IBM's System Storage™ TS7680 ProtecTIER® Deduplication Gateway for System z®: A mainframe powerhouse

By Deni Connor
Principal Analyst, Storage Strategies NOW
February 2010

The need to extend processing windows, allow for rapidly increasing data growth and provide higher reliability top the list of data protection challenges that IT managers are facing. A new storage system from IBM solves these issues with a system designed to simplify tape processing operation and improve tape application performance with high-speed inline data deduplication.

The TS7680 ProtecTIER Deduplication Gateway

The IBM System Storage TS7680 ProtecTIER Deduplication Gateway for System Z is designed to combine high-performance computing with inline data deduplication. The system scales to as much as 1PB of physical storage capacity in a highly available clustered configuration that provides redundancy through automatic failover.

The TS7680 ProtecTIER Deduplication Gateway delivers the ease of management, performance and scalability required to meet the data storage needs of today's data centers. By reducing storage infrastructure costs and providing inline data deduplication and simplified storage management, the TS7680 ProtecTIER can store up to 25 times more data to disk, resulting in reduced off-peak production batch windows and allowing for increased data retention on disk.

At its roots, the IBM TS7680 is a gateway-based virtual tape library (VTL) used to perform backup to disk operations. ProtecTIER software runs on the TS7680 server emulating one or more physical tape drives and a single tape library. For existing tape infrastructures, there is no need to change existing software, policies or procedures. The TS7680 gateway repository resides within disk pools made up of Fibre Channel attached disk storage arrays.

In addition, although virtual tape libraries accept and process all tape operation commands, they are not hit with the delays associated with using real tape media such as robotic movements, media mounting, media positioning or media eject operations. In a tape-based environment, these delays can last from a few seconds to a few minutes per tape.

The TS7680 ProtecTIER for System Z can support up to 256 virtual drives with up to one million virtual volumes. The system provides high-end performance using the IBM System Storage DS8000, DS5000 or XIV System Storage disk storage systems but also has the ability to support third party disk systems unlike its competition.

The extreme flexibility of IBM TS7680-based Virtual Tape Library Solutions also means the number of virtual drives and the size of the virtual tape cartridges can be defined to provide optimal backup and restore performance for the amount of data being stored and the number of clients being backed up. So even faster restore speeds can be achieved since the IBM TS7680 ensures that the data is written sequentially to a single piece of virtual media.

Understanding deduplication

Data deduplication works by eliminating duplicate instances of identical data as the data is being ingested into the system (inline). Eliminating duplicate data reduces the storage footprint needed to store the data. The inline deduplication algorithm inside ProtecTIER is called HyperFactor®, which uses an industry first "all in-memory resident" index to keep track of duplicated data and to maintain a single repository for its global file system.

To better understand the concept of data deduplication technology, consider the example of a Microsoft Excel file attached to an e-mail. If the e-mail is sent to multiple recipients and then forwarded to yet another set of recipients, data deduplication technology can be used to store the spreadsheet only once. This is an illustration of data deduplication technology working at the file level. Next, let's consider what happens when one of the e-mail recipients modifies a few cells in the Excel file and forwards it to a group of colleagues. Data deduplication at the byte level, like those incorporated by ProtecTIER's HyperFactor, can be used to store only the new, unique data associated with the changed Excel file as shown in Figure 1.
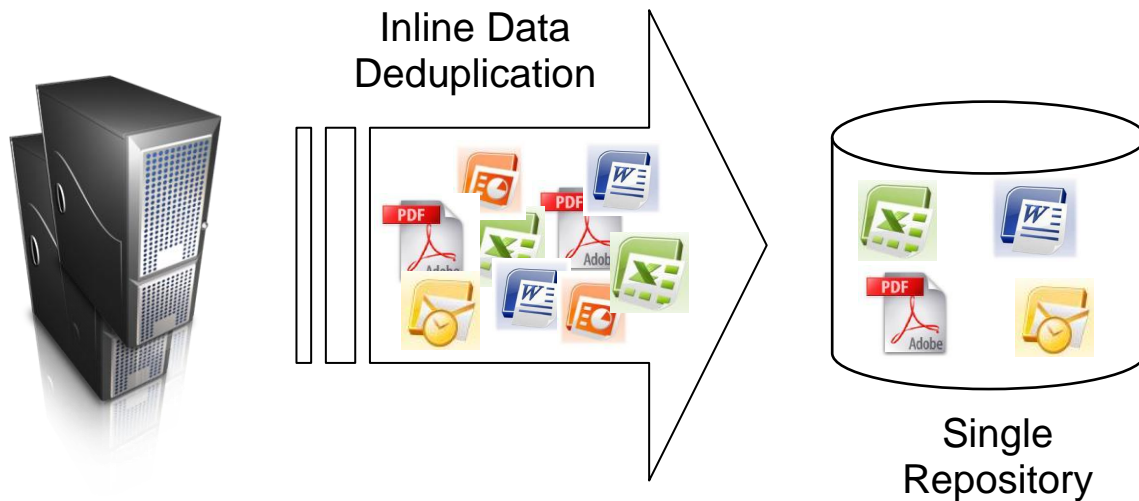
Figure 1. ProtecTIER's HyperFactor Inline Data Deduplication

Disk capacity savings that can be achieved with data deduplication varies according to a number of factors. Factors include policies in use, retention policy and the rate of change within the data. The ratio of capacity ingested versus capacity stored on disk after duplication is generally known as the data deduplication factor. A deduplication ratio of 10:1 reduces disk storage capacity requirements by 90%.

Planned advanced disaster recovery will feature native replication functionality that will electronically transmit data to remote locations for rapid cost-effective disaster recovery. With IBM advanced disaster recovery, only unique deduplicated data will be transmitted to the DR site. The benefit of advanced disaster recovery will greatly reduce network bandwidth, eliminate the need to transport tape media offsite and lower overall disaster recovery costs. Since the TS7680 greatly reduces the amount of physical storage needed by a factor of up to 25 times or more, the major cost of the WAN bandwidth required is also significantly decreased. The TS7680 is fully integrated with z/OS. The virtual tape library appears to the mainframe host as an automated tape library with standard 3592 Model J1A devices in 3590 emulation mode. No host application or tape management changes are required.

SSG-NOW Assessment

Enterprise-class data protection challenges are driving IT administrators to adopt a number of technologies including backup to disk, virtual tape libraries and data deduplication solutions. The IBM System Storage TS7680 ProtecTIER Deduplication Gateway for System Z is an enterprise-class disk based virtual tape solution that enables data centers to effectively reduce the required storage footprint by built-in data deduplication, up to 25x. ProtecTIER, powered by HyperFactor, eliminates data redundancy as data is being ingested into the system – without performance, capacity or data integrity compromise.

With this powerful combined solution, IT organizations no longer need to choose between optimizing for backups and optimizing for restores. The IBM TS7680 makes the entire mainframe based tape processing environment faster, simpler, more reliable, verifiable and more cost-effective, enabling administrators to meet increasingly stringent service levels and reduce overall costs.