IBM TotalStorage SAN Volume Controller

Planning Guide

Version 1.2.1
Errata January 24, 2005

GA22-1052-03 Errata

Contents

About this guide

This guide provides errata information that pertains to release 1.2.1 of the IBM TotalStorage SAN Volume Controller Planning Guide.

This guide contains the corrections and additions on a per chapter basis. The chapter numbers in this guide correspond directly with the chapter numbers in the Planning Guide supplied with your SAN Volume Controller.

Who should use this guide

Before using the IBM TotalStorage SAN Volume Controller, you should review the errata contained within this guide and note the details with respect to the copy of the Planning Guide supplied with you SAN Volume Controller.

Last Update

This document was last updated: January 24, 2005

Chapter 4. Planning guidelines for using your SAN Volume Controller in a SAN environment

The following addition should be noted.

Switch zoning for the SAN Volume Controller

Page 37 Insert this chapter after Switch operations over long distances

Operation of an SAN Volume Controller cluster in SAN fabrics with long distance fibre links

An SVC cluster may be connected, via the SAN fabric switches, to application hosts, storage controllers or other SVC clusters, via short wave or long wave optical fibre channel connections with a distance of up to 300m (short wave) or 10 km (long wave) between the cluster and the host, other clusters and the storage controller. Longer distances are supported between SVC clusters when using inter cluster PPRC.

1. A cluster should be regarded as a single entity for disaster recovery purposes. This includes the backend storage that is providing the quorum disks for that cluster. This means that the cluster and the quorum disks should be co-located. Locating the components of a single cluster in different physical locations for the purpose of disaster recovery is not recommended, as this may lead to issues over maintenance, service and quorum disk management, as described below.
2. All nodes in a cluster should be located close to one another, within the same set of racks and within the same room. There may be a large optical distance between the nodes in the same cluster. However, they must be physically co-located for convenience of service and maintenance.
3. All nodes in a cluster must be on the same IP subnet. This is because the nodes in the cluster must be able to assume the same cluster or service IP address.
4. A node must be in the same rack as the UPS from which it is supplied.

Whilst splitting a single cluster into two physical locations might appear attractive for disaster recovery purposes, there are a number of practical difficulties with this approach. These difficulties, which do not apply in the case of the standard, two cluster solution, largely arise over the difficulty of managing a single quorum disk in a cluster that is distributed over two different physical locations. Consider the following configuration:

```
  Site 1                              Site 2
 -----------          5km           -----------
|  Node A1  |----FC switches----|  Node A2  |
 -----------                        -----------
MDisk group X                     MDisk group Y
```

Nodes A1 and A2 form an I/O group

There is a node from each I/O group at both sites with remote copy relationships set up so that the primary VDisks at site 1 come from MDisks in group X (i.e. MDisks at site 1) and the secondary VDisks at site 2 come from MDisks in group Y (i.e. MDisks at site 2). It would appear that this arrangement will provide a means of recovering from a disaster at one or other site i.e. if site 1 fails, there will be a live I/O group (albeit it in degraded mode) at site 2 to perform the I/O workload. There are however a number of issues with this arrangement:

1. If either site fails, there is a degraded I/O group at the other site with which to continue I/O. Performance therefore during a disaster recovery is significantly impacted, since throughput of the cluster is reduced and the cluster caching is disabled.
2. The disaster recovery solution is asymmetric. Thus, it is not possible to run applications on both sites and allow either to suffer a failure. One site must be regarded as the primary site and the other is there to provide a recovery site. Consider the situation where the quorum disk is at site 2 (i.e. in MDisk group Y). If site 1 fails, then site 2 retains quorum and can proceed and act as a disaster recovery site. However, if site 2 were to fail, then site 1 cannot act as a disaster recovery site, since site 1 will only see half the nodes in the cluster and will not be able to see the quorum disk. The cluster components at site 1 will no longer form an active cluster (error code 550). It is not possible to communicate with the nodes at site 1 in this state and all I/O will immediately cease. An active cluster can only start operating at site 1 if the quorum disk re-appears or if a node from site 2 becomes visible. And in that case, it is likely, or at least possible, that site 2 might be able to resume operations anyway.

From the discussion above, it can be seen that the split cluster configuration can only provide asymmetric disaster recovery facilities, with substantially reduced performance. This is unlikely to be satisfactory for most production disaster recovery situations.

Splitting a cluster might be thought to be useful if the quorum disks are at a third "site", such that a disaster will only take down one of these three sites. However, even a three site configuration will have significant limitations, since SVC will not be able to change the path it uses to communicate with a quorum disk under all circumstances. Therefore, to be tolerant of a single site failure, it is necessary to ensure that the path to the quorum disk from a node in one site does not go through a switch in the second site before reaching the quorum disk in the third site as shown in the diagram below.

```
  Site 1                         Site 2
 -----------         5km        -----------
|  Node A1  |----FC switches----|  Node A2  |
 -----------          |         -----------
 -----------          |         -----------
|  Node B1  |         |        |  Node B2  |
 -----------          |         -----------
MDisk group X    Quorum Disk    MDisk group Y
                   Site 3
```

A1/A2 are one I/O group and B1/B2 are the other

Quorum device is at a third site

IBM recommends that split site operation is achieved by using a separate cluster at each site. For this reason, IBM does not test all possible failure modes within a split cluster configuration, as the characteristics of these configurations are unlikely to provide satisfactory disaster recovery capabilities in all failure modes. Customers who wish to use split cluster operation should take this it account when deciding if split cluster operation will meet their needs.

Alternatively, the following configuration is unlikely to perform satisfactorily.

```
  Site 1            5km          Site 2
  FC switch -------------------- FC switch
      |-----------         ----------|
 -----------      \       /      -----------
|  Node A1  |      \     /      |  Node A2  |
 -----------        \   /        -----------
 -----------         | |         -----------
|  Node B1  |        | |        |  Node B2  |
 -----------         | |         -----------
MDisk group X    Quorum Disk    MDisk group Y
                   Site 3
```

A1/A2 are one i/o group and B1/B2 are the other

Nodes at site 1 are connected to one switch

Nodes at site 2 are connected to another switch

Quorum device is at a third site and is connected to both sites

Note: All of the above discussion applies to a split cluster where there is significant optical distance between the nodes within a single cluster. Long distance (up to 10 km) connection of remote hosts or remote controllers are supported in an SVC cluster, as the issues of quorum disk and inter I/O group links mentioned above are not relevant. Thus, the following configuration is acceptable:

```
  Site 1                         Site 2
 -----------         10km        -------------------
|  Node A1  |----FC switches----| Application hosts |
 -----------                     -------------------
 -----------
|  Node A2  |
 -----------
```

IBM recommends the use of two cluster configurations for all production disaster recovery systems. Customer who wish to use split cluster operation should contact their IBM Regional Advanced Technical Specialist for specialist advice relating to their particular circumstances.

Chapter 6. Planning for configuring the SAN Volume Controller

The following corrections should be noted.

Maximum configuration

Pages 69 and 70Corrections to Copy Services Properties in Table 18

Table 18: SAN Volume Controller maximum configuration values

Objects	**Maximum number**	**Comments**
Copy Services Properties		
Remote Copy Relationships	1024	– – –
Remote Copy Consistency groups	256	– – –
Remote Copy VDisk per I/O group	16 TB	– – –
FlashCopy mappings	2048 (See Note.)	– – –
FlashCopy consistency groups	128	– – –
FlashCopy VDisk per I/O group	16 TB	– – –
Note: SAN Volume Controller supports up to 512 FlashCopy mappings per consistency group.		