Data Mobility Group

datamobilitygroup.com

White Paper
April 26, 2005
Dianne McAdam

Data Protection—
Putting the 'T' in 'D2D2T'

 76 NorthEastern Blvd. Suite 29A Nashua, NH 03062 Phone: 877.463.6456 Fax: 603.563.8282

While the decision to protect your data is not a choice, as all data must be protected, the strategy you choose to protect your data should be given thoughtful consideration.

Every company has numerous applications that support business functions. These applications range from the highly critical, such as customer support or order processing, that must be available 24 by 7, to those, such as human resources, that are necessary but do not have the same stringent availability requirements.

Thus, choosing the right data protection strategy and technology for a particular data type is not an easy one-size-fits-all decision. Each application must be carefully evaluated to determine the data protection solution that best addresses the business availability and recovery requirements while keeping costs under control.

Tape has been a central fixture in corporate data centers for the last 50 years, serving as the primary data backup-and-restore medium. However, shrinking backup windows, the need to recover critical applications more quickly, and declining disk costs have led to the increased consideration of disk based data protection. Vendors are responding to customers business pains related to shrinking backup windows and more stringent restore requirements through developing and delivering disk-to-disk-based solutions to the market place. While D2D solutions can help reduce the backup window and provide improved restore times, one must realize that not every application should use this type of strategy. Protecting all data with D2D solutions does have a very expensive price to pay as the TCO is considerably higher than tape based solutions.

While disk has received a great deal of attention lately as a backup-and-restore medium, it should not be viewed as the only way to backup data. Rather, it should often be viewed as adding an additional level of data protection to the existing backup infrastructure. D2D should be considered as part of a larger backup strategy of primary-disk-to-secondary-disk-to-tape, or D2D2T. While disk can be used to store the most current versions of backups or for caching, tape continues to maintain its place in the data center both for retention of long-term backups and for archival storage. Today's environment demands multi-level data protection.

Matching Business Requirements to Data Protection

Data is one of the corporation's most important assets. Many customers are reporting data growth rates that exceed 50 percent each year and this trend is expected to continue. The loss or unavailability of data that drives revenue-generating applications harms a company's bottom line. Protecting this growing amount of data, while providing high levels of availability within budget constraints, can be a challenging task.

Protecting data through a tape-based backup process has been a standard part of IT operations for years. Tape backups are used to restore data that has been corrupted or accidentally deleted. Traditional tape backup is not without its flaws. These backups are usually executed nightly and if corruption occurs minutes prior to the next backup cycle, the restored data can be 24 hours old. Furthermore, restoring large databases that span several volumes can take hours depending on what tape technology is used. These databases frequently contain an enterprise's most critical data.

Two major metrics are used to quantify acceptable levels of the backup/restore process—the measurable goals to which the process aspires. The Recovery Time Objective (RTO) is a goal for how long the restore process will take. The Recovery Point Objective (RPO) is a goal for the maximum age of the data which will be used for a restore operation.

Low-priority applications, such as Human Resources, may have an RTO and RPO of 24 hours. Here, tape backups can easily satisfy these objectives. On the other hand, online transaction applications, such as order processing, typically have more stringent business objectives. RTO and RPO for these applications are defined in terms of hours or minutes, not days. Tape backups must be augmented to meet these business objectives.

Disk-to-Disk Technologies

Storage vendors have responded to the demand for shorter RTO and RPO by developing disk-to-disk backup/restore solutions. Backup applications write directly to disk, rather than to tape, in an effort to speed up the backup and, more importantly, the restore process.

Tape is—and will continue to be—a perfect, and most cost effective technology for reading and writing large streams of data, such as those from satellite feeds, due to its streaming, sequential nature. But legacy tape does not typically perform nearly as well when dealing with small blocks of data. Backup applications can not always stream data to tape. The result is that tape backups can take too long. Disk systems, on the other hand, perform well when presented with small blocks of data.

Storage vendors have designed disk-to-disk backup/restore solutions in very different ways. With some D2D implementations, disk emulates tape and appears to the operating system as if it were a physical tape drive. With other implementations, disk does not masquerade as tape, but responds to the operating system as disk.

Disk As Tape, or "Virtual Tape"

The use of disk in the guise of a tape drive dates back to the late 1990s, when IBM and StorageTek introduced the concept of "virtual tape" for mainframes. In these implementations, disk subsystems were placed in the backup stream between tape libraries and mainframes. In reality, virtual tape solutions were the first "D2D2T" type solutions.

A disk buffer appeared to the operating system as a tape drive and responded to the operating system's standard tape commands, hence the term "virtual tape." These virtual tape systems were not stand-alone disk systems; they were tightly coupled disk caches in front of tape libraries. Data was written to disk first, then to tape later as the disk cache filled up.

This disk-emulating-tape concept has now been applied to the open systems market. In these implementations, the target device for backups is usually an ATA-based disk array that responds to software commands just as if it were a tape drive. Some integrated disk/tape implementations are similar to the original mainframe solutions. Some vendors use software that works only with their own disk subsystems while other vendors use host-based software that works with any vendor's disk array. Many of these implementations are not integrated with tape drives and libraries. In fact, several of these solutions do not require tape at all. Eliminating tape from these solutions, however, introduces risk to the organization and is an incomplete solution. Tape provides a second level of data protection against the failure of the disk-based appliance.

Customers implementing software-only and disk/software solutions can, however, continue to use existing tape drives and libraries. The backup application that initially creates the disk backup image can be used later to copy the image to physical tape for longer-term storage or offsite disaster recovery. The cartridges remain under the control of the backup application. One caveat: This extra step consumes additional I/O bandwidth and backup server cycles.

With integrated disk/tape solutions, sometimes called true virtual tape solutions, the capacity of the disk and tape layers is not presented separately but as one logical pool. The backup application writes to disk without actually knowing where the backup copy resides (i.e., on disk or tape). The virtual tape software keeps track of the physical location of the backed up data.

Integrated disk/tape solutions can automatically migrate disk backup images to tape as business requirements demand, independent of the backup application. In these cases, the physical cartridges remain under the control of the virtual tape, not the backup, application. This approach does not require the additional bandwidth and backup server cycles that the non-integrated approaches require.

Implementing virtual tape generally requires few changes to an existing backup infrastructure. Usually, the disk array and software can easily be plugged into existing backup products such as Atempo's TimeNavigator, CommVault's Galaxy, Computer Associates' ArcServe, Legato's Networker, and Veritas's NetBackup. The normal process of directing the backup stream to tape is simply redirected to disk.

The advantages of disk-as-tape solutions are many. Backups on disk can be restored more quickly than tape-resident backups, improving RTO. And since few changes are required of the backup applications, these solutions can be implemented more quickly. Integrated disk/tape solutions view tape and disk resources as one large pool, simplifying management and providing an automated mechanism to direct older backups to more cost-effective tape.

In summary, virtual tape can be very effective and complements tape by providing the necessary two levels of data protection, with the primary backup copy stored on disk and the secondary copy on tape.

Disk-As-Disk

Unlike disk-as-tape systems, disk-as-disk systems do not emulate tape. Backed up data is written in disk, not tape, format. These products are typically not integrated with drives and libraries, so additional processes or steps are required to copy the disk image to tape.

Disk-as-disk backup/restore has been implemented in several ways. One approach uses traditional snapshot or copy technology to send a copy of data to a secondary disk array. The snapshot process is invoked again to restore data. Other approaches provide integrated hardware and software appliances that can be targets for backup within a storage area network (SAN) or over a LAN. Some vendors have developed specialized appliances to address specific problems that have plagued IT administrators for years. For example, if the file system currently being backed up is corrupted, the backup will simply mirror the file system's corrupted state. One appliance runs consistency checks against the file systems to flush out integrity problems when the backup process finishes. Another appliance stores data as objects, then uses a filtering technology to find and eliminate redundant objects, reducing storage requirements. Other appliances store data in compressed format to save space.

Disk-as-disk backup solutions can be integrated into existing backup processes to improve the overall process. However, storage administrators must be aware of the implementation changes, such as those required within backup applications to change the target device from tape to disk. And disk-as-disk solutions require additional steps to clone the disk images to physical tape. Furthermore, disk-as-disk solutions are likely to be considerably more expensive than traditional tape backup, thus customers should think hard about the amount of data stored on this type of solution.

The Incompleteness of Disk-to-Disk Technology

D2D solutions are effective for reducing the backup window through the use of non-disruptive disk copy functions. They are also a key enabler of business continuity plans that call for rapid data recovery in the event of a failure. Nevertheless, D2D clearly doesn't replace tape completely. Tape still plays a very important role in the data center and complements a D2D backup solution.

D2D solutions are prone to the same vulnerabilities as any disk subsystem. As such, they should be RAID-protected and backed up to tape as well. Many D2D backup products reside in the same storage area network (SAN) as primary disk, rendering them ineffective for disaster recovery. A data center outage, such as a power failure, would make both the primary disk and the D2D backup inaccessible. Disaster recovery best practices dictate that a copy of data be stored at an offsite location. Tape is both removable for offsite storage and is the least expensive storage medium.

Database corruption can remain undetected for days or even weeks; therefore, recovery may require restoring data from a backup that is several days old. Some D2D products do not have sufficient capacity to hold weeks' worth of backups; adding extra disks to this solution also adds to the cost. A better solution is to write several days of backup to disk and then copy all older backups from disk to tape.

Ninety percent of data recovery operations occur within the first day; about 95% of data recovery operations occur within the first week.[1]Tape is the most cost-effective medium for storing infrequently-accessed data for long periods.

Data volumes are growing exponentially, and the capacity needed to back up these volumes is increasing at the same rate. Unlike disk products, which require additional controllers, new software, and more floor space as they scale, tape libraries can easily scale to accommodate data growth by simply adding tape cartridges. There has been a great amount of press coverage about the erosion in the cost of disk over the last few years, fueling speculation that disk will soon cost less than tape. However, increases in tape cartridge capacity now and in the future will ensure that tape remains the least expensive solution for storing large quantities of data by about an order of magnitude.

New regulations are forcing IT organizations to retain data for longer periods. HIPAA regulations, for example, require healthcare organizations to keep certain patient information for the life of the patient or longer. Since much of the data will never need to be retrieved, spinning all of it on disk for years is unnecessary and, more importantly, costly.

The Sarbanes-Oxley Act requires publicly-held companies to retain e-mail and other documents in an unalterable format for seven years. WORM tape not only prevents data tampering but also is cost-effective for long-term retention.

[1] Source: Horison Information Strategies

Designing a Complete Data Protection Solution

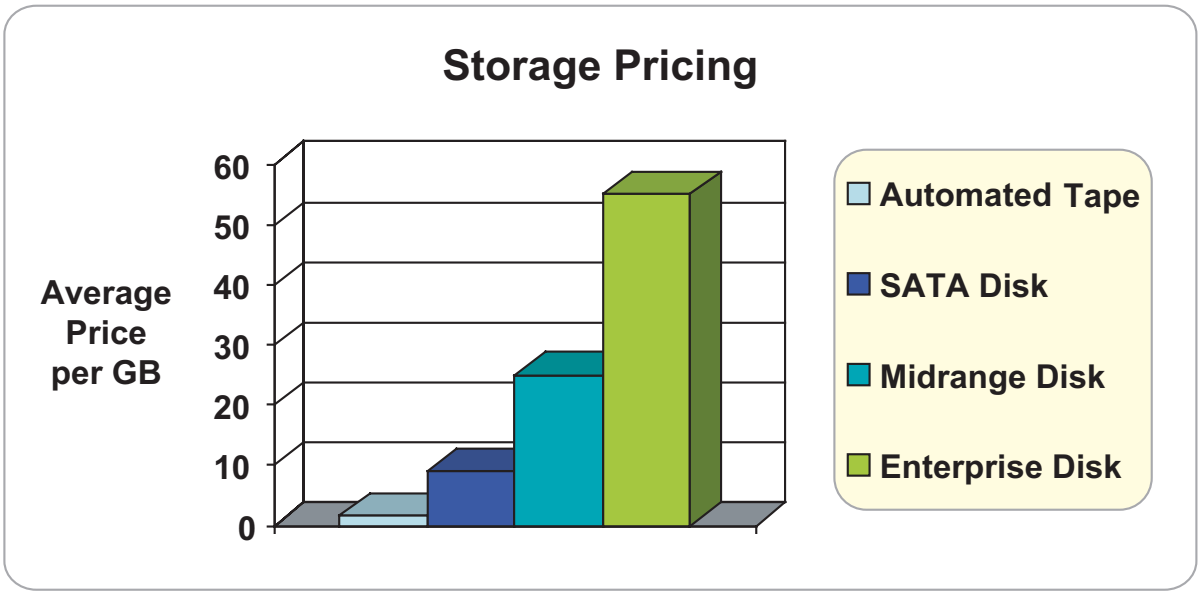
D2D solutions are certainly effective for reducing the backup and restore windows, and some have received considerable press coverage lately. But D2D solutions only offer one level of data protection. Enterprises today need to design a multi-tiered protection approach, and should match the RIGHT data to the RIGHT data protection strategy, whether that be virtual tape, disk-to-disk, D2D2T solutions, or traditional tape backup. This is not a new idea. IBM developed it in the 1980s for MVS users looking to migrate data off expensive high-performance disk onto less-expensive disk and later to tape. The company called this process hierarchical storage management (HSM). HSM saved customers money by integrating disk and tape with software that managed the migration of data based on established policies. Today's disk-to-disk-to-tape products for open-systems storage are an extension of this same concept.

The advantages of an integrated disk/tape solution are compelling. Adding tape to D2D solutions provides an additional level of protection, ensuring that additional backup copies are available if, for example, the disk-based system fails. Tapes that are sent offsite, either electronically or physically, provide recovery for a local data center disaster.

Integrated disk/tape virtual tape solutions provide greater flexibility when designing a data protection solution for the enterprise. Disk and tape resources are managed as one large pool, providing high levels of scalability. Unlike some disk-as-disk solutions that require constant monitoring to determine if there is enough capacity to store the next series of backups, additional tape cartridges can be added to the virtual tape library to accommodate growth while keeping costs under

control. Virtual tape policies can automatically migrate backups between disk and tape storage as dictated by business objectives, without intervention by IT staff or by backup applications. These policies can also automatically export tapes from the library to be sent offsite for disaster recovery protection.

Finally, tapes are still and will continue to be less expensive than disk storage for long-term retention. According to a recent study by Horison Information Strategies, enterprise class (FICON, Fibre Channel) disk costs about $40-$70 per GB, midrange (SCSI, Fibre Channel) disk about $15-$35 per GB, SATA disk about $3-$15 per GB, and automated tape about $.50 to $3.00 per GB. Disk capacities are expected to increase from 300-400 GB today to 4-5 TB by 2013. Tape capacity will increase at a greater amount—from 320-500 GB today to 6-8 TB in 2013.This predicted increase in capacity will allow tape to remain less expensive than disk in the future.



Source: Horison Information Strategies

Customers evaluating an integrated disk/tape solution need to consider cost, scalability, and performance. But they must also find the right vendor, one that will help them to design the right solution for their unique business requirements.

Disk and Tape—A Good Marriage

Though disk as a replacement for tape is central to many discussions about backup today, we argue that disk should not replace tape in the data center but complement it. Users should consider D2D as part of a larger backup strategy of primary disk-to-secondary-disk-to-tape (D2D2T). While disk can be used to store the most current versions of some application backups, tape continues to find value in the data center as a medium for the retention of long-term backups and for archival storage for many years, in some cases for decades.

Disk and tape make a good marriage. They're better together than apart.