to control which factors and their settings users can specify, select **Security** from the IBM Security Verify administrator dashboard.

After you have added all users to IBM Security Verify, each user receives an email notification with instructions on setting up the second factors for their account. When these users log in to the management interfaces that are configured to use multifactor authentication, they are directed to IBM Security Verify to present the second factor.

Configuring multifactor authentication with Duo Security

The system integrates with Duo Security to provide multifactor authentication for system users.

With Duo Security, security administrators can configure the system as an application that requires two factors for users and user groups to access the system with either the management GUI or CLI.

You can manage every aspect of your Duo two-factor authentication system from the Duo Admin Panel including creating and managing applications, enrolling and activating users, issuing and managing SMS passcodes and bypass codes, managing mobile devices, fine-tuning the user experience of your Duo installation.

Multifactor authentication can be used to protect both local users, including superuser, and remote users.

For information on how to configure directory synchronization for remote users with Duo Security, see Duo Directory Synchronization.

Duo Security configures the management GUI and the command-line interface as separate API clients that require separate credentials. For CLI access, Duo Security communicates with the system through standard REST API requests. For GUI-based logins, the system communicates with Duo Security through the OpenID Connect (OIDC) protocol.

Prerequisites

Ensure that the following prerequisite tasks are completed on the system before you configure multifactor authentication on Duo Security:

1. Configure a DNS server. To create a DNS server, select **Settings** > **Network** > **DNS**.
2. Configure an HTTP proxy server or configure your firewall to access Duo Security. To create an HTTP proxy server, **Settings** > **Network** > **Internal Proxy Server**. If your system does not directly connect to the Internet, you can create a firewall exception to allow your system access to Duo Security.

   Important: If the proxy server is not configured correctly, the system cannot communicate with Duo Security and the login to the system fails.
3. For the management GUI and the command-line interface, ensure that the inactivity logout is equal to or greater than the time it takes for a user to receive a one-time passcode (OTP) from the authentication service. The default value for the inactivity timeout is 30 minutes for the management

GUI and 15 minutes for the CLI. To set the inactivity timeout for both interfaces in the management GUI, select **Settings** > **Security** > **Inactivity Logout**.

4. Ensure that the SSH grace time for the command-line interface is equal to or greater than the time it takes for a user to receive a one-time passcode (OTP) from the authentication service. The default value for the SSH grace time period is 60 seconds. To set SSH grace time on the system in the command-line interface, use the **chsecurity -sshgracetime** command.

The following prerequisite steps on Duo Security must be completed before you can configure multifactor authentication on the system:

1. Create a subscription for Duo Security. For more information, see Editions & Pricing page on Duo Security website. During subscription creation, you specify a tenant that is used to create a URL to access the Duo Security dashboard.
2. To create a Duo Security application, enter the following:

   ```
   admin-XXXXXXXX.duosecurity.com
   ```

   where XXXXXXXX is unique to your subscription that you specified when you created your subscription.
3. On **Applications** page, click **Protect an Application** and select **Web SDK**. The credentials on this page are the OpenID Client ID and Secret required for GUI access.
4. On **Applications** page, click **Protect an Application** and select **UNIX Application**. The credentials on this page are the API Client Integration Key and Secret Key required for CLI access.

Note: Duo Security does not support the use of IP addresses when accessing your system by using a web browser. Once multifactor authentication is configured, make sure that a hostname is used when accessing your system by using a web browser.

Using the management GUI

To configure multifactor authentication on Duo Security, complete these steps:

1. Select **Settings** > **Security** > **Multifactor Authentication**.
2. On **Multifactor Authentication** page, select **Duo Security** authentication provider.
3. Enter the hostname and port of the authentication server. For Duo Security, enter the following:

   ```
   api-XXXXXXXX.duosecurity.com
   ```

   Where XXXXXXXX is unique to your subscription that you specified when you created your subscription.
4. For the **OpenID Credentials**, complete the following steps in the Duo Security interface to add the **Client ID** and **Client Secret**:
   a. Select **Dashboard** > **Applications** > **Protect an Application**.
   b. Search **Web SDK**, click **Protect**.
   c. Copy **Client ID** and **Client Secret**.
5. For the **API Client Credentials**, complete the following steps in the Duo Security interface to add the **Integration Key** and **Secret Key**:
   a. Select **Dashboard** > **Applications** > **Protect an Application**.
   b. Search **UNIX Application**, click **Protect**.
   c. Copy **Integration Key** and **Secret Key**.
6. Click **Save**.
7. On the confirmation page, click **Confirm** to enable multifactor authentication for the system.

Multifactor authentication is enabled for the system. You can configure user groups to use multifactor authentication. Click **Navigate** to launch the **User Groups** page.

Using the CLI

You can enable multifactor authentication on the system by using the following command:

- To enable multifactor authentication with Duo Security, enter the following command:

```
chauthmultifactorduo -enable -hostname api-XXXXXXXX.duosecurity.com -openidclientid
openid_client_id -openidclientsecret openid_client_secret -integrationkey integration_key
-secretkey secret_key
```

  where parameter **-enable** enables multifactor authentication service and **-hostname** is the hostname of the Duo Security tenant. The **-openidclientid** and the **-openidclientsecret** are the Open ID Client and Open ID Secret for the system, required to enable multifactor authentication for login to the storage system GUI. The **-integrationkey** and the **-secretkey** are the integration key and secret key for the system, required to enable multifactor authentication for login to the Storage Virtualize CLI.

Configuring user groups for multifactor authentication with Duo Security

After you configure multifactor authentication on the system, you must enable multifactor authentication for user groups and add users to those groups in Duo Security.

As part of multifactor authentication configuration, you must enable the function per user group. The system supports enabling multifactor authentication for local and remote user groups.

The security administrator must define all local users manually in Duo Security. When users log into the system with multifactor authentication, their username is used to look up the required second factor in Duo Security. Users can set up multiple second factors to avoid getting locked out of the system.

Note: When you configure multifactor authentication for the first time, follow these guidelines to avoid getting locked out of the system unintentionally:

- Ensure at least one user with Security Administrator role does not have multifactor authentication enabled. If the security administrator gets locked out of the system because of errors in the multifactor authentication set up, the additional user can still access the system.
- Enable multifactor authentication on a user group without logged in users or leave an SSH session active to avoid locking out users on the system.

For information on how to configure directory synchronization for remote users with Duo Security, see Duo Directory Synchronization.

Using the management GUI

To enable multifactor authentication on user groups, complete these steps:

For existing user groups:

1. In the management GUI, select **Access** > **Users by Group**.
2. Select the user group from the left navigation and select **User Group Actions** > **Properties**.
3. On the **User Group Properties** page, select **On** under **Multifactor Authentication** to enable second-factor authentication for all users with the user group. These users authenticate with the first factors that are stored on the local system and then are required to provide a second factor to access the system through a supported authentication service.
4. Click **OK**.

For new user groups for local users:

1. In the management GUI, select **Access** > **Users by Group** > **Create User Group**.
2. On the **Create User Group** page, enter the following information:

   Group Name
   Enter a name of the user group.

   Ownership Group
   If ownership groups are configured on your system, you can select an ownership group for the user group.

Multifactor Authentication
Select **On** to enable second-factor authentication for local users on the system. These users authenticate with the first factors that are stored on the local system and then are required to provide a second factor to access the system through a supported authentication service.

Role
Select a role that for the user group.

3. Click **Create**.
4. Select **Access** > **Users by Group** > **Create User**.
5. On the **Create User** page, enter the following information:

Name
Enter a user name for the user. This user name must match the user name that is added on Duo Security.

Authentication mode
Select **Local**.

User Group
Select the name of the user group that the local user belongs to.

Password
Enter a password that is used as the first factor for management GUI access.

SSH key
For CLI users, include a public SSH key that is used as the first factor for CLI access.

6. Click **Create**. Repeat these steps for all local users.

For new user groups for remote users:

1. Select **Access** > **Users** > **Create User Groups**.
2. On the **Create User Group** page, enter the following information:

Group Name
Enter the name of the group that is on the remote LDAP server. The name of the group on the system must match.

Ownership Group
If ownership groups are configured on your system, you can select an ownership group for the user group.

Remote Authentication
Select **LDAP**.

Multifactor authentication
Select **On** to enable second-factor authentication for remote users on the system. These users authenticate with the first factors that are stored on the remote LDAP server and then are required to provide a second factor to access the system through a supported authentication service.

3. Click **Create**.
4. Remote users are defined on the remote authentication service. Only remote users who require access to the command-line interface using a Secure Shell (SSH) key, need to be created. To create remote users, who require access to the command-line interface, select **Access** > **Users by Groups** > **Create User**.
5. On the **Create User** page, enter the following information:

Name
Enter a name for the user.

Authentication mode
Select **Remote**.

SSH key
For remote users who require to access the system, select the public SSH key which stored locally on the system.

6. Click **Create**. Repeat these steps for all remote users.

Using the CLI

To enable multifactor authentication on user groups, enter the following commands:

For existing user groups

```
chusergrp -multifactor yes <id or name of group>
```

For new user groups for local users

```
mkusergrp -name name -role role -multifactor yes
```

For new user groups for remote users

```
mkusergrp -name name -role role -multifactor yes -remote
```

Add local users to Duo Security
After user groups are enabled for multifactor authentication on the system, add local users in Duo Security.

To add local users in Duo Security complete these steps:

1. Log into Duo Security with your Duo Administrator account.
2. Complete the steps described in the Duo Security documentation. For more information, see Enrolling users.

Add remote users to Duo Security
After user groups are enabled for multifactor authentication on the system, add remote users in Duo Security.

For remote users that authenticate with LDAP servers, Duo Security supports automatic enrollment. Duo Security's Directory Sync feature duplicates any users and groups that are defined on the source LDAP server. Any subsequent changes that are made to the source LDAP server are copied automatically to Duo Security. To add remote users in Duo Security, complete these steps:

1. Log in to Duo Security with your Duo Administrator account.
2. Configure the Directory Sync feature on Duo Security.

   For information on how to configure directory synchronization for remote users with Duo Security, see Duo Directory Synchronization.

Changing MFA credentials of system

You can change multifactor authentication credentials for the system with the management GUI or command-line interface (CLI).

If credentials are updated on the multifactor authentication provider, these new credentials must be changed on the system. If the credentials are not updated on the system, the system cannot communicate with the authentication providers and user cannot be authenticated to access the system.

Using the management GUI

To update multifactor authentication credentials on the system, complete these steps:

1. Select **Settings** > **Security** > **Multifactor Authentication**.
2. Select the **Authentication provider** to change the credentials.

3. In **Authentication server** enter the correct host name and port number.
4. For changing credentials for IBM Security Verify on Storage Virtualize GUI, enter the new Client ID and Client secret under the **OpenID Credentials** heading.
5. For changing credentials for IBM Security Verify on Storage Virtualize CLI, enter the new Client ID and Client secret under the **API Client Credentials** heading.
6. For changing credentials for Duo Security on Storage Virtualize GUI, enter the new Client ID and Client secret under the **OpenID Credentials** heading.
7. For changing credentials for Duo Security on Storage Virtualize CLI, enter the new Integration key and Secret key under the **API Client Credentials** heading.
8. Click **Save**.

Using the command-line interface

- Use the command **chauthmultifactorverify** to change the credentials for IBM Security Verify. By running the following command you can change the credentials.

```
chauthmultifactorverify -openidclientid openid_client_id
                        -openidclientsecret openid_client_secret
                        -cliclientid cli_client_id
                        -cliclientsecret cli_client_secret
```

  where *openid_client_id* specifies the IBM Security Verify OpenID Connect client ID for the system. It is required to enable the multifactor authentication (MFA) for login to the Storage Virtualize GUI. The *openid_client_secret* specifies the IBM Security Verify OpenID Connect client secret for the system. The *cli_client_id* specifies the IBM Security Verify CLI client ID for the system. It is required to enable the multifactor authentication (MFA) for login to the Storage Virtualize CLI. The *cli_client_secret* specifies the IBM Security Verify CLI client secret for the system.

- Use the command **chauthmultifactorduo** to change the credentials for Duo Security. By running the following command you can change the credentials.

```
chauthmultifactorduo -openidclientid openid_client_id
                     -openidclientsecret openid_client_secret
                     -integrationkey integration_key
                     -secretkey secret_key
```

  where *openid_client_id* specifies the Duo Security OpenID Connect client ID for the system. It is required to enable the multifactor authentication (MFA) for login to the Storage Virtualize GUI. The *openid_client_secret* specifies the Duo Security OpenID Connect client secret for the system. The *integration_key* specifies the Duo Security CLI integration key for the system. It is required to enable the multifactor authentication (MFA) for login to the Storage Virtualize CLI. The *secret_key* specifies the Duo Security CLI secret key for the system.

Changing the failmode parameter for multifactor authentication

The failmode parameter determines how the system responds if a user logs into the system when multifactor authentication is enabled, but the authentication server is unavailable. Failmode setting can be managed by command-line interface (CLI).

Use the failmode parameter in the **chauthmultifactorverify** command to set one of the following values:

secure
: When the authentication server is unavailable, specify this value to prevent any user with multifactor authentication enabled from accessing the system. When a user logs into the system, the login attempt fails.

insecure
: When the authentication server is unavailable, specify this value to allow any user with multifactor authentication enabled to access the system. When a user logs into the system, the login attempt succeeds.

Enter the following command to prevent access to the system for users with multifactor authentication enabled:

```
chauthmultifactorverify -failmode secure
```

Enter the following command to allow access to the system for user with multifactor authentication enabled:

```
chauthmultifactorverify -failmode insecure
```

User authentication

You can configure authentication and authorization for users of the system. The system supports multifactor authentication, single sign-on, local user authentication and LDAP authentication for remote users.

The system supports both local users, and remote users who are authenticated to the system through a remote authentication service. You can create local users who can access the system. These user types are defined based on the administrative privileges that they have on the system. Local users must provide either a password, a Secure Shell (SSH) key, or both. Local users are authenticated through the authentication methods that are configured on the system. If the local user needs access to the management GUI, a password is needed for the user. If the user requires access to the command-line interface (CLI) through SSH, either a password or a valid SSH key file is necessary. Local user passwords are securely stored by using the PBKDF2 hashing algorithm. Local users must be part of a user group that is defined on the system. User groups define roles that authorize the users within that group to a specific set of operations on the system.

In addition to these first factor authentication methods for local users, you can enable the system to require multifactor authentication for all local users defined in a user group. To use multifactor authentication on the system, you need to configure a supported authentication service and enable multifactor authentication on the system and user groups. For example, the system integrates with IBM Security Verify that provides second factor authentication. Local users must be also manually added to the supported authentication service and set up their second factors.

A remote user is authenticated on a remote LDAP server. A remote user does not need to be added to the list of users on the system, although they can be added to configure optional SSH keys. For remote users, an equivalent user group must be created on the system with the same name and role as the group on the remote LDAP server. Remote users cannot access the system when the remote LDAP server is down. In that case, a local user account must be used until the LDAP service is restored. Remote users have their groups that are defined by the remote authentication server.

The system also supports requiring multifactor authentication for remote users. As with local users, you must configure a supported authentication service and enable the feature on the system and remote user groups. However, remote users can be automatically managed with IBM Security Verify Bridge for Directory Sync. For remote users that authenticate with LDAP servers, install and configure IBM Security Verify Bridge for Directory Sync on your LDAP server, such as Windows Active Directory. IBM Security Verify Bridge for Directory Sync duplicates any users and groups that are defined on the source LDAP server into the Cloud Directory in IBM Security Verify. Any subsequent changes that are made to the source LDAP server are copied automatically to the Cloud Directory in IBM Security Verify.

For information on how to configure directory synchronization for remote users with Duo Security, see Duo Directory Synchronization.

In addition to multifactor authentication, the system also supports single sign-on for remote users only. Remote users are authenticated to all applications through a single set of credentials. Single sign-on requires that the feature is enabled at the system level and on user groups.

The system supports several authentication methods. However, it's important to understand the key differences and limitations of each authentication method. The following table summarizes the differences and limitations of each supported authentication method:

Table 79. Authentication methods comparison			
	Multifactor Authentication	**Single Sign-on**	**LDAP authentication**
First factor handled by	IBM Storage Virtualize	Identity Provider (IdP)	LDAP server
Second factor handled by	Multifactor authentication service (for example, IBM Security Verify)	Identity Provider (IdP)	NA
CLI login support	Yes	No	Yes
GUI login support	Yes	Yes	Yes
REST API support	No	No	Yes
Local user support	Yes	No	No
Remote user support	Yes	Yes	Yes
Supported services	• IBM Security Verify • Duo Security	• IBM Security Verify • Microsoft Active Directory Federation Services • Microsoft Azure Active Directory • Okta	• Microsoft Active Directory • IBM Security Directory Server • Others (for example, OpenLDAP)
Protocol	OpenID Connect (OIDC)	OpenID Connect (OIDC)	LDAP

Related information

"Managing users and user groups" on page 230
You can use the management GUI or the command-line interface to manage user and user groups on the system.

Remote Authentication

Remote authentication allows users to authenticate to the system using credentials that are stored on an external authentication service. When you configure remote authentication, you do not need to configure users on the system or assign more passwords. Instead, you can use your existing passwords and user groups that are defined on the remote service to simplify user management and access to enforce password policies more efficiently, and to separate user management from storage management.

A remote user is authenticated on a remote LDAP server. A remote user does not need to be added to the list of users on the system, although they can be added to configure optional SSH keys. For remote users, an equivalent user group must be created on the system with the same name and role as the group on the remote LDAP server. Remote users cannot access the system when the remote LDAP server is down. In that case, a local user account must be used until the LDAP service is restored. Remote users have their groups that are defined by the remote authentication server.

Using the management GUI

To configure remote authentication with LDAP, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Remote Authentication**.
2. Select **Configure Remote Authentication**.
3. Select **LDAP**.
4. Select the type of LDAP server that is used for authentication.

5. Select one of the following security options:

   LDAP with StartTLS
   Select this option to configure extensions that upgrade the standard LDAP port (389) to an encrypted port that uses TLS or SSL. The initial connection to the directory server is decrypted but can be used on systems that do not have port 636 available.

   LDAPS
   Select this option to secure LDAP communication by using the default secure port (636). The connections for all transactions with the directory server are encrypted.

   LDAP with no security
   Select this option to transport data in clear text format without encryption.

6. Specify optional service credentials or modify advanced LDAP settings. The following LDAP attributes can be configured:

   User attribute
   For all server types, users are authenticated with a username that is defined with the LDAP user attribute. This attribute must exist in your LDAP schema and must be unique for each of your users. Active Directory users can also authenticate by using their user principal names (UPN) or NT login names.

   Group attribute
   Authenticated users are assigned roles according to their LDAP group memberships. The groups to which a user belongs are stored in the LDAP group attribute. This attribute value can be the distinguished name of each group, or a colon-separated list of user group names.

   Audit log attribute
   If an LDAP user completes an audited action, the contents of the audit log attribute are recorded in the audit log.

7. Click **Next**.
8. Define up to six LDAP servers to use for authentication. Multiple servers can be configured to provide access to different sets of users for redundancy. You can also configure which servers are preferred to authenticate users. You can specify either IP addresses or domain name for the LDAP servers. If you specify a fully qualified domain name, a DNS server must be configured on your system. To configure a DNS server for the system, select **Settings** > **Network** > **DNS**.

   Note: The system does not support using LDAP referrals to find related LDAP servers. Each required LDAP server must be explicitly configured on the system.

9. Configure user groups on the system to match groups that are configured on the remote authentication service. For each user group on the authentication service, a corresponding user group must be created with the same name. In the management GUI, complete these steps:

   a. Select **Access** > **Users** > **Create User Groups**.
   b. On the **Create User Group** page, enter the following information:

      Group Name
      Enter the name of the group that is on the remote LDAP server. The name of the group on the system must match.

      Ownership Group
      If ownership groups are configured on your system, you can select an ownership group for the user group.

      Remote Authentication
      Select **LDAP**.

      Multifactor authentication
      Select **On** to enable second-factor authentication for remote users on the system. These users authenticate with the first factors that are stored on the remote LDAP server and then are required to provide a second factor to access the system through a supported authentication service.

   c. Click **Create**.

d. Remote users are defined on the remote authentication service. Only remote users who require access to the command-line interface using a Secure Shell (SSH) key, need to be created. To create remote users, who require access to the command-line interface, select **Access** > **Users by Groups** > **Create User**.

e. On the **Create User** page, enter the following information:

Name
Enter a name for the user.

Authentication mode
Select **Remote**.

SSH key
For remote users who require to access the system, select the public SSH key which stored locally on the system.

f. Click **Create**. Repeat these steps for all remote users.

10. Verify your LDAP configuration. To test authentication to the LDAP servers, select **Test LDAP Authentication** and enter corresponding credentials for the user.

Using the command-line interface

To enable user authentication with LDAP by using the command-line interface, follow these steps:

1. Configure LDAP by entering the **chldap** command.

   This command provides default settings for both IBM Security Directory Server and Microsoft Active Directory. To configure authentication with IBM Security Directory Server schema defaults and Transport Layer Security (TLS), for example, enter the following command:

   ```
   chldap -type itds -security tls
   ```

   LDAP configuration can be inspected with the **lsldap** command.

   Note: Use TLS so that transmitted passwords are encrypted.

2. Specify the **mkldapserver** command to define up to six LDAP servers to use for authentication.

   Multiple servers can be configured to provide access to different sets of users or for redundancy. All servers must share the settings that are configured with **chldap**.

   Note: The system does not support using LDAP referrals to find related LDAP servers. Each LDAP server that is required must be explicitly configured on the system.

   To configure an LDAP server with an SSL certificate and users in the `cn=users,dc=company,dc=com` subtree, for example, enter the following command:

   ```
   mkldapserver -ip 9.71.45.108 -basedn cn=users,dc=company,dc=com -sslcert /tmp/sslcert.pem
   ```

   If you have a DNS server configured on the system, you can also specify a domain name in the **-ip** parameter. For example:

   ```
   mkldapserver -ip myldap.myco.com -basedn cn=users,dc=company,dc=com -sslcert /tmp/sslcert.pem
   ```

   You can also configure which servers are preferred to authenticate users.

   Specify **lsldapserver** for LDAP server configuration information. Specify **chldapserver** and **rmldapserver** to change the configured LDAP servers.

3. Configure user groups on the system by matching those user groups that are used by the authentication service.

   For each group of interest that is known to the authentication service, a system user group must be created with the same name and with the remote setting enabled. If members of a group that is called `sysadmins`, for example, require the system administrator (admin) role, enter the following command:

```
mkusergrp -name sysadmins -remote -role Administrator
```

If none of the user groups match a system user group, the user cannot access the system.

4. Verify your LDAP configuration by using the **testldapserver** command.

   To test the connection to the LDAP servers, enter the command without any options. A user name can be supplied with or without a password to test for configuration errors. To process a full authentication attempt against each server, enter the following commands:

   ```
   testldapserver -username username -password 'password'
   ```

5. Enter the following command to enable LDAP authentication:

   ```
   chauthservice -type ldap -enable yes
   ```

6. Configure users who do not require Secure Shell (SSH) key access.

   Delete system users who must use the remote authentication service and do not require SSH key access.

   Remember: A superuser cannot be deleted or use the remote authentication service.

7. Configure users who require SSH key access.

   All system users who use the remote authentication service and require SSH key access must have remote settings that are enabled and a valid SSH key that is configured on the system.

8. Specify the type of security to use when communicating with LDAP servers.

   Specify *tls* to enable TLS. Select this option to configure extensions that upgrade the standard LDAP port (389) to an encrypted port that uses TLS. The initial connection to the directory server is unencrypted but can be used on systems that do not have port 636 available.

   Specify *ssl* to enable SSL security. This option secures LDAP communication by using the default secure port (636). The connections for all transactions with the directory server are encrypted. The default value is none.

Single Sign-on

Single Sign-on (SSO) authentication requires users to register their credentials only once when the user signs on to the application for the first time. The user information is stored at the Identity Provider (IdP) that manages the user credentials and determines whether the user is required to authenticate again or not.

SSO is a feature that delegates all authentication to a trusted Identity Provider (IdP). The applications are configured with the IdP. When the user logs in to the application, the IdP validates the registered information and enables the user to log in to the application. In this method, both the first and second factor authentications are delegated to a trusted Identity Provider, which is useful if you do not want IBM Storage Virtualize to do any of the authentications. Also, additional authentication factors can be configured to provide multifactor authentication for users.

IBM Storage Virtualize uses the OpenID Connect (OIDC) protocol to communicate with the SSO service when a user logs in to the GUI. SSO is supported only for web browsers. SSO can be used only to protect user login to the management GUI. The SSO feature is not supported for the command-line interface. To provide multifactor authentication for CLI access, the system must be integrated natively with IBM Security Verify. For more information, refer to “Multifactor authentication” on page 399. The SSO service is only supported for remote users that are defined on a remote LDAP server, and does not support local users that are defined locally within the IBM Storage Virtualize system, including the superuser.

Remote users are users who are defined on a remote LDAP server. For remote users that authenticate with LDAP servers, install and configure IBM Security Verify Bridge for Directory Sync on your LDAP server, such as Windows Active Directory. IBM Security Verify Bridge for Directory Sync duplicates any users and groups that are defined on the source LDAP server into the Cloud Directory in IBM Security Verify. Any subsequent changes that are made to the source LDAP server are copied automatically to the Cloud

Directory in IBM Security Verify. For more information, see IBM Security Verify Bridge for Directory Sync in the IBM Security Verify documentation.

To protect CLI access for local users or remote users with SSO, local and remote user groups can be configured with the password and SSH key authentication method, which requires users to authenticate with both a password and an SSH key when a user logs in through the CLI.

When SSO is configured and enabled for the user groups, the GUI displays a new **Sign In with SSO** option on the login window. The user is redirected to do the authentication with the configured SSO service. When the authentication is successful using SSO, the user does not have to reauthenticate on subsequent login attempts.

MFA versus SSO

Multifactor authentication (MFA) and Single Sign-on (SSO) can both be configured to provide multiple authentication factors when a user logs in to the system. However, it's important to understand the key differences and limitations of each authentication method. The following table summarizes the difference and limitations of both the authentication methods and enables a user to select a method based on their needs.

Table 80. Authentication methods summary

	Multifactor Authentication	**Single Sign-on**
First factor handled by	IBM Storage Virtualize	Identity Provider (IdP)
Second factor handled by	Multifactor authentication service (IBM Security Verify)	Identity Provider (IdP)
CLI login support	Yes	No
GUI login support	Yes	Yes
Local user support	Yes	No
Remote user support	Yes	Yes
Supported services	IBM Security Verify	Microsoft Active Directory Federation Services
Protocol	OpenID Connect (OIDC)	OpenID Connect (OIDC)

LDAP versus SSO

LDAP (Lightweight Directory Access Protocol) or Single Sign-on (SSO) can be configured for authenticating remote users when a user logs in to the system. However, it's important to understand the key differences and limitations of each authentication method. The following table summarizes the difference and limitations of both the authentication methods and enables a user to select a method based on their needs.

	LDAP authentication	**Single Sign-on**
First factor handled by	LDAP server	Identity Provider (IdP)
Second factor handled by	NA	Identity Provider (IdP)
CLI login support	Yes	No
GUI login support	Yes	Yes
Local user support	No	No
Remote user support	Yes	Yes

Supported services	• Microsoft Active Directory • IBM Security Directory Server • Others (for example, OpenLDAP)	Microsoft Active Directory Federation Services
Protocol	LDAP	OpenID Connect (OIDC)

Configuring single sign-on

Single sign-on delegates all authentication to a trusted Identity Provider (IdP).

With single sign-on, users need to provide their credentials once when they log in to an application or system, rather than repeatedly providing the credentials for every individual application or system. Each individual IBM Storage Virtualize system is considered a separate application and must be added to the IdP. When single sign-on is enabled on the system, both first-factor and second-factor authentication are delegated to the IdP. For more information on supported single sign-on providers, see .

In the management GUI, users select **Sign In with SSO** on the log in prompt and are redirected to complete authentication through the configured IdP. You can configure the IdP to provide first factor authentication to your users. Several multifactor authentication cloud-based providers can be added to the single sign-on configuration to require additional user authentication if necessary. When authentication is completed successfully, the uses are redirected to management GUI.

Prerequisites

Ensure that the following prerequisite tasks are completed on the system before you configure single sign-on:

1. Ensure that the system is updated to 8.5.0 or later release.
2. Configure a DNS server. To create a DNS server, select **Settings** > **Network** > **DNS**.
3. If your authentication server is outside of your private network, you must configure an HTTP proxy server or configure your firewall to access your authentication server.

   Note: If you plan to use a proxy server to access the Identity Provider, ensure that you select **Yes** on the **Use proxy** option when you configure single sign-on in the management GUI.

   To create an HTTP proxy server, **Settings** > **Network** > **Internal Proxy Server**. If your authentication server is within your private network, proxy or firewall changes are not required.
4. For the management GUI, ensure that the inactivity logout is equal to or greater than the time it takes for a user to receive a one-time passcode (OTP) from the authentication service. The default value for the inactivity timeout is 30 minutes for the management GUI. To set the inactivity timeout in the management GUI, select **Settings** > **Security** > **Inactivity Logout**. To set the GUI inactivity timeout on the command-line interface, use the **`chsecurity -guitimeout`** command.

Choose an authentication provider

Configure single sign-on using a supported authentication provider. Once single sign-on has been configured, you must then decide which user groups should be enabled for single sign-on.

Configuring single sign-on with PingOne

PingOne can be configured as the authentication provider for the system.

Note: Ensure that the prerequisite tasks are completed on the system before you configure single sign-on. For more information, see the prerequisites in “Configuring single sign-on” on page 427.

Prerequisites

The following prerequisite steps on PingOne must be completed before you can configure single sign-on on the system:

1. Create a subscription for PingOne. A free trial subscription is also available. For more information, see Ping Single Sign-On. During subscription creation, you specify a tenant that is used to create a URL to access the PingOne dashboard.
2. Access the PingOne administrator dashboard by entering the following URL in a web browser (this is usually received in your welcome email). The sign-on URL is specific to your admin console and includes your environment ID. For example, the URL might look like this:

   ```
   http://console.pingone.asia/index.html?env=cde34209-9366-45aa-9e5f-a19d3ba3eada
   ```

3. In the PingOne administrator console, from the menu select **Connections** > **Applications**.
4. Select **Application +** to add a new application.

   Note: Each system must be added as a separate application.

   The following table shows the required fields and actions for the **Register an application** page in the PingOne interface.

   Table 81. Register an application page

Field	Action
Application Name	Enter a name to identify the system on PingOne. If you are adding multiple systems, enter a unique name.
Application Type	Select OIDC Web App.

5. Click **Save**. After the system is saved as a new app registration, the application reloads with the **Overview** page selected.
6. To find the OpenID Configuration Endpoints URL for the tenant, on the Configuration page expand URLs and copy the OIDC Discovery Endpoint URL. This value must be specified as the OpenID Configuration Endpoint URL on the Single sign-on page in the management GUI.
7. On the Configuration page, expand **General** and copy the Client ID. This value must be specified as the Client ID on the Single sign-on page in the management GUI.
8. On the Configuration page, expand **General** and copy the Client Secret. This value must be specified as the Client ID on the Single sign-on page in the management GUI.

   Note: The system does not support client secret values that begin with a hyphen character. Use the reveal button to inspect the Client Secret automatically generated by PingOne. If the Client Secret begins with a hyphen character, click the **Generate New Secret** button.
9. On the Configuration page, click the **Edit** button. For **Response Type**, ensure that **Code** is selected. For **Grant Type**, ensure that **Authorization Code** is selected. For **Redirect URLs**, multiple redirect URIs can be specified for the management GUI. For management GUI access, the redirect URI consists of the management IP address or hostname followed by /sso. For example, https://hostname.com/sso. For **Token EndPoint Authentication Method**, select **Client Secret Post**. Click the **Save** button.
10. On the Resources page, click **Edit** on the Allowed Scopes and ensure that **openid** and **profile** are selected. Click the **Save** button.
11. Close the Application window to return to the Applications page. Click the toggle to enable user access to this application.
12. In the PingOne administrator console, from the menu select **Connections** > **Resources** and then select **OpenID Connect**.
13. Ensure the authentication provider is configured to send back the group claim to the system. The group claim identifies which groups the authenticating user belongs to. Some authentication

providers do not send the group claim by default, so this typically requires some configuration on the authentication provider. On the **Attributes** tab, click the **Edit** button. Click **+ Add** to add a new attribute for the group claim. Ensure that the **Attributes** field contains **group** and the **PingOne Mappings** field contains a value of **Group Names**. Configure the **ID Token** and **UserInfo** values according to your organization's requirements; this determines how the **group** attribute is sent back to the application. Click the **Save** button.

14. If you require the group claim to be sent back to the application in the ID Token, open the **Scopes** page and **Edit** the **profile** scope. Click inside the **Mapped Attributes** area and type **group** to add the group attribute. Click the **Save** button.
15. On the Attributes page, ensure that the existing `preferred_username` attribute has a **PingOne Mappings** value of **Username** and that **ID Token** is selected.
16. In the PingOne administrator console, from the menu select **Connections** > **Applications**.
17. Select your application. On the Attribute Mappings page, click **Edit** on **Custom Attributes**. Click **+ Add** to add a new attribute. For the **Attributes** field, enter groups. For the **PingOneMappings** field, select **GroupNames**. Click **Save**.

Using the management GUI

To configure single sign-on with PingOne, complete these steps:

1. Select **Settings** > **Security** > **Single Sign-on**.
2. Enter the OpenID Configuration Endpoint URL of the authentication server that you obtained from the application details earlier. For PingOne, the format is:

   ```
   https://auth.pingone.region/environment-id/as/.well-known/openid-configuration
   ```

   Where region is the region that the PingOne tenant is hosted in, and environment-id is the Environment ID specific to this application. This can be found by selecting from the menu **Connections** > **Applications** > **Click your application** > **Configuration** > **OIDC Discovery Endpoint**.
3. For the **OpenID Credentials**, add the **Client ID** and **Client Secret** that you copied on from the Configuration page in the PingOne interface.
4. For the User claim, the value to enter depends on how your authentication provider is configured. The User claim must match the name that the authentication service uses to specify the username attribute in the ID Token it sends to the system. Typically this value is preferred_username, but can be customized on the authentication provider.
5. For the Group claim, the value to enter depends on how your authentication provider is configured. The Group claim must match the name that the authentication service uses to specify the group attribute, either in the ID Token it sends to the system, or in the UserInfo endpoint. Typically this value is groups, but can be customized on the authentication provider.
6. For Proxy server, consider how the system accesses the authentication provider. For an authentication provider within your network, a proxy server usually isn't needed. If you connect to the authentication provider through the Internet, check the box and ensure a proxy server is defined on the system.
7. Click **Save**. On the confirmation page, click **Confirm** to enable single sign-on for the system.

Single sign-on is enabled for the system. You can configure user groups to use single sign-on. Click **Navigate** to launch the **User Groups** page.

Using the command-line interface

To enable single sign-on, enter the following command:

```
chauthsinglesignon -oidcconfigurationendpoint https://auth.pingone.region/environment-
id/as/.well-known/openid-configuration -clientid xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx
-clientsecret xxxxxxxx -userclaim preferred_username -groupclaim groups -enable
```

In the example, region is the tenant region that is associated with your subscription, and environment-id is the Environment ID specific to this application. The values for the **`-clientid`** and the

-clientsecret are the Client ID and Client Secret that are automatically generated when you created your system as an application in PingOne, and can be obtained from the Configuration page of your application in the PingOne interface. The values for the **-userclaim** and the **-groupclaim** should match the name of the claims configured for the ID Token on the authentication provider.

Configuring single sign-on with IBM Security Verify

IBM Security Verify can be configured as the authentication provider for the system.

Note: Ensure that the prerequisite tasks are completed on the system before you configure single sign-on. For more information, see the Prerequisites section in "Configuring single sign-on" on page 427.

Prerequisites

The following prerequisite steps on IBM Security Verify must be completed before you can enable single-sign on the system:

1. Create a subscription for IBM Security Verify. You need an IBMid to create a subscription. A 90-day free trial subscription is also available. For more information, see Cloud identity and access management (IAM) solutions. During subscription creation, you specify a tenant that is used to create a URL to access the IBM Security Verify dashboard.
2. Access the IBM Security Verify administrator dashboard by entering the following URL in a web browser:

   ```
   https://tenant.verify.ibm.com/ui/admin
   ```

   Where `tenant` is the name of the tenant that you specified when you created your subscription. Usually this tenant name is associated with your company or organization.
3. In the IBM Security Verify interface, select **Applications** > **Applications** > **Add application**.
4. Select **IBM Storage Virtualize** > **Add Application**.

   Note: Each system must be added as a separate application.

   The following table shows the required fields and actions for the **General** tab in the IBM Security Verify interface.

Table 82. General tab	
Field	**Action**
Name	Enter a name to identify the system on IBM Security Verify. If you are adding multiple systems, enter a unique name.
Description	Enter a brief description of the system.
Company name	Name of organization or company.

   The following table shows the required fields and actions for the **Sign-on** tab in the IBM Security Verify interface. The Sign-on tab is used to add the management GUI as an API-based client.

Table 83. Sign-on tab		
Field	**Action**	**Details**
Application URL	Enter the URL for your system.	Enter the URL that is used to access the management GUI.

Table 83. Sign-on tab (continued)

Field	Action	Details
Grant type	Select **Authorization code** and **JWT bearer**.	Two grant types are required for setting up SSO for the system. **Authorization code** indicates that the client can request access to protected resources on behalf of users.
Client ID	This value is automatically generated when the system is saved as an application.	This value must be entered to the Single Sign-on page in the management GUI under **OpenID Credentials**.
Client secret	This value is automatically generated when the system is saved as an application.	This value must be entered to the Single Sign-on page in the management GUI under **OpenID Credentials**.
User consent	Select **Do not ask for consent**.	
Redirect URIs	Enter the locations where the authorization server sends users after they are successfully authorized and granted an authorization code or access token.	Multiple redirect URIs can be specified for the management GUI. The redirect URI is comprised of the management IP address or hostname followed by `/sso`. For example, https://hostname.com/sso
JWT bearer user identification	Select **Username**.	Indicates that the username field in the JWT bearer is used to find users in the Cloud Directory and determines what second factors IBM Security Verify presents to users when they log into the system.
JWT bearer default identity source	Ensure **Cloud Directory** is selected.	Indicates that the IBM Security Verify Cloud Directory is used to look up the second factor for the username. After you configure multifactor authentication on the system, users and user groups must be added to the Cloud Directory.
Generate refresh token	Ensure that this option is unchecked.	
Send all known user attributes in the ID token	Ensure that this option is checked.	

Table 83. Sign-on tab (continued)		
Field	**Action**	**Details**
Access policies	Complete these steps: a. Deselect **Use default policy**. b. Click the **Edit** icon. c. Select **Always require 2FA in all devices**. d. Click **OK**.	This action creates an access policy which controls the authentication steps for system access. Access policies can specify different authentication requirements based on properties of the user or connection. In this case, all users must complete a second factor authentication every time they access the system from all devices.
Restrict custom scopes	Ensure this option is unchecked.	

There is no action required for the **API access** tab in IBM Security Verify.

The following table shows the required fields and actions for the **API access** tab in the IBM Security Verify. Click **Save**. After the system is saved as a new application, the **Custom Application** reloads with the **Entitlements** tab selected.

5. On the **Entitlements** tab, select **Automatic access for all users and groups**.
6. Click **Save**.
7. Select **Applications** and select the application name that represents the system.
8. On the **Sign-on** tab, copy the **Client ID** and the **Client secret**. These values must be specified as the OpenID credentials on the Single sign-on page in the management GUI
9. Ensure the authentication provider is configured to send back the group claim in the ID Token sent to the system. The group claim identifies which groups the authenticating user belongs to. Some authentication providers do not send the group claim by default, so this typically requires some configuration on the authentication provider.

Using the management GUI

1. Select **Settings** > **Security** > **Single Sign-on**.
2. Enter the OpenID Configuration Endpoint URL of the authentication server. For IBM Security Verify, enter the following:

```
https://tenant.verify.ibm.com/oidc/endpoint/default/.well-known/openid-configuration
```

where `tenant` is the name that is associated with your subscription.

3. For the **OpenID Credentials**, add the **Client ID** and **Client Secret** that you copied on from the Sign-on tab in the IBM Security Verify interface.
4. For the User claim, the value to enter depends on how your authentication provider is configured. The User claim must match the name that the authentication service uses to specify the username attribute in the ID Token it sends to the system. Typically this value is preferred_username, but can be customized on the authentication provider.
5. For the Group claim, the value to enter depends on how your authentication provider is configured. The Group claim must match the name that the authentication service uses to specify the group attribute in the ID Token it sends to the system. Typically, this value is groupIds, but can be customized on the authentication provider.
6. For Proxy server, consider how the system accesses the authentication provider. For an authentication provider within your network, a proxy server usually is not needed. If you connect to the authentication provider through the internet, check the box and ensure a proxy server is defined on the system.

7. Click **Save**. On the confirmation page, click **Confirm** to enable single sign-on for the system.

Single sign-on is enabled for the system. You can configure user groups to use single sign-on. Click **Navigate** to launch the **User Groups** page.

Using the command-line interface

To enable single sign-on, enter the following command:

```
chauthsinglesignon -oidcconfigurationendpoint https://tenant.verify.ibm.com/oidc/endpoint/
default/.well-known/openid-configuration -clientid xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx
-clientsecret xxxxxxxx -userclaim preferred_username -groupclaim groupIds -enable
```

In the example, tenant is the tenant name that is associated with your subscription. The values for the **-clientid** and the **-clientsecret** are the Open ID Client and Open ID Secret that are automatically generated when you created your system as a custom application in IBM Security Verify on the Sign-on tab in the IBM Security Verify interface. The values for the **-userclaim** and the **-groupclaim** should match the name of the claims configured for the ID Token on the authentication provider.

Configuring single sign-on with Microsoft Active Directory Federation Services (AD FS)

Microsoft Active Directory Federation Services (AD FS) can be configured as the authentication provider for the system.

Note: Ensure that the prerequisite tasks are completed on the system before you configure single sign-on. For more information, see the prerequisites in “Configuring single sign-on” on page 427.

Prerequisites

The following prerequisite steps on AD FS must be completed before you can enable single-sign on the system:

1. Ensure that you are running at least AD FS 5.0 on a Microsoft Windows 2019 server.
2. Access the Server Manager dashboard.
3. From right menu, select **Tools** > **AD FS Management**.
4. From left menu, choose **AD FS** > **Application Groups**. Select **Add Application Group...** from the **Actions** menu.
5. The **Add Application Group Wizard** displays. The following table shows the required fields and actions for the **Welcome** step:

Table 84. Welcome

Field	Action
Name	Enter a name to identify the system on the AD FS instance. If you are adding multiple systems, each name must be unique. For example, specify the actual name of the IBM Storage Virtualize system.
Description	Enter a brief description of the system.
Template	Select **Server application accessing a web API**.

6. Click **Next**.
7. The following table shows the required fields and actions for the **Server application** step:

Table 85. Server application

Field	Action	Details
Name	The wizard automatically adds "Server application" to the name that you specified on the **Welcome** page. If multiple systems need to be added, you can change this to include other identifying information.	
Client Identifier	This value is automatically generated when the system is saved as an application. Copy this value and save it. You need to add this value on the **Configure Web API** page in the wizard and in the management GUI.	This value must be entered to the **Single Sign-on** page in the management GUI under **OpenID Credentials** as the **Client ID**.
Redirect URI	Enter the redirect URI for the management GUI. For management GUI access, the redirect URI is consisted of the host name, fully qualified domain name, or management IP address followed by `/sso`. Depending on how you access the system, multiple redirect URIs can be specified. For example: **Hostname** `https://hostname/sso` **Fully-qualified domain name** `https://hostname.com/sso` **IP address** `https://1.2.3.4/sso` **Note:** Single sign-on is not supported on the service assistant interface.	The redirect URI is the location where AD FS sends users after they are successfully authenticated.

8. Click **Next**.
9. The following table shows the required fields and actions for the **Configure Application Credentials** step.

Table 86. Configure Application Credentials		
Field	**Action**	**Details**
Generate a shared secret	a. Select **Generate a shared secret**. b. After the shared secret is automatically generated, click **Copy to clipboard**. **Note:** This shared secret must be added to the management GUI when you enable single sign-on for the system. After it is generated, you cannot display this secret in the application properties after the system is added to AD FS. Ensure that you retain the copied version of the client secret. The client secret is sensitive so keep it safe.	This value must be entered on the **Single Sign-on** page in the management GUI under **OpenID Credentials** as the **Client secret**.

10. Click **Next**.
11. The following table shows the required fields and actions for the **Configure Web API** step.

Table 87. Configure Web API		
Field	**Action**	**Details**
Name	The wizard automatically adds "Web API" to the name that you specified on the **Welcome** page. If multiple systems need to be added, you can change this to include other identifying information.	
Identifier	a. Copy the automatically generated value in the **Client Identifier** field from the **Server Application** page. b. Click **Add**	

12. Click **Next**.
13. The following table shows the required fields and actions for the **Choose Access Control Policy** step.

Table 88. Choose Access Control Policy		
Field	**Action**	**Details**
Choose an access control policy	From the options, select one of the following supported access policies: **Permit everyone** Select this option if you are only configuring single sign-on without any additional factors to authenticate users. **Permit everyone and require MFA** Select this option if you are configuring single sign-on with additional factors to authenticate users.	These two options are not the only valid access control policies available to select for the system. These options are suggested since they cover most use cases; however, custom access policies can also be created to suit your needs and security policies.

14. Click **Next**.
15. The following table shows the required fields and actions for the **Configure Application Permissions** step.

Table 89. Configure Application Permissions		
Field	**Action**	**Details**
Permitted scopes	Select the following scopes: • Select **allatclaims**. • Verify that **openid** is selected.	

16. Click **Next**.
17. On the **Summary** page, verify the settings and click **Next**.
18. Click **Close** on the confirmation message. The **Application Groups** page displays. The system is now configured as an application group in AD FS.
19. Right click the new application group that you created and select **Properties**.
20. On the **Properties** page, select the name of the system under the Web API section and click **Edit...**.
21. On the **Web API Properties** page, select **Issuance Transform Rules** tab.
22. On the **Issuance Transform Rules** tab, click **Add Rule...**.
23. On the **Choose Rule Type** page, select **Send LDAP Attributes as Claims** as the claim rule template and click **Next**.
24. On the **Choose Claim Rule** page, enter the following values on the page:

    Claim rule name:
    Enter a name for the rule that you are setting.

    Attribute store:
    Select **Active Directory**.

    Mapping of LDAP attributes to outgoing claim types:
    In the **LDAP Attributes** column, select **Token-Groups - Unqualified Names**.

    In the **Outgoing Claim Type** column, select **Group**.
25. Click **Finish**. The **Web API Properties** page displays.
26. On the **Web API Properties** page, click **Apply**.

Before you enable single sign-on on the system, ensure that you have copied and saved the following information:

Client Identifier
This value is automatically generated on the **Server Application** task of the **Add Application Group** wizard. This value must be entered on the **Single Sign-on** page in the management GUI under **OpenID Credentials** as the **Client ID**.

Shared secret
This value is automatically generated on the **Configure Application Credentials** task of the **Add Application Group** wizard. This value must be entered on the **Single Sign-on** page in the management GUI under **OpenID Credentials** as the **Client secret**.

OpenID configuration endpoint
The administrator of your single sign-on server can provide this information. However you can display the OpenID configuration endpoint by selecting **Server Manager** > **AD FS Management** > **AD FS** > **Service** > **Endpoints**

Using the management GUI

To configure single sign-on on the system, complete these steps:

1. Select **Settings** > **Security** > **Single Sign-on**.
2. For the **Authentication server**, enter the OpenID Connect Discovery endpoint that is associated with the Identity Provider. For example, for an AD FS endpoint, enter the following:

   ```
   https://<AD FS server URL>/adfs/.well-known/openid-configuration
   ```

   where <AD FS server URL> is the fully-qualified domain name for the single sign-on server.
3. Under **OpenID Credentials**, paste the **Client Identifier** that you copied from the **Server application** task from the **Add Application Group** wizard in the **Client ID** field. Paste the client secret that was generated on the **Configure Application Credentials** task in the **Add Application Group** wizard in the **Client secret** field.
4. Under **Claims**, the system automatically sets the **User claim** to upn and **Group claim** to group.
5. Click **Save**. On the confirmation page, click **Confirm** to enable single sign-on for the system.

Single sign-on is enabled for the system. You can configure user groups to use single sign-on. Click **Navigate** to launch the **User Groups** page.

Using the command-line interface

To enable single sign-on, enter the following command:

```
chauthsinglesignon -oidcconfigurationendpoint https://authentication-server-url/adfs/.well-
known/openid-configuration -clientid xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx -clientsecret
xxxxxxxx -groupclaim group -userclaim upn -enable
```

In the example, authentication-server-url is the URL to your authentication server. The values for the **-clientid** and the **-clientsecret** are **Client Identifier** and the client secret that was generated in the **Add Application Group** wizard in the AD FS management dashboard.

Configuring single sign-on with Microsoft Entra ID

Microsoft Entra ID can be configured as the authentication provider for the system.

Note: Ensure that the prerequisite tasks are completed on the system before you configure single sign-on. For more information, see the prerequisites in “Configuring single sign-on” on page 427.

Prerequisites

The following prerequisite steps on Microsoft Entra ID must be completed before you can configure single sign-on on the system:

1. Create a subscription for Microsoft Entra ID. A free trial subscription is also available. For more information, see Microsoft Entra ID. During subscription creation, you specify a tenant that is used to create a URL to access the Microsoft Entra ID dashboard.
2. Access the Microsoft Entra ID administrator dashboard by entering the following URL in a web browser:

   ```
   https://portal.azure.com
   ```

3. In the Microsoft Entra ID interface, from the menu select **Entra ID** > **App registrations** > **New registration**.
4. Select **IBM Storage Virtualize** > **Add application**.

   Note: Each system must be added as a separate application.

   The following table shows the required fields and actions for the **Register an application** page in the Microsoft Entra ID interface.

   Table 90. Register an application page

Field	**Action**
Name	Name Enter a name to identify the system on Microsoft Entra ID. If you are adding multiple systems, enter a unique name.
Supported account types	Select which user accounts should be able to use this application.
Redirect URI	Select Web. Multiple redirect URIs can be specified for the management GUI. The redirect URI is consisted of the management IP address or hostname followed by `/sso`. For example, `https://hostname.com/sso`

5. Click **Register**. After the system is saved as a new app registration, the application reloads with the **Overview** page selected.
6. To find the OpenID Configuration Endpoints URL for the tenant, select **Overview** > **Endpoints** and copy the OpenID Connect metadata document URL. This value must be specified as the OpenID Configuration Endpoint URL on the Single sign-on page in the management GUI.
7. On the **Overview** page, copy the **Application (client) ID**. This value must be specified as the Client ID on the Single sign-on page in the management GUI.
8. Go to **Manage** > **Certificates & secrets** and click **New client secret**. Enter a description and expiry value. Click Add. After the client secret has been added, the Certificates & secrets page appears. Copy the **Secret ID**. This value must be specified as the Client secret on the Single sign-on page in the management GUI.
9. Ensure the authentication provider is configured to send back the group claim in the ID Token sent to the system. The group claim identifies which groups the authenticating user belongs to. Some authentication providers do not send the group claim by default, so this typically requires some configuration on the authentication provider. Go to **Manage** > **Token configuration** and click **Add groups claim**. In the Add groups claim page, select **Customize token properties by type** for the ID token, then choose your preferred method of identifying group names. For example:
   a. Select the types of groups as **Security groups**.
   b. Select the ID type, Access type, and SAML type as **sAMAccountName**.
   c. Click **Add**.

10. Enable ID Tokens for the application. Select **Manage** > **Authentication**. For **Implicit grant and hybrid flows**, ensure **ID tokens** (used for implicit and hybrid flows) is selected. Click **Save**.

Using the management GUI

To configure single sign-on with Microsoft Entra ID, complete these steps:

1. Select **Settings** > **Security** > **Single Sign-on**.
2. Enter the OpenID Configuration Endpoint URL of the authentication server. For Microsoft Entra ID, enter the following:

   ```
   https://login.microsoftonline.com/tenant/v2.0/.well-known/openid-configuration
   ```

   Where `tenant` is the ID of the tenant. This can be found by selecting from the menu **Entra ID** > **Click your application** > **Overview** > **Endpoints** > **OpenID Connect metadata document**.
3. For the **OpenID Credentials**, add the **Client ID** and **Client Secret** that you copied on from the Overview page in the Microsoft Entra ID interface.
4. For the User claim, the value to enter depends on how your authentication provider is configured. The User claim must match the name that the authentication service uses to specify the username attribute in the ID Token it sends to the system. Typically this value is preferred_username, but can be customized on the authentication provider.
5. For the Group claim, the value to enter depends on how your authentication provider is configured. The Group claim must match the name that the authentication service uses to specify the group attribute in the ID Token it sends to the system. Typically this value is groups, but can be customized on the authentication provider.
6. For Proxy server, consider how the system accesses the authentication provider. For an authentication provider within your network, a proxy server usually isn't needed. If you connect to the authentication provider through the Internet, check the box and ensure a proxy server is defined on the system.
7. Click **Save**. On the confirmation page, click **Confirm** to enable single sign-on for the system.

Single sign-on is enabled for the system. You can configure user groups to use single sign-on. Click **Navigate** to launch the **User Groups** page.

Using the command-line interface

To enable single sign-on, enter the following command:

```
chauthsinglesignon -oidcconfigurationendpoint https://login.microsoftonline.com/tenant/
v2.0/.well-known/openid-configuration-clientid xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx
-clientsecret xxxxxxxx -userclaim preferred_username -groupclaim groups -enable
```

In the example, `tenant` is the tenant name that is associated with your subscription. The values for the **`-clientid`** and the **`-clientsecret`** are the Open ID Client and Open ID Secret that are automatically generated when you created your system as an app registration in Microsoft Entra ID can be obtained from the Overview page in the Microsoft Entra ID interface. The values for the **`-userclaim`** and the **`-groupclaim`** should match the name of the claims configured for the ID Token on the authentication provider.

Configuring single sign-on with Okta

Okta can be configured as the authentication provider for the system.

Note: To configure single sign-on with Okta, you must upgrade the system to 8.5.3 or later release.

Note: Ensure that the prerequisite tasks are completed on the system before you configure single sign-on. For more information, see the Prerequisites section in "Configuring single sign-on" on page 427.

Prerequisites

The following prerequisite steps on Okta must be completed before you can configure single sign-on on the system:

1. Create a subscription for Okta. A free trial subscription is also available. For more information, see Okta . During subscription creation, you specify a tenant that is used to create a URL to access the Okta dashboard.
2. Access the Okta administrator dashboard by entering the following URL in a web browser:

   ```
   https://tenant.okta.com/admin/dashboard
   ```

   Where tenant is the name of the tenant that you specified when you created your subscription. Usually this tenant name is associated with your company or organization.
3. In the Okta interface, from the menu select **Applications** > **Applications** > **Create App Integration**.
4. For the Sign-in method, select **OIDC** – **OpenID Connect**. For the Application type, select **Web Application**. Click **Next**.

   Note: Each system must be added as a separate application.

   The following table shows the required fields and actions for the **General Settings** page in the Okta interface

Table 91. General settings on the app integration page	
Field	**Action**
App integration name	Enter a name to identify the system on Okta. If you are adding multiple systems, enter a unique.
Grant type	Ensure **Authorization Code** is selected.
Sign-in redirect URIs	Multiple redirect URIs can be specified for the management GUI. The redirect URI is consisted of the management IP address or hostname followed by `/sso`. For example, `https://hostname.com/sso`.

   The following table shows the required fields and actions for the **Assignments** page in the Okta interface.

Table 92. Assignments settings on the app integration page	
Field	**Action**
Controlled access	Select which users or groups can access the app. If you are unsure, select **Allow everyone** .
Enable immediate access	Ensure **Enable immediate access with Federation Broker Mode** is disabled

   Click **Save**. After the system is saved as a new app integration, the application reloads with the General page selected.
5. On the **General** page, copy the **Client ID**. This value must be specified as the Client ID on the Single sign-on page in the management GUI.
6. On the **General** page, ensure **Client authentication** is set to **Client secret**.
7. On the **General** page, ensure **Proof Key for Code Exchange (PKCE)** is disabled. The system does not support PKCE.
8. On the **General** page, copy the **Client Secret** or create a new one. This value must be specified as the Client ID on the Single sign-on page in the management GUI.

9. Ensure the authentication provider is configured to send back the group claim in the ID Token sent to the system. The group claim identifies which groups the authenticating user belongs to. Some authentication providers do not send the group claim by default, so this typically requires some configuration on the authentication provider. From the menu, select **Security** > **API** > **Authorization Servers**. Edit the default authorization server. Under the Claims tab, edit the claim type **ID**. Add a new claim called **groups** with **Include in token type** set to **ID Token / Always**, a **Value type** of Groups, **a Filter** of **Matches regex** with value .* and an **Include in** of **Any scope**.
10. Click **Save**.

Using the management GUI

To configure single sign-on with Okta, complete these steps:

1. Select **Settings** > **Security** > **Single Sign-on**.
2. 2. Enter the OpenID Configuration Endpoint URL of the authentication server. For Okta, enter the following:

   ```
   https://tenant.okta.com/oauth2/default/.well-known/openid-configuration
   ```

   where `tenant` is the hostname of the tenant.
3. For the **OpenID Credentials**, add the **Client ID** and **Client Secret** that you copied on from the Overview page in the Okta interface.
4. For the User claim, the value to enter depends on how your authentication provider is configured. The User claim must match the name that the authentication service uses to specify the username attribute in the ID Token it sends to the system. Typically this value is preferred_username, but can be customized on the authentication provider.
5. For the Group claim, the value to enter depends on how your authentication provider is configured. The Group claim must match the name that the authentication service uses to specify the group attribute in the ID Token it sends to the system. Typically this value is groups, but can be customized on the authentication provider.
6. For Proxy server, consider how the system accesses the authentication provider. For an authentication provider within your network, a proxy server usually is not needed. If you connect to the authentication provider through the internet, check the box and ensure a proxy server is defined on the system.
7. Click **Save**. On the confirmation page, click **Confirm** to enable single sign-on for the system.

Single sign-on is enabled for the system. You can configure user groups to use single sign-on. Click **Navigate** to launch the **User Groups** page.

Using the command line interface

To enable single sign-on, enter the following command:

```
chauthsinglesignon -oidcconfigurationendpoint https://tenant.okta.com/oauth2/default/.well-
known/openid-configuration-clientid xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx -clientsecret xxxxxxxx
-userclaim preferred_username -groupclaim groups -enable
```

In the example, `tenant` is the tenant name that is associated with your subscription. The values for the **`-clientid`** and the **`-clientsecret`** are the Open ID Client and Open ID Secret that are automatically generated when you created your system as an app integration in Okta can be obtained from the General page in the Okta. The values for the **`-userclaim`** and the **`-groupclaim`** should match the name of the claims configured for the ID Token on the authentication provider.

Configuring single sign-on with Duo Security

Duo Security can be configured as the authentication provider for the system.

Duo Security does not support acting as an Identity Provider (IdP) for single sign-on. To use Duo for single sign-on, an alternative authentication source must be configured. Refer to the Duo

Security documentation for Configure Your Authentication Source. You might also need to refer to the documentation for your chosen authentication source in order to integrate it with Duo Security.

Note: Ensure that the prerequisite tasks are completed on the system before you configure single sign-on. For more information, see the Prerequisites section in "Configuring single sign-on" on page 427.

Prerequisites

The following prerequisite steps on Duo Security must be completed before you can configure single sign-on on the system:

1. Create a subscription for Duo Security. A 30-day free trial subscription is also available. For more information, see Sign Up for a Free Trial | Duo Security. During subscription creation, you specify a tenant that is used to create a URL to access the Duo Security dashboard.
2. Access the Duo Security administrator dashboard by entering the following URL in a web browser:

   ```
   https://admin-tenant.duosecurity.com
   ```

   where `tenant` is the name of the tenant that you specified when you created your subscription. Usually, the tenant name is associated with your company or organization.
3. In the Duo Security interface, select **Applications** > **Protect an Application**.
4. Select **Generic OIDC Relying Party**.

   Note: Each system must be added as a separate application.

   The new application loads. The following table shows the required fields and actions for the **Metadata** section in the Duo Security interface.

Table 93. Metadata section

Field	Action	Details
Client ID	This value is automatically generated when the system is saved as an application.	This value must be entered on the Single Sign-on page in the management GUI under **OpenID Credentials**.
Client Secret	This value is automatically generated when the system is saved as an application.	This value must be entered on the Single Sign-on page in the management GUI under **OpenID Credentials**.
Discovery URL		Indicates the OpenID Connect configuration end point URL of the authentication server. This value must be entered on the Single Sign-on page in the management GUI under **Authentication server**.

The following table shows the required fields and actions for the **Relying Party** section in the Duo Security interface.

Table 94. Relying Party section

Field	Action	Details
Grant Type	Select **Authorization Code**.	**Authorization code** indicates that the client can request access to protected resources on behalf of users.

Table 94. Relying Party section (continued)		
Field	**Action**	**Details**
Sign-in Redirect URLs	Enter the locations where the authorization server sends users after they are successfully authorized and granted an authorization code or access token.	Multiple redirect URIs can be specified for the management GUI. For management GUI access, the redirect URI consists of the management IP address or hostname followed by `/sso`. For example, `https://hostname.com/sso`

Note: Duo Security does not support the use of IP addresses when accessing your system by using a web browser. Once single sign-on is configured, make sure that a hostname is used when accessing your system by using a web browser.

The following table shows the required fields and actions for the **OIDC Response** section in the Duo Security interface. In this section, you configure the OpenID Connect response from Duo Security to the system during authentication. Depending on the name of the IdP attributes on your authentication source, you create mappings between the IdP attributes (for example a SAML attribute sent back from your authentication source) and the claims that Duo Security sends back in the OIDC response.

Table 95. OIDC Response section		
Field	**Action**	**Details**
Scopes	Select **openid** and **profile**.	After selecting profile, a list of IdP Attribute and Claim mappings will appear. Ensure at least two attributes (username and group) are mapped correctly to claims in the OIDC response. For example, to map an IdP Attribute called **Username** from your authentication source, enter **Username** into the IdP Attribute field. You can choose the name of the claim that gets sent back in the OIDC response to the system. For example, to map the IdP Attribute called **Username** to an OIDC claim called **groups**, enter groups in the Claim field.
Additional Scopes	No action is required.	Additional scopes are not required.

Optionally, configure the policy settings for this application in the **Policy** section.

The following table shows the required fields and actions for the **Settings** section in the Duo Security interface.

Table 96. Settings section		
Field	**Action**	**Details**
Name	Enter a name.	Enter a name to identify the system on Duo Security. If you are adding multiple systems, enter a unique name.
Additional Scopes	No action is required.	Additional scopes are not required.

5. Click **Save**. After the system is saved as a new application, the application reloads with the chosen settings.

Using the management GUI

To configure single sign-on with Duo Security, complete these steps:

1. Select **Settings** > **Security** > **Single Sign-on**.
2. Enter the OpenID Configuration Endpoint URL of the authentication server. This is the Discovery URL from the Metadata settings of the application you created earlier in Duo Security. For Duo Security, it is in the following format:

   ```
   https://sso-tenant.sso.duosecurity.com/oidc/clientid/.well-known/openid-configuration
   ```

   where `tenant` is the name that is associated with your subscription, and `clientid` is the client ID that is associated with your application.
3. For the **OpenID Credentials**, add the **Client ID** and **Client Secret** that you copied on from the Sign-on tab in the Duo Security interface.
4. For the User claim, the value to enter depends on how your authentication provider is configured. The User claim must match the name that the authentication service uses to specify the username attribute in the ID Token of the OIDC response it sends to the system. Use the **Claim** name that maps to your username **IdP Attribute** from the OIDC Response section of your application.
5. For the Group claim, the value to enter depends on how your authentication provider is configured. The Group claim must match the name that the authentication service uses to specify the group attribute in the ID Token of the OIDC response it sends to the system. Use the **Claim** name that maps to your group **IdP Attribute** from the OIDC Response section of your application.
6. For Proxy server, consider how the system accesses the authentication provider. For an authentication provider within your network, a proxy server usually is not needed. If you connect to the authentication provider through the Internet, check the box, and ensure that a proxy server is defined on the system.
7. Click **Save**. On the confirmation page, click **Confirm** to enable single sign-on for the system.

Single sign-on is enabled for the system. You can configure user groups to use single sign-on. Click **Navigate** to launch the **User Groups** page.

Using the command-line interface

To enable single sign-on, enter the following command:

```
chauthsinglesignon -oidcconfigurationendpoint https://sso-tenant.sso.duosecurity.com/oidc/
clientid/.well-known/openid-configuration -clientid xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx
-clientsecret xxxxxxxx -userclaim username -groupclaim groups -enable
```

In the example, `tenant` is the tenant name that is associated with your subscription, and `clientid` is the client ID that is associated with your application. The values for the **`-clientid`** and the **`-clientsecret`** are the Open ID Client and Open ID Secret that are automatically generated when you created your system as an application in Duo Security, and are displayed in the Metadata section of the application on the Duo Security interface. The values for the **`-userclaim`** and the **`-groupclaim`**

must match the name of the claims that are configured for the ID Token in the **OIDC Response** section of the application on the authentication provider.

Configuring user groups for single sign-on

After you have configured the system to use single sign-on, you can configure user groups to use single sign-on. Like LDAP, you must create the remote groups with names which match the name of a group on the Identity provider.

Using the management GUI

To enable single sign-on for user groups, complete these steps:

For existing user groups:

1. In the management GUI, select **Access** > **Users by Group**.
2. Select the user group from the left navigation and select **User Group Actions** > **Properties**. For each user group on the authentication service, a corresponding user group must be created with the same name.
3. On the **User Group Properties** page, select **Single sign-on** under **Remote Authentication**.
4. Click **OK**.

For new user groups

1. Select **Access** > **Users by Group** > **Create User Group**.
2. On the **Create User Group** page, enter the following information:

   Group Name
   Enter the name of the group that is from AD FS. The name of the group on the system must match.

   Ownership Group
   If ownership groups are configured on your system, you can select an ownership group for the user group.

   Remote Authentication
   Select **Single Sign-on**.

   Role
   Select a role for the user group. The role determines privileges that users are granted when they are assigned to the user group.
3. Click **Create**.

Using the CLI

To configure user groups for single sign-on, complete these steps:

For existing user groups

```
chusergrp -remote yes <id or name of group>
```

For new user groups for remote users

```
mkusergrp -name <name> -role <role> -remote
```

Supported authentication providers

The system uses the OpenID Connect (OIDC) protocol to communicate with supported authentication providers. The system does not currently support the Security Assertion Markup Language (SAML) protocol.

Multifactor authentication
The system supports the following MFA providers:

- IBM Security Verify
- Duo Security

Single Sign-on
The system supports the following SSO providers:

- IBM Security Verify
- Microsoft Active Directory Federation Services (AD FS)
- Microsoft Azure Entra ID
- Okta
- Duo Security
- PingOne

Single sign-on providers are often used to provide multifactor authentication in addition to single sign-on. For example, Microsoft AD FS supports a range of third-party authentication providers for additional multifactor authentication capability such as Duo Security, RSA and Okta. For more information, see Configure Additional Authentication Methods for AD FS.

Note: Connections to authentication providers might break if you change the SSL security level. This happens if the authentication providers do not support any of the ciphers and protocols allowed at the set level, particularly TLS 1.3. In this case, you can revert to a lower level of security to restore connectivity.

Managing proxy server for SSO login

When using single sign-on, the system must be able to contact the single sign-on service.

Consider your network settings and whether the single sign-on service is accessible to the system. If the single sign-on service is not accessible, then firewall rules need to be added to allow communication, or alternatively a proxy server can be used. In this case, a proxy server can be used to permit traffic between the system and the single sign-on service.

Note: If the firewall rules or proxy server is not configured correctly, the system is not able to access the single sign-on service and users are not able to authenticate to the system by using single sign-on.

Using the management GUI

Complete the following steps:

1. Select **Settings** > **Security** > **Single-Sign on**.
2. In **Proxy server**, check the box **Use proxy server** so single sign-on can use proxy server.
3. Click **Save**.

Using the command-line interface

Use the command **chauthsinglesignon** to specify whether the proxy server should be used.

```
chauthsinglesignon -proxy yes/no
```

If proxy value is set to *yes* then the traffic is routed through the proxy defined on the system. If proxy value is set to *no* then the traffic is not routed through the proxy server. The proxy server can be created through **mkproxy**, for more information refer to **mkproxy command**.

Changing SSO credentials of system

Changing credentials for the Storage Virtualize application can be completed through management GUI or command-line interface (CLI).

If the security administrator changes the single sign-on credentials used for the application, the new credentials must be updated on the system. If the credentials are not updated, the system will not be able

to communicate with the authentication service and users will not be able to authenticate successfully to the system.

Using the management GUI

Complete the following steps:

1. Select **Settings** > **Security** > **Single Sign-on**.
2. In **Authentication server** enter the OpenID configuration URL for the SSO provider.
3. Select **OpenID Credentials** and enter the new client ID and client secret.
4. Click **Save**.

Using the command-line interface

Use the command **chauthsinglesignon** to change the credentials. By running the following command you can change the credentials.

```
chauthsinglesignon -clientid client_id
                   -clientsecret client_secret
```

where *clientid client_id* specifies the OpenID Connect client ID for the system, required to authenticate with the single sign-on provider. The *clientsecret client_secret* specifies the OpenID Connect client secret for the system, required to authenticate with the single sign-on provider.

Encryption

To use encryption on the system you must purchase and activate encryption licenses, set up your method of key management, and then create encrypted objects.

Encryption overview

Encryption is a technology that uses cryptography to help ensure confidentiality of sensitive information. Encryption uses keys to encode information so that it cannot be understood by unauthorized parties. Depending on your model, the system supports both encryption of data-at-rest and encryption of data-in-flight.

To use encryption of data-at-rest on the system, you need to:

- Purchase and activate licenses for the encryption feature.
- Decide which methods to use for managing the main keys.
- Configure encryption with the chosen key management methods.
- Create encrypted objects such as arrays, pools, or cloud accounts.

To use encryption of data-in-flight on the system, you need to:

- Purchase and activate licenses for the encryption feature.
- Create secure IP partnerships between systems.

Configuring encryption is a nondisruptive procedure. During the procedure, the system continues to process I/O operations normally and the existing storage objects are not impacted.

Encryption of Data at Rest (EDaR)

EDaR is the encryption of static data that is being stored on an internal drive or external storage. Data is automatically encrypted as it is written to the storage, and automatically decrypted as it is read from the storage. This feature protects against the potential exposure of sensitive user data and user metadata that is stored on discarded, lost, or stolen storage devices. Depending on your model, the system supports data encryption that uses encryption-capable hardware and software.

To use EDaR the encryption feature must be licensed and configured on the system.

EDaR is performed that uses the symmetric Advanced Encryption Standard (AES) algorithm with 256-bit encryption keys in XTS mode (XTS-AES-256), as defined in the IEEE 1619-2007 standard and NIST Special Publication 800-38E. The data encryption keys are protected by use of NIST's AES key wrap that uses an intermediate 256-bit wrapping key. This intermediate key is in turn protected that uses NIST's AES key wrap that uses an intermediate 256-bit wrapping key. The access keys are managed that uses whichever key management methods are enabled on the system. The wrapped keys are stored securely on the system in nonvolatile memory so that they can be accessed when the system starts up. During normal system operation, all unwrapped keys are stored securely in volatile memory, the contents of which are securely discarded on system shutdown or power loss. All data encryption keys (DEKs) and key encryption keys (KEKs) on the system are AES-256 bit keys, and all keys are protected by using an AES wrap-key operation.

All NVMe drives that are supported by the system, including IBM FlashCore Modules (FCMs) and a range of other third-party drives, are self-encrypting drives (SEDs) that encrypt data within the electrical circuit of each individual drive, with no performance penalty. Depending on the model, IBM FlashCore Modules are either FIPS 140-2 Level 2 or FIPS 140-3 Level 3 validated. For drives that support FIPS 140-3, the system must be running software version 8.6.2 or later to enable this.

When drives are connected through a Serial Attached SCSI (SAS) network, the SAS protocol chip provides data encryption capabilities, with no performance penalty. The SAS chip uses data encryption algorithms that are FIPS 140-2 Level 1 compliant.

The system software also can apply encryption to data on external storage devices that do not support built-in encryption. In this scenario, the software offloads the job of data encryption to the Advanced Encryption Standard New Instructions (AES-NI)-capable CPU within the hardware, with a small performance penalty, which depends on your configuration. The software uses algorithms that are FIPS 140-2 Level 1 compliant.

EDaR is always applied after data reduction technologies, such as compression and deduplication.

Encryption of Data in Flight (EDiF)

EDiF is the encryption of data that is being transmitted from one location to another. Data is encrypted before it is sent over the link and is decrypted when it is received on the other side. This feature protects against threats such as eavesdropping and unauthorized proxy attacks. Depending on your model, the system supports data encryption that uses encryption-capable hardware and software.

To use EDiF the encryption feature must be licensed, but there is no need to configure encryption on the system.

The system supports EDiF for data that is being replicated between two systems that are connected by Ethernet and configured in a secure Internet Protocol (IP) partnership. When a secured IP partnership is created, for example, between a production system and a recovery system, the data is secured as it travels through the network between the production system and the recovery system. Secured IP partnerships use a combination of IPsec and IKEv2 to secure data in flight. IKEv2 is an IPsec-based tunnelling protocol that uses secure key exchange algorithms to establish a secure connection to the partner system. IPsec is a suite of security protocols that helps ensure packets that are transmitted over the network are authenticated and encrypted.

In secured IP partnerships, the partner systems authenticate with each other, negotiate the security parameters, exchange encryption keys, and establish secured network tunnels through which encrypted data travels. Partner systems are authenticated by certificates that are issued by either the system's internal root CA or a trusted third-party root CA or intermediate CA.

The system does not support EDiF between host servers and the system, the system and back-end storage devices, or systems that are connected by FC that are configured in an FC replication partnership.

Licensing encryption

To use encryption, an encryption feature license must be purchased and activated for every machine in the system. A machine is defined to be either a control enclosure or a node, depending on the platform and model used. As an example, an eight-node system consisting of four IBM Storage FlashSystem

control enclosures would require four feature licenses. Alternatively, an eight-node system composing of eight IBM SAN Volume Controller nodes would require eight feature licenses. Licenses for the encryption feature are only made available in countries that permit the use of encryption technologies. Encryption can be configured once all required feature licenses have been activated. See for more details.

Key management methods

To configure encryption on the system, the user must have the SecurityAdmin role. When configuring encryption on the system, one or more key management methods must be used to manage the main keys for the system.

The system supports the following key management methods:

•

•

The system also supports an encryption recovery key, which can be used as a backup method alongside any other key management methods, to help ensure that encrypted data is available when there is a problem accessing the main keys.

For organizations with strict security policies regarding USB flash drives, the system supports disabling the USB ports to prevent unauthorized transfer of system data to portable media devices. If you have such security requirements, consider that uses key servers to manage encryption keys instead.

It is possible to simultaneously configure USB flash drives and key servers to help ensure that access to encrypted data is retained if either method is inaccessible, or if the keys are permanently lost for one of the methods.

Note:

- To protect against permanent key loss for one of the methods, a simultaneous configuration must be planned. It is not permitted to enable another key method when the keys for an existing method have already been lost.

If your system contains existing volumes in nonencrypted pools, you can migrate these volumes to encrypted pools after encryption has been configured.

Using encryption

Once the encryption feature has been configured, the system allows logical configuration objects to be created as encrypted, meaning that the system will automatically encrypt and decrypt the data being stored for that object.

Depending on your configuration, the system will automatically apply the right method of encryption (hardware-based or software-based). Encrypted arrays consisting of internal drives will always use hardware-based encryption and will only use the encryption key belonging to the array, even if the array is part of an encrypted pool. Encrypted pools consisting of external storage will always use software-based encryption and will only use the encryption key belonging to the pool.

Unlocking an encrypted system

When the system is unlocked and running normally, each node in the system keeps a copy of the main key in secure volatile memory. When a node restarts, the in-memory copy of the key is discarded and another copy is fetched from a partner node, without needing to retrieve the main key from elsewhere. If all the nodes in the system have lost the in-memory copy of the main key simultaneously (in scenarios such as a full system shut down or unexpected power loss), the system tries to retrieve the main key from one of the configured key management methods.

If key servers are configured, the key is retrieved automatically from any key servers that are online and available to the system. If USB flash drives are configured, the key is automatically retrieved from any USB flash drives that are locally installed. Should any of these methods be unavailable at the time the

system requires a master key. The system is locked and all encrypted arrays, pools, and cloud accounts are held offline.

Note: If all copies of the main key are unavailable when the system is locked, and your system contains encrypted arrays with SAS drives, then all SAS drives connected to the enclosure go offline, even drives belonging to unencrypted arrays.

If the encryption recovery key is configured, the recovery key can be supplied to unlock the system to bring encrypted data online.

The system requires a main encryption key to be available during the following operations:

- System power-on
- System restart
- User-initiated rekey operations
- System recovery

Planning for encryption

To use encryption on the system, you must purchase an encryption license, upload certificates, activate the license on the system, set up your method of key management, and create copies of the keys. If you use secured IP partnerships to secure connections between partnered systems, you also require an encryption license. If you have not purchased a license, contact a customer representative to purchase an encryption license.

To encrypt data that is stored on drives, the control enclosure on which they are connected must contain an active license and be configured to use encryption. For systems that also support self-encrypting drives, the drives also benefit from obtaining a license to protect access to these drives. When encryption is activated and enabled on the system, valid encryption keys must be present on the system when the system unlocks the drives or the user generates a new key. If key server encryption is enabled on the system, the key is retrieved from the key server. If USB encryption is enabled on the system, the encryption key must be stored on USB flash drives that contain a copy of the key that was generated when encryption was enabled.

If you are using encryption to protect data that is copied to cloud storage, the cloud account is always synchronized with the system encryption settings. If both USB flash drives and key servers are configured, the cloud account that is created supports both of these methods. If just one encryption method is configured and the other is disabled, the cloud account supports encryption with the remaining configured encryption method. To ensure that the cloud account supports encryption, one or both methods must be configured with active keys when the cloud account is created.

If a cloud account is created with one encryption method, you can configure the second method later, but the cloud account must be online while the configuration occurs. After the second method is configured, the cloud account will support both key providers.

If you use secured IP partnerships to secure connections between partnered systems, you also require an encryption license. If you have not purchased a license, contact a customer representative to purchase an encryption license.

Before you activate and enable encryption, you must determine the method of accessing key information during times when the system requires an encryption key to be present. The system requires an encryption key to be present during the following operations:

- System power-on
- System restart
- User initiated rekey operations
- System recovery

Several factors must be considered when planning for encryption.

- Physical security of the system

- Additional security requirements for USB ports and portable media. If your environment requires USB ports to be disabled, use key servers to manage encryption keys.
- Need and benefit of manually accessing encryption keys when the system requires
- Availability of key data
- Encryption license is purchased, activated, and enabled on the system
- If you are using IBM Security Guardium Key Lifecycle Manager to create and manage keys, ensure that you are using version 2.7.0 or later. If you are using version 2.7, the system supports one master (primary) key server and secondary key servers. However, replication is a manual process and during rekey operations, keys are not available until replication is completed. If you use version 3.0 or higher, the system supports multiple master key servers, which automatically replicates keys to all configured key servers.
- If you are using Gemalto SafeNet KeySecure key servers to create and manage keys, determine whether the system needs a user name and password to authenticate to the KeySecure key servers. If you plan to use a username and password to authenticate the system to these key servers, you must configure user credentials for authentication in the key server management interface. For KeySecure versions of 8.10 and up, administrators can configure a username and password to authenticate the system when it connects. Before version KeySecure 8.10, the use of a password is optional.
- If you are using Thales CipherTrust Manager key server to create and manage keys, determine whether the system needs a user name and password to authenticate to the CipherTrust Manager key servers. If you plan to use a username and password to authenticate the system to these key servers, you must configure user credentials for authentication in the key server management interface. After configuration, you can select your username from the interface. For more information on migrating key servers, refer to “Migrating from Gemalto SafeNet KeySecure to Thales CipherTrust Manager key servers” on page 465.

Key server encryption

Key servers provide useful features that make them desirable to use such as being responsible for encryption key generation, backups, and following an open standard that aids interoperability. When planning for key server encryption, the following items are important to consider.

SSL certificates

- IBM Storage Virtualize and key servers use SSL certificates to authenticate each other and create a secure connection.
- The system supports certificates signed by both the internal root certificate authority and a trusted thrid-party CA.
- The correct certificates and their certificate authority certificates must be installed in IBM Storage Virtualize and on the key servers to establish a secure connection.
- The connection between IBM Storage Virtualize and key servers becomes interrupted if certificates are renewed or expire. Installing the certificate authority certificates prevents this interruption by allowing the endpoint certificates to be renewed without disrupting the connection.
- IBM Security Guardium Key Lifecycle Manager key servers do not currently support chain of trust checking with IBM Storage Virtualize. The new system certificates generated must be exported to the key servers to re-establish a secure connection.
- If you are using multifactor authentication with IBM Security Verify, the management GUI are unavailable when you update the certificate. The new certificate generated must be exported using the CLI and added as a new signer certificate to IBM Security Verify for successful authentication.

Although both Thales CipherTrust Manager and Gemalto KeySecure key servers support the same type of configurations, you need to ensure that you complete the prerequisites on these key servers before you can enable encryption on the system. To configure the Thales CipherTrust Manager key servers, complete the following tasks before you enable encryption:

1. The server certificate that is selected for use with the KMIP interface in Thales CipherTrust Manager should be downloaded to your local machine, ready to upload to the system.

2. By default, Thales CipherTrust Manager is configured to require a username in the 'common name' field of the Storage Virtualize certificate. This username must be added when generating a new certificate or a new certificate request. You must also create a user with the same name in Thales CipherTrust Manager. This user owns the encryption keys for the system and must be added to the Key Users group in Thales CipherTrust Manager.
3. The Storage Virtualize certificate must be trusted by the Thales CipherTrust Manager key servers. Thales CipherTrust Manager does not support self-signed certificates. Thus, the Storage Virtualize certificate must be signed by a certificate authority (CA). If the system's root Certificate Authority (CA) is used to sign the certificate, then the system's root certificate must be installed as an External CA in Thales CipherTrust Manager and added to the list of CAs that can be used for KMIP. Alternatively, the system certificate can be signed by a trusted third-party CA. The third-party root certificate must be installed as an External CA in Thales CipherTrust Manager and added to the list of CAs that can be used for KMIP.
4. If you currently have encryption that is enabled with USB flash drives, at least one of the USB flash drives must be inserted into the system before key servers can be configured for managing keys.

For SafeNet KeySecure key servers, ensure that you complete the following tasks before you enable encryption:

1. Each key server must be configured to allow TLS 1.2 for secure communications.
2. Ensure that a valid SSL certificate from each KeySecure key server is installed on the system and in use. Either add the server certificate for each KeySecure key server, or add the root CA certificate that was used to sign each server certificate.
3. If you plan to use a username and password to authenticate the system to these key servers, you must configure user credentials for authentication in the key server management interface. For KeySecure versions of 8.10 and up, administrators can configure a username and password to authenticate the system when it connects. Before version KeySecure 8.10, the use of a password is optional. To set up authentication with a username and password between the system and KeySecure key servers, disable global keys on the **High Security** menu in the SafeNet KeySecure interface. When global keys are disabled, key servers cannot authenticate clients to create or access keys without valid credentials.
4. The Storage Virtualize certificate must be trusted by the SafeNet KeySecure key servers. If the system's root CA is used to sign the certificate, then the system's root certificate must be installed as an External CA in SafeNet KeySecure and added to the list of CAs that can be used for KMIP. Alternatively, the system certificate can be signed by a trusted third-party CA. The third-party root certificate must be installed as an External CA in SafeNet KeySecure and added to the list of CAs that can be used for KMIP. If Storage Virtualize certificate is self-signed, then the self-signed certificate must be installed as an External CA in SafeNet KeySecure and added to the list of CAs that can be used for KMIP. It is recommended to not use a self-signed certificate, as the connection between Storage Virtualize and the key servers is interrupted when the certificate is renewed, until the new certificate is added to the key servers.
5. If you currently have encryption that is enabled with USB flash drives, at least one of the USB flash drives must be inserted into the system before key servers can be configured for managing keys.

Note: For more information on the supported Generic KMIP key servers, see https://www.ibm.com/support/pages/ibm-storage-virtualize-supported-key-servers.

System requirements

Only one type of key server is supported at this time. The system implements the Key Management Interoperability Protocol (KMIP) that is sent over an SSL connection between the client and the server. Support is provided for internal-signed and external CA-signed certificates. The system validates the server's SSL certificate and conforms to the KMIP standard. Existing SAS hardware needs access to at least one master key to unlock and needs to be able to respond to key server master keys. Enabling key servers for the first time is a simple procedure. Once key server encryption is enabled, the type can be configured and enabled, server end-points can be created, and then the keys can be prepared and committed.

Security requirements

Security of all key server communications is governed by TLS 1.2 and TLS 1.3 protocols. Encryption keys are distributed between nodes in the system using TLS 1.2 and TLS1.3. The system uses AES-256 encryption that uses OpenSSL library interfaces. To establish a connection between the key server and the system, the key server or services must support the configured TLS version.

IP addresses and ports

All nodes that want to communicate with key servers must have their service IP address configured. A node must have its full service IP stack configured (address, gateway, mask) in order for that node to be a candidate for attempting to contact the key server. Key servers are typically set up on a private LAN, and this requires enforcement of service IP addresses. If only a subset of nodes have service IP addresses set, then those nodes without a service IP address log an error. The IP address that the user supplies must be the one that the system uses to communicate with the key server.

Each key server has a TCP port associated with its access. Since a key server serves multiple clients, the system allows the user to use a different port for each server and enables access for this port when required. KMIP server conformance mandates that TCP port 5696 be supported, so this is the default port for the server end point.

Key generation policy and key database

If key server encryption is enabled, then the key server generates and manages the master keys. The node generates all other keys.

The key database can be clustered or unclustered depending on the type of key server that is used. For unclustered key servers, the user needs to consider backup and replication of the key database. IBM Security Guardium Key Lifecycle Manager is an example of a key server product where replication must be configured in order for encryption keys to be shared automatically between IBM Security Guardium Key Lifecycle Manager instances. Without replication configured, manual backup and restore operations must be used. Other products might self-replicate, so other key server instances automatically have any new keys created. For IBM Security Guardium Key Lifecycle Manager, complete backups and restores by following the IBM Security Guardium Key Lifecycle Manager user guide. The SafeNet KeySecure server and Thales CipherTrust Manager support both clustered and unclustered key servers configuration.

Update requirements for key server encryption

If encryption was enabled on a pre-7.8.0 code level system and the system is updated to code level 7.8.x or above, you must run a USB rekey operation to enable key server encryption. Use the management GUI or run the **chencryption** command before you enable key server encryption. To perform a rekey operation, use the management GUI or run the following commands.

```
chencryption -usb newkey -key prepare
chencryption -usb newkey -key commit
```

The rekey operation must be run after the update is completed to the 7.8.x code level or higher and before you attempt to enable key server encryption.

Configuring encryption

If you use secured IP partnerships to secure connections between partnered systems, you also require an encryption license. If you have not purchased a license, contact a customer representative to purchase an encryption license.

The system supports both encryption of data at rest and encryption of data in flight. The system supports encryption of data in flight for replicated data between systems in an IP partnership. The secured IP partnership also requires an encryption license on each node or node canister in the partnership. This support protects against the potential exposure of sensitive user data and user metadata that is stored on discarded, lost, or stolen storage devices. Depending on your model, the system also supports self-encrypting drives, where data encryption is completed in the drive itself. Data encryption keys remain on the drive without being stored in system memory. In addition, the system supports automatic locks of encrypted drives when the system or drive is powered down. When the drive or system restarts, a

master key, which is created when you enable encryption on the system, is required to unlock the drive and continue I/O operations. If you add a new control enclosure to a system that has encryption that is already enabled, the control enclosure must also be licensed.

For systems that support more than one control enclosure, a licensed key for the encryption function must be added to all the control enclosures in the system. To obtain license keys, you need the machine type and model (MTM), serial number (S/N), and machine signature to manually activate the keys. Before you can obtain MTM, S/N, and machine signature, ensure that the control enclosure has been added to the system. These values are required if you are activating keys manually on the system.

If all copies of the encryption key are lost and your system contains encrypted arrays with of SAS drives, then all SAS drives connected the enclosure go offline, even those drives that belong to unencrypted arrays.

Licensing encryption

Before you can configure encryption on the system, you must purchase and activate encryption licenses. If you intend to use encryption of data-in-flight to secure IP connections between partnered systems, you also require an encryption license. If you have not already purchased a license, contact a customer representative.

Before you begin

For systems that support more than one control enclosure, a licensed key for the encryption function must be added to all the control enclosures in the system. To obtain license keys, you need the machine type and model (MTM), serial number (S/N), and machine signature to manually activate the keys. Before you can obtain MTM, S/N, and machine signature, ensure that the control enclosure has been added to the system. These values are required if you are activating keys manually on the system.

Complete the following steps to find machine type and model (MTM), serial number (S/N), and machine signature:

1. In the management GUI, select **Monitoring** > **System Hardware**. The system automatically detects if there is a second control enclosure candidate available.

   Click **Add Enclosure** and complete the wizard to add the second control enclosure to the system.
2. After the second control enclosure is added to the system, both control enclosures display. For each control enclosure, select the expand icon to open the System Hardware - Enclosure Details page.
3. On the System Hardware - Enclosure Details page, select **Enclosure Actions** > **Properties**. The machine type and model (MTM), serial number (S/N), and machine signature display on the Properties page. Complete this for both control enclosures.

Using the management GUI

Within the management GUI, there are two ways to activate an encryption license on the system. During system setup, you are prompted to either manually or automatically activate the license on the system. Automatic activation requires that the notebook that is being used to activate the license is connected to an external network.

If you purchased a license after system setup is completed, go to **Settings** > **Systems** > **Licensed Functions** and click to expand **Encryption Licenses**. These instructions assume that system setup is completed.

If you completed system setup and want to activate an encryption license, complete these steps:

1. In the management GUI, select **Settings** > **Systems** > **Licensed Functions**.
2. Click to expand **Encryption Licenses** and select the control enclosure on which to activate the license. You can select manual activation of encryption.
3. To activate encryption manually, complete these steps:

a. Select the control enclosures on which encryption will be activated, and select **Actions** > **Activate License Manually**.

b. On the Activate License Manually page, you must retrieve license keys by completing the form at https://www.ibm.com/servers/eserver/ess/landing/index.html . To complete the form by selecting your product and entering machine type and model, serial number, machine signature, and authorization code that was sent in your license agreement. Copy or download the keys.

c. Click **Activate**.

Using the command-line interface

You can use the CLI command to either activate the key directly or provide a path to the file where the key resides. As with activation through the management GUI, you need to use the authorization code that you received with your purchase agreement to obtain the key. If you purchased a license after system setup is completed, use the `Licensed Functions` option to activate the license. You can also use either of the following commands to activate an encryption license on the system:

1. To activate the license by using the key directly, enter the **`activatefeature -licensekey`** command, where key is the license key to activate a feature. The key consists of 16 hexadecimal characters that are organized in four groups of four characters with each group separated by a hyphen (such as 0123-4567-89AB-CDEF).
2. To activate the license with a file path that stores the key, complete these steps:
   a. Use scp to copy the license key file (2076_XXXXXXX.xml) to the /tmp directory.
   b. Using the command-line interface, enter the **`activatefeature -licensekeyfile filepath`**, where filepath is full path-to-file that contains all required license information (such as /tmp/keyfile.xml).

For more information, see **`Licensing and featurization commands`**.

Encryption with key servers

A key server is a centralized system that generates and manages encryption keys that are used by the system. Key servers are ideal in environments with many systems, since key servers send keys to the system automatically over the network without requiring physical access to the systems.

Some key servers support replication of keys among multiple key servers. If multiple key servers are supported, you can specify up to four key servers that connect to the system over both a public network or a separate private network.

The system supports the Key Management Interoperability Protocol (KMIP), which is a standard for encryption of stored data and management of cryptographic keys.

The system supports the following key servers to handle key management on the system.

- IBM Security Guardium Key Lifecycle Manager key server
- Generic KMIP key servers
  - Thales CipherTrust Manager key server
  - Gemalto SafeNet KeySecure key server
  - Fortanix DSM key server
  - Utimaco ESKM key server
  - Entrust CSP Vault key server

Note: For more information on the supported key servers, see IBM Storage Virtualize Supported Key Servers.

One of the key servers defined on the system must be designated as the primary key server. The primary key server is used by the system to create new encryption keys during a rekey operation. All key servers defined on the system are used to fetch the current encryption key when required. When using key

servers to manage the master key for the system, a copy of the master key is stored on each defined key server. The master key is a 256-bit AES key, generated by the key server.

When key management is configured using key servers, the system automatically fetches the key from each key server when required. Additionally, every 30 minutes the system will automatically validate that each configured key server is accessible and able to provide the current key to the system. The validation happens on every node in the system, to every key server defined on the system. A key server can have one of three statuses:

- **Online**: the key server is accessible and able to provide the current encryption key to all nodes in the system
- **Degraded**: the key server is accessible and able to provide the current encryption key to only some nodes in the system
- **Offline**: the key server is not accessible and cannot provide the current encryption key to any node in the system

Using key servers with IBM Security Guardium Key Lifecycle Manager

The system supports different types of key server configurations on IBM Security Guardium Key Lifecycle Manager. The following configurations are supported:

- IBM Security Guardium Key Lifecycle Manager key servers designate one primary key server, which can have up to three secondary key servers (also known as clones) defined. These additional key servers support more paths when it delivers keys to the system. However, during rekey operations, only the path to the primary key server is used. When the system is rekeyed, secondary key servers are not used until the primary key server replicates the new keys to these secondary key servers. Replication must be complete before keys can be used on the system. You can either schedule automatic replication or complete it manually with IBM Security Guardium Key Lifecycle Manager. During replication, key servers are not available to distribute keys or accept new keys. The total time that it takes for a replication to complete on the IBM Security Guardium Key Lifecycle Manager depends on the number of key servers that are configured as clones. If replication is triggered manually, the IBM Security Guardium Key Lifecycle Manager issues a completion message when the replication completes. Verify that all key servers contain replicated key and certificate information before keys are used on the system.
- Key servers can also be configured with multiple primary key servers where each key server can create new encryption keys. In this instance, any server can be set as the primary key server. The primary key server is the key server that the system uses when you create any new key server encryption keys. If multiple primary servers are enabled on the IBM Security Guardium Key Lifecycle Manager, the key is immediately replicated to the other key servers in the configuration.

For more information about the supported versions, see the IBM Documentation for IBM Security Guardium Key Lifecycle Manager.

When you create key server objects on the system for IBM Security Guardium Key Lifecycle Manager key servers, you must create a device group, in addition to name, IP address, port, and certificate information. The *device group* is a collection of security credentials (including keys and groups of keys) that allows for restricted management of subsets of devices within a larger pool. The system must be defined on the key server to the **SPECTRUM_VIRT** device group if you are using the default settings. If the **SPECTRUM_VIRT** device group does not exist on the key server, it must be created based on the **GPFS device family**. If you are configuring multiple key servers, the **SPECTRUM_VIRT** device group must be defined on the primary and all additional key servers. If you are configuring IBM Security Guardium Key Lifecycle Manager 5.0 or later, you need to create a catalog for the specified device group in FlashSystem (for example, **SPECTRUM_VIRT**) under **Endpoint Management** > **Catalog**. This catalog must be based on the GPFS catalog.

If you are using IBM Security Guardium Key Lifecycle Manager to create and manage keys, ensure that you are using version 2.7.0 or later. If you are using version 2.7, the system supports one master (primary) key server and secondary key servers. Keys are not available until replication is completed between the key servers. If you use version 3.0 or higher, the system supports multiple master key servers, which automatically replicates keys to all configured key servers.

Using Generic KMIP key servers

Encryption key servers create and manage encryption keys that are used by the system. In environments with many systems, key servers distribute keys remotely without requiring physical access to the systems.

The system supports up to four key servers. If the system is accessing multiple key servers, they need to belong to the same cluster of key servers.

Using key servers with Thales CipherTrust Manager and Gemalto SafeNet KeySecure

Thales CipherTrust Manager and Gemalto SafeNet KeySecure key servers also supports KMIP and creates keys on demand, sharing with the other clustered servers, providing redundant access. The system supports different types of configurations on key servers. The following configurations are supported:

- Thales CipherTrust Manager and KeySecure key servers use an active-active model, where multiple key servers are used to provide redundancy. In these configurations one key server must be specified as the primary key server. The primary key server is the key server that the system uses when you create any new encryption keys. The key is immediately replicated to the other key servers in the cluster. All of the key servers that are defined on the system can be used to retrieve keys. Although it is possible to configure a single key server instance, two key servers are recommended to ensure availability of keys, if one key server experiences an outage.
- The system supports up to four key servers. If the system is accessing multiple key servers, they need to belong to the same cluster of key servers.

If you are using Gemalto SafeNet KeySecure key servers to create and manage keys, determine whether the system needs a user name and password to authenticate to the KeySecure key servers. If you plan to use a username and password to authenticate the system to these key servers, you must configure user credentials for authentication in the key server management interface. For KeySecure versions of 8.10 and up, administrators can configure a username and password to authenticate the system when it connects. Before version KeySecure 8.10, the use of a password is optional.

If you are using Thales CipherTrust Manager key server to create and manage keys, determine whether the system needs a username and password to authenticate to the CipherTrust Manager key servers. If you plan to use a username and password to authenticate the system to these key servers, you must configure user credentials for authentication in the key server management interface. After configuration, you can select your username from the interface. For more information on migrating key servers, refer to .

Enabling encryption with key servers

You can use the management GUI or the command-line interface (CLI) to enable encryption with key servers.

Configuring encryption with IBM Security Guardium Key Lifecycle Manager key servers

Ensure that you complete the following tasks on the IBM Security Guardium Key Lifecycle Manager before you enable encryption:

1. The system supports only TLS versions 1.2 and 1.3. In the IBM Security Guardium Key Lifecycle Manager, specify **TLSv1.2** to use the TLS1.2 protocol, or TLSv1.3 to use the TLS1.3 protocol.
2. Ensure that the DB2 database service is started automatically on startup.
3. Ensure that a valid SSL certificate from IBM Security Guardium Key Lifecycle Manager is installed on the system and in use. If automatic replication is configured on IBM Security Guardium Key Lifecycle Manager, then this certificate needs to be uploaded to the system one time. However, if automatic replication is not configured on the IBM Security Guardium Key Lifecycle Manager, a certificate for each stand-alone key server must be uploaded to the system.
4. Ensure that a device group exists on IBM Security Guardium Key Lifecycle Manager called **SPECTRUM_VIRT**, which is based on the GPFS family. If you are configuring multiple key servers, the **SPECTRUM_VIRT** device group must be defined on the primary and all secondary key servers.

5. If encryption is enabled with USB flash drives, insert at least one of the USB flash drives into the system before key servers can be configured for managing keys.

For more information about completing these tasks, see the IBM Documentation for IBM Security Guardium Key Lifecycle Manager.

To enable encryption with a IBM Security Guardium Key Lifecycle Manager key server in the management GUI, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Encryption**.
2. Click **Enable Encryption**.
3. On the **Welcome** page, select **Key Servers**. Click **Next**.

   Note: You can also select both **Key Servers** and **USB flash drives** to configure both methods to manage encryption keys. If either method becomes disabled, you can use the other method to access encrypted data on your system.
4. Select IBM Security Guardium Key Lifecycle Manager **(with KMIP)** for the key server type.
5. Enter the name, IP address or domain name, and port for each key server. If you are configuring multiple key servers, the first key server that you specify is the primary key server. If you specify a fully qualified domain name, a DNS server must be configured on your system. To configure a DNS server for the system, select **Settings** > **Network** > **DNS**.
6. Select **SPECTRUM_VIRT** for the device group for the key servers. This device group must also be configured on each of the key servers for the system.
7. On the **Key Server Certificate** page, you must upload all the necessary key server certificates to the system. The key server certificate can be the key server endpoint certificate, the root CA certificate, or a file that contains all CA certificates within that chain. This file does not need to include the key server certificate, it should only have the intermediate and root CA certificates. In case both the endpoint certificate and the CA certificate of key servers are installed on the system, the endpoint certificate takes priority over the CA certificate. If the key servers are configured for automatic replication, the certificate is copied from the primary key server to all secondary key servers. All the IBM Security Guardium Key Lifecycle Manager instances are connected over secure connections with the same key server certificate. If replication is used on the IBM Security Guardium Key Lifecycle Manager, only one key server certificate needs to be installed on the system. The IBM Security Guardium Key Lifecycle Manager uses this single certificate to replicate keys with each other. If only one certificate is used and automatically replicated to all configured key servers, select the certificate that you downloaded to the system in the **Certificate** field. If automatic replication is not configured, select all the valid certificates that you downloaded to the system for each of the configured key servers. Click **Next**.
8. On the **System Certificate** page, click **Export system certificate** and download the system certificate. Copy the system certificate to the truststore for the **SPECTRUM_VIRT** device group on each configured key server. You must not upload the root certificate to the key servers, as IBM Security Guardium Key Lifecycle Manager does not support chain of trust checking for the SPECTRUM_VIRT device group. For more information, see online documentation of IBM Security Guardium Key Lifecycle Manager.
9. If you have USB flash drives configured as your encryption method, the **Disable USB Encryption** page displays. If you want to migrate to key servers and disable USB flash drives, select **Yes**. If you want keep both encryption methods, click **No**.
10. Click **Next**.
11. On the **Summary** page, verify the configuration for the key servers and click **Finish**.

To enable encryption with an IBM Security Guardium Key Lifecycle Manager key server in the command-line interface, complete the following steps:

1. To enable encryption on your system, see **chencryption** command.
2. To enable the key server type and supply the certificates of the key server, see **chkeyserverisklm** command.

3. To use an internally signed certificate or an externally signed certificate, see **satask exportrootcertificate** command and **chsystemcert** command.
4. To create the primary key server and up to three more secondary key servers and specify the key server certificate, see **mkkeyserver** command.
5. To verify that the system is prepared, see **lsencryption** command.

Configuring encryption with Thales CipherTrust Manager or Gemalto SafeNet KeySecure key servers

For SafeNet KeySecure key servers, ensure that you complete the following tasks before you enable encryption:

1. Each key server must be configured to allow TLS 1.2 for secure communications.
2. Ensure that a valid SSL certificate from each KeySecure key server is installed on the system and in use. Either add the server certificate for each KeySecure key server, or add the root CA certificate that was used to sign each server certificate.
3. If you plan to use a username and password to authenticate the system to these key servers, you must configure user credentials for authentication in the key server management interface. For KeySecure versions of 8.10 and up, administrators can configure a username and password to authenticate the system when it connects. Before version KeySecure 8.10, the use of a password is optional. To set up authentication with a username and password between the system and KeySecure key servers, disable global keys on the **High Security** menu in the SafeNet KeySecure interface. When global keys are disabled, key servers cannot authenticate clients to create or access keys without valid credentials.
4. The Storage Virtualize certificate must be trusted by the SafeNet KeySecure key servers. If the system's root CA is used to sign the certificate, then the system's root certificate must be installed as an External CA in SafeNet KeySecure and added to the list of CAs that can be used for KMIP. Alternatively, the system certificate can be signed by a trusted third-party CA. The third-party root certificate must be installed as an External CA in SafeNet KeySecure and added to the list of CAs that can be used for KMIP. If Storage Virtualize certificate is self-signed, then the self-signed certificate must be installed as an External CA in SafeNet KeySecure and added to the list of CAs that can be used for KMIP. It is recommended to not use a self-signed certificate, as the connection between Storage Virtualize and the key servers is interrupted when the certificate is renewed, until the new certificate is added to the key servers.
5. If you currently have encryption that is enabled with USB flash drives, at least one of the USB flash drives must be inserted into the system before key servers can be configured for managing keys.

To enable encryption with a Thales CipherTrust Manager or KeySecure key server using the management GUI, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Encryption**.
2. Click **Enable Encryption**.
3. On the **Welcome** page, select **Key Servers**. Click **Next**.

   Note: You can also select both **Key Servers** and **USB Flash Drives** to configure both methods to manage encryption keys. If either method becomes unavailable, you can use the other method to access encrypted data on your system.
4. Select **Thales CipherTrust Manager** or **Gemalto SafeNet KeySecure** for the key server type.
5. Enter the name, IP address or domain name, and port for each key server. If you are configuring multiple key servers, the first key server that you specify is the primary key server. If you specify a fully qualified domain name, a DNS server must be configured on your system. To configure a DNS server for the system, select **Settings** > **Network** > **DNS**.
6. On the **Key Server Credentials** page, enter a user name and password that is used to authenticate the system to the key servers.
7. On the **Key Server Certificate** page, you must upload all the necessary key server certificates to the system. The key servers can only use a certificate from a trusted third party, or a system signed

certificates. All instances are connected over secure connections with the same key server certificate. Either the server certificate for each key server, or the root CA certificate or a file that contains all CA certificates within that chain. This file does not need to include the key server certificate, only the intermediate and root CA certificates. Any server certificates take priority over any CA certificate that is installed on the system for the key servers. Click **Next**.

8. If you are using Thales CipherTrust Manager, the Storage Virtualize certificate must be signed by a CA.
   - If the Storage Virtualize certificate is signed by the system's root CA, click **Export Root Certificate**. Install the root certificate as an external CA on the Thales CipherTrust Manager key servers and add it to the list of external CAs that can be used for KMIP.
   - If the Storage Virtualize certificate is signed by a trusted third-party CA, install the third-party CA's root certificate as an external CA on the Thales CipherTrust Manager key servers and add it to the list of external CAs that can be used for KMIP.

   If you are using SafeNet KeySecure:

   - If the Storage Virtualize certificate is signed by the system's root CA, click **Export Root Certificate**. Install the root certificate as an external CA on the SafeNet KeySecure key servers and add it to the list of external CAs that can be used for KMIP.
   - If the Storage Virtualize certificate is signed by a trusted third-party CA, install the third-party CA's root certificate as an external CA on the SafeNet KeySecure key servers and add it to the list of external CAs that can be used for KMIP.
   - If the Storage Virtualize certificate is self-signed, click **Export System Certificate**. Install the Storage Virtualize certificate as an external CA on the SafeNet KeySecure key servers and add it to the list of external CAs that can be used for KMIP.
9. Select **The system's public key certificate has been transferred to each configured key server**.
10. If you have USB flash drives configured as your encryption method, the **Disable USB Encryption** page displays. If you want to migrate to key servers and disable USB flash drives, select **Yes**. If you want both encryption methods that are configured simultaneously, click **No**.
11. Click **Next**.
12. On the **Summary** page, verify the configuration for the key servers and click **Finish**.

To enable encryption with a Thales CipherTrust Manager or a KeySecure key server in the command-line interface, complete the following steps:

1. To enable encryption on your system, see **chencryption** command.
2. To enable the key server type and supply the root certificate authority (CA) certificate, see **chkeyserverciphertrustmanager** command.
3. To use an internally signed certificate or an externally signed certificate, see **satask exportrootcertificate** command and **chsystemcert** command.
4. To create the primary key server and up to three more secondary key servers and specify the key server certificate, see **mkkeyserver** command.
5. To verify that the system is prepared, see **lsencryption** command.

Configuring encryption with Generic KMIP key servers

For Generic KMIP key servers, ensure that you complete the following tasks before you enable encryption:

1. Each key server must be configured to allow TLS 1.2 for secure communications.
2. Ensure that a valid SSL certificate from each Generic KMIP key server is installed on the system and in use. Either add the server certificate for each Generic KMIP key server, or add the root CA certificate that was used to sign each server certificate.
3. If you plan to use a username and password to authenticate the system to these key servers, you must configure user credentials for authentication in the key server management interface. When

global keys are disabled, key servers cannot authenticate clients to create or access keys without valid credentials.

Note: Secret size (password) up to 32 bytes is supported and 64 bytes secret size is not supported by IBM Storage FlashSystem.

4. The IBM Storage FlashSystem certificate must be trusted by the Generic KMIP key servers. All Generic KMIP key servers does not support self-signed certificates. Thus, the Storage Virtualize certificate must be signed by a certificate authority (CA). If the system's root Certificate Authority (CA) is used to sign the certificate, then the system's root certificate must be updated in the Generic KMIP key server that can be used for KMIP. Alternatively, the system certificate can be signed by a trusted third-party CA. The third-party root certificate must be updated in Generic KMIP key server that can be used for KMIP.

   Note: For various KMIP key servers, certificate requirements are different and as mentioned:

   - For Fortanix DSM key server, the certificate contains "UUID" in the CN field.
   - For Thales CipherTrust Manager and Utimaco key servers, the certificate contains "username" in the CN field.
5. If you currently have encryption that is enabled with USB flash drives, at least one of the USB flash drives must be inserted into the system before key servers can be configured for managing keys.

Note: For more information on the supported Generic KMIP key servers, see IBM Storage Virtualize Supported Key Servers.

The management GUI steps are similar to the Thales CipherTrust Manager or Gemalto SafeNet KeySecure key servers.

Rekeying a system with key servers

The key server master key for the system can be changed by performing a rekey operation. During the rekey process, the key server generates a new master key and the existing master key becomes obsolete.

Consider rekeying the key server master key periodically according to your organization’s security policy, or when you think the existing key has been compromised or is known by an unauthorized party.

Only a single encryption method can be rekeyed at once. If you have multiple methods of encryption configured on your system, ensure that the current rekey operation is completed before starting another rekey operation.

Ensure that all key servers are online before starting, as the current encryption key will need to be supplied to the system during the rekey process. The key server master key cannot be rekeyed unless the current key is available.

Note: If all key servers have lost the current key, you must disable and then re-enable encryption using key servers. For more information, see

If the system is using an encrypted cloud account for Transparent Cloud Tiering, the cloud account must be online during the rekey operation.

If you are using multiple master or active-active key servers, new keys are automatically replicated to all configured key servers. In configurations with a single primary key server and multiple secondary key servers, only the primary key server is updated during the rekey operation. Any additional key servers go offline, and the system reports an error against those key servers until the new key is replicated from the primary to the secondary key servers.

Note: To avoid data loss, back up your data on the key server management application every time that you rekey.

Using the management GUI

Before creating a new key on all configured key servers, the key servers must be online and connected to the system. In the management GUI, select **Settings** > **Security** > **Encryption**. Expand **Key Servers** to

display details on all the configured key servers on the system. Verify that the status of the key servers is online and available to the system.

To rekey the system that uses key server encryption, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Encryption**.
2. Expand **Key Servers** to display all the configured key servers on the system and select **Rekey**.
3. Click **OK** on the message dialog. The encryption key is generated by the primary key server and is copied to the primary key server. If errors occur during the rekey process, status messages display problems with the copy or creation of a new key. To determine and fix other possible errors, select **Monitoring** > **Events**.

After the rekey operation completes, the new keys are replicated instantly if you have multiple primary or active-active key server configurations.

Using the command-line interface

Before creating a new key on all configured key servers, the key servers must be online and connected to the system. In the command-line interface, run the **lskeyserver** command to verify whether the key servers are online and available to the system.

To rekey a system with encryption using key servers, complete these steps:

1. To start the rekey process, enter the following command to prepare a new master key:

   ```
   chencryption -keyserver newkey -key prepare
   ```

   The new key is created successfully when the **keyserver_pmk_rekey_uid** field shows the name of the new key.
2. Commit the new key by running the following command:

   ```
   chencryption -keyserver newkey -key commit
   ```

   For more information, see **chencryption** command.

   The master key has been rekeyed successfully when the **keyserver_pmk_uid** field shows the name of the new key, and the **keyserver_pmk_rekey_uid** field is blank. For more information, see **lsencryption** command.

Disabling encryption with key servers

Encryption with key servers can be disabled using either the management GUI or the command-line interface.

Note: For security, encryption methods (including the recovery key) can only be disabled when physically connected to the technician port on the configuration node.

Using the management GUI

When disabling encryption using the management GUI, encryption using key servers is automatically disabled in the process. See for instructions on disabling encryption using the management GUI.

To disable only key servers, refer to the instructions described in Using the command-line interface.

Using the command-line interface

Follow these steps to disable encryption using key servers:

1. Identify the configuration node of the system. For more information, see “Configuration node” on page 487.
2. Connect your computer to the technician port of the configuration node. For more information, see the Node canisters page in the relevant Hardware Guide.

3. In a terminal window, use Secure Shell (SSH) software to connect to the cluster IP address of the system and authenticate using the credentials of any user with the SecurityAdmin role:

   ```
   ssh username@cluster_ip
   ```

   For more information, see **Connecting to the CLI with OpenSSH**.
4. To disable the encryption using key server, enter the following command:

   ```
   chencryption -keyserver disable
   ```

   For more information, see **chencryption** command.

   Encryption using key server has been disabled successfully when the **keyserver_pmk_uid** field is blank and the **keyserver_status** field is licensed. For more information, see **lsencryption** command.

Changing the primary key server

Using the management GUI

To change which key server is the primary key server, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Encryption**.
2. Expand **Key Servers** and right-click a key server that is not marked as the primary and select **Make Primary**.
3. The selected key server is now the primary key server.

Using the command-line interface

Migrating between key management methods

The migration between different key server types cannot be performed directly from one type of key server to another. USB flash drives encryption must be used to facilitate this process. If you want to migrate from one type of key server to another, you first must migrate from your current key servers to USB encryption, and then, migrate from USB to the other type of key servers.

To migrate from key servers to USB flash drives, use the command-line interface only. During migration, the system supports simultaneous configuration of both key management methods. After the migration completes, you can disable the old key management method.

Using the management GUI

During migration, the system does not disable the currently configured key management method until the new method is configured completely. Therefore, encrypted data can still be accessed with the current key until the migration is completed. For example, if you were migrating from USB flash drives to key servers, the old keys on the USB flash drive are still available until the key server encryption is configured. However, at least one of the USB flash drives with the current encryption key must be inserted into the system before you can migrate to the key server. After the key servers are configured, the old keys on the USB flash drive can no longer decrypt data on the system. Dispose of any old USB flash drives according to your recommended procedures for disposal of sensitive information. For organizations that require strict security policies regarding USB flash drives, the system supports disabling these ports to prevent unauthorized transfer of system data to portable media devices. After keys are migrated to the key servers, you can use the command-line interface to disable USB ports.

Note: The management GUI supports migration from USB flash drives to a key server encryption method only. To migrate from key servers to USB flash drives, you must use the command-line interface.

Before migrating to key server-based encryption, ensure that at least one USB flash drive with the current encryption key is inserted into the system. To migrate encryption from a USB flash drive to key servers, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Encryption**.
2. On the **Encryption** page, verify the following information:
   a. Expand **USB Flash Drives** and verify that USB flash drives are configured and detected in the system.
   b. Expand **Key Servers** and verify that key servers are not configured on the system.
3. Under **Key Servers**, click **Configure**.

   Note: To enable key servers to manage encryption keys on versions 8.1.0.0 or later, a USB rekey operation must be performed.
4. Select the type of key server that you are using to manage encryption keys. The system supports IBM Security Guardium Key Lifecycle Manager, Thales CipherTrust Manager, Gemalto SafeNet Key Secure, and Generic KMIP key servers to handle key management on the system. If you currently use Gemalto SafeNet Key Secure key servers, you can migrate to Thales CipherTrust Manager.

   Note: If you want to migrate to a Generic KMIP key server, select Thales CipherTrust Manager key server. For more information on the supported Generic KMIP key servers, see IBM Storage Virtualize Supported Key Servers.
5. Enter the name, IP address or fully qualified domain name, and port for each key server. If you are configuring multiple key servers, the first key server that you specify is the primary key server. If you specify a fully qualified domain name, a DNS server must be configured on your system. To configure a DNS server for the system, select **Settings** > **Network** > **DNS**.
6. If you selected **IBM SGKLM (with KMIP)**, the **Key Server Options** page displays. Select **SPECTRUM_VIRT** for the device group for the key servers. This device group must also be configured on each of the key servers for the system.
7. If you selected **Thales CipherTrust Manager** or **Gemalto SafeNet KeySecure**, the **Key Server Credentials** page displays. If you enabled authentication to the key server that uses a username and password, enter that username and password. These credentials are used to authenticate the system each time it connects to the key server. The username and password must match the credentials that are configured on the key server.
8. On the **Key Server Certificate** page, you must upload all the necessary key server certificates to the system. The key servers can use self-signed certificates or certificates that are signed by a certificate authority. You can also use both these types of certificates on the key servers. If all key server certificates are signed by the same CA, upload the root CA certificate. If the key servers use self-signed certificates, the certificates must be uploaded separately to the system. Any self-signed certificates take priority over any CA-signed certificate that is installed on the system for the key servers. If the key server certificates are signed by a chain of CAs, then a certificate chain file must be created and installed.

   Note: Generic KMIP does not support a self-signed system certificate. Only CA-signed certificates can be used by Generic KMIP key servers.
9. If using IBM Security Guardium Key Lifecycle Manager, export the system certificate and install that on the key servers.
10. If you are using Thales CipherTrust Manager, then the IBM Storage Virtualize certificate must be signed by a CA.
    - If IBM Storage Virtualize certificate is signed by the system's root CA, complete these steps:
      a. Select **Settings** > **Security** > **System Certificates**.
      b. Click **Export Root Certificate**.
      c. Install the root certificate as an external CA on the Thales CipherTrust Manager key servers and add it to the list of external CAs that can be used for Key Management Interoperability Protocol (KMIP).

- If IBM Storage Virtualize certificate is signed by a trusted third-party CA, install the third-party CA's root certificate as an external CA on the Thales CipherTrust Manager key servers and add it to the list of external CAs that can be used for KMIP.

11. If you are using Gemalto SafeNet Key Secure, then the IBM Storage Virtualize certificate can be CA-signed or self-signed.
    - If IBM Storage Virtualize certificate is signed by the system's root CA, complete these steps:
        a. Select **Settings** > **Security** > **System Certificates**.
        b. Click **Export Root Certificate**.
        c. Install the root certificate as an external CA on the SafeNet KeySecure key servers and add it to the list of external CAs that can be used for KMIP.
    - If the IBM Storage Virtualize certificate is signed by a trusted third-party CA, install the third-party CA's root certificate as an external CA on the SafeNet KeySecure key servers and add it to the list of external CAs that can be used for KMIP.
    - If the Storage Virtualize certificate is self-signed, complete these steps:
        a. Select **Settings** > **Security** > **System Certificates**.
        b. Click **Export System Certificate**.
        c. Install the certificate as an external CA on the SafeNet KeySecure key servers and add it to the list of external CAs that can be used for KMIP.
12. Return to the **System Encryption Certificate** page and select **The system's public key certificate has been transferred to each configured key server**.
13. On the **Disable USB Encryption** page, select **Yes** and click **Next**.
14. On the **Summary** page, verify the configuration for the key servers and click **Finish**. After the configuration completes, keys that were stored on the USB flash drives are not valid. Ensure that you dispose of all USB flash drives securely.

Using the CLI

Migrating from Gemalto SafeNet KeySecure to Thales CipherTrust Manager key servers

You can migrate from Gemalto SafeNet KeySecure key servers to Thales CipherTrust Manager key servers non-disruptively with the management GUI. The command-line interface must be used to view the unique ID of the encryption key.

Prerequisites

Before you migrate key servers, ensure that you complete the following tasks.

Update system to supported level
The system must be updated to a level that includes support for Thales CipherTrust Manager key servers.

Migrate encryption keys from Gemalto SafeNet KeySecure to Thales CipherTrust Manager key servers
The encryption key that the system uses must be migrated from the SafeNet KeySecure servers to the Thales CipherTrust Manager key servers. Use the **`svcinfo lsencryption`** to view the **`keyserver_pmk_uid`** that identifies the encryption key. The Thales CipherTrust Manager documentation covers migrating encryption keys from SafeNet KeySecure servers toThales CipherTrust Manager servers. You must use these instructions to migrate these encryption keys before you can complete the rest of the migration.

Configure system certificates on the system

IBM Storage Virtualize uses certificates to establish a secure connection to encryption key servers. Certificates must be configured before migrating to Thales CipherTrust Manager.

SafeNet KeySecure supports self-signed IBM Storage Virtualize certificates. Thales CipherTrust Manager does not support self-signed certificates, so the IBM Storage Virtualize certificate must be signed by a certificate authority (CA).

The IBM Storage Virtualize certificate can be signed by the system's root CA or by a trusted third-party CA. If the certificate is signed by the system's root CA, the root certificate must be added as an external certificate authority in Thales CipherTrust Manager. If the certificate is signed by a trusted third-party CA, the third-party root certificate must be added as an external certificate authority.

The root certificate must also be installed on the SafeNet KeySecure servers to ensure that the system can communicate with the SafeNet KeySecure servers when the new signed system certificate is installed.

If Thales CipherTrust Manager is configured to require a username in KMIP client certificates, then the username should be included in the IBM Storage Virtualize certificate.

Download the key server certificate
Download the key server certificate that is used with the KMIP interface in Thales CipherTrust Manager to your local workstation. You need to upload this certificate to the system during migration.

Create a Thales CipherTrust Manager username
By default, Thales CipherTrust Manager is configured to require a username in the 'common name' field of the client's SSL certificate. On Thales CipherTrust Manager, ensure that the following tasks are completed for this username.

1. Create a user with this username.
2. Ensure that this user owns the encryption key that was migrated to the Thales CipherTrust Manager key servers.
3. Ensure that this user is added to the **Key Users** group.

Using the management GUI

To migrate key servers, complete the following steps.

Note: Do not regenerate new encryption keys on the system until after the migration is completed.

1. In the management GUI, select **Settings** > **Security** > **Encryption**.
2. On the **Encryption** page, verify that all the SafeNet KeySecure key servers are online.
3. On the **Certificate** page, under **Key Server certificate authority** , click **Update Certificate**, and select the key server certificate. This key server certificate was downloaded to your local system from the KMIP interface in Thales CipherTrust Manager as part of the prerequisites.
4. Right-click one of the non-primary SafeNet KeySecure key servers and select **Remove**.
5. Select **Add Key Server** and add the key server details for the first CipherTrust Manager key server. Ensure that **Make Primary Key Server** is selected.

   Note: A certificate does not need to be added if the certificate authority was updated in Step 3.
6. Repeat steps 4 and 5 until all of the SafeNet KeySecure key servers are replaced byThales CipherTrust Manager key servers.

   Note: Do not select **Make Primary Key Server** for the remaining server.
7. The migration is complete. If necessary, you can rekey encryption keys on the system.

Using the CLI

Migrating between different key server types

The migration between different key server types cannot be performed directly from one type of key server to another. USB flash drives encryption must be used to facilitate this process. If you want to migrate from one type of key server to another, you first must migrate from your current key servers to USB flash drives encryption, and then, migrate from USB flash drives to the other type of key servers. For

example, migrating an IBM Storage Virtualize system that is servers that are configured with IBM Security Guardium Key Lifecycle Manager key servers to SafeNet KeySecure key servers.

Adding nodes or enclosures to an encryption-enabled system

You can add nodes or enclosures to the IBM Storage Virtualize system to increase its capacity. If encryption is enabled on the system, you must manually copy the current encryption key.

The system automatically shares the current encryption key in memory with nodes or enclosure when they are added, but does not automatically write the key file to any new USB flash drives. The current encryption key file is automatically written to new USB flash drives only when you enable encryption for the first time or rekey the system. To copy the current encryption key, you can use either of the following methods:

- Manually copy the current encryption key file from an existing USB flash drive to any new USB flash drives that were received with the new node or enclosure.
- Rekey the system.

Copying the current encryption key by using a USB flash drive

If encryption is configured with USB flash drives, you can copy the existing key file from the USB flash drive that contains the current encryption key for the system. You can use either the management GUI or the command-line interface to identify the USB flash drive.

Identifying the USB flash drive with the current encryption key

If you have multiple USB flash drives, you must first identify the USB flash drive that has a valid copy of the current encryption key for the system.

Using the management GUI

To identify the USB flash drive with the current encryption key by using the management GUI, complete these steps:

1. Insert a USB flash drive into an available USB port.
2. In the management GUI, select **Settings** > **Security** > **Encryption**
3. On the **Encryption** page, expand **USB flash drives**.
4. Check the value for **State**. If the key on the USB flash drive is valid, **validated** is shown. If a different value is shown, the USB flash drive does not have the right key.

Using command-line interface

To identify the USB flash drive with current encryption key from command-line interface, enter the following command:

```
lsportusb
```

The output shows the details of the USB port where the USB flash drive is inserted and the state of the USB flash drive. If the key on the USB flash drive is valid, **validated** is shown for **state**. If a different value is shown, the USB flash drive does not have the right key. For more information, see **lsportusb**command.

Note: Insert only one USB flash drive at a time and repeat the procedure to copy the key for each USB flash drive separately.

Copying the encryption key to a new USB flash drive

1. To identify the file name of the current encryption key for the system, enter the following command:

```
lsencryption
```

The key file name is displayed in the **usb_key_filename** field:

```
encryptionkey_000000E03BA000F0_000701DD00000001_my-storage-system-name
```

2. Insert the USB flash drive with the current encryption key into another system, such as your personal workstation or a server in your environment. Ensure that the drive is mounted.
3. Navigate to the file system of the USB flash drive with the current encryption key, locate the file with file name identified using **lsencryption** command on the device, and copy the located file.
4. Insert the new USB flash drive into the same system where you inserted the USB flash drive with the current encryption key. Ensure that the drive is mounted.
5. Navigate to the file system of the new USB flash drive, copy the encryption key, and safely unmount and remove the drive from the system.
6. Insert the new USB flash drive into the encryption-enabled system and run the **lsportusb** command to ensure that the copy of the key is valid and copied correctly.

Rekeying the system

You can rekey the system and manage the encryption keys when you use key server, USB flash drives, or both. Rekeying writes a new encryption key to all the USB flash drives currently inserted into the system. If encryption is configured with USB flash drives, it is possible to create new keys and store them on USB flash drives. If you configured key servers to manage encryption keys, you can generate new keys with the encryption key servers. If you use both a key server and USB flash drives, you must rekey each of them individually considering one at a time.

Tip: Before you rekey the system, ensure that the new nodes or enclosures are added to the system.

Key server
If you use a key server for encryption, see "Rekeying a system with key servers" on page 461.

USB flash drive
If you use USB flash drives for encryption, see "Rekeying a system with USB flash drives" on page 472.

Certificates that are used for encryption key servers

When you enable encryption with key servers, two types of certificates are required to ensure secure communication between the system and the encryption key server.

Certificates are the primary method that is used by the key servers to authenticate the system and for the system to authenticate to the key servers. The exchange of these certificates verifies that access to the encryption keys that are stored on the key servers is allowed. The authentication of the system ensures that the key servers do not give access to keys to an untrusted party. The authentication of the key servers ensures that the system does not ask for sensitive keys to be stored by an untrusted party. Security of the system relies on two factors. First, the public certificates of the key servers and the system must be exchanged securely so that each device can trust the other. Second, the key servers and the system must keep their private key, which is associated with the certificate, secure.

The key server certificates, which are used by the system to verify the key servers, must be transferred to the system and installed. When you are adding key servers to the system, either a certificate that is signed by a trusted third party, or a self-signed certificate can be used. A combination of these certificates can be used if you are adding more than one key server. If a combination of self-signed and CA-signed certificates are installed, then the self-signed key server certificates that are installed on the system take priority over any CA-signed key server certificates that are installed on the system.

Note: Generic KMIP does not support a self-signed system certificate. Only CA-signed certificates can be used by Generic KMIP key servers.

If all key servers are using a certificate that is signed by the same certificate authority, it is recommended that the CA certificate is installed in the key server type's settings (for example, **chkeyserverisklm -sslcert** ...) rather than the key server endpoints (for example, **mkkeyserver -sslcert** ...). The CA certificate for the key server type's settings is used only when you connect to a key server if the key server

endpoint does not have a certificate that is installed. If you are using individual self-signed certificates for each key server, the certificates are installed against each key server endpoint. If you are using key server certificates that are part of a chain of trust, see “Using certificate chains for key servers” on page 469.

If the key servers are part of the same system and are configured to replicate their keys and certificates with each other, then the key server certificates are copied to all of the key servers. The system can connect to all key servers with the same server certificate. In addition, the system’s certificate must be installed on each of the configured key servers. The key server administrator accepts the system certificate to grant access to the key servers. To configure the system certificate for secure communications, select **Settings > Security > Secure Communications**.

Note: For more information on the supported Generic KMIP key servers, see IBM Storage Virtualize Supported Key Servers.

Using certificate chains for key servers

In order for the system to authenticate by using SSL with a key server that uses signed certificate chains, the system requires multiple certificates to be installed. All certificates in the certificate chain (excluding the server certificate itself, that is, the leaf certificate) must be installed. For example, if a key server X’s certificate is signed by certificate authority Y, where certificate authority Y’s certificate is signed by Root certificate authority Z, then the system requires the certificates Y and Z to be installed on the system to successfully communicate with key server X. To install this certificate, create a single certificate file that contains the full chain of certificates, excluding the key server's certificate. The first certificate in the file is the CA certificate, which was used to sign the leaf of the certificate chain (the key server certificate). The last certificate in the file is the root CA certificate. In between the first certificate in the file and the root certificate, all of the intermediate CA certificates are specified in order. If multiple intermediate certificates exist, include these certificates in order from leaf to root.

The resulting certificates file must contain the first intermediate certificate at the top of the file, followed by subsequent intermediate CA certificates in order, one-by-one. Finally, the root CA certificate is last at the bottom of the file. The resulting file contains only the PEM data for each certificate with anything else, such as comments, removed from the file. Each certificate in the file is separated by a new line and each certificate begins with the header “-----BEGIN CERTIFICATE-----” and end with “-----END CERTIFICATE-----”.

An example certificate file, for a certificate chain of depth 3, might look like the following example. The example includes two certificates. The included certificates are the intermediate CA certificate Y first, followed by the root CA certificate Z. The key server certificate X is not included. If a bigger certificate chain is used, then more intermediate CA certificates would be included in the certificate file.

```
-----BEGIN CERTIFICATE-----
MIIF4DCCA8igAwIBAgICEAAwDQYJKoZIhvcNAQELBQAwgYUxCzAJBgNVBAYTAkdC
MRAwDgYDVQQIDAdFbmdsYW5kMRMwEQYDVQQHDApNYW5jaGVzdGVyMQwwCgYDVQQK
DANJQk0xDDAKBgNVBAsMA1NURzEPMA0GA1UEAwwGcm9vdENBMSIwIAYJKoZIhvcN
AQkBFhNqYW1pZXByeUB1ay5pYm0uY29tMB4XDTE4MTEwNzIxNTg1OVoXDTI4MTEw
NDIxNTg1OVoweDELMAkGA1UEBhMCR0IxEDAOBgNVBAgMB0VuZ2xhbmQxDDAKBgNV
BAoMA0lCTTEMMAoGA1UECwwDU1RHMRcwFQYDVQQDDA5pbnRlcm1lZGlhdGVDQTEi
MCAGCSqGSIb3DQEJARYTamFtaWVwcnlAdWsuaWJtLmNvbTCCAiIwDQYJKoZIhvcN
AQEBBQADggIPADCCAgoCggIBAKbE3NRBJHGcOLBdRRe9InwQqw2kIQ7dDMyAOFbh
/EcuZ03ex0q8HYgJmzGaSwjFZlFBOOZ3eZGWxMgFzKljX8Q7LuQBMi/t7Vxh9Qec
TDcjJ4jtClgFFzlYHFNk0dFJ9sryrhuNMeel64nafJGQQ+dy+vONbqBZfgaMXSbZ
DgQUmmtVqALn+PgFKOjRIMSTpm4hddxn0yg+Av1rAYlRkbpE0B69ktLRZMtnb0s/
dXE73LTMjA7+62P4/DHlU6cBs3ljURsV59OpYrGnTDO1mFIo8Ot89aHQotW/SWCm
RYIsxZ7Hfh6P/fdhScSdJXyyz8Nfs8qr2Ec3Vz2dss0k3MI0E4L2lyKcYaPgUmxF
DlO39ZEcffLrUsfI2GN9RuvE2umqk6FgSCbDP3mxjMxUs3rB8oItU2g5sJ1MVGqb
gYUtS/19hPpjDB/yAjI5Dlv0uUlefBTMPASpLEzF8AKkkVg98Vmkkx5eYjuu3Qpr
rHXN8CVGvhw9TUYzA9Pt9POmq6icMoBHpzujGcDrMflM8aE3sI+EmsFnbZnESeje
CWpc6Q8EZaxnnnHixfMBzX7Al8IIhnzeNUL5x46oUaFzpthYN3FlwWN/W8UDMPPr
h9mh82/CfZqol7suOh5WFWBbA2YGqRQacezMwxteAmvoCvxXhjducU1AzPLkwyPd
xXUBAgMBAAGjZjBkMB0GA1UdDgQWBBSXo1kphy8JAvB25abXPxEDUgHccTAfBgNV
HSMEGDAWgBQ6vl82Zxr3V7UJHz0StpCWmfakQzASBgNVHRMBAf8ECDAGAQH/AgEA
MA4GA1UdDwEB/wQEAwIBhjANBgkqhkiG9w0BAQsFAAOCAgEAIPyFwnJrVyEuC+Hi
8tj0L/aIJabuT8QNmZcqcSW4dfvDDupp7VUiheexCr9btuYIUGrL59clrrJwPFRw
iZTDq3sOQQQz8keYEwXGsjTFafFcdkFIKg1rD+JZojrCvMvpGn1Zt0eqCkla7PsB
pbaXRIbh/ybyafxzPhBYeFDdl9BQQf9vsa1DkNABsx80hy/lHjohaLlG5C+aG+5r
2bUXRNhgl3lSToQkD/JgdvBbgycIcriiL4u6mTJbnpiCB2JbBdyuAzTkdNAZStCv
kB/N+E+7agjx32IxHwjr1FjQZuyqRbHcsvz3f5VB4pGX/6+xeFQn6DlUewvB6CBg
OqsUddu+FKgFw4f3l59UVC1KicQ4Hokbst9WJGkSMgdvQOXyxwapHrsGiTTbzekG
```

```
dD7XUdUiLGcT1CDB3NrcHsJ1a2PCUgMQpeEsWDyUspBrvo40w3jfSoKtFmwp3o6a
vx9esOWzevLzAbv7YGZD/iXTQbUk2iNJjt5lRIee20GxZK+/BV36378U0CNKwdUt
mKAaOHY/DsGzOghGdvCz8/g5zhPN+bw9ZXm116cAygIJt2xytVy1cLvtbY2Vzgd9
zE1hCdm6oOfijOQ6F1JAzE/vewTxxOuNzp/T4LLGmrXtWwqhkrFrhRUiwsRb5zjY
XIh94QnyIz09Oj/W0qxllTPIGxk=
-----END CERTIFICATE-----
-----BEGIN CERTIFICATE-----
MIIF8jCCA9qgAwIBAgIJAMmFTDgLwpj8MA0GCSqGSIb3DQEBCwUAMIGFMQswCQYD
VQQGEwJHQjEQMA4GA1UECAwHRW5nbGFuZDETMBEGA1UEBwwKTWFuY2hlc3RlcjEM
MAoGA1UECgwDSUJNMQwwCgYDVQQLDANTVEcxDzANBgNVBAMMBnJvb3RDQTEiMCAG
CSqGSIb3DQEJARYTamFtaWVwcnlAdWsuaWJtLmNvbTAeFw0xODExMDcyMTEwMzda
Fw0zODExMDIyMTEwMzdaMIGFMQswCQYDVQQGEwJHQjEQMA4GA1UECAwHRW5nbGFu
ZDETMBEGA1UEBwwKTWFuY2hlc3RlcjEMMAoGA1UECgwDSUJNMQwwCgYDVQQLDANT
VEcxDzANBgNVBAMMBnJvb3RDQTEiMCAGCSqGSIb3DQEJARYTamFtaWVwcnlAdWsu
aWJtLmNvbTCCAiIwDQYJKoZIhvcNAQEBBQADggIPADCCAgoCggIBAM1wapd+YEWk
1/INfxjA+0mwwmz4fFgbyIjLtPjh6Z9KMRx8x368PXZ7Y1br/Zt9rbMx0cwxEO6b
U6XG/h91Fk4mCPKsjFRImlrFEHWDZ+/mv38SgF1yqTJ4Cb1ZjZS6bY0VwLo0gRiu
hUjwMUvoIksUc9+dauuQU8TtllqmXfOCVdsONiixlTCQy+sqbceYucEZItXkFWCH
6VAKrbyQS2QZP8KAoCw9j1skiUAqJ9hVwryQ602chqqzUGoAJbSBL2Xn8JQM2LEn
2KlVoIQFASTmnyKuaqeBHcIB5KCSDCQcy8TtybO/X38YmpYzfHC+L7Tr3gcgfEZf
vAitoPTFjqqBlXbNR5Vl8t8aMNq++WP42iI5sv40NWrSqsURXMxFFkMV++YYaeq9
kra5WCfeZMSYOzolM5fNBLrkA1sfFWqRH6IMGFSxlg+GRlbXS47RwpgncI/GSStN
RR9mfYiFU05VbnZSlLMhYZZv3KoZars+V5KNGfi95zBcMocinq/SVFXCRtIZfp3A
M6/OKL7mziA4NgjaaybmNbQOZGD4/eZDRSljSoMpL4xgs9BMEv4QyfKbOJXMQHqi
to7jNJRkpz3yxlwkLAGkigyCKmQFAzj0ykB3V+NwmLTJezKBolOvdjda+2rkP2/b
EvCBHIqgAP9t9EOLMDke8v0dyryN+ESJAgMBAAGjYzBhMB0GA1UdDgQWBBQ6vl82
Zxr3V7UJHz0StpCWmfakQzAfBgNVHSMEGDAWgBQ6vl82Zxr3V7UJHz0StpCWmfak
QzAPBgNVHRMBAf8EBTADAQH/MA4GA1UdDwEB/wQEAwIBhjANBgkqhkiG9w0BAQsF
AAOCAgEAiGo8x8nGlOVXQhkVzLzwoXMHNDHTpn13u8SGkX68zVaulvwuXOXl4d6b
QMXrMLQ3KonrKSpBCgTbyjDHp+6/NiFUs0t+O4gZczQHx8CgyjkRYmhrtcNW/RGb
66jYTkZpcTbJsrxOyzJY04Z0cr5Qxi9lgj+MHFPcSj3WJxWidurkQCh/AOSERuYg
LhNxNihomfKplKGUBKAhwTjBW7abIN4u/15iNaM/67X+/MOF1IKjB9U6ikgIjY5D
VetEEJhnYLjuzHpCmNeFZ1hbySMaiKP3kUNKSidEh5Oi+UP2UujgZi8VMPZApStW
AmRVCWtpiwcBD2WqsUuL+XcpqWBZ7GcNa89maMcybTjPAl0y2phi27ZXm78LmpCS
uK/UbRfpUpAsxiIPK1tnFkKcqdWBYNuOt1GIW9ADsYytDjo14X+i7axJYPtiqW+H
UxY/6b16c2/YOeCwIq+1g4882VtAoQfGmODC2AhhFqHrFhfOQ3Ag86x9scUjQl+J
4uYQkKsR1xW/y9Tx9dpqZZrZEu3B91SrtvA/6GtGFFs7jyeRCPdhb8+KF2NBonNl
FZC+p8qEeMSaCbBdI+ptL7vKQzzwKT/72Ts5BsABn4ZHXsD/39x0HPqSfWcEKR1e
s8hisZjl8mRRb5+sHRsmNoh9ki1AjIUTSjn/O+94Xahc0LvMWJk=
-----END CERTIFICATE-----
```

Encryption with USB flash drives

USB flash drives are low-cost storage devices that can be used to manage the master encryption key for the system. You can configure encryption and use USB flash drives store local copies of the master encryption key for the system, which can be provided to the system when required by installing them in USB ports.

Note: According to your organization's security policy, the USB ports on the system can be disabled to prevent them from being used to store encryption key files.

During configuration of encryption, the system must store a minimum of three copies of the master key on USB flash drives. It is possible to create more than three copies of the master key during the preparation phase of configuration, by physically swapping out the current USB flash drives with new USB flash drives until the wanted number of key copies have been written successfully. It is also possible to make more copies of the master key after configuration, by making a copy of the encryption key file from one of the USB flash drives. It can then be stored on other storage media, or managed using secret management software.

Note: The encryption key file written to USB flash drives is highly sensitive and should be managed securely. If copying encryption key files, take care not to make the file accessible to unauthorized parties.

Two options are available for accessing key information on USB flash drives:

USB flash drives are left inserted in the system at all times

If you want the system to bring encrypted data online automatically after a system restart, a USB flash drive must be left installed in one or more canisters in the system. When the system powers on, the encryption key is read from any available USB flash drive installed in the system. This method requires that the physical environment where the system is located is secure. If the location is secure, it prevents an unauthorized person from stealing the USB flash drives to make copies of the encryption keys, as well as stealing system components such as drives. The risk with this approach is that if the

location isn't secure and the master encryption keys are attached to the system, then stealing both the system and the USB flash drives means the encrypted data at rest can be accessed.

USB flash drives are not left inserted in the system

For the most secure operation, do not keep the USB flash drives inserted into the canisters in the system. However, this method requires that you manually install the USB flash drives containing the encryption master key in the canisters during certain operations when the system requires an encryption key to be present. USB flash drives that contain the current master key must be stored securely to prevent theft or loss. During operations where the system requires an encryption key to be present, one or more USB flash drives must be installed manually into any canister so data can be accessed. After the system reads the encryption key from USB flash drives, the system will be unlocked and encrypted data will be accessible. After the system is online, the USB flash drives must be removed and stored securely to prevent theft or loss. The advantage of this approach is that even if the system or drives are stolen, the encrypted data at rest cannot be accessed because the master keys are not present.

Enabling encryption with USB flash drives

The system supports enabling encryption that uses USB flash drives to store encryption keys. USB flash drive-based encryption requires physical access to the systems and is effective in environments with a minimal number of systems. For organizations that require strict security policies regarding USB flash drives, the system supports disabling these ports to prevent unauthorized transfer of system data to portable media devices.

If you have such security requirements, use key servers to manage encryption keys. In addition, if you are using USB flash drives to manage encryption keys but want to disable access to these ports for security reasons, you can migrate to encryption that uses key servers. For more information, see .

If the location is not secure, all USB flash drives with the key can be removed from the system and be stored securely. Extra copies of the key must be created and stored securely to ensure access to the system if the USB flash drives become damaged or stolen. During these operations, you are responsible for ensuring the security of the system. Use these general guidelines when you enable encryption and manage flash drives that contain an encryption key.

1. In addition to the copies that are generated on the USB flash drives when encryption is enabled on the system, make at least one more copy on another USB flash drive and store it in a secure location.
2. In addition, copy the encryption key to other forms of storage to provide resiliency and to mitigate risk, if, for example, the USB flash drives are from a faulty batch of drives.
3. Ensure that each copy of the encryption key is valid before writing any user data to the system. The system validates any key material on a USB flash drive when it is inserted into the canister. If the key material is not valid, the system logs an error. If the USB flash drive is not usable or failed, the system does not display the drive as an active USB flash drive.
4. Securely store all copies of the encryption key. As an example, any USB flash drives that are not left inserted into the system might be locked in a safe. Comparable precautions should be taken to securely protect any other copies of the encryption key stored on other forms of storage.

Using the management GUI

To enable USB flash drive encryption that uses the management GUI, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Encryption**.
2. Click **Enable Encryption**.
3. On the **Welcome** window, select **USB Flash Drives**, then click **Next**.

   Note: You can also select both **Key Servers** and **USB Flash Drives** to configure both methods to manage encryption keys. If either method becomes unavailable, you can use the other method to access encrypted data on your system.

4. In the wizard, you are prompted to insert the required number of USB flash drives into the system. If the system has three or more USB ports, at least three USB flash drives must be entered. If the system has less than three USB ports, a USB flash drive must be entered into every port on the system.
5. When the system detects the USB flash drives, the encryption key is automatically copied to the USB flash drives. Ensure that you create any required extra copies for backups.
6. If the system has less than three USB ports, the USB flash drives containing the new key must be removed from the system and new drives added until the number of key copies is three or more. When a new USB flash drive is added while the rekey is in progress, the key is copied onto it automatically.
7. After all copies are completed, click **Confirm** to complete the enablement process.

   Note: You can leave the USB flash drives inserted into the system. However, the area where the system is located must be secure to prevent the USB flash drives from being lost or stolen. If the area where the system is located is not secure, remove all the USB flash drives from the system and store them in a secure location.

Using the command-line interface

1. To enable the encryption that uses USB flash drives, enter the following command:

   ```
   chencryption -usb enable
   ```

   Encryption that uses USB flash drives has been enabled successfully when the status field shows enabled upon executing **lsencryption** command.
2. Insert the required number of USB flash drives into the system. If the system has three or more USB ports, at least three USB flash drives must be entered. If the system has less than three USB ports, a USB flash drive must be entered into every port on the system.
3. Enter the following command to prepare a new USB master key:

   ```
   chencryption -usb newkey -key prepare
   ```

   The new key has been created successfully when the `usb_rekey_filename` field shows the name of the new key upon executing **lsencryption** command.

   If the system has less than three USB ports, the USB flash drives containing the new key must be removed from the system and new drives added until the number of key copies is three or more. When a new USB flash drive is added while the rekey is in progress, the key is copied onto it automatically.
4. After all copies are completed, run the following command to complete the enablement process:

   ```
   chencryption -usb newkey -key commit
   ```

   The master key has been rekeyed successfully when the `usb_key_filename` field shows the name of the new key and the `usb_rekey_filename` is blank, upon executing **lsencryption** command. For more information, see **lsencryption** command.
5. If an error occurs during the creation of the initial master key creation process, it can be cancelled by running the following command:

   ```
   chencryption -usb newkey -key cancel
   ```

   If the key creation is cancelled, the process must be re-attempted before encryption with USB flash drives is fully enabled on the system. For more information, see **chencryption**.

Rekeying a system with USB flash drives

Rekeying is the process of creating a new key for the system. To create a new key, encryption must be enabled on the system; however, the rekey operation works whether there are encrypted objects or not. Only a single encryption method can be rekeyed at once.

If you have multiple methods of encryption configured on your system, ensure that the current rekey operation is completed before starting another rekey operation. If you are generating new keys for cloud

storage, the cloud account must be online during the rekeying operation. Rekeying is a non-disruptive process and can be completed without any impact to availability of encrypted or non-encrypted storage.

Before creating a new key, ensure that at least one USB port contains a USB flash drive that contains the current key. During the rekey process, a new key is generated and copied to the USB flash drives. The new key is then used instead of the current key. The rekey operation fails unless at least one USB flash drive contains the current key. To complete a rekey of the system, you need at least three USB flash drives to store the copied key material.

Once a rekey of the system is complete, the old key material will no longer be of use and cannot be used to unlock the encrypted storage. If any USB flash drives which contained a copy of the old key were not plugged in during the rekey scenario, they will not contain a copy of the new master key and the key will have to be copied to the drive manually.

Using the management GUI

To rekey the system with encryption using USB flash drives in the management GUI, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Encryption**.
2. Expand **USB Flash Drives** to display all the detected USB flash drives on the system and select **Rekey**.
3. In the wizard, you are prompted to insert the required number of USB flash drives into the system. If the system has three or more USB ports, at least three USB flash drives must be entered. If the system has less than three USB ports, a USB flash drive must be entered into every port on the system.
4. When the system detects the USB flash drives, the encryption key is automatically copied to the USB flash drives. Ensure that you create any required extra copies for backups.
5. If the system has less than three USB ports, the USB flash drives containing the new key must be removed from the system and new drives added until the number of key copies is three or more. When a new USB flash drive is added while the rekey is in progress, the key is copied onto it automatically.
6. After all copies are completed, click **Confirm** to complete the rekey process.
7. If errors occur during the rekey process, status messages display problems with the copy or creation of a new key. For example, if the minimum number of USB drives are inserted but none of them have an existing encryption key, the rekey operation fails. To determine and fix other possible errors, select **Monitoring** > **Events**.

Using the command-line interface

To rekey the system with encryption using USB flash drives on the command-line interface, complete these steps:

1. Insert the required number of USB flash drives into the system. If the system has three or more USB ports, at least three USB flash drives must be entered. If the system has less than three USB ports, a USB flash drive must be entered into every port on the system.
2. To start the rekey process, run the following command to prepare a new master key:

   ```
   chencryption -usb newkey -key prepare
   ```

   The new key has been created successfully when the **usb_rekey_filename** field shows enabled upon executing **lsencryption** command.
3. If the system has less than three USB ports, the USB flash drives containing the new key must be removed from the system and new drives added until the number of key copies is three or more. When a new USB flash drive is added while the rekey is in progress, the key is copied onto it automatically.
4. After all copies are completed, run the following command to complete the rekey process:

   ```
   chencryption -usb newkey -key commit
   ```

The master key has been rekeyed successfully when the **usb_key_filename** field shows the name of the new key and the **usb_rekey_filename** is blank upon executing the **lsencryption** command. For more information, see **lsencryption** command.

5. If an error occurs during the rekey process, it can be cancelled by running the following command:

```
chencryption -usb newkey -key cancel
```

   If the rekey is cancelled, the system will continue to use the existing master key. For more information, see **chencryption**.

Disabling encryption with USB flash drives

Encryption with USB flash drives can be disabled using the management GUI.

Note: For security, encryption methods (including the recovery key) can only be disabled when physically connected to the technician port on the configuration node.

Using the management GUI

When disabling encryption using the management GUI, encryption using USB flash drives is automatically disabled in the process. See for instructions on disabling encryption using the management GUI.

Using the command-line interface

Follow these steps to disable encryption using USB flash drives:

1. Identify the configuration node of the system. For more information, see Configuration node.
2. Connect your computer to the technician port of the configuration node. For more information, see the Node canisters page in the relevant Hardware Guide.
3. In a terminal window, use Secure Shell (SSH) software to connect to the cluster IP address of the system and authenticate using the credentials of any user with the SecurityAdmin role:

```
ssh username@cluster_ip
```

   For more information, see **Connecting to the CLI with OpenSSH**.
4. To disable the encryption using USB flash drives, enter the following command:

```
chencryption -usb disable
```

   For more information, see **chencryption** command.

   Encryption using USB flash drives has been disabled successfully when the **usb_key_filename** field is blank and the **status** field is licensed. For more information, see **lsencryption** command.
5. The encryption key files remain on the USB flash drives but are not used again. The security administrator of the system is responsible for removing unused or expired encryption key files from the USB flash drives.

Encryption recovery key

The system supports enablement of an encryption recovery key to supplement an existing encryption method (such as USB flash drives or key servers). The encryption recovery key can be used to bring the system's encrypted storage back online after an outage.

The encryption recovery key is an ASCII string, which should be copied down and stored in a safe location, such as a password manager. Once enabled, the encryption recovery key can be supplied in place of a USB flash drives encryption key or key server encryption key to bring the system's encrypted storage back online after an outage. The encryption recovery key is intended to supplement an existing encryption method (such as USB flash drives encryption or key server encryption) and should not be used to 'stand-alone'.

Enabling the encryption recovery key

You can use the management GUI or the command-line interface (CLI) to enable the encryption recovery key.

Using the management GUI

To enable the encryption recovery key while enabling encryption on the system for the first time, complete these steps:

1. In the management GUI, select **Settings** > **Security** > **Encryption** and click **Enable Encryption**.
2. Once enabled, the encryption recovery key page will be displayed. Click **Generate recovery key** to generate key and begin the enablement process.

   Note: The encryption recovery key will not be redisplayed after completing the wizard.
3. Enter the encryption recovery key into the input field and click **Finish**.
4. If entered correctly, an information message should appear to state the recovery key has been entered correctly. Click **Close**. The encryption recovery key is enabled and can be used to unlock the system.

To enable the encryption recovery key while also enabling a second encryption method (such as USB flash drives encryption or key server encryption), toggle the corresponding drop-down menu on the **Settings** > **Security** > **Encryption** page and click **Configure**. Follow the wizard to enable the selected encryption method.

Using the command-line interface

To enable the encryption recovery key while both USB flash drives encryption and key server encryption are already enabled, you must use the command-line interface to configure the recovery key. Refer to the **chencryption** command.

Follow these steps to the enable encryption recovery key:

1. Enter the following command to enable the recovery key:

   ```
   chencryption -recoverykey enable
   ```
2. Enter the following command to prepare a new recovery key:

   ```
   chencryption -recoverykey newkey -key prepare
   ```

   The new recovery key will be displayed on screen.

   Note: The recovery key is sensitive and must be stored in a safe location.
3. To confirm that the recovery key has been stored correctly, the system requires the recovery key to be confirmed. To confirm the recovery key, run the following command and enter the new recovery key when prompted:

   ```
   chencryption -recoverykey newkey -key confirm
   Enter the new recovery key for the system:
   ```

   A confirmation message is displayed when the recovery key has been entered correctly.
4. Commit the recovery key by running the following command:

   ```
   chencryption -recoverykey newkey -key commit
   ```

Rekeying the encryption recovery key

The encryption recovery key for the system can be changed by performing a rekey operation. During the rekey process, the system generates a new encryption recovery key and the existing recovery key becomes obsolete.

Consider rekeying the encryption recovery key periodically according to your organization's security policy, or when you think the existing recovery key has been compromised or is known by an unauthorized party.

Only a single encryption method can be rekeyed at once. If you have multiple methods of encryption configured on your system, ensure that the current rekey operation is completed before starting another rekey operation.

Ensure that the existing recovery key is available, as it needs to be supplied to the system during the rekey process. The recovery key cannot be rekeyed unless the current recovery key is known and available.

Note: If you have lost the recovery key, you must disable and then re-enable the recovery key.

The new encryption recovery key for the system is sensitive, so make sure that the new recovery key is stored securely in a safe location. It is suggested that the rekey procedure is used in a private location, where your browser or terminal window cannot be seen by others.

Using the management GUI

The system does not currently support rekeying the encryption recovery key by using the management GUI. Refer to the instructions described in "Using the command-line interface" section.

Using the command-line interface

Follow these steps to rekey the encryption recovery key:

1. The current recovery key must be supplied to the system within 30 minutes of starting a rekey. To validate the current recovery key, enter the following command:

   ```
   chencryption -recoverykey validate
   Enter the recovery key for the system:
   ```

   An interactive prompt is displayed on the screen. When the correct recovery key has been entered, a success message appears.
2. To start the rekey process, enter the following command to prepare a new recovery key:

   ```
   chencryption -recoverykey newkey -key prepare
   ```

   The new recovery key will be displayed on screen.

   Note: The recovery key is sensitive and must be stored in a safe location.

   The system creates an identifier for the new recovery key, which can be used as a label when storing the key in a safe location (for example, in a password manager).
3. To confirm that the recovery key has been stored correctly, the system requires the recovery key to be confirmed. To confirm the recovery key, run the following command and enter the new recovery key when prompted:

   ```
   chencryption -recoverykey newkey -key confirm
   Enter the new recovery key for the system:
   ```

   A confirmation message is displayed when the recovery key has been entered correctly.
4. Commit the recovery key by running the following command:

   ```
   chencryption -recoverykey newkey -key commit
   ```

   For more information, see **chencryption** command.

The recovery key has been rekeyed successfully when the **recovery_key_name** field shows the name of the new recovery key, and the **recovery_key_rekey_name** field will be blank. For more information, see **lsencryption** command.

Note: A rekey operation needs to be performed after the upgrade from 8.7.0 to a later software version so that rekey time can be updated in GUI or CLI.

Unlocking the system using the encryption recovery key

Systems with encryption configured must be unlocked by using an encryption key to bring encrypted storage online.

During startup, the system automatically attempts to fetch an encryption key to unlock the system. Depending on the configuration, the system automatically looks for encryption keys that are stored on locally attached USB flash drives or network-attached key servers. When certain events or outages occur, such as a full system power down, the system must retrieve an encryption key that can unlock the system. If the system is not able to automatically retrieve an encryption key then the encrypted arrays, pools, and cloud accounts are held offline and data are inaccessible.

If the system has configured an encryption recovery key, this key can be manually supplied to the system to bring the encrypted storage back online. After entering the recovery key, monitor the system to confirm that the encrypted storage is now online and that encrypted data is accessible to host applications.

Using the management GUI

The system does not currently support unlocking the system with the encryption recovery key by using the management GUI. Refer to the instructions described in "Using the command-line interface" section.

Using the command-line interface

If the system is online, but encrypted storage is held offline, the recovery key can be supplied by running the following command on the configuration node:

```
chencryption -recoverykey validate
```

For more information see **chencryption** command.

In rare scenarios, the system is not able to come online because all quorum disks are offline. This can happen when all quorum disks are configured on Serially Attached SCSI (SAS) drives, and the SAS adapter is locked by a missing encryption key. In this scenario, nodes are held in service state (reporting node error 550) and cannot join the system. The recovery key can be supplied by running the following command on any node in service state:

```
satask unlockencryption -recoverykey
```

Enter the recovery key that uses the interactive prompt. A confirmation message is displayed when the recovery key has been entered correctly.

Disabling the encryption recovery key

The encryption recovery key can be disabled using either the management GUI or the command-line interface.

Note: For security, encryption methods (including the recovery key) can only be disabled when physically connected to the technician port on the configuration node.

Using the management GUI

When disabling encryption using the management GUI, encryption recovery key is automatically disabled in the process. See for instructions on disabling encryption using the management GUI.

To disable only the encryption recovery key, refer to the instructions described in Using the command-line interface.

Using the command-line interface

Follow these steps to disable the encryption recovery key:

1. Identify the configuration node of the system. For more information, see Configuration node.
2. Connect your computer to the technician port of the configuration node. For more information, see the Node canisters page in the relevant Hardware Guide.
3. In a terminal window, use Secure Shell (SSH) software to connect to the cluster IP address of the system and authenticate using the credentials of any user with the SecurityAdmin role:

   ```
   ssh username@cluster_ip
   ```

   For more information, see **Connecting to the CLI with OpenSSH**.
4. To disable the encryption recovery key, enter the following command:

   ```
   chencryption -recoverykey disable
   ```

   For more information, see **chencryption** command.
5. The recovery key has been disabled successfully when the **recovery_key_name** field is blank and the **recovery_key_status** field is licensed. For more information, see **lsencryption** command.

Decommissioning encryption

If encryption is no longer needed, you can disable the function on the system. This might be required when all encryption master keys for the system have been lost, when returning a loan system, or when repurposing a storage system for a different use. After decommissioning encryption on the system, all encrypted data and encryption configuration is removed.

During the disablement process, if the system is using key servers to manage keys, the master encryption keys remain on the key server but are not used again. The administrator of the key server is responsible for removing unused or expired keys from the key server. If the system is using USB flash drives to manage keys, the encryption key files remain on the USB flash drives but are not used again. The security administrator of the system is responsible for removing unused or expired encryption key files from the USB flash drives. If the system is using an encryption recovery key, any stored recovery keys are no longer used and can be discarded.

Note: All encrypted arrays, pools, and cloud accounts must be removed before disabling encryption.

You can use the management GUI to fully disable encryption on the system.

Note: For security, encryption methods (including the recovery key) can only be disabled when physically connected to the technician port on the configuration node.

Before you begin

Follow these steps to connect to the technician port on the configuration node:

1. Identify the configuration node of the system. See Configuration node for more details.
2. Connect your computer to the technician port of the configuration node. For more information, see the Node canisters page in the relevant Hardware Guide.

Using the management GUI

To fully disable encryption in the management GUI, complete the following steps:

1. In the management GUI, select **Settings** > **Security** > **Encryption**.
2. On the Encryption page, change the State to **Disabled**.

Using the command-line interface

To fully disable encryption using the command-line interface, complete the following steps:

1. In a terminal window, use Secure Shell (SSH) software to connect to the configuration node and authenticate using the credentials of any user with the SecurityAdmin role:

   ```
   ssh username@192.168.0.1
   ```

   For more information, see **Connecting to the CLI with OpenSSH**.
2. To disable encryption on a system that uses an encryption recovery key, enter the following command:

   ```
   chencryption -recoverykey disable
   ```

   The recovery key has been disabled successfully. For more information, see **chencryption** command.
3. To verify whether recovery key has been disabled successfully, run **lsencryption** command and check that the `recovery_key_name` field is blank and the `recovery_key_status` is licensed. For more information, see **lsencryption** command.
4. To disable encryption on a system that uses key servers to manage encryption keys, enter the following command:

   ```
   chencryption -keyserver disable
   ```

   The encryption key that uses the key server has been disabled successfully. For more information, see **chencryption** command.
5. To verify whether encryption key that uses the key server has been disabled successfully, run **lsencryption** command and check that the `keyserver_pmk_uid` field is blank and the `keyserver_status` is licensed. For more information, see **lsencryption** command.
6. To disable encryption on a system that uses USB flash drives to manage encryption keys, enter the following command:

   ```
   chencryption -usb disable
   ```

   The encryption key that uses the USB flash drives has been disabled successfully. For more information, see **chencryption** command.
7. To verify whether encryption key that uses the USB flash drives has been disabled successfully, run **lsencryption** command and check that the `usb_key_filename` field is blank and the `status` is licensed. For more information, see **lsencryption** command.

Secure data deletion

The system provides methods to securely erase data from a drive or boot drive when a node or node canister is decommissioned.

Secure data deletion effectively erases or overwrites all traces of existing data from a data storage device. The original data on that device becomes inaccessible and cannot be reconstructed. You can securely delete data on individual drives and on a boot drive of a control enclosure. The methods and commands that are used to securely delete data enable the system to be used in compliance with European Regulation EU2019/424.

Deleting data from SAS drives

You can use the **chdrive -task erase** *drive_id* command to securely delete data from a SAS solid-state drive (SSD) or SAS hard disk drive (HDD) drive. The secure erase data functionality supports most devices. However, it is not supported on a few older drive models. When the command is used against the unsupported drive models, the `CMMVC6626E: The task was not initiated because a command was rejected by the drive that you have specified` error appears. Also, the

command fails if the drive is an active member of an array, the drive is in auto manage mode, or a firmware upgrade is in progress on the drive or the enclosure. If the command process did not complete or fails, you cannot access the data on the drive; you can only enter the **chdrive -task erase** command again.

To see the status of the data deletion task and the estimated completion time, you can use the **lsdriveprogress** command. During the data deletion process, the drive activity LED indicator flashes green. When the process completes, the drive activity LED indicator becomes solid green.

Note: Do not power off the enclosure or remove the drive until the data deletion process completes. Otherwise, you might not see any progress from the **lsdriveprogress** command and the drive remains offline. If the drive is offline, wait until the activity LED indicator is a solid green before you try to use or remove the drive again.

The below table compares the methods that the system uses to securely delete data from the drives. The method varies according to the commands that each type of drive can support. The completion time for the erase procedure also varies, depending on the amount of data and the method that is used to delete the data. In each case, when the operation completes, the result is that the data on the drive effectively becomes impossible to access.

Table 97. Comparison of methods to securely delete data from drives

Priority	Deletion type	Method	Completion time
1	Cryptographic erase	Changes the encryption key and makes the data inaccessible.	Instant
2	Block erase	Quickly raises and lowers the voltage level of the storage element. Physical blocks are altered with a vendor-specific value.	Fast
3	Data overwrite	Replaces the existing data with random data.	Slow

The methods that are used to securely delete data vary according to manufacture, drive type, and drive firmware. For more information, see the documentation that is provided by the drive manufacturer.

If a drive supports more than one data deletion method, the system uses the highest-priority method. Consider the following examples:

- If a SAS solid-state drive (SSD) supports all of the methods, cryptographic erase is used to delete the data.
- If a SAS SSD supports block erase and data overwrite, the system uses the block erase method to delete the data.
- If a SAS hard disk drive (HDD) supports cryptographic erase and data overwrite, the system uses cryptographic erase to delete the data.

For more information about the drives that are supported by the system, see the Control enclosure replaceable units page in the relevant Hardware Guide.

Deleting data from NVMe drives

You can use the **chdrive -task erase** command to delete data from an NVM Express (NVMe) drive. The drive is marked as `offline` and the deletion process starts. Use the **lsdriveprogress** command to monitor the status of the data deletion process and the estimated completion time.

You can also securely delete data from NVMe drives by using the **chdrive -task format** *drive_id* command. When you issue **chdrive -task format**, the system changes the internal key in the NVMe drive and then erases the user data. All previous data in the drive becomes unreadable.

Before an NVMe drive is added to an encryption pool, the drive is encrypted with a new key. Because that encryption key is removed during the formatting process, the data on the NVMe drive can no longer be read.

Deleting data from boot drives

You can also securely delete the data on a boot drive of a control enclosure. To do so, use the **satask rescuenode -secureerase -force** command. The system then uses the block erase method to delete most of the vital product data (VPD) from the boot drive. Only the serial number, FRU part number, and MTM information remain in the VPD.

Password policy

With password policy support, system administrators can set security requirements that are related to password creation and expiration, timeout for inactivity, and actions after failed logon attempts.

Password policy support allows administrators to set security rules that are based on their organization's security guidelines and restrictions. The system supports the following password and security-related rules with this support.

Password creation rules

Administrator can set and manage the following rules for all passwords that are created on the system:

- Specify password length requirements for all users.
- Require passwords to use uppercase and lowercase characters.
- Require passwords to contain special characters.
- Prevent users from reusing recent passwords. This parameter is not supported on FlashSystem 5015, FlashSystem 5035 and FlashSystem 5045.
- Require users to change password on next login under any of these conditions:
  - Their password expired.
  - An administrator created new accounts with temporary passwords.

Password expiration and rules for locking accounts

The administrator can create the following rules for password expiration:

- Set password expiration limit.
- Set a password to expire immediately.
- Set number of failed login attempts before the account is locked.
- Set time for locked accounts.
- Automatic log out for inactivity.
- Locking superuser account access.

  Note: The superuser account is the default user that can complete installation, initial configuration, and other service-related actions on the system. If the superuser account is locked, service tasks cannot be completed.

Two person integrity

Use two person integrity (TPI) to prohibit critical and risky tasks in the system from being executed by a single security administrator and by requiring the involvement of two security administrators.

TPI requires two security administrators to work together to complete certain tasks. Protecting data is an important part of IBM Storage Virtualize, and TPI helps mitigate the chance of data loss, prevent inadvertent mistakes on operations, and enhance security.

Requirements for enabling TPI:

- Ensure to have two users with the security administrator role.
- The two users can be local, remote, or a combination of both.
- If using remote users, a remote user group of security administrator role must be defined on the system and the remote authentication service must be enabled.

Requirements for disabling TPI:

After TPI is enabled, a user with an approved TPI request can disable TPI.

When you enable TPI, the users that belong to user groups of security administrator role are assigned the restricted security administrator role instead. However, their user groups retain their security administrator role.
Once TPI is enabled, a role elevation request and approval process is required to perform certain sensitive tasks:

- The restricted security administrator can issue a role elevation request on its own behalf to complete certain tasks in the system.
- Another restricted security administrator or a security administrator must approve the role elevation request.
- For example, this role elevation request and approval process is required to remove a Safeguarded snapshot.
- The restricted security administrators or security administrators can approve or deny role elevation requests, cancel role elevation requests, or revoke a role elevation request that was approved.

Available actions for a restricted security administrator that has an approved role elevation request.

- Create, change, or remove security administrator user groups.
- Create, change, or remove locally defined remote users.
- Change the non security administrator user group attribute on an existing local user to a security administrator user group.
- Modify attributes on existing local users that are members of the security administrator user groups.
- Change the role of existing non security administrator user groups to the security administrator role.
- Change the security administrator role of an existing user group to a non security administrator role.
- Remove and change Safeguarded backups and Safeguarded backup locations.
- Delete Safeguarded snapshots.
- Use a provisioning policy to define a set of rules that are applied when volumes are created within a storage pool or child pool.
- Change the single sign-on credentials that are used for the system.
- Remove the Safeguarded snapshot policy association from a volume group.

Configuring two person integrity (TPI)

Enable TPI when you want to prohibit critical and risky tasks in the system from being executed by a single security administrator and by requiring the involvement of two security administrators.

When TPI is enabled, the users that belong to user groups of security administrator role are assigned the restricted security administrator role. However, their user groups retain their security administrator role.

The restricted security administrator can submit a TPI request on its own behalf to complete certain tasks in the system. Another restricted security administrator or a security administrator must approve the TPI request. For example, this request and approval process is required to remove a Safeguarded snapshot.

Prerequisites

Before you enable TPI, ensure that you have at least two users in the system with the security administrator role. The users can be local, remote, or a combination of both.

Using the management GUI

To enable TPI, complete the following steps:

1. In the management GUI, select **Settings** > **Security** > **User Access** > **Two person integrity**.
2. Select **Enabled**.
3. Click **Save**.

4. To apply the TPI changes, click **Logout**.
5. After you sign in, the page displays that the current role is updated to restricted security administrator, and **Manage Role Elevation Requests** and **Request Elevated role** are displayed in the user menu list

Click **Snooze for 5 minutes** to stay on the page for 5 more minutes to review the settings.

Note: Snooze for 5 minutes is available only if user role is changed from security administrator to restricted security administrator.

To disable TPI, complete the following steps:

1. In the management GUI, select **Settings** > **Security** > **User Access** > **Two person integrity**.
2. Toggle the **Enabled** switch to the off position.
3. Click **Save**.

Using command-line interface

To enable TPI, enter the following command. Any security administrator can enable TPI.

```
chsecurity -twopersonintegrity yes
```

To disable TPI, enter the following command. After TPI is enabled, a user with an approved TPI request can disable TPI.

```
chsecurity -twopersonintegrity no
```

Managing two person integrity (TPI)

Use two person integrity (TPI) to prohibit critical and risky tasks in the system from being executed by a single security administrator and by requiring the involvement of two security administrators.

TPI requires two security administrators to work together to complete certain tasks. Protecting data is an important part of IBM Storage Virtualize, and TPI helps mitigate the chance of data loss, prevent inadvertent mistakes on operations, and enhance security.

When you enable TPI, the users that belong to user groups of security administrator role are assigned the restricted security administrator role instead. However, their user groups retain their security administrator role.
Once TPI is enabled, a role elevation request and approval process is required to perform certain sensitive tasks:

- The restricted security administrator can issue a role elevation request on its own behalf to complete certain tasks in the system.
- Another restricted security administrator or a security administrator must approve the role elevation request.
- For example, this role elevation request and approval process is required to remove a Safeguarded snapshot.
- The restricted security administrators or security administrators can approve or deny role elevation requests, cancel role elevation requests, or revoke a role elevation request that was approved.

Viewing elevation requests

To access the Manage Role Elevation Requests page, select **Access** > **Manage Role Elevation Requests**.

- You must have the security administrator role or restricted security administrator role to access the page.
- After your request is approved, you must log out of the management GUI. Then, log in again for the setting to take effect.
- The maximum number of active role elevation requests is four.

- You can't approve your own requests.
- Approved requests are displayed until their duration expires.
- Denied, revoked, and canceled requests are automatically removed from the page.
- If no requests are pending, information about how to request an elevated role is displayed.

Viewing the audit Log

To view a list of completed actions, select **Access** > **Audit Log**. If TPI is enabled, the Elevated Role column is displayed; if TPI is disabled, the Elevated Role column is hidden.

Chapter 19. System

A system consists of one or more I/O groups. Each I/O group contains two nodes. In enclosure-based systems, a control enclosure is an I/O group.

I/O group

The pair of nodes within a single enclosure is known as an *input/output (I/O) group*.

Note: The current version supports a single I/O group in all systems. FlashSystem grid can be used to scale out a storage system.

When a write operation is performed to a volume, the node that processes the I/O duplicates the data onto the partner node that is in the I/O group. After the data is protected on the partner node, the write operation to the host application is completed. The data is physically written to the disk later. *Volumes* are logical disks that are presented to the system by nodes. Volumes are also associated with the I/O group.

When an application server processes I/O to a volume, it can access the volume with either of the nodes in the I/O group. When you create a volume, you can specify a preferred node. Many of the multipathing driver implementations that the system supports use this information to direct I/O to the preferred node. The other node in the I/O group is used only if the preferred node is not accessible.

If you do not specify a preferred node for a volume, the system selects the node in the I/O group that has the fewest volumes to be the preferred node. After the preferred node is chosen, it can be changed only when the volume is moved to a different I/O group.To view the current preferred node for a volume, select **Volumes** > **All Volumes** in the management GUI. Right-click the volume and select **Properties**.

To access information about I/O groups in the management GUI, select **Settings** > **System** > **I/O Groups**. In the **I/O Groups** page, you can view the configured I/O groups on the system. The **I/O Groups** page displays all the hardware that is assigned to the I/O groups that are configured on the system. Use the directional arrow to expand details on the I/O groups or its related hardware.

Read I/O is processed by referencing the cache in the node that receives the I/O. If the data is not found, it is read from the disk into the cache. The read cache can provide better performance if the same node is chosen to service I/O for a particular volume.

I/O traffic for a particular volume is, at any one time, managed exclusively by the nodes in a single I/O group. Although a system can have multiple nodes within it, the nodes manage I/O in independent pairs.

Figure 47 on page 486 shows a write operation from a host (1) that is targeted for volume A. This write is targeted at the preferred node, Node 1 (2). The write operation is cached and a copy of the data is made in the partner node, the cache for Node 2 (3). The host views the write as complete. Later, the data is written, or de-staged, to storage (4).

© Copyright IBM Corp. 2024

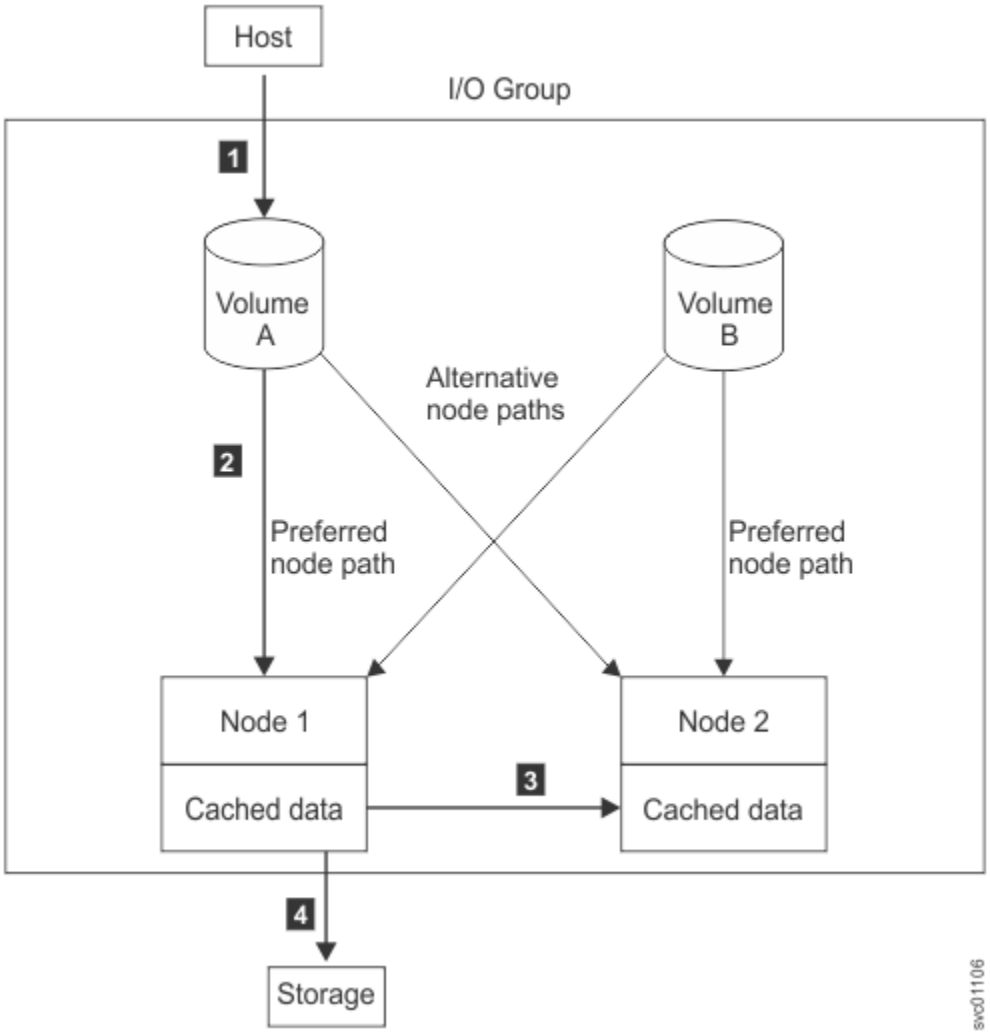


Figure 47. Example of a write operation in an I/O group

When a node fails within an I/O group, the other node in the I/O group assumes the I/O responsibilities of the failed node. Data loss during a node failure is prevented by mirroring the I/O read and write data cache between the two nodes in an I/O group.

If only one node is assigned to an I/O group or if a node fails in an I/O group, the cache is flushed to the disk and then goes into write-through mode. Therefore, any writes for the volumes that are assigned to this I/O group are not cached; they are sent directly to the storage device. If both nodes in an I/O group go offline, the volumes that are assigned to the I/O group cannot be accessed.

When a volume is created, the I/O group to provide access to the volume must be specified. However, volumes can be created and added to I/O groups that contain offline nodes. I/O access is not possible until at least one of the nodes in the I/O group is online.

Note: You cannot create arrays that span I/O groups.

Nodes

A node is a single processing unit within a system. For redundancy, nodes are deployed in pairs.

Each pair of nodes is known as an I/O group. Each node can be in only one I/O group.

In enclosure-based systems, a node canister is equivalent to a node. The two node canisters in a control enclosure make an I/O group. Within a control enclosure, nodes communicate with each other via the enclosure hardware. For more details, see **Expansion enclosures** topic in the relevant hardware guide.

Nodes use Fibre Channel to communicate with hosts, external storage, other nodes (outside the enclosure) and Ethernet ports to communicate with hosts and external storage or other systems. Each node has a service IP address. The system also has one or two management IP addresses.

If Ethernet ports are used for I/O, nodes are configured with additional IP addresses.

Configuration node

At any one time, a single node in the system provides a focal point for configuration commands. If the configuration node fails, another node in the system takes over its responsibilities.

The configuration node also handles the following functions:

- Accepts user logins to the management GUI and CLI via the IP address that is bound to Ethernet port 1 and optionally to the second management IP address that is bound to Ethernet port 2.
- Sends system configuration changes to the other nodes in the system.
- Sends call home and system notifications on behalf of the system.

If the configuration node fails or is taken offline, the system chooses a new configuration node. This action is called configuration node failover. The new configuration node takes over the management IP addresses. Thus, you can access the system through the same IP addresses although the original configuration node failed. During the failover, there is a short period when you cannot use the command-line tools or management GUI.

For correct failover operation, all nodes of the system must be connected to the same subnet and the management IP configuration must be valid on the subnet.

Ethernet Link Failures

If the Ethernet link to the system fails because of an event that is unrelated to the system, the system does not attempt to fail over the configuration node to restore management IP access. For example, the Ethernet link can fail if a cable is disconnected or an Ethernet router fails. To protect against this type of failure, the system provides the option for two Ethernet ports that each have a management IP address. If you cannot connect through one IP address, attempt to access the system through the alternative IP address.

Identify the configuration node using the GUI

To identify the configuration node, follow these steps:

1. In a web browser, navigate to the Service Assistance Tool (SAT) of any node in the system, by entering the following address:

   ```
   https://<service_ip>/service
   ```

   Where `service_ip` is the service IP address of the chosen node.
2. Log into the SAT by authenticating using the superuser credentials.
3. Once logged in, the dashboard displays a list of all the nodes in the system. The configuration node is labelled as **CONFIG**.
4. Click the **LED** toggle to turn on the identification LED on the node canister.
5. Check for the node canister with the identification LED lit and perform the required service action.
6. When finished, click the LED toggle to turn off the identification LED on the node canister.

Identifying the configuration node using the command-line interface

To identify the configuration node, follow these steps:

1. In a terminal window, use Secure Shell (SSH) software to connect to the service IP address of any node in the system and authenticate using the superuser credentials:

```
ssh superuser@service_ip
```

Where `service_ip` is the service IP address of the chosen node.

2. Once logged in, run **`sainfo lsservicenodes`** command to display a list of all the nodes that can be serviced by using the service assistant CLI. From the list of nodes, check the `panel_name` field to determine the list of all nodes in the same system as the node that is running the command.
3. For each of the nodes above, run the following command to display the service status of the node:

```
sainfo lsservicestatus
```

The configuration node is the one which reports a `config_node` value of Yes.

4. Turn on the identification LED for the configuration node canister by running the following command:

```
satask chnodeled -on panel_name
```

Where `panel_name` is the panel name of the configuration node.

5. Check for the node canister with the identification LED lit and perform the required service action.
6. When finished, turn off the identification LED for the configuration node canister by running the following command:

```
satask chnodeled -off panel_name
```

Where `panel_name` is the panel name of the configuration node.

Node status

The *node status* identifies the status of the node within the IBM Storage Virtualize system.

Table 98 on page 488 describes the operational status of a node or node canister. To display the node status on the GUI, select **Monitoring** > **System Hardware**. You can also see the node status by executing the **`lsnode`** or **`lsnodecanister`** command or using the service assistant GUI.

Table 98. Node status

Status	Description
Active	The node is online and can process I/O from hosts.
Adding	The node or node canister was added to the clustered system but is not yet synchronized with the system status. The node or node canister status changes to Online after synchronization is complete. **Note:** A node can stay in the Adding status for a long time. Wait at least 30 minutes before you take further action. If after 30 minutes the node status is still Adding, you can delete the node and add it again. If the node that was added is at a lower code level than the rest of the system, the node is upgraded to the system code level, which can take up to 20 minutes. During this time, the node is shown as Adding.
Candidate	The node is not defined in an I/O group; however, it can be added as one of the node pairs in an I/O group.
Deleting	The node or node canister is in the process of being deleted from the system.
Flushing	The system is removing data from the cache for the node.
Online	The node canister is operational and is assigned to a cluster.
Online spare	The node has been defined to be a hot-spare node and it is being used as a surrogate for another node in the I/O group.
Offline	The node canister is not operational. Run the fix procedures to determine the problem.

Table 98. Node status (continued)	
Status	**Description**
Pending	The status of the node or node canister is in transition. In a few seconds, the status of the node or node canister will change.
Service	The node or node canister has been placed in this state to allow it to be serviced or it has a fatal node error, which means that it is not safe for this node to be part of the system.
Spare	The node has been defined to be a hot-spare node; however, it is not being used by the system. The node has a valid cluster ID and a valid node ID.
Starting	The node has been restarted. It is attempting to join the system cluster however, it cannot yet perform I/O.

Hot-spare node

When you add nodes, you can specify up to four of them as *hot-spare nodes*. A hot-spare node can become online (handling I/O operations) automatically if needed.

For example, if a node fails, an available hot-spare node that matches the failed node is activated automatically and moves to the **Online Spare** state. The hot-spare node handles I/O operations for the failed node until it comes back online. After the node returns to the system, the hot-spare node returns to the **Spare** state, which indicates it can be automatically swapped for other failed nodes on the system.

The loss of a node, either for unplanned reasons, such as hardware failure, or planned outages, such as upgrades, can result in loss of redundancy or degraded system performance. To reduce this possibility, a hot-spare node is kept powered on and visible on the system. A hot-spare node has active system ports, but no host I/O ports, and is not part of any I/O group. If a node fails or is upgraded, this spare node joins the system and assumes the place of the failed node, restoring redundancy. Only host connection on Fibre Channel ports that support node port virtualization (NPIV) can be used for hot-spare nodes. The hot-spare node uses the same N_Port ID Virtualization (NPIV) worldwide port names (WWPNs) for its Fibre Channel ports as the failed node, so host operations are not disrupted. The hot-spare node retains its node identifier when it was the spare. During an upgrade, the spare node is added to the system when a node is removed. The hot-spare node replaces each node that shuts down for upgrade in a system.

You can assign node pairs to specific I/O groups and then assign the extra nodes as hot-spare nodes. When a hot-spare node is added to the system, it is in **Spare** state, which indicates that it is not part of an I/O group. If a node in an I/O group fails, a hot-spare node automatically replaces that node and becomes a part of the I/O group. While the hot-spare node is in the I/O group, it is in the **Online Spare** state and returns to the **Spare** state when the original node rejoins the I/O group. A system can contain up to four spares at any time, which includes any hot-spare nodes that are online as spare nodes. Ensure that all cabling is correct to ensure that the system detects the nodes. If a node is not detected, review the installation information that was included with the system.

The following restrictions apply to using hot-spare node on the system:

-
- Hot-spare nodes can be used with Fibre Channel-attached external storage only.
- Hot-spare nodes cannot be used with SAS-attached storage.
- Hot-spare nodes cannot be used with iSCSI-attached storage.
- Hot-spare nodes cannot be used with storage that is directly attached to the system.

Criteria that requires an exact match

- Memory capacity
- Fibre Channel port ID
- Compression support
- Site

Criteria that is recommended to match, but can be different

- Hardware type
- CPU count
- Number of Fibre Channel ports

If the nodes on the system support and are licensed to use encryption, the hot-spare node must also support and be licensed to use encryption.

If an adapter PCI slot location on the spare node does not match the active nodes, an active node cannot be replaced by a spare node by using the command. If the user encounters the error , it means that the command failed because the specified node does not have a status of *candidate*. It is recommended to have adapters in the same slot for spare nodes and the active node for replace command to work.

When the online spare node is put into the Service state, it is immediately removed back to spare and 5 minutes later rejoin as online spare, in the cluster. Instead of putting the online spare into service, wait for the original to come back or remove the online spare and then perform their maintenance.

Adding hot-spare nodes by using the command-line interface

To add a spare node to the system, enter the following command, where *panel_name* is the name of the node that is displayed in the service assistant or in the output of command.

```
addnode -panelname panel_name -spare
```

For enhanced stretched configurations, use the following command to specify the name and site location of the spare node.

```
addnode -panelname panel_name -spare -site site_id
```

Adding a node or enclosure to increase the size of the system

You can add a node or enclosure to the system by using the CLI or management GUI.

An enclosure can be added to the system if the enclosure previously failed and is being replaced with a new enclosure or if a repair action causes the enclosure to be unrecognizable by the system. Adding an enclosure to the system increases the capacity of the entire system. When you add an enclosure to a system, check that the licensed functions of the system support the additional enclosure. For more information, see “Licensed functions” on page 496.

You can use either the management GUI or the command-line interface to add a node to the system. Some models might require you to use the front panel to verify that the new node was added correctly.

Before you add a node to a system, you must make sure that the switch zoning is configured such that the node that is being added is in the same zone as all other nodes in the system. If you are replacing a node and the switch is zoned by worldwide port name (WWPN) rather than by switch port, make sure that the switch is configured such that the node that is being added is in the same VSAN or zone.

Note: Use a consistent method (either only the management GUI, or only the CLI) when you add, remove, and re-add nodes. If a node is added by using the CLI and later re-added by using the GUI, it might get a different node name than it originally had.

You can use either the management GUI or the command-line interface to add an enclosure to the system. If you are adding an enclosure that was previously removed from a system, ensure that these conditions are met:

- All hosts that accessed the removed enclosure through its WWPNs are reconfigured to use the WWPN for the new enclosure or to no longer access the enclosure. Failure to do so can result in data corruption.
- Ensure that the system ID is reset on the new control enclosure. You can use either the command-line interface or the Service Assistant to verify the system ID on the new control enclosure.

To use the command-line interface, enter the following command:

```
chvpd -resetclusterid
```

To use the Service Assistant, complete the following steps:

- Connect to the service assistant on either of the nodes in the control enclosure.
- Select **Configure Enclosure**.
- Select the **Reset the system ID** option. Do not make any other changes on the panel.
- Click **Modify** to make the changes.

You can use either the **addcontrolenclosure** command or the **Add Enclosure** wizard in the management GUI. To access the **Add Enclosure** wizard, select **Monitoring** > **System**. On the **System -- Overview** page, select **Add Enclosure** to start the wizard. Complete the wizard and verify the new enclosure. If **Add Enclosure** is not displayed, it indicates a potential cabling issue. Check the installation information to ensure that the enclosure was cabled correctly.

Complete these steps to add an enclosure to the system by using the command-line interface:

1. Using the **sainfo lsservicestatus** command (on the service CLI of the new enclosure), record the WWNN of the new enclosure.
2. Record the serial number of the enclosure, which is needed in later steps.
3. Enter the following command to verify that the enclosure is detected on the fabric:

   ```
   # svcinfo lscontrolenclosurecandidate
   serial_number product_MTM
   66G00C2       2076-624
   ```

4. Enter this command to determine the I/O group where the enclosure must be added:

   ```
   lsiogrp
   ```

5. Record the name or ID of the first I/O group that has a node count of zero. You will need the ID for the next step.
6. Enter this command to add the enclosure to the system:

   ```
   addcontrolenclosure  -iogrp iogrp_name | iogrp_id -sernum enclosureserialnumber
   ```

   where *iogrp_name* | *iogrp_id* is the name or ID of the I/O group and *enclosureserialnumber* is the serial number of the enclosure.

   Note: This step might take a considerable amount of time.
7. Record this information for future reference:
   - Serial number.
   - Worldwide node name of both node canisters.
   - All of the worldwide port names.
   - The name or ID of the I/O group that contains the enclosure.
8. Enter this command to verify that the node canisters in the enclosure are online:

   ```
   lsnodecanister
   ```

Quorum

A quorum device is used to break a tie when a SAN fault occurs, when exactly half of the nodes that were previously a member of the system are present. A quorum device is also used to store a backup copy of important system configuration data.

Just over 256 MB is reserved to store backup copy of important system configuration data on each quorum device.

It is possible for a system to split into two groups where each group contains half the original number of nodes in the system. A quorum device determines which group of nodes stops operating and processing I/O requests. In this tie-break situation, the first group of nodes that accesses the quorum device is marked as the owner of the quorum device and as a result continues to operate as the system, handling all I/O requests. If the other group of nodes cannot access the quorum device or finds that the quorum device is owned by another group of nodes, it stops operating as the system and does not handle I/O requests.

A system can have only one active quorum device that is used for a tie-break situation. However, the system uses up to three quorum devices to record a backup of system configuration data to be used in the event of a disaster. The system automatically selects one quorum device to be the active quorum device.

The other quorum devices provide redundancy if the active quorum device fails before a system is partitioned. To avoid the possibility of losing all the quorum devices with a single failure, assign quorum disk candidates on multiple storage systems or run IP quorum applications on multiple servers.

Single site configurations

The normal configuration is to use a managed drive or an MDisk as the quorum device when the system is not configured as a stretched system. A system automatically assigns quorum disk candidates. When you add new storage to a system or remove existing storage, however, it is a good practice to review the quorum disk assignments. Optionally an IP quorum device can be configured either as an alternative to using quorum disks or to provide additional redundancy.

Stretched configurations

To provide protection against failures that affect an entire location, such as a power failure, you can use a configuration that splits a single system across three physical locations.

A stretched system has system nodes divided between two sites. If a SAN fault causes loss of connectivity between sites or a fault causes a site wide outage then the quorum configuration determines which site continues operating and processing I/O requests. A high availability solution has the active quorum device configured at a third site so that the system will continue to operate after any single-site failure.

Generally, when the nodes in a system are split among sites, configure the system this way:

- Site 1: Half of system nodes + one quorum device
- Site 2: Half of system nodes + one quorum device
- Site 3: Active quorum device

Typically the quorum devices at site 1 and site 2 are quorum disks and the quorum device at site 3 is an IP quorum application. However, the system can be configured to use either quorum disks or IP quorum applications at any site. This configuration ensures that a quorum device is always available, even after a single-site failure.

When you are using an IP quorum application at a third site, you can configure a preference for which site continues operation if there is a loss of connectivity between the two sites. If only one site runs critical applications, you can configure this site as preferred. If a preferred site is configured and a failure causes an outage at the preferred site, the other site wins the tie-break and continues operating and processing I/O requests.

A stretched system can be configured without a quorum device at a third site. If there is no third site, then quorum must be configured to select a site to always win a tie-break. If there is a loss of connectivity between the sites, then the site that is configured as the winner continues operating and processing I/O requests and the other site stops until the fault is fixed. If there is a site outage at the wining site, then the system stops processing I/O requests until this site is recovered or the manual quorum override procedure is used.

Generally, when the nodes in a system are split between two sites and there is no third site quorum, configure the system this way:

- Site 1: Half of system nodes + one or two quorum devices
- Site 2: Half of system nodes + one quorum device

Typically, the quorum devices at site 1 and site 2 are both quorum disks and are automatically configured by the system. It is possible to configure IP quorum applications as an alternative to using quorum disks. When a winner site has been configured and both sites are operational, there is no active quorum device. The quorum devices at site 1 and site 2 are only used to retain a backup copy of important system configuration data. If a failure results in just the nodes at the winner site continuing operation, then the system automatically selects one of the quorum devices at that site to be the active quorum device to protect against further failures.

IP quorum application

The IP quorum application is a Java™ application that runs on a server that is separate from the storage system.

There are two types of IP quorum application:

- A **system** IP quorum application is only used for single-system tie-break if half of the nodes are unavailable, and optionally for metadata storage
- A **partnership** IP quorum application is only used to arbitrate between systems in a policy-based high availability partnership. The IP quorum application can be generated on either system in the partnership. The partnership must be created before generating and deploying the IP quorum application.

The maximum number of IP quorum applications that can be deployed on a single system is five. This enables multiple servers to be used to provide redundancy. Only one instance of the IP quorum application per server, per system, is supported. For example, a server can run two IP quorum instances if each instance is connected to a different nonpartnered storage system. Ensure that bandwidth is available to support multiple IP quorum instances.

Do not deploy the IP quorum application on a server that depends on storage that is presented by the system. This action can result in a situation where the nodes need to detect the IP quorum application to process I/O, but cannot because the IP quorum application cannot access storage.

An Ethernet connectivity issue can prevent an IP quorum application from accessing a node that is still online and an event is raised on the system if this occurs.

Refer to the following support article for the IP quorum application requirements: https://www.ibm.com/support/pages/node/7013877

Warning: For the IP quorum application to be able to connect using TLS 1.3 (SSL Security Levels 6 and 7), the version of Java running the application must also support TLS 1.3.

Configuring IP Quorum on the storage system

You can configure IP quorum on the storage system by using the management GUI or CLI:

For partnership IP quorum applications:

Using the management GUI
In the management GUI, the Policy-based replication setup procedure includes generating the IP quorum application.

For system IP quorum applications:

Using the management GUI
In the management GUI, select **Settings** > **System** > **IP Quorum**.

Deploying the IP Quorum Application on the server

Follow these steps to deploy the IP quorum application on the server:

1. Create a separate directory that is dedicated to the IP quorum application.

2. Transfer the IP quorum application from the storage system to the dedicated directory.
3. Use the **ping** command on the server to verify that it can establish a connection with the service IP address of each node in both systems.
4. Enter the command **java -jar ip_quorum.jar** to initialize the IP quorum application.

   Note: The IP quorum application needs to be running always.

The options available when running the IP quorum application can be listed with **-help**:

```
1. $ java -jar ip_quorum.jar -help
2. -name (optional) to help identify the IP quorum app instance in the GUI/lsquorum.  Can only
contain 1-20 characters that are A-Z, a-z or 0-9.
3. -debug (optional) run the app in debug mode to display verbose messages on stdout and in the
log file.
4. -emit (optional) display T3 metadata header information
5. -location (optional) default = ip_quorum.log.16845080043810. Specify the full, or relative
path to store your log files. Allowed characters: [a-z A-Z 0-9 .-_/]
6. -rotation (optional) default = 5. Specify the maximum number of log files. Allowed values:
[1-10]
7. -size (optional) default = 5120. Specify the log file size in kb. Allowed values:
[1024-10240]
8. -version / -v (optional) display the vrmf of the system when the app was generated,
timestamp and generation id.
```

Reviewing and monitoring the IP Quorum Application

You can monitor and review the IP quorum application from each system by using the management GUI or CLI:

Using the management GUI

In the management GUI, select **Settings** > **System** > **IP Quorum**.

System topology

The system topology can be set up in following ways.

- *Standard* topology, where all nodes in the system are at the same site.

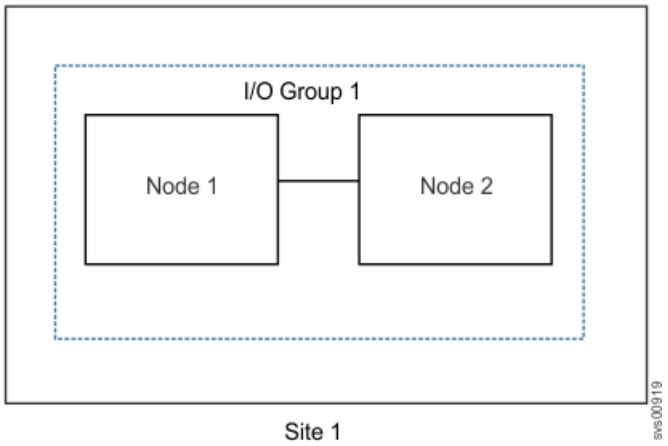


Figure 48. Example of a standard system topology

- *Stretched* topology, where each node of an I/O group is at a different site. When one site is not available, access to a volume can continue but with reduced performance. This topology is only available on SAN Volume Controller.

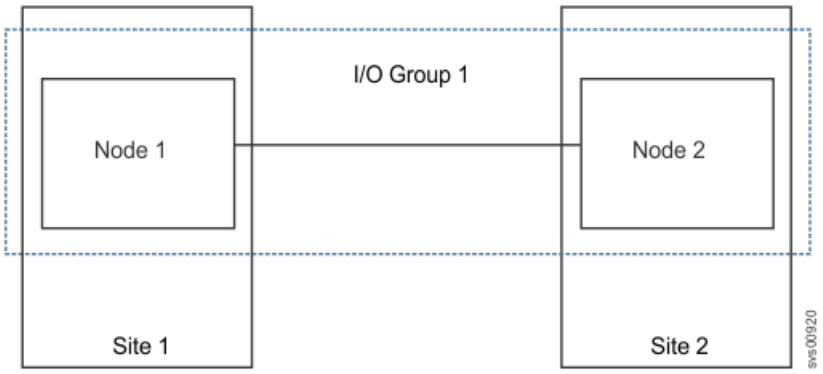


Figure 49. Example of a stretched system topology

Summary of system topology and volumes

Table 99 on page 495 summarizes the types of volumes that can be associated with each system topology.

Table 99. System topology and volume summary

Topology	Volume Type		
	Basic	**Stretched**	**Custom**
Standard	X		X
Stretched	X	X	X

System layers

The *system layer* affects how the system interacts with another system and other external systems that run IBM Storage Virtualize software.

Systems that run IBM Storage Virtualize software can either be in the *replication layer* or the *storage layer*.

Differences between the storage layer and the replication layer

This table compares the characteristics and requirements of storage layer and replication layer systems.

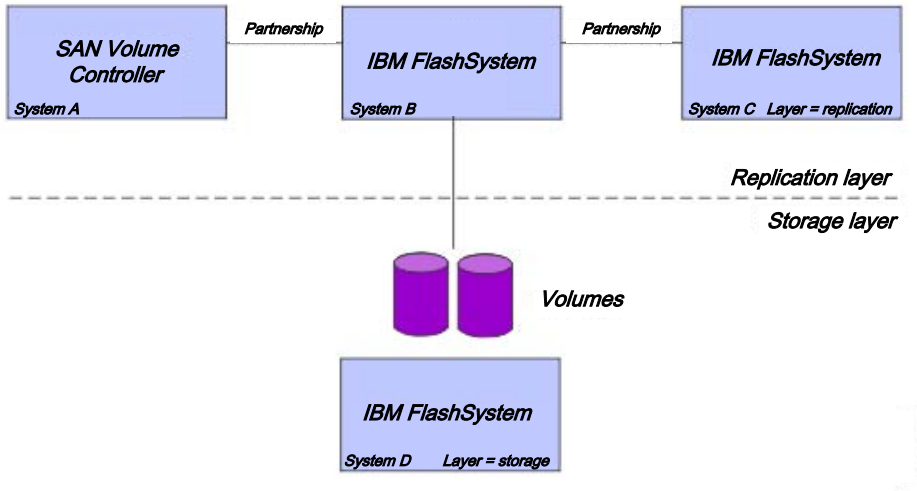
Table 100. Comparison of storage layer and replication layer systems

Storage layer	Replication layer
• The system can complete replicate with other storage-layer systems. • The system can provide external storage for replication-layer systems. • The system cannot use another Storwize family or FlashSystem family system that is configured with storage-layer as external storage.	• The system can complete replicate with other replication-layer systems. • The system cannot provide external storage for other IBM Storage Virtualize systems. • The system can use another Storwize family or FlashSystem family system that is configured with storage-layer as external storage.

A FlashSystem family system is in the replication layer by default, but the layer can be changed. For example, you might want to change a system to a replication layer to replicate with another system in the storage layer. Systems in a FlashSystem grid must be in the replication layer.

To create a partnership between systems, both systems must be in the same layer.

Figure 50. Example configuration with systems in different layers



Changing the layer of a FlashSystem family system

To change the storage layer of an existing system with internal storage before you add a second system into the SAN zone, you do not need to stop I/O operations.

However, if the system has Fibre Channel (FC) connections to another FlashSystem family, or system in the SAN fabric, I/O operations must be stopped temporarily. In this scenario, the FC ports must be disabled (for example, by unplugging all the FC ports, changing zoning, disabling switch ports) before you change the system layer. Then, you must re-enable the FC ports.

Before you change the system layer, the following conditions must be met:

- No host object can be configured with worldwide port names (WWPNs) from other IBM Storage Virtualize systems.
- No system partnerships can be defined.
- No IBM Storage Virtualize systems can be visible on the SAN fabric.

To change the layer of a system, follow these general steps:

1. Use the **lssystem** command to check the current system layer.
2. Configure SAN zoning so that the system whose layer you are changing is not visible to any other system.
3. Ensure that all other configuration requirements are met.
4. Run the **detectmdisk** command to ensure that the system detects any fabric changes that were done in step "3" on page 496.
5. Enter the one of the following commands to switch the system to the appropriate layer.
   - To switch to the replication layer, enter the **chsystem -layer replication** command.
   - To switch to the storage layer, enter the **chsystem -layer storage** command.
6. If the SAN configuration was changed in step "2" on page 496, restore the normal SAN configuration.
7. Run the **detectmdisk** command to update the list of MDisks.

Licensed functions

The base license that is provided with your system includes the use of its basic functions. However, there are also extra licenses that can be purchased to expand the capabilities of your system. Administrators are responsible for purchasing extra licenses and configuring the systems within the license agreement, which includes configuring the settings of each licensed function on the system.

The system supports differential, capacity, and key-based licensing. For example, with external virtualization differential licensing charges different rates for different types of storage, which provides cost-effective management of capacity across multiple tiers of storage. Licensing for these functions is based on the number of storage capacity units (SCUs) purchased. With other functions, like remote mirroring and FlashCopy, the license grants a specific number of terabytes of capacity for that function. Key-based licenses require an authorization code and key to be downloaded to the system before the function can be used.

Differential licensing is the term that is used to refer to the measurement of external virtualization and compression licenses, where different types of storage are charged at different rates. Differential licensing uses a storage capacity unit (SCU)-based license scheme. For each SCU, a client is entitled to manage different amounts of usable capacity based on a storage category classification. Table 101 on page 497 identifies the storage categories and SCU ratios for each category.

Table 101. Supported SCU licenses, drive classes, and ratios

License	Drive Classes	SCU Ratio
SCM	Storage Class Memory (SCM) devices	1 SCU equates to 1.00 TB usable of Category 1 storage
Flash	All flash devices, other than SCM drives	1 SCU equates to 1.18 TB usable of Category 1 storage
Enterprise	10K or 15K RPM drives	1 SCU equates to 2.00 TB usable of Category 2 storage
Nearline	Nearline Serial ATA (SATA) drives	1 SCU equates to 4.00 TB usable of Category 3 storage

Table 102 on page 497 provides an example of how SCU licensing might work based on category types.

Table 102. Example of SCU usage

Storage category	TB usable	SCU ratio	Total SCUs
1	42	1.18	35
2	400	2.00	200
3	800	4	200
Totals	1242		435

The system supports the following licensed functions:

External Virtualization

Enter the total number of storage capacity units (SCUs) you are licensed to virtualize across tiers of storage on your system. The system supports virtualization licensing for different tiers of storage. A license can be purchased for a specific quantity of storage capacity units (SCU) which can be used to virtualize a designated number of tebibytes (TiB) of storage. The used capacity for each tier can be monitored to learn how the virtualization license is distributed across the following tiers of storage:

- Storage class memory (SCM)
- Tier 0 Flash
- Tier 1 Flash
- Enterprise
- Nearline

FlashCopy

A FlashCopy base license provides you with an amount of capacities equal to the internal storage. Beyond this capacity, FlashCopy requires additional licensed capacity, which is charged per terabyte. The FlashCopy function copies the contents of a source volume to a target volume. The FlashCopy

function is also used to create cloud snapshots of volumes in systems that have transparent cloud tiering enabled. The used capacity for FlashCopy mappings is the sum of all of the volumes that are the source volumes of a FlashCopy mapping and volumes with cloud snapshots. The Used TiB value indicates the amount of TiB that this system has used of the total amount of licensed TiBs for all FlashCopy operations on the system.

Note: For the 4983-AH8 model, FlashCopy license is applicable to external storage only and is charged per terabyte. FlashCopy operations that are used on internal storage are part of the base license.

Encryption
In addition to these licensed functions, you can also purchase a separate license to use software-based encryption. Software-based encryption protects data at rest and protects other keys that are used on self-encrypting drives. For each enclosure in the system, you must have an encryption license.

Bitmap space configuration

As part of Copy Services, the system supports policy-based replication. If you are currently using relationships to replicate data between systems, you may need to reallocate bitmap space to use policy-based replication.

Before you begin

The total I/O group memory that can be dedicated to these functions is not defined by the physical memory in the system. The I/O group memory is constrained by the software functions that use the memory.

To use policy-based replication on the system requires a set amount of allocated I/O group memory.

If you plan to use policy-based replication with other functions, use the following table to determine the required allocated memory for other features:

Table 103. Product comparison of required I/O group memory for policy-based replication

Product	Total I/O group memory (excluding FlashCopy functions)	Required I/O group memory for policy-based replication	Available memory for other features (RAID and volume mirroring)
IBM Storage FlashSystem 9500	5920 MiB	4609 MiB	1311 MiB
IBM Storage FlashSystem 9200	3628 MiB	2561 MiB	1067 MiB
IBM Storage FlashSystem 7300	3628 MiB	2561 MiB	1067 MiB
IBM Storage FlashSystem 7200	3628 MiB	2561 MiB	1067 MiB
IBM SAN Volume Controller	3628 MiB	2561 MiB	1067 MiB
IBM Storage FlashSystem 5045 (requires a minimum of 32 GiB memory in each node canister)	896 MiB	713 MiB	183 MiB
IBM Storage FlashSystem 5200	2088 MiB	1537 MiB	551 MiB

Table 103. Product comparison of required I/O group memory for policy-based replication (continued)			
Product	**Total I/O group memory (excluding FlashCopy functions)**	**Required I/O group memory for policy-based replication**	**Available memory for other features (RAID and volume mirroring)**
IBM Storage Virtualize for Public Cloud	2088 MiB	1537 MiB	551 MiB

For example, an IBM Storage FlashSystem 9500 system has a maximum of 5920 MiB memory that is used by replication, volume mirroring, RAID, and policy-based replication. If policy-based replication is enabled, it uses 4609 MiB, the remaining 1311 MiB can be allocated between replication, volume mirroring, and RAID with maximum limits as 1024, 512 and 800 MiB each in order. If policy-based replication is disabled, a maximum of 1824 MiB can be allocated between replication, volume mirroring, and RAID with maximum limits as 1024, 512 and 800 MiB each in order.

In planning the installation for a system, consider the future requirements for the advanced functions.

This table describes the configuration of the bitmap space in a system.

Table 104. Bitmap space default configuration				
Copy Service	**Minimum allocated bitmap space**	**Default allocated bitmap space**	**Maximum allocated bitmap space**	**Minimum[1] functionality when using the default values**
FlashCopy[3]	0	20 MiB	20 GiB	40 TiB of FlashCopy target provisioned capacity 5 TiB of incremental FlashCopy target provisioned capacity
Volume mirroring	0	20 MiB	512 MiB	40 TiB of mirrored volumes, or 40 TiB of non-mirrored formatting standard-provisioned volumes, or 20 TiB of mirrored formatting standard-provisioned volumes
RAID	0	40 MiB	800 MiB	80 TiB writable capacity using RAID 0, 1, or 10. Slightly less than 120 TiB writable capacity in five-disk RAID 6 array.

Table 104. Bitmap space default configuration (continued)				
Copy Service	**Minimum allocated bitmap space**	**Default allocated bitmap space**	**Maximum allocated bitmap space**	**Minimum[1] functionality when using the default values**
[1] The sum of all bitmap memory allocation for all functions except FlashCopy must not exceed 1824 MiB. The actual amount of functions might increase based on settings such as grain size and strip size. RAID is subject to a 15% margin or error. [3] FlashCopy includes the FlashCopy function configured with change volumes, active-active relationships. Volume Group Snapshots can use up to 20GiB of the FlashCopy bitmap memory, all other features using FlashCopy can use up to 4GiB of the 20GiB.				

Note: For IBM Storage FlashSystem 9500, snapshots are supported by the total of 20 GiB bitmap space. FlashCopy functions other than snapshots are limited to a maximum of 4 GiB of bitmap space.

The following tables describe the amount of bitmap space necessary to configure the various Copy Services functions and RAID:

This table provides an example of the amount of memory that is required for remote mirroring functions, FlashCopy functions, and volume mirroring.

Table 105. Examples of memory required

Function	**Grain size**	**1 MiB of memory provides the following provisioned capacity for the specified I/O group**
FlashCopy	256 KiB	2 TiB of total FlashCopy source provisioned capacity
FlashCopy	64 KiB	512 GiB of total FlashCopy source provisioned capacity
Incremental FlashCopy	256 KiB	1 TiB of total incremental FlashCopy source provisioned capacity
Incremental FlashCopy	64 KiB	256 GiB of total incremental FlashCopy source provisioned capacity
Volume mirroring or format	256 KiB	2 TiB of mirrored provisioned capacity

Table 105. Examples of memory required (continued)		
Function	**Grain size**	**1 MiB of memory provides the following provisioned capacity for the specified I/O group**
Notes: • For multiple FlashCopy targets, you must consider the number of mappings. For example, for a mapping with a grain size of 256 KiB, 8 KiB of memory allows one mapping between a 16 GiB source volume and a 16 GiB target volume. Alternatively, for a mapping with a 256 KiB grain size, 8 KiB of memory allows two mappings between one 8 GiB source volume and two 8 GiB target volumes. • When creating a FlashCopy mapping, if you specify an I/O group other than the I/O group of the source volume, the memory accounting goes toward the specified I/O group, not toward the I/O group of the source volume. • For volume mirroring, the 512 MiB of memory space enables 1 PiB of total provisioned capacity. • When creating new FlashCopy relationships or mirrored volumes, additional bitmap space is allocated automatically by the system if required. • Provisioned capacity and formatting of standard-provisioned volumes both use the mirroring bitmap space.		

Before you specify the configuration changes, consider the following factors:

- For FlashCopy mappings, only one I/O group consumes bitmap space. By default, the I/O group of the source volume is used.
- When you create a reverse mapping; for example, to run a restore operation from a snapshot to its source volume; a bitmap is also created for this reverse mapping.
- The smallest possible bitmap is 4 KiB; therefore, a 512 byte volume requires 4 KiB of bitmap space.

On existing systems, also consider these factors:

- When you create FlashCopy mappings and mirrored volumes, or formatted, standard-provisioned volumes, the system attempts to automatically increase the available bitmap space. You do not need to manually increase this space.

This table shows the distributed RAID requirements for bitmap memory.

Table 106. Distributed RAID requirements	
RAID level	**Approximate required bitmap memory (MiB)**
RAID 1	Drive capacity * (count + 2 * stripe) / (stripe * 256 * 8)
RAID 6	Drive capacity * (count + 3 * stripe) / (stripe * 256 * 8)
Note: Drive capacity is in GiB, count = drive count, Stripe = stripe width, strip = strip length (in KiB). • For FlashCore Modules, the effective capacity should be used and not the physical capacity	

You can also use the **lsiogrp** and **chiogrp** command-line interface (CLI) commands to modify the settings.

Throttles

Throttles are a mechanism to control the amount of resources that are used when the system is processing I/Os on supported objects. The system supports throttles on hosts, host clusters, volumes, copy offload operations, and storage pools. If a throttle limit is defined, the system either processes the I/O for that object, or delays the processing of the I/O to free resources for more critical I/O operations.

When you configure throttles on the system, keep in mind the following guidelines:

- The throttle limit is a per node limit. For example, if a throttle limit is set for a volume at 100 IOPS, each node on the system that has access to the volume allows 100 IOPS for that volume. Any I/O operation that exceeds the throttle limit are queued at the receiving nodes.
- The multipath policies on the host determine how many nodes receive I/O operations and the effective throttle limit.
- If more than one throttle applies to an I/O operation, the lowest and most stringent throttle is used.

For each type of throttle, you can create either a IOPS or bandwidth limit or both. The throttle limit controls whether the processing of the host I/O continues or if it is delayed for later processing. Several reasons can exist to delay processing for specific hosts and host clusters. For example, if a host or host cluster has I/O intensive workloads, such as data mining systems, you can create a throttle limit that automatically delays the processing of I/O operations that exceed the configured throttle. Each host in the system can have a throttle defined or you can define throttle limits for a host cluster. In a host cluster throttle, all the hosts in the host cluster share the throttle limit.

Note: If you are creating a host cluster, any hosts that are added cannot have throttles configured. Throttles can be applied to the entire host cluster.

The system also supports throttles to delay processing of I/O operations for volumes. If storage systems provide storage to a wide variety of applications, then production volumes with more critical I/O can be competing with volumes that have lower priority operations. For example, volumes that are used for backup or archive operations can have I/O intensive workloads, potentially taking bandwidth from production volumes. Volume throttle can be used to limit I/Os for these types volumes so that I/O operations for production volume are not affected.

You can also create throttles for systems that have copy offload features enabled, such as offloaded data transfer (ODX) on Microsoft Windows Server 2012 or for XCOPY/WRITESAME features on VMware hosts. Copy offload frees up hosts and speeds the copy process by offloading the operations of certain types of hosts to a storage system. For systems with these features enabled, administrators can define throttles to delay processing for copy offloads to free bandwidth for other more critical operations. When a throttle for copy offload is defined, the throttle applies across the entire system. Like other throttles on the system, you can set IOPS throttles, bandwidth throttles, or both; however, bandwidth throttles are more effective for copy offload operations.

Throttles can be defined for storage pools to control I/O operations on back-end storage systems. Storage pool throttles can be used to avoid overwhelming the back-end storage and be used with virtual volumes. You can set throttles for both parent and child pools. Throttles work in a hierarchy when they are applied to both the parent and child pool. Throttles that are defined in the parent pool applies to volumes in both parent and child pools. However, throttles that are defined on the child pool only apply to the volumes in that child pool. For mirrored volumes that are in different pools, only the throttling that applies to the primary copy is considered, regardless of which copy the I/O is directed to. In this case, any throttles defined on the secondary copy of the volume are ignored. Only throttles on the primary copy are used, whether the pool is parent or child pool. When throttles are defined in both parent and child pools, the more restrictive throttle applies to the volume in the child pool, but the I/O is counted against both pools. Consider these examples:

- A parent pool is configured with a throttling limit of 5K IOPS and its single child pool had a limit at 3K IOPS. When I/O operations to a volume in the child pool reaches the throttling limit, the volume cannot receive any more I/O during that time. However, volumes in the parent pool can still use 2K of IOPS to process during that time.
- A parent pool is configured with a throttling limit of 4K IOPS . The parent pool also has two child pool and each child pool has a throttling limit of 3K of IOPS.
- A parent pool is configured with a throttling limit of 5K IOPS . The parent pool has also has two child pool without a throttle defined. All in all, the volumes in each pool cannot handle more than 4K IOPS but each of these pools can could serve its IOPS limit if the volumes in the other pools consume at a lower rate of IOPS In this example, the volumes in the parent and both child pools can reach the 5K IOPS together, regardless of which pool the volumes are located.

Throttles for volumes, hosts, host clusters, copy offload, and storage pools can be configured in the management GUI and by using the command-line interface (CLI).

Chapter 20. Support

The system provides several options to help you to identify and resolve problems that may require assistance from service support personnel.

Call Home

Call Home connects the system to support personnel who can monitor and respond to system events to ensure that your system remains up and running.

If a software or hardware error occurs, Call Home notifies the support center, and automatically opens a service request. Call Home sends service-related information to inform support personnel so they can quickly act to resolve the problem.

Cloud Services

Prerequisites

If you are using Call Home with cloud services to send notifications to the IBM Support, you need to ensure that the system can connect to the IBM Support.

Before you configure Call Home with cloud services, ensure that the following prerequisites are configured on the system:

- Ensure that all the nodes on the system have internet access.
- Ensure that a valid service IP address is configured on each node on the system.

Call Home with cloud services support the following configurations that have extra network requirements.

Table 107. Supported network configurations for Call Home with cloud services

Supported configuration	DNS configuration	Connection requirements
Call Home with cloud services with an internal proxy server	Required	Configure proxy to allow outbound traffic on port 443 to `esupport.ibm.com`
Call Home with cloud services with a DNS server	Required	Configure firewall to allow outbound traffic on port 443 to `esupport.ibm.com`.

Call Home with cloud services sends notifications directly to a centralized file repository that contains troubleshooting information that is gathered from your system. Support personnel can access this repository and be assigned issues automatically as problem reports. This method of transmitting notifications from the system to support removes the need for you to create problem reports manually.

As part of Call Home with cloud services configuration, you can define an internal proxy server within your network to manage connections between the system and the IBM Support. If an SSL inspection proxy is in use, you must define the proxy server and add its SSL certificate to enable Call Home to work correctly. For more information, see .

To configure Call Home with cloud services, use the **Settings** > **Support** > **Call Home** panel in the management GUI.

Email Notifications

For redundancy email notifications can also be configured. Call Home with email notifications sends notifications through an email server to support, and local users or services that monitor activity on the system. Call Home with email notifications requires configuration of at least one email server and local user.

© Copyright IBM Corp. 2024

To configure Call Home with email notifications, use the **Settings** > **Support** > **Call Home** panel or the **Settings** > **Notifications** > **Email** panel in the management GUI.

Inventory Information and Heartbeats

Inventory data is a summary of the hardware components and configuration of a system and can be sent by using Call Home with cloud services or Call Home with email notifications. Service personnel can use this information to contact you when relevant software updates are available or when an issue that can affect your configuration is discovered. It is a good practice to enable inventory reporting.

If you are using Call Home with cloud services, the system sends a heartbeat notification to support as part of the Call Home function every 24 hours. Heartbeat notifications contain the same information about the system as inventory notifications, but they are not sent to any local user.

Because inventory information is sent by using the call home email function, you must enable call home email before you can send an inventory information email.

You can adjust the contact information, sensor sensitive information, adjust the frequency of inventory emails, and enable or disable the sending of inventory data/heartbeats by using the management GUI. Use the **Settings** > **Support** > **Call Home** panel in the management GUI.

Storage Insights

You can also use Call Home with cloud services to send configuration, capacity, and performance metadata about your IBM Storage Virtualize systems to IBM Storage Insights for detailed analysis, monitoring, and reporting.

For more information on IBM Storage Insights, see .

Support assistance

Support assistance enables support personnel to access the system to complete troubleshooting and maintenance tasks.

You can configure either local support assistance, where support personnel visit your site to fix problems with the system, or local and remote support assistance. Remote support assistance allows support personnel to access the system remotely from the support center. Both local and remote support assistance use secure connections to protect data exchange between the support center and system. All actions that are completed with support assistance are recorded for auditing purposes. Local support assistance must be configured before remote support assistance is enabled.

Local support assistance

When you enable local support assistance, you can specify the IP address or domain name for the support connections. When support personnel log on to the systems with local support assistance, they are assigned either the Monitor role or the Restricted Administrator role.

The Monitor role can view, collect, and monitor logs and errors to determine the solution to problems on the system.

The Restricted Administrator role gives support personnel access to administrator tasks to help solve problems on the system. However, this role restricts these users from deleting volumes or pools, unmapping hosts, or creating, deleting, or changing users.

Roles limit access of the assigned user to specific tasks on the system. Users with the service role can set the time and date on the system, delete dump files, add and delete nodes, apply service, and shut down the system. They can also view objects and system configuration settings but cannot configure, modify, or manage the system or its resources. They also cannot read user data.

Remote support assistance

With remote support assistance, support personnel can access the system remotely through a secure connection from the support center. However, before you enable remote support assistance between the system and support, you first need to configure local support assistance. You must have either Call Home with cloud services or Call Home with email notifications configured. For more information about Call Home, see "Call Home" on page 503

During system initialization, you can optionally set up a service IP address and remote support assistance. If you did not configure a service IP address, go to **Settings** > **Network** > **Service IPs** to configure a service IP for each node on the system. If you use a firewall to protect your internal network you can configure a remote proxy server to allow access.

To prevent connection errors, ports 22 and 443 must be configured to support the service IP addresses remote support assistance.

When you enable remote support assistance, you can specify either IP address or domain name for support. If you specify a fully qualified domain name, a DNS server must be configured on your system. To configure a DNS server for the system, select **Settings** > **Network** > **DNS**. In addition, you can define a shared-token that will be generated by the system and sent to the support center. If the system needs support services, support personnel can be authenticated onto the system with a challenge-response mechanism. After support personnel obtain the response code, it is entered to gain access to the system. Service personnel have three attempts to enter the correct response code. After three failed attempts, the system generates a new random challenge and support personnel must obtain a new response code.

Remote code load

Remote code load (RCL) is a service that allows remote support engineers to complete code updates on the storage system.

RCL is the process of having IBM support personnel securely connect to and update the microcode on the storage system. The RCL service is the preferred code delivery method, which proves to be both efficient and secure for IBM clients. RCL is fast and easy to coordinate because it does not require an onsite visit of an IBM services technician.

Prerequisites

If you are configuring remote support assistance, the following prerequisites are required for all configurations.

- Call home must be configured and functioning with a valid email server. To configure call home, select **Settings** > **Notifications** > **Email** in the management GUI or through system setup.
- Service IP addresses must be configured on each node on the system. To configure service IP addresses, select **Settings** > **Network** > **Service IPs** in the management GUI.
- A DNS server must be configured on your system. To configure a DNS server, select **Settings** > **System** > **DNS** in the management GUI.

  Note: DNS of your local system should allow for local and remote servers. It should not be configured to allow only a single external DNS server like Google 8.8.8.8.
- You can configure your firewall to allow traffic to pass directly from the system or you can route traffic through an HTTP proxy server within your environment.
- With the addition of the HTTP proxy support, Remote Support Proxy servers are no longer necessary, but they are still fully supported for existing configurations. Optionally, a Remote Support Proxy can be configured to consolidate firewall traffic from a number of storage systems. Remote upgrades cannot be completed through the Remote Support Proxy server.

Firewall configuration

The following network connections between IBM and the system are required to enable support assistance.

esupport.ibm.com

The esupport.ibm.com network connection is used to for the following actions:

- Uploading logs to the IBM Enhanced Customer Data Repository (ECUREP)
- Connecting to Call home with cloud services
- Downloading software from FixCentral

Note: The esupport.ibm.com network connection is fully certified to securely transmit data for Blue Diamond (HIPPA) users and General Data Protection Regulation (GDPR) protected users.

If you are using a firewall to route traffic instead of an HTTP proxy server, use the following information to configure a firewall rule.

Table 108. Firewall rule configuration

Source	Target	Port	Protocol	Direction
The service IP address of every node or node canister.	esupport.ibm.com	443	https	Outbound only

Remote Access

Use the following information to configure a firewall rule for remote access.

Table 109. Firewall rule configuration for remote access

Source	Target	Port	Protocol	Direction
The service IP address of every node or node canister	170.225.126.11, 170.225.126.12, 170.225.127.11, and 170.225.127.12	443 or 22	https or ssh	Outbound only

For remote access, port 443 is suggested as it provides more security. If the connection from port 443 remote access is unsuccessful, then port 22 is used as a backup. When you configure the firewall rules with port 443, the fallback of remote access from port 443 to port 22 is not applicable.

Configure support assistance

To configure support assistance, use the **Settings** > **Support** > **Support Assistance** > **Set Up Support Assistance** panel in the management GUI.

Support packages

If support assistance is configured on your system, you can either automatically or manually upload new support packages to the support center to help analyze and resolve errors on the system. To upload support packages, you must have at least one DNS server.

You can select individual logs to either download to review or send directly to the support center for analysis.

You can select from the following types of support packages:

Type 1: Standard logs

This option contains many different log files for both the underlying operating system and the software that includes these critical log files:

- Event log
- Audit logs

- Linux based logs that include `/var/log/messages`

These logs are sufficient to diagnose many problems, especially simple hardware replacement procedures.

Type 2: Standard logs plus one existing statesave
This option contains the standard logs for the system and an existing statesave from any of the nodes in the system. Statesaves can either be a *dump* or a *livedump*. A *dump* file is collected when the software restarts for any reason. This file can be used to understand why the software restarted. A *livedump* collects the current state of the software with minimal impact to I/O operations. The contents of a livedump are similar to the contents of a dump.

Type 3: Standard logs plus the most recent statesave from each node
Contains the standard logs for the system and the most recent statesave from each of the nodes on the system.

Type 4: Standard logs plus new statesave
This option contains the standard logs and a new statesave (livedump) for all the nodes in the system and packages them with the most recent logs.

When you are deciding what type of package to send, consider the speed of the upload and if more information is necessary later to resolve the issue. The standard logs are uploaded more rapidly since the support package is smaller than the other options. However, the standard logs might not include all the information that is needed to resolve the problem on the first attempt. Support personnel can require more data to resolve the problem. If a new statesave (livedump) is not generated shortly after the standard log, the livedump might not contain the data necessary to resolve the issue. When you generate the standard logs with the new statesave, you have the greatest chance of resolving the issue without more information. However, the support package is larger and can take longer to upload to the support center.

The following table shows the four types of support packages and the approximate size for those support packages for a basic configuration. The table describes common troubleshooting scenarios that require that type of support package. Actual sizes of support packages vary based on the number of volumes and MDisks, the software level, and the compressibility of the logs.

Table 110. Support packages

Description	Approximate Support Package Size: System with 1 I/O group and 30 volumes)	Approximate Support Package Size: System with 4 I/O groups and 250 volumes	Common Scenarios
Standard logs	10 MB	340 MB	• Simple hardware replacement • Critical performance issues[1] • All other issues
Standard logs plus one existing statesave	50 MB	520 MB	
Standard logs plus most recent statesave from each node	90 MB	790 MB	2030, 1196 or 1195 errors

Table 110. Support packages (continued)

Description	Approximate Support Package Size: System with 1 I/O group and 30 volumes)	Approximate Support Package Size: System with 4 I/O groups and 250 volumes	Common Scenarios
Standard logs plus new statesaves	90 MB	790 MB	• Compressed volumes issues • Host issues • External storage issues • General performance issues • Critical performance issues[1]

1. Critical performance issues require that both standard logs and a new statesave are generated and uploaded to the support center. Generate and upload the standard logs first and then create the new statesaves.

Uploading support packages

To upload support packages, use the **Settings** > **Support** > **Support Package** panel in the management GUI.

Chapter 21. Software update

To obtain fixes or new features, you need to update the system software or drive firmware.

The system update procedure upgrades the software and hardware components. This procedure restarts each node in turn to enable host I/O to continue through the partner node during the update.

The drive firmware is updated by using a drive-specific procedure.

Some of the components of the system can be updated via a patch. The patch installation procedure does not perform any node restarts, so minimizes impact to the system environment.

Prior to updating the system or drives, the latest upgrade test utility must be installed and run. These activities indicate whether the system has issues that need to be resolved before the update.

Obtaining packages

The update packages are provided on **FixCentral**.

Package types

The following table shows different packages that are published to **FixCentral**:

*Table 111. Packages published to **FixCentral***

Fix Type	Example FixID	Description
Main product software	Storage-IBM_FlashSystem9100-8.6.0.2 This Fix contains the 8.6.0.2 release of IBM Storage Virtualize	This fix contains a full image of software that can be used to update your system. **Note:** • Before installing this software, run the upgrade test utility. • This package does not contain any drive firmware.
Drive Bundle	Storage-IBM_FlashSystem9100-DriveMicrocode-231019 This bundle contains firmware for all supported drives and was created on 19 October 2023.	This fix contains the latest firmware for all supported drives. The fix contains various file bundles. The exact file bundle that you need to use depends on your system, and is described in the “Which Drive File Should I use” text file. **Note:** Before installing this firmware, run the upgrade test utility.

© Copyright IBM Corp. 2024

*Table 111. Packages published to **FixCentral** (continued)*		
Fix Type	**Example FixID**	**Description**
Software Upgrade Test Utility	Storage-IBM_FlashSystem9100-SwUpgradeTestUtility **Note:** This FixID never changes.	This fix is a lightweight utility that performs a validation of planned update (either drive update or software update) to make sure that there are no known issues with the update. It is essential to always use the latest upgrade test utility to protect against known issues. When downloading either the drive bundle or the main firmware from **FixCentral**, the upgrade test utility is automatically included in the download.
Patches	These fixes are not available from **FixCentral** without a direct download link.	Under some scenarios, IBM releases a patch that fixes or updates a small part of the product, and can be updated without any interruption to I/O processing.
Generally available patches	These fixes are available from FixCentral without a direct download link.	Under some scenarios, IBM releases a patch that fixes or updates a part of the product, and can be updated without any interruption to I/O processing.

Transferring packages

If Call Home with cloud services is enabled, the system connects to Fix Central and directly transfer the latest software update available to the storage system. The user can also transfer specific packages directly to the system by using the GUI. For more information, see “Call Home” on page 503.

To transfer a specific package to the system, select **Transfer** on the **Update System** panel in the GUI. The package can also be obtained by using the **`satask downloadsoftware`** CLI command.

Note: Fix ID or Patch ID can be provided with access key, if one is required. For example, specify Storage_Disk-2076-8.6.2.0-ifix4:17241343604005011 to install package Storage_Disk-2076-8.6.2.0-ifix4 with access key 17241343604005011.

If patch auto-update is enabled, select **View available patches** on the **Update System** panel in the GUI to view and transfer generally available patches.

To obtain a package from **FixCentral** without using Call Home, access the following site http://www.ibm.com/support.

The packages are available under **Downloads** > **Fixes, updates and drivers**.

Note: The *md5sum* files are available on **FixCentral** to enable users to confirm a successful file transfer.

Once the update package is on your machine, select **Upload** on the **Update System** panel in the GUI to transfer the update package to the system. If using the CLI, copy the relevant file to the `/update` directory on the config node.

System update

During the system update, the nodes are updated sequentially. Each node update involves a restart, during which time the partner node in the I/O group processes I/O. After all the nodes in the system are successfully restarted, the new code level is automatically committed.

After the commit, the system can automatically update firmware levels on enclosure hardware components. Similar to the node updates, these updates are performed in a non-disruptive manner.

Updating the system by using the GUI

Before updating the system, ensure that the multi-pathing driver on connected hosts is correctly configured and working.

The update is started on the **Update System** panel in the GUI. The GUI guides the user through the procedure, including the preinstall checks that are performed by the upgrade test utility.

Updating the system by using the CLI

To update the system by using the CLI:

- Download the system code file to the `/update` directory.
- Download and install the upgrade test utility by using the **`applysoftware`** command.
- To perform preinstall checks, use the following command:

```
svcupgradetest -v <v.r.m.f>
```

  Note: For example, if you are planning to update the system to SVC 8.6.0.4, use the command **`svcupgradetest -v 8.6.0.4`**
- Install the system code by using the **`applysoftware`** command.

Note: For more information on the use of **`svcupgrade`** test, see https://www.ibm.com/support/pages/software-upgrade-test-utility.

User-driven system configuration actions are restricted while the update is in progress.

Once the system update is complete, a notification is sent on the GUI. The update status, including the estimated completion time, is available in the **Update System** GUI panel or the **`lsupdate`** CLI command.

The following table shows the estimated time for an update on a four-node system:

Table 112. Upgrade time for a four-node cluster

Steps	Time for step (in minutes)	Minimum time (in minutes)	Maximum time (in minutes)
Get ready	1	1	1
Upgrade node A2	9-24	10	25
Upgrade node B4	9-24	19	49
Wait for 10 minutes (for multi-pathing recovery)	10	29	59
Upgrade node A1	9-24	38	83
Upgrade node B3	9-24	47	107
Commit	10	57	117
Total	-	57	117

Updating the system manually

During a manual system upgrade, each node is removed, individually updated, and added back to the cluster in turn.

Manual upgrade allows users to directly control the removal and upgrade of individual nodes. It is also used when automatic CCU has encountered problems midway through upgrade and requires additional intervention from the user. Some of the checks from automatic CCU are removed, so manual upgrade should be used with caution.

As each node is removed and then updated, which also includes a restart, the partner node in the I/O group processes I/O. Only one node should be updated at a time. After the final node is removed for upgrade, the system automatically commits the upgrade.

Start the update on the **Settings** > **System** > **Update System** in the management GUI by using the **Alternative Update Options** tab. Select the appropriate test utility, upgrade package, and code level. When asked, select **Service Assistant Manual Update**.

Remove a node from the cluster, use the **Update Manually** panel to upload the necessary file, and start the update. After the node update is complete, it is added back to the system automatically. The update status indicates that a manual upgrade is in progress. Completion estimate is not provided because it depends on the pause between each node being removed and upgraded.

Update each node, one at a time, other than the configuration node by using this method. After all nodes, other than the configuration node, are updated and successfully returned to the cluster, remove the configuration node. Removal of the configuration node causes the remaining nodes, which are all on the uplevel version of the code, to commit the upgrade. One of the remaining nodes is elected as the new configuration node. After the commit is complete, a notification is sent to the GUI.

After the commit is complete, the removed node can be added back to the system. This node identifies that it is still on the downlevel build and it automatically updates the code level to conform with the rest of the cluster.

You can also start the manual update by using the **`applysoftware`** CLI command by using the **`-prepare`** option. Ensure that the **`installsoftware`** command is run with the **`-pacedccu`** option. Continue to run the upgrade test utility before proceeding with the update.

The user-driven system configuration actions are restricted while the update is in progress.

After the commit, the system might automatically update firmware levels on enclosure hardware components. Similar to the node updates, these updates are performed in a nondisruptive manner.

Patch installation

Under some scenarios, IBM releases a patch that fixes or updates a part of the product. Patch installation does not cause any interruption to the IO processing.

Patches are made available on **FixCentral** with the help of direct download link (for more information, see Obtaining packages) or as part of a general collection of available patches.

For patches that are made available by a direct download link

To start the installation process:

1. Go to **Settings** > **System** > **Update System**.
2. To install the patch, under **Alternative update options** section, check **Use an existing package**.
3. Select the patch from the list of available packages on the system and click **Install Patch**.

For patches that are released as part of a general collection of available patches

To start the installation process:

1. Go to **Settings** > **System** > **Update System**.
2. To install the patch, select **View available patches**, and click **Install** from the overflow menu.

3. To view or remove the patches in the GUI, select **View installed patches** on the **Update System** panel.

Note: The **View available patches** link only works correctly if the patch auto-update service is enabled. To enable the patch-auto update service use the **Settings** > **System** > **Patch auto update** panel in the management GUI.

Also, the **View available patches** link is displayed only when minimum of one patch is installed on the system or at least one patch is available for patch installation.

Drive update

You can update specific drives or all drives in the system. The update procedure determines whether any drives are already at the code level that is provided in the package and does not update such drives.

Drive updates by using packages

The software upgrade test utility reports if any drives need updating when run as a part of the system update or drive update procedure.

The latest drive firmware package is available on **FixCentral**. For more information, see Obtaining packages.

To update drives using GUI, select **Pools > Internal Storage > Actions > Upgrade All** or **Upgrade**.

To monitor the progress of the upgrade, select **Monitoring > Background Tasks**.

To update drives by using the CLI:

- Download the target-drive firmware file to the `/update` directory, and download and install the upgrade test utility by using command.
- To perform preinstall checks, use the following command:

```
svcupgradetest -w <firmware_package>
```

Note: For more information on the use of **`svcupgrade`** test, see https://www.ibm.com/support/pages/software-upgrade-test-utility.

Chapter 22. Management GUI

The system includes an easy-to-use management GUI to help you to monitor, manage, and configure your system.

You can access the management GUI by opening any supported web browser and entering the management IP address. In addition to simple setup, configuration, and management functions, the management GUI provides several functions that help filter and sort data on the system:

Automatic and manual resizing of main navigation
The main navigation menu automatically shrinks or expands based on the width of the browser viewport. If sufficient space is available, the navigation menu automatically expands to show the main navigation item names. When in the expanded mode, a toggle button is available to shrink or expand the navigation menu manually.

Filtering and sorting objects
On windows that include columns, you can sort each column by clicking the column heading. You can use the filtering feature to display items that include only text that you specify.

Selecting multiple objects
You can use the Ctrl key to select multiple items and choose actions on those objects by right-clicking them to display the actions menu. You can right-click any column heading to add or remove columns from the table.

Using preset options
The management GUI includes several preestablished configuration options to help you save time during the configuration process. For example, you can select from several preset options when you create a new volume. These preset options incorporate commonly used parameters.

Viewing help
Another useful interface feature is integrated help. You can access help for some fields and objects in the management GUI by hovering over the question mark icon near the object. Panel-specific help is available by clicking **Need Help** or by using the **Help** link in the upper-right corner of the management GUI.

Checking your web browser settings for the management GUI

You must have the latest supported web browser to access the management GUI. The management GUI supports the following HTML5-compliant browsers:

- Mozilla Firefox
- Mozilla Firefox Extended Support Release (ESR)
- Microsoft Edge
- Google Chrome

Note: The minimum viewport for browser window for the management GUI is 1024 x 768. Scaling browser view to less than those dimensions affects readability of data on management GUI pages.

IBM supports higher versions of the browsers if the vendors do not remove or disable function that the product relies upon. For browser levels higher than the versions that are certified with the product, customer support accepts usage-related and defect-related service requests. If the support center cannot re-create the issue, support might request the client to re-create the problem on a certified browser version. If a problem is identified with the browser, IBM might investigate potential solutions or work-arounds that the client can implement until a permanent solution becomes available.

The GUI requires the following configuration settings:

- JavaScript enabled
- Cookies enabled
- File Download enabled

© Copyright IBM Corp. 2024

- Disable or Replace Context menus (Mozilla Firefox only)

Refer to browser documentation on how to configure these settings.

When to use the management GUI

The management GUI is the primary tool that is used to service your system.

Regularly monitor the status of the system using the management GUI. If you suspect a problem, use the management GUI first to diagnose and resolve the problem.

Use the views that are available in the management GUI to verify the status of the system, the hardware devices, the physical storage, and the available volumes. The **Monitoring** > **Events** panel provides access to all problems that exist on the system. Use the **Recommended Actions** filter to display the most important events that need to be resolved.

If there is a service error code for the alert, you can run a fix procedure that assists you in resolving the problem. These fix procedures analyze the system and provide more information about the problem. They suggest actions to take and step you through the actions that automatically manage the system where necessary. Finally, they check that the problem is resolved.

If there is an error that is reported, always use the fix procedures within the management GUI to resolve the problem. Always use the fix procedures for both system configuration problems and hardware failures. The fix procedures analyze the system to ensure that the required changes do not cause volumes to be inaccessible to the hosts. The fix procedures automatically perform configuration changes that are required to return the system to its optimum state.

Chapter 23. Ansible collection

The Ansible collection provides a series of Ansible modules and plugins for interacting with the IBM Storage Virtualize family products.

For more information, refer to https://galaxy.ansible.com/ui/repo/published/ibm/storage_virtualize/docs/.

© Copyright IBM Corp. 2024

Chapter 24. Storage Virtualize RESTful API

The Storage Virtualize Representational State Transfer (REST) model application programming interface (API) consists of command targets that are used to create, read, update, and delete system resources. These command targets allow command parameters to pass through unedited to the Storage Virtualize command-line interface, which handles parsing parameter specifications for validity and error reporting. Use Hypertext Transfer Protocol Secure (HTTPS) to successfully communicate with the RESTful API server.

The RESTful API server does not consider transport security (such as SSL), but instead assumes that requests are initiated from a local, secured server. The HTTPS protocol provides privacy through data encryption. The RESTful API provides security by requiring command authentication by using tokens. The token timeout is a user configurable option and the default value is 60 minutes. However, the token timeout value can be viewed and changed by using **lssecurity** or **chsecurity** commands.

Uniform Resource Locators (URLs) target different node objects on the system. The HTTPS POST method acts on command targets that are specified in the URL. To make changes or view information about different objects on the system, you must create and send a request to the system. You are required to provide certain elements for the RESTful API server to receive and transform the request into a command, as described in the next section.

Making an HTTPS request

To interact with the system by using the RESTful API, make an HTTPS command request with a valid configuration node URL destination. Open TCP port 7443 and include the keyword rest. For IBM Storage Virtualize for Public Cloud on Amazon Web Services (AWS), you must manually open TCP port 7443 in the AWS security group before you use the RESTful API. The following URL formats are used for all requests:

```
https://system_ip:7443/rest/v1/target/ID
```

Where:

- *system_ip* is the IP address to which you are sending requests.
- The port number is always 7443 for the Storage Virtualize RESTful API.
- *rest* is a keyword.
- *v1* is version 1 of API.
- *target* is the target object of commands, which includes any object IDs, names, and parameters.
- *ID* is the object ID.

```
https://system_ip:7443/rest/v1/command
```

Where:

- *system_ip* is the system IP address, which is the address that is taken by the configuration node of the system.
- The port number is always 7443 for the Storage Virtualize RESTful API.
- rest is a keyword.
- *v1* is the version 1 of the API.
- *command* is the target command object (such as **auth** or **lseventlog** with any parameters).

  The *command* specification follows this format:

```
command_name,method="POST",headers={'parameter_name': 'parameter_value',
'parameter_name': 'parameter_value',...}
```

© Copyright IBM Corp. 2024

Why we use POST exclusively: All of the Storage Virtualize RESTful API command targets are named after Storage Virtualize commands, with names that already reflect Create, Read, Update, and Delete actions. MK commands make (create) resources, LS commands list (read) resources, CH commands change (update) resources, and RM commands remove (delete) resources. Using HTTP methods on such commands is redundant.

All commands, including LS commands, accept at least one named parameter because the Storage Virtualize RESTful API is POST method only. It implements a convention of appending positional parameters to the URI, and packing named parameters into the request body as JavaScript Object Notation (JSON) strings.

Why we do not use GET: Most HTTP servers reject GET requests with body data. To support GET, means appending named parameters to the URI as a query string. A query string does not support arbitrary data. A valid URI cannot contain any spaces and other reserved characters that might be present in the data. Therefore, you must URL-encode named parameter data before you append it to the URI. In addition to that inconvenience, the key-value structure of named parameter data must be conveyed in some way. The query string would likely be a URL-encoded JSON object. Supporting the GET method means that parameter data represented in your programming language of choice must pass through two separate encoders before the data is sent as part of an HTTP request.

The more elegant method for the command targets was to adopt POST exclusively.

As mentioned, in addition to the URL and the name of the command target, other information is required in the request line and in the body of the HTTP request that regards the action to take on the specified object. In the request line, include the POST HTTP method. Include any required parameters (such as RAID level or IP address) in the body of the request. For more information, refer to `https://<system-ip>:7443/rest/explorer/` for OpenApi explorer guide.

Provide any required parameters as valid JSON in the HTTP body, such as shown in the following example:

```
{'X-Auth-Username': 'superuser'}
```

The request is routed to port 7443 on the specified destination (which must be the configuration node of the system), where the request is received by the RESTful API server. The server runs the command, collects any resulting output, then creates an HTTP response like the following example:

Response header

```
content-length: 273
content-type: application/json;  charset=UTF-8
date: Mon,21 Jun 2021 10:47:18 GMT
server: nginx
strict-transport-security:  max-age=31536000; includeSubDomains
```

Response body

```
{
  "key": "value"
}
```

API limits

To handle the incoming requests from client, the web server is configured with some rate limits so that API does not fail. The API returns 429 HTTP error code to the client if that limit is reached. The following table lists the rate limits that are configured on the web server.

Table 113.

Rate Limit	Type	Requests
Max active connections per cluster	RESTful API	4

Table 113. (continued)

Rate Limit	Type	Requests
Max requests/sec to auth endpoint	RESTful API	3
Max requests/sec to command endpoints	RESTful API	10
Number of simultaneous CLIs in progress	System	1

Authentication overview

Aside from data encryption, the HTTPS server requires authentication of a valid username and password for each API session. Use two authentication header fields to specify your credentials: X-Auth-Username and X-Auth-Password.

The **/auth** endpoint uses the POST method for authentication request. The **/auth** endpoint returns JWT(JSON Web Token) authentication token that authenticates every command that a user runs. The JWT token has the token expiry encoded into it.

Initial authentication requires that you POST the authentication target (/auth) with the username and password. The RESTful API server returns a JWT token. The system supports configuring a session between 10 minutes to 2 hours. However, the default single session lasts a maximum of 1 hour. If the maximum allotted time is reached, an error code 403 occurs that indicate the loss of authorization. Also, the token expiry is configurable and can be set between 10 - 120minutes. Use the **/auth** command target to authorize again with the username and password.

For example, the following command passes the authentication command to target node IP 192.168.10.109 at port 7443:

```
https://192.168.10.109:7443/rest/v1/auth, method="POST",
   headers={'X-Auth-Username': 'superuser', 'X-Auth-Password': 'passw0rd'}
```

The HTTP request that was sent to the API server looks like:

```
:authority: 192.168.10.109:7443
:method: POST
:path: /rest/v1/auth
:scheme: https
accept: application/json
content-length: 0
x-auth-password: passw0rd
x-auth-username: superuser
```

Where:

- The host is directing the HTTP request to the port (7443) and IP address (192.168.10.109) on the system .
- POST is the only HTTPS method that the Storage Virtualize RESTful API supports
- *rest* is a keyword.
- *v1* is the version 1 of the API.
- **/auth** is the authentication request command.
- The *accept* specifies the content type (application/json).
- The content length specifies the message size.
- The *x-auth-password* and *x-auth-username* field is used to authenticate and retrieve token.

The successful **auth** command returns a token similar to the following example:

```
{"token":
eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzUxMiJ9.eyJpYXQiOjE2MTg4NDc4ODAsImV4cCI6MTYxODg1
MTQ4MCwianRpIjoiYTMzY2VlYmRkZGM3YmFhOTUwMmZmNGQ5ZmEwMzBhZTIiLCJzdiI6eyJ1c2VyI
joic3VwZXJ1c2VyIn19.1SXD8DoUYY-BO_cjGL9XZSnjTyBSRfp25YSWbr0Y9k5UaQrmz8xtp2XU
TR5l0PZKhqii4eU2AYf7gaTd3kfxWw}
```

The returned token string can be decoded by using any JWT decode library freely available on the web. The following is the example of the decoded JWT token string:

```
HEADER:ALGORITHM & TOKEN TYPE
{
  "typ": "JWT",
  "alg": "HS512"
}
PAYLOAD:DATA
{
  "iat": 1618847880,
  "exp": 1618851480,
  "jti": "a33ceebdddc7baa9502ff4d9fa030ae2",
  "sv": {
    "user": "superuser"
  }

}
```

API authentication management

To authenticate, use the user ID and the authentication (auth) target.

Use the authentication command target (auth) to authenticate by using the POST method. In this process, you provide a username and password through the X-Auth-Username and X-AuthPassword header fields. The beginning of the session is the only instance where you must enter the username and password. Upon successfully entering the username and password, an authentication JWT token is displayed, which has an expiry encoded into it. The token is a base64 encoded string that you include in future commands by using the X-Auth-Token header. The X-Auth-Token header in combination with the JWT authentication token replaces the username and password for each action. The JWT token authorizing hours are user configurable. The token remains authorized only for the configured time even if the token is unused. The default authorized time for the JWT token is one hour. After the configured time is reached, an error code of 403 occurs that indicates the loss of authorization. Use the **/auth** command target to reauthenticate with the username and password.

The system supports configuring a session between 10 minutes to 2 hours. However, the default single session lasts a maximum of 1 hour. If the maximum allotted time is reached, an error code 403 occurs that indicate the loss of authorization.

REST response

The following example shows the REST response for **lsuser** command:

```
[
    {
        "id": "0",
        "name": "superuser",
        "password": "yes",
        "ssh_key": "no",
        "remote": "no",
        "usergrp_id": "0",
        "usergrp_name": "SecurityAdmin",
        "owner_id": "",
        "owner_name": "",
        "locked": "no",
        "password_change_required": "no"
    }
]
```

The following example shows **lsuser** command that uses the SSH protocol :

```
lsuser
```

The resulting output:

```
id name password ssh_key remote usergrp_id usergrp_name owner_id owner_name locked password_change_required
0 superuser yes   no      no     0          SecurityAdmin                    no        no
```

Command target parameters

Refer to `https://<system_ip>:7443/rest/explorer/` for OpenApi explorer for command targets parameters.

Storage Virtualize Performance Statistics

The IBM Storage Virtualize REST model API replaces the Common Information Model (CIM) methods to extract general performance statistics from a Storage Virtualize system for new users.

Overview

Application programmers, who are migrating away from CIM to REST API, and new REST API users can complete the following tasks to extract system performance statistics:

- Understand the now decommissioned CIM-based performance statistics for the IBM Storage Virtualize systems.
- Discover and connect to the REST API to list and download performance data files.
- Retrieve and extract metrics and apply calculations that were previously applied by the CIM agent.

Introduction

The CIM agent is a set of standards that are developed by the Distributed Management Task Force (DMTF). CIM includes an open approach to the design and implementation of storage systems, applications, databases, networks, and devices.

Newer systems increasingly rely on the REST architecture to manage and monitor systems, as it optimally utilises HTTP, thereby providing an advantage over CIM. The HTTP-based infrastructure is widely available and comes with built-in features, including caches, proxies, and scalability. REST systems are stateless, making REST operations simple and performance efficient.

IBM Storage Virtualize support a REST API implementation that is tuned for performance and scalability. Therefore, existing systems that rely on CIM agent functionality must upgrade to use the REST interface as the underlying delivery mechanism to obtain performance metrics.

The following sections provide information related to the migration process, calculations that are performed internally by the CIM agent, and coding examples. This information aims at simplifying the migration process such that the level of complexity involved in implementing and managing the existing scripts is greatly reduced.

Performance data files

The performance statistics of the IBM Storage Virtualize are stored in the performance data files.

The IBM Storage Virtualize system collects performance statistics and stores them on the management node for a certain time, usually statistics are updated at 5 minutes interval which is configurable via the chsystem command and can be viewed using the lssystem command. The performance data files are accessible over the REST API endpoints. You can list and download the files by using a scripting language, for example Python or Linux utility like cURL. The REST API accepts response type JSON (JavaScript Object Notation) which is a language-independent data format and has built-in support in

various scripting languages. You can filter the JSON response from the REST API server on the client and enable the client end programs to remove any noise from the data.

Note: The following examples use Python programming language, and the code and variables that are declared in previous examples.

Authenticating

Acquire a token to complete operations on the system by using the REST API. To request an API token, make an HTTP POST call to `/auth` endpoint and provide a username and password in the request header. See the following example.

Request

```
import requests

url = "https://example.com:7443/rest/v1"
username = "myuser"
password = "mypassword"

headers = {
    "X-Auth-Username": username,
    "X-Auth-Password": password
}

response = requests.post(url + "/auth", headers=headers, verify=False)
token = response.json()['token']

response.json()
```

Response

```
{"token":
"eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzUxMiJ9.eyJpYXQiOjE3MDg1MjY4NzYsImV4cCI6MTcwODUzMDQ3NiwianRpIjoiZT
EzMzQ4YmUzZGRkMWRjZTQ4YzhkYzE5Y2RhMTFjYzMiLCJzdiI6eyJ1c2VyIjoic3VwZXJ1c2VyIn19.I--
KYnXUsFQ4wbtw_zmfkHrZ6V4bsWJgZA1j3WeHchw_kJVFRrPAhrjzJlYybIIDBYA05OSnmIZ9YWjpFY9uUQ"}
```

Listing files

The XML statistics files that contain the performance data are stored in `/dumps/iostats/` directory on the config node. The `/lsdumps` API endpoint identifies the files that are accessible for download through the API. To list the contents of `/dumps/iostats/`, use the `prefix` parameter in the request body. See the following example.

Request

```
headers = {
    "X-Auth-Token": token,
}
data = {
    "prefix": "/dumps/iostats/"
}
response = requests.post(url + "/lsdumps", json=data, headers=headers, verify=False)
response.json()
```

Response

```
[{'id': '0', 'filename': 'Nm_stats_F304745-2_240221_151627'},
 {'id': '1', 'filename': 'Ng_stats_F304745-2_240221_151627'},
 {'id': '2', 'filename': 'Nn_stats_F304745-2_240221_151627'},
 {'id': '3', 'filename': 'Nv_stats_F304745-2_240221_151627'},
…
]
```

Filtering on file names

The system collects statistics over an interval and creates files that can be downloaded. For each collection interval, the system creates five statistics files:

- Nm_stats for managed disks (MDisks)
- Nv_stats for volumes and volume copies
- Ng_stats for volume groups
- Nn_stats for nodes
- Nd_stats for drives

The file name consists of several parts, as shown in the following description:

<category>_<panel_name>_<yymmdd>_<hhmmss>

For example, Nv_stats_F304745-2_240221_151627

The naming convention enables a filtering capability that a client-side script might use to download:

- Specific type of files
- Files that were created at a specific time interval.

See the following example.

Request

```
import re

headers = {
    "X-Auth-Token": token,
}
data = {
    "prefix": "/dumps/iostats/"
}
response = requests.post(url + "/lsdumps", json=data, headers=headers, verify=False)
files = [row['filename'] for row in response.json()]

# Filter on category
nm_files = [name for name in files if re.match("Nm_stats", name)]
nv_files = [name for name in files if re.match("Nv_stats", name)]
ng_files = [name for name in files if re.match("Ng_stats", name)]
nn_files = [name for name in files if re.match("Nn_stats", name)]
nd_files = [name for name in files if re.match("Nd_stats", name)]
# Filter on time interval
tm_files = [name for name in files if re.match(".*_142127$", name)]
# Filter on panel name
pn_files = [name for name in files if re.match(".*F304745-2.*", name)]
# Make a selection
target_file = nm_files[0]
# Print
pn_files[:2]
```

Response

```
['Nm_stats_F304745-2_240222_143127', 'Nn_stats_F304745-2_240222_143127']
```

Downloading files

You can download the files that can be listed by using the /lsdumps API endpoint, through the /download endpoint. The /download interface accepts two parameters:

- prefix: Name of the target directory
- filename: Name of the target file

The contents of the downloaded file can be processed either in memory or stored on the host for post processing.

Note: The performance data files are constantly updated on the system. If an error occurs during the file download that means the file is rotated. Usually, you can fetch the latest list of files, followed by another download request. If not, reevaluate statistics intervals and collection intervals to ensure that the list and download is completed within the network and system limits.

See the following example.

Request

```
headers = {
    "X-Auth-Token": token,
}
data = {
    "prefix": "/dumps/iostats",
    "filename": target_file
}
response = requests.post(url + "/download", json=data, headers=headers, verify=False)

if response.status_code != 200:
    raise Exception("File not found")

response.content[:100]
```

Response

```
b'<?xml version="1.0" encoding="utf-8" ?>\n<diskStatsColl\nxmlns="http://ibm.com/storage/
management/perf'
```

Getting performance statistics

You can extract various data metrics from the system by using code.

The following code snippet provides a reference for authenticating a user session, getting a list of the statistics files, and downloading a statistics file for individual calculations. The example code is also used in later sections to perform the actual calculations for various types of metrics that can be pulled from the system.

Note: The following examples use Python programming language, and the code and variables that are declared in previous examples.

```
import re
import xml.etree.ElementTree as ET
import requests
import json
import pprint

def get_token(url, username, password):
    response = requests.post(
        url + "/auth",
        headers={"X-Auth-Username": username, "X-Auth-Password": password},
        verify=False)
    return response.json()['token']

def get_namespace(element):
    m = re.match(r'\{.*\}', element.tag)
    return m.group(0) if m else ''

def get_directory_listing(url, token):
    response = requests.post(
        url + "/lsdumps",
        json={"prefix": "/dumps/iostats"},
        headers={"X-Auth-Token": token},
        verify=False)
    if response.status_code != 200:
        raise Exception("Invalid directory listing")
    return response.json()

def get_file_contents(url, token, filename):
    response = requests.post(
        url + "/download",
        json={"prefix": "/dumps/iostats", "filename": filename},
        headers={"X-Auth-Token": token},
        verify=False)
    if response.status_code != 200:
        raise Exception("File not found")
    return response.content

def get_latest_stats(url, username, password):
    token = get_token(url, username, password)
    files = [row['filename'] for row in get_directory_listing(url, token)]
    nm_files = [name for name in files if re.match("Nm_stats", name)]
    nv_files = [name for name in files if re.match("Nv_stats", name)]
```

```
    ng_files = [name for name in files if re.match("Ng_stats", name)]
    nn_files = [name for name in files if re.match("Nn_stats", name)]
    nd_files = [name for name in files if re.match("Nd_stats", name)]

    nm_xml = ET.fromstring(get_file_contents(url, token, nm_files[0]))
    nv_xml = ET.fromstring(get_file_contents(url, token, nv_files[0]))
    ng_xml = ET.fromstring(get_file_contents(url, token, ng_files[0]))
    nn_xml = ET.fromstring(get_file_contents(url, token, nn_files[0]))
    nd_xml = ET.fromstring(get_file_contents(url, token, nd_files[0]))

    xml = {
        "Nm_stats": nm_xml,
        "Nv_stats": nv_xml,
        "Ng_stats": ng_xml,
        "Nn_stats": nn_xml,
        "Nd_stats": nd_xml
    }
    return xml

def populate_fields(xml, fields):
    for item in fields.items():
        item[1]['value'] = int(
            xml[item[1]['filename']].
            find(get_namespace(xml[item[1]['filename']]) + item[1]['element']).
            attrib[item[1]['field']]
        )
```

For details on different types of metrics, see the following sections:

Node statistics

The following table lists:

- Performance statistics calculations previously provided by the CIM agent
- Respective fields in the statistics file that is reported for individual nodes.

Table 114. Calculations

CIMOM field	Calculation
TotalIOs	READ_OPERATION_TAG + WRITE_OPERATION_TAG
ReadIOs	READ_OPERATION_TAG
ReadHitIOs	READ_HITCACHE_TAG
WriteIOs	WRITE_OPERATION_TAG
WriteHitIOs	WRITE_HITCACHE_TAG

Var	File name	XML field	Description
DEVICE_ID_TAG	Nn_stats	"idx"	Indicates the enclosure or node ID
READ_OPERATION_TAG	Nn_stats	"ro"	Indicates the number of Read I/Os
WRITE_OPERATION_TAG	Nn_stats	"wo"	Indicates the number of Write I/Os
READ_HITCACHE_TAG	Nv_stats	"ctrh"	Indicates the number of reads received from other components and processed in read-cache hit
WRITE_HITCACHE_TAG	Nv_stats	"ctwfw"	Indicates the number of writes received

Var	File name	XML field	Description
			from other components and processed in write-cache hit

The following code example shows an illustration of the calculations in a scripting form.

Request

```
def get_latest_cimom_sample_node_stats(url, username, password):
    xml = get_latest_stats(url, username, password)

    fields = {
        "READ_OPERATION_TAG": {
            "filename": "Nn_stats",
            "element": "node",
            "field": "ro",
            "value": 0},
        "WRITE_OPERATION_TAG": {
            "filename": "Nn_stats",
            "element": "node",
            "field": "wo",
            "value": 0},
        "READ_HITCACHE_TAG": {
            "filename": "Nv_stats",
            "element": "vdsk",
            "field": "ctrh",
            "value": 0},
        "WRITE_HITCACHE_TAG": {
            "filename": "Nv_stats",
            "element": "vdsk",
            "field": "ctwfw",
            "value": 0}
    }
    populate_fields(xml, fields)
    cimom_fields = {
        "TotalIOs": fields['READ_OPERATION_TAG']['value'] + fields['WRITE_OPERATION_TAG']
['value'],
        "ReadIOs": fields['READ_OPERATION_TAG']['value'],
        "ReadHitIOs": fields['READ_HITCACHE_TAG']['value'],
        "WriteIOs": fields['WRITE_OPERATION_TAG']['value'],
        "WriteHitIOs": fields['WRITE_HITCACHE_TAG']['value']
    }
    return cimom_fields

pprint.pprint(get_latest_cimom_node_stats(url, username, password))
```

Response

```
{'ReadHitIOs': 0,
 'ReadIOs': 19291610,
 'TotalIOs': 39009988,
 'WriteHitIOs': 0,
 'WriteIOs': 19718378}
```

Backend (MDisk) volumes

The following table lists:

- Performance statistics calculations accomplished by the CIM agent
- Respective fields in the statistics file that is reported for individual mdisks.

Table 115. Calculations

CIMOM field	Calculation
deviceId	DEVICE_ID_TAG
TotalIOs	READ_OPERATION_TAG + WRITE_OPERATION_TAG

Table 115. Calculations (continued)	
CIMOM field	**Calculation**
KBytesTransferred(sizeUnits)	(READ_BLOCK_TAG + WRITE_BLOCK_TAG)*sizeUnits/1024
IOTimeCounter	READ_TIME_TAG + WRITE_TIME_TAG
ReadIOs	READ_OPERATION_TAG
ReadIOTimeCounter	READ_TIME_TAG
KBytesRead(sizeUnits)	READ_BLOCK_TAG*sizeUnits/1024
WriteIOs	WRITE_OPERATION_TAG
WriteIOTimeCounter	WRITE_TIME_TAG
KBytesWritten(sizeUnits)	WRITE_BLOCK_TAG*sizeUnits/1024

Var	File name	XML field	Description
DEVICE_ID_TAG	Nm_stats	"idx"	Indicates the name of the MDisk for which the statistics apply
READ_OPERATION_TAG	Nm_stats	"ro"	Indicates the cumulative number of MDisk read operations that were processed (since the node started)
WRITE_OPERATION_TAG	Nm_stats	"wo"	Indicates the cumulative number of MDisk write operations that were processed (since the node started)
READ_BLOCK_TAG	Nm_stats	"rb"	Indicates the cumulative number of blocks of data that was read (since the node started)
WRITE_BLOCK_TAG	Nm_stats	"wb"	Indicates the cumulative number of blocks of data written (since the node started)
READ_TIME_TAG	Nm_stats	"re"	Indicates the cumulative read external response time in milliseconds for each MDisk. The cumulative response time for disk reads is calculated by starting a timer when a host read command is issued and stopped when the command completes successfully. The elapsed time is

Var	File name	XML field	Description
			added to the cumulative counter
WRITE_TIME_TAG	Nm_stats	"we"	Indicates the cumulative write external response time in milliseconds for each MDisk. The cumulative response time for disk writes is calculated by starting a timer when a host write command is issued and stopped when the command completes successfully. The elapsed time is added to the cumulative counter

The following code example shows an illustration of the calculations in a scripting form.

Request

```
def get_latest_cimom_sample_mdisk_stats(url, username, password):
    xml = get_latest_stats(url, username, password)

    fields = {
        "READ_OPERATION_TAG": {
            "filename": "Nm_stats",
            "element": "mdsk",
            "field": "ro",
            "value": 0},
        "WRITE_OPERATION_TAG": {
            "filename": "Nm_stats",
            "element": "mdsk",
            "field": "wo",
            "value": 0},
        "READ_BLOCK_TAG": {
            "filename": "Nm_stats",
            "element": "mdsk",
            "field": "rb",
            "value": 0},
        "WRITE_BLOCK_TAG": {
            "filename": "Nm_stats",
            "element": "mdsk",
            "field": "wb",
            "value": 0},
        "READ_TIME_TAG": {
            "filename": "Nm_stats",
            "element": "mdsk",
            "field": "re",
            "value": 0},
        "WRITE_TIME_TAG": {
            "filename": "Nm_stats",
            "element": "mdsk",
            "field": "re",
            "value": 0}
    }
    populate_fields(xml, fields)
    cimom_fields = {
        "TotalIOs": fields['READ_OPERATION_TAG']['value'] + fields['WRITE_OPERATION_TAG']
['value'],
        "KBytesTransferred": (fields['READ_BLOCK_TAG']['value'] + fields['WRITE_BLOCK_TAG']
['value']) / 1024,
        "IOTimeCounter": fields['READ_TIME_TAG']['value'] + fields['WRITE_TIME_TAG']['value'],
        "ReadIOs": fields['READ_OPERATION_TAG']['value'],
        "ReadIOTimeCounter": fields['READ_TIME_TAG']['value'],
        "KBytesRead": fields['READ_BLOCK_TAG']['value'] / 1024,
        "WriteIOs": fields['WRITE_OPERATION_TAG']['value'],
        "WriteIOTimeCounter": fields['WRITE_TIME_TAG']['value'],
        "KBytesWritten": fields['WRITE_BLOCK_TAG']['value'] / 1024
```

```
    }
    return cimom_fields
pprint.pprint(get_latest_cimom_sample_mdisk_stats(url, username, password))
```

Response

```
{'IOTimeCounter': 0,
 'KBytesRead': 0.0,
 'KBytesTransferred': 1632.0,
 'KBytesWritten': 1632.0,
 'ReadIOTimeCounter': 0,
 'ReadIOs': 0,
 'TotalIOs': 18,
 'WriteIOTimeCounter': 0,
 'WriteIOs': 18}
```

Volume (VDisk) statistics

The following table lists:

- Performance statistics calculations accomplished by the CIM agent
- Respective fields in the statistics file that is reported for individual vdisks.

Table 116. Calculations

CIMOM field	Calculation
deviceId	DEVICE_ID_TAG
TotalIOs	READ_OPERATION_TAG + WRITE_OPERATION_TAG
KBytesTransferred(sizeUnits)	(READ_BLOCK_TAG + WRITE_BLOCK_TAG)*sizeUnits/1024
IOTimeCounter	READ_TIME_TAG + WRITE_TIME_TAG
ReadIOs	READ_OPERATION_TAG
ReadHitIOs	READ_HITCACHE_TAG
ReadIOTimeCounter	READ_TIME_TAG
KBytesRead(sizeUnits)	READ_BLOCK_TAG*sizeUnits/1024
WriteIOs	WRITE_OPERATION_TAG
WriteHitIOs	WRITE_HITCACHE_TAG
WriteIOTimeCounter	WRITE_TIME_TAG
KBytesWritten(sizeUnits)	WRITE_BLOCK_TAG*sizeUnits/1024

Var	File name	XML field	Description
DEVICE_ID_TAG	Nv_stats	"idx"	Indicates the volume for which the statistics apply
READ_OPERATION_TAG	Nv_stats	"ro"	Indicates the cumulative number of volume read operations that were processed (since the node started)
WRITE_OPERATION_TAG	Nv_stats	"wo"	Indicates the cumulative number of volumes

Var	File name	XML field	Description
			write operations that were processed (since the node started).
READ_BLOCK_TAG	Nv_stats	"rb"	Indicates the cumulative number of blocks of data read (since the node started)
WRITE_BLOCK_TAG	Nv_stats	"wb"	Indicates the cumulative number of blocks of data written (since the node started)
READ_TIME_TAG	Nv_stats	"rl"	Indicates the cumulative read response time in milliseconds for each volume. The cumulative response time for volume reads is calculated by starting a timer when a SCSI read command is received and stopped when the command completes successfully. The elapsed time is added to the cumulative counter.
WRITE_TIME_TAG	Nv_stats	"wl"	Indicates the cumulative write response time in milliseconds for each volume. The cumulative response time for volume writes is calculated by starting a timer when a SCSI write command is received and stopped when the command completes successfully. The elapsed time is added to the cumulative counter.
READ_HITCACHE_TAG	Nv_stats	"ctrh"	Indicates the number of reads received from other components and processed in read-cache
WRITE_HITCACHE_TAG	Nv_stats	"ctwfw"	Indicates the number of writes received from other components and processed in write-cache

The following code example shows an illustration of the calculations in a scripting form.

Request

```
def get_latest_cimom_sample_vdisk_stats(url, username, password):
    xml = get_latest_stats(url, username, password)

    fields = {
        "READ_OPERATION_TAG": {
            "filename": "Nv_stats",
            "element": "vdsk",
            "field": "ro",
            "value": 0},
        "WRITE_OPERATION_TAG": {
            "filename": "Nv_stats",
            "element": "vdsk",
            "field": "wo",
            "value": 0},
        "READ_BLOCK_TAG": {
            "filename": "Nv_stats",
            "element": "vdsk",
            "field": "rb",
            "value": 0},
        "WRITE_BLOCK_TAG": {
            "filename": "Nv_stats",
            "element": "vdsk",
            "field": "wb",
            "value": 0},
        "READ_TIME_TAG": {
            "filename": "Nv_stats",
            "element": "vdsk",
            "field": "rl",
            "value": 0},
        "WRITE_TIME_TAG": {
            "filename": "Nv_stats",
            "element": "vdsk",
            "field": "wl",
            "value": 0},
        "READ_HITCACHE_TAG": {
            "filename": "Nv_stats",
            "element": "vdsk",
            "field": "ctrh",
            "value": 0},
        "WRITE_HITCACHE_TAG": {
            "filename": "Nv_stats",
            "element": "vdsk",
            "field": "ctwfw",
            "value": 0}
    }
    populate_fields(xml, fields)
    cimom_fields = {
        'TotalIOs': fields['READ_OPERATION_TAG']['value'] + fields['WRITE_OPERATION_TAG']['value'],
        'KBytesTransferred': (fields['READ_BLOCK_TAG']['value'] + fields['WRITE_BLOCK_TAG']['value'])
* 1024 / 1024,
        'IOTimeCounter': fields['READ_TIME_TAG']['value'] + fields['WRITE_TIME_TAG']['value'],
        'ReadIOs': fields['READ_OPERATION_TAG']['value'],
        'ReadHitIOs': fields['READ_HITCACHE_TAG']['value'],
        'ReadIOTimeCounter': fields['READ_TIME_TAG']['value'],
        'KBytesRead': fields['READ_BLOCK_TAG']['value'] * 1024 / 1024,
        'WriteIOs': fields['WRITE_OPERATION_TAG']['value'],
        'WriteHitIOs': fields['WRITE_HITCACHE_TAG']['value'],
        'WriteIOTimeCounter': fields['WRITE_TIME_TAG']['value'],
        'KBytesWritten': fields['WRITE_BLOCK_TAG']['value'] * 1024 / 1024
    }
    return cimom_fields

pprint.pprint(get_latest_cimom_sample_vdisk_stats(url, username, password))
```

Response

```
{'IOTimeCounter': 0,
'KBytesRead': 0.0,
'KBytesTransferred': 1632.0,
'KBytesWritten': 1853.0,
'ReadHitIOs': 0,
'ReadIOTimeCounter': 3,
'ReadIOs': 178,
'TotalIOs': 178,
'WriteHitIOs': 54,
'WriteIOTimeCounter': 2,
'WriteIOs': 54}
```

Disk (drive) statistics

The following table lists:

- Performance statistics calculations accomplished by the CIM agent
- Respective fields in the statistics file that is reported for individual drives.

Table 117. Calculations

CIMOM field	Calculation
deviceId	DEVICE_ID_TAG
TotalIOs	READ_OPERATION_TAG + WRITE_OPERATION_TAG
KBytesTransferred(sizeUnits)	(READ_BLOCK_TAG + WRITE_BLOCK_TAG)*sizeUnits/1024
IOTimeCounter	READ_TIME_TAG + WRITE_TIME_TAG
ReadIOs	READ_OPERATION_TAG
ReadIOTimeCounter	READ_TIME_TAG
KBytesRead(sizeUnits)	READ_BLOCK_TAG*sizeUnits/1024
WriteIOs	WRITE_OPERATION_TAG
WriteIOTimeCounter	WRITE_TIME_TAG
KBytesWritten(sizeUnits)	WRITE_BLOCK_TAG*sizeUnits/1024

Var	File name	XML field	Description
DEVICE_ID_TAG	Nd_stats	"idx"	Indicates the name of the MDisk for which the statistics apply
READ_OPERATION_TAG	Nd_stats	"ro"	Indicates the cumulative number of MDisk read operations that were processed (since the node started)
WRITE_OPERATION_TAG	Nd_stats	"wo"	Indicates the cumulative number of MDisk write operations that were processed (since the node started)
READ_BLOCK_TAG	Nd_stats	"rb"	Indicates the cumulative number of blocks of data that was read (since the node started).
WRITE_BLOCK_TAG	Nd_stats	"wb"	Indicates the cumulative number of blocks of data written (since the node started)
READ_TIME_TAG	Nd_stats	"re"	Indicates the cumulative read external response time in milliseconds

Var	File name	XML field	Description
			for each MDisk. The cumulative response time for disk reads is calculated by starting a timer when a SCSI read command is issued and stopped when the command completes successfully. The elapsed time is added to the cumulative counter
WRITE_TIME_TAG	Nd_stats	"we"	Indicates the cumulative write external response time in milliseconds for each MDisk. The cumulative response time for disk writes is calculated by starting a timer when a SCSI write command is issued and stopped when the command completes successfully. The elapsed time is added to the cumulative counter

The following code example shows an illustration of the calculations in a scripting form.

Request

```
def get_latest_cimom_sample_drive_stats(url, username, password):
    xml = get_latest_stats(url, username, password)

    fields = {
        "READ_OPERATION_TAG": {
            "filename": "Nd_stats",
            "element": "mdsk",
            "field": "ro",
            "value": 0},
        "WRITE_OPERATION_TAG": {
            "filename": "Nd_stats",
            "element": "mdsk",
            "field": "wo",
            "value": 0},
        "READ_BLOCK_TAG": {
            "filename": "Nd_stats",
            "element": "mdsk",
            "field": "rb",
            "value": 0},
        "WRITE_BLOCK_TAG": {
            "filename": "Nd_stats",
            "element": "mdsk",
            "field": "wb",
            "value": 0},
        "READ_TIME_TAG": {
            "filename": "Nd_stats",
            "element": "mdsk",
            "field": "re",
            "value": 0},
        "WRITE_TIME_TAG": {
            "filename": "Nd_stats",
            "element": "mdsk",
            "field": "we",
```

```
            "value": 0}
    }
    populate_fields(xml, fields)
    cimom_fields = {
        'TotalIOs': fields['READ_OPERATION_TAG']['value'] + fields['WRITE_OPERATION_TAG']['value'],
        'KBytesTransferred': (fields['READ_BLOCK_TAG']['value'] + fields['WRITE_BLOCK_TAG']['value'])
* 1024 / 1024,
        'IOTimeCounter': fields['READ_TIME_TAG']['value'] + fields['WRITE_TIME_TAG']['value'],
        'ReadIOs': fields['READ_OPERATION_TAG']['value'],
        'ReadIOTimeCounter': fields['READ_TIME_TAG']['value'],
        'KBytesRead': fields['READ_BLOCK_TAG']['value'] * 1024 / 1024,
        'WriteIOs': fields['WRITE_OPERATION_TAG']['value'],
        'WriteIOTimeCounter': fields['WRITE_TIME_TAG']['value'],
        'KBytesWritten': fields['WRITE_BLOCK_TAG']['value'] * 1024 / 1024

    }
    return cimom_fields

pprint.pprint(get_latest_cimom_sample_drive_stats(url, username, password))
```

Response

```
{'IOTimeCounter': 409969,
'KBytesRead': 356062720.0,
'KBytesTransferred': 356064872.0,
'KBytesWritten': 2152.0,
'ReadIOTimeCounter': 409940,
'ReadIOs': 695437,
'TotalIOs': 695451,
'WriteIOTimeCounter': 29,
'WriteIOs': 14}
```

Fibre Channel (FC) statistics

The following table lists:

- Performance statistics calculations accomplished by the CIM agent
- Respective fields in the statistics file that is reported for individual FC ports.

Table 118. Calculations

CIMOM field	**Calculation**
portType	PORT_TYPE_TAG
deviceId	PORT_FC_WWPN_TAG
portId	PORT_ID_TAG
nodeId	NODE_ID
LinkFailures	LINK_FAILURES
LossOfSignalCounter	LOSS_SIGNAL
LossOfSyncCounter	LOSS_SYNC
PrimitiveSeqProtocolErrCount	PRIMITIVE_SEQ_PROTOCOL_ERRCOUNT
CRCErrors	INVALID_CRC
InvalidTransmissionWords	INVALID_WORD
BBCreditZeroTime	ZERO_BUFFER_CREDIT_COUNTER
writeBytesTransmitted	BYTES_TO_CONTROLLER + BYTES_TO_CLUSTERNODE + BYTES_TO_HOST + BYTES_TO_OTHER_CLUSTERNODE

Table 118. Calculations (continued)

CIMOM field	Calculation
BytesReceived	BYTES_FROM_CLUSTERNODE + BYTES_FROM_CONTROLLER + BYTES_FROM_HOST + BYTES_FROM_OTHER_CLUSTERNODE
DataTransmissionDelayCount	DATA_TRANS_DELAY_COUNT
DataTransmissionDelayTime	DATA_TRANS_DELAY_TIME
DataTransmissionDelayMeasured	DATA_TRANS_DELAY_MEASURED

Var	File name	XML field	Description
PORT_ID_TAG	Nn_stats	"id"	Indicates the port identifier for the node
PORT_TYPE_TAG	Nn_stats	"type"	
PORT_FC_WWPN_TAG	Nn_stats	"wwpn"	Indicates the worldwide port name for the node
LINK_FAILURES	Nn_stats	"lf"	Indicates a link failure count
LOSS_SYNC	Nn_stats	"lsy"	Indicates the loss-of-synchronization count
LOSS_SIGNAL	Nn_stats	"lsi"	Indicates the lost-of-signal count
PRIMITIVE_SEQ_PROTOCOL_ERRCOUNT	Nn_stats	"pspe"	Indicates the primitive sequence-protocol error count
INVALID_WORD	Nn_stats	"itw"	Indicates the number of transmission word counts that are not valid
INVALID_CRC	Nn_stats	"icrc"	Indicates the number of CRC that is not valid
ZERO_BUFFER_CREDIT_COUNTER	Nn_stats	"bbcz"	Indicates the total time in microseconds for which the buffer credit counter was at zero. This statistic is only reported by 8 Gbps Fibre Channel ports. For other port types, this statistic is 0
BYTES_FROM_CLUSTERNODE	Nn_stats	"lnbr"	Indicates the bytes received to other nodes in the same cluster
BYTES_FROM_CONTROLLER	Nn_stats	"cbr"	Indicates the bytes received from controllers
BYTES_FROM_HOST	Nn_stats	"hbr"	Indicates the bytes received from hosts

Var	File name	XML field	Description
BYTES_FROM_OTHER_CLUSTERNODE	Nn_stats	"rmbr"	Indicates the bytes received to other nodes in the other clusters
BYTES_TO_CLUSTERNODE	Nn_stats	"lnbt"	Indicates the bytes transmitted to other nodes in the same cluster
BYTES_TO_HOST	Nn_stats	"hbt"	Indicates the bytes transmitted to hosts
BYTES_TO_CONTROLLER	Nn_stats	"cbt"	Indicates the bytes transmitted to disk controllers
BYTES_TO_OTHER_CLUSTERNODE	Nn_stats	"rmbt"	Indicates the bytes transmitted to other nodes in the other clusters
DATA_TRANS_DELAY_MEASURED	Nn_stats	"dtdm"	Indicates the number of transfers that had their data transmission delay measured
DATA_TRANS_DELAY_COUNT	Nn_stats	"dtdc"	Indicates the number of transfers that experienced excessive data transmission delay
DATA_TRANS_DELAY_TIME	Nn_stats	"dtdt"	Indicates the total time in microseconds for which data transmission was excessively delayed

The following code example shows an illustration of the calculations in a scripting form.

Request

```
def get_latest_cimom_sample_fc_port_stats(url, username, password):
    xml = get_latest_stats(url, username, password)

    fields = {
        "LINK_FAILURES": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "lf",
            "value": 0},
        "LOSS_SYNC": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "lsy",
            "value": 0},
        "LOSS_SIGNAL": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "lsi",
            "value": 0},
        "PRIMITIVE_SEQ_PROTOCOL_ERRCOUNT": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "pspe",
            "value": 0},
        "INVALID_WORD": {
            "filename": "Nn_stats",
            "element": "port",
```

```
            "field": "itw",
            "value": 0},
        "INVALID_CRC": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "icrc",
            "value": 0},
        "ZERO_BUFFER_CREDIT_COUNTER": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "bbcz",
            "value": 0},
        "BYTES_FROM_CLUSTERNODE": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "lnbr",
            "value": 0},
        "BYTES_FROM_CONTROLLER": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "cbr",
            "value": 0},
        "BYTES_FROM_HOST": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "hbr",
            "value": 0},
        "BYTES_FROM_OTHER_CLUSTERNODE": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "rmbr",
            "value": 0},
        "BYTES_TO_CLUSTERNODE": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "lnbt",
            "value": 0},
        "BYTES_TO_HOST": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "hbt",
            "value": 0},
        "BYTES_TO_CONTROLLER": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "cbt",
            "value": 0},
        "BYTES_TO_OTHER_CLUSTERNODE": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "rmbt",
            "value": 0},
        "DATA_TRANS_DELAY_MEASURED": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "dtdm",
            "value": 0},
        "DATA_TRANS_DELAY_COUNT": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "dtdc",
            "value": 0},
        "DATA_TRANS_DELAY_TIME": {
            "filename": "Nn_stats",
            "element": "port",
            "field": "dtdt",
            "value": 0}
    }
    populate_fields(xml, fields)
    cimom_fields = {
        'LinkFailures': fields['LINK_FAILURES']['value'],
        'LossOfSignalCounter': fields['LOSS_SIGNAL']['value'],
        'LossOfSyncCounter': fields['LOSS_SYNC']['value'],
        'PrimitiveSeqProtocolErrCount': fields['PRIMITIVE_SEQ_PROTOCOL_ERRCOUNT']['value'],
        'CRCErrors': fields['INVALID_CRC']['value'],
        'InvalidTransmissionWords': fields['INVALID_WORD']['value'],
        'BBCreditZeroTime': fields['ZERO_BUFFER_CREDIT_COUNTER']['value'],
        'writeBytesTransmitted': fields['BYTES_TO_CONTROLLER']['value'] +
fields['BYTES_TO_CLUSTERNODE']['value'] + fields['BYTES_TO_HOST']['value'] +
fields['BYTES_TO_OTHER_CLUSTERNODE']['value'],
        'BytesReceived': fields['BYTES_FROM_CLUSTERNODE']['value'] + fields['BYTES_FROM_CONTROLLER']
['value'] + fields['BYTES_FROM_HOST']['value'] + fields['BYTES_FROM_OTHER_CLUSTERNODE']
```

```
['value'],
        'DataTransmissionDelayCount': fields['DATA_TRANS_DELAY_COUNT']['value'],
        'DataTransmissionDelayTime': fields['DATA_TRANS_DELAY_TIME']['value'],
        'DataTransmissionDelayMeasured': fields['DATA_TRANS_DELAY_MEASURED']['value']
    }
    return cimom_fields

pprint.pprint(get_latest_cimom_sample_fc_port_stats(url, username, password))
```

Response

```
{'BBCreditZeroTime': 0,
'BytesReceived': 1710209,
'CRCErrors': 0,
'DataTransmissionDelayCount': 32,
'DataTransmissionDelayMeasured': 20098258,
'DataTransmissionDelayTime': 998,
'InvalidTransmissionWords': 7,
'LinkFailures': 6,
'LossOfSignalCounter': 0,
'LossOfSyncCounter': 6,
'PrimitiveSeqProtocolErrCount': 0,
'writeBytesTransmitted': 628974}
```

Chapter 25. Conventions used in this IBM Documentation

Ensure that you are familiar with the notational conventions that are used in this IBM Documentation.

Notifications

This topic describes the meaning of notices and how to search, synchronize, and navigate by keyboard within the IBM Documentation.

Notices

The following notices are used in the product documentation and convey the following meanings:

Note: These notices provide important tips, guidance, or advice.



Attention: These notices indicate possible damage to programs, devices, or data. An attention notice is displayed before the instruction or situation in which damage could occur.



CAUTION: These notices indicate situations that can be potentially hazardous to you. A caution notice is displayed before the description of a potentially hazardous procedural step or situation.



DANGER: These notices indicate situations that can be potentially lethal or extremely hazardous to you. A danger notice is displayed before a description of a potentially lethal or extremely hazardous procedural step or situation.

Emphasis

Different typefaces are used in this guide to show emphasis.

The following typefaces are used to show emphasis.

Emphasis	Meaning
Boldface	Text in **boldface** represents menu items.
`Bold monospace`	Text in **`bold monospace`** represents command names.
Italics	Text in *italics* is used to emphasize a word. In command syntax, it is used for variables for which you supply actual values, such as a default directory or the name of a system.
`Monospace`	Text in `monospace` identifies the data or commands that you type, samples of command output, examples of program code or messages from the system, or names of command flags, parameters, arguments, and name-value pairs.

Numbering conventions

A specific numbering convention is used for this product.

The following numbering conventions are used:

- 1 kilobyte (KB) is equal to 1024 bytes.
- 1 megabyte (MB) is equal to 1 048 576 bytes.

© Copyright IBM Corp. 2024

- 1 gigabyte (GB) is equal to 1 073 741 824 bytes.
- 1 terabyte (TB) is equal to 1 099 511 627 776 bytes.
- 1 petabyte (PB) is equal to 1 125 899 906 842 624 bytes.

Note: The management GUI uses the same numerical values, but displays those values in the KiB, MiB, GiB, TiB, or PiB style of presentation.

