Data Mobility Group

datamobilitygroup.com

White Paper
March 25, 2005
Dianne McAdam

Disk-to-Disk Backup—It's Also About Tape

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

Tape has been a central fixture in corporate data centers for the last 50 years, serving as the primary data backup-and-restore media. But shrinking backup windows, the need to recover critical applications quickly, the emergence of Serial ATA technology, and declining disk costs have combined to push tape out of the spotlight in favor of disk. After all, backing up directly to disk may speed up the backup process and also reduce restore time for individual files.

However, while disk has received a great deal of attention lately as the new backup-and-restore media, it should not be viewed as an all-out replacement for tape but as a complement to the existing backup process.

What Tape Does Better Than All Other Storage Mediums

Those who argue that users should replace all tape cartridges within their data centers with hard disk drives fail to remember the superior attributes of tape drives and media.

First, tape is removable. Backups can be directed to tape and later removed from the system to help prevent intentional or unintentional corruption from a virus or system error.

Second, tape is transportable. It can be taken to an offsite location without falling victim to the shock and vibration restrictions that can destroy a disk drive. This provides a second level of data protection. If the main data center suffers an outage, data can be restored from offsite tapes.

Third, tape media has been constantly improved since the appearance of round tape reels in the 1950s. Round tape reels were manually fed into tape drives that maintained tension through vacuum columns. Today's media are contained within cartridges safe from operator fingerprints and tension is more gently regulated within the drives. Early tape media based on iron oxide technology was subject to oxidation and high failure rates. New media makes use of magnetic particle technology commonly known as Metal Particulate or MP, a very reliable technology that improves the shelf life of tape to 30 years. This improvement in shelf life makes tape a very suitable media for both current and older backup copies. Regulatory requirements require companies to retain data for many years, in some cases for decades. This long shelf life makes tape an excellent choice for long term archival of data.

Fourth, tape is still the most cost effective storage media—not just in a $/GB comparison. Tape requires 1/10th the power consumption of a low-duty-cycle disk system (an important consideration for companies looking to contain soaring electrical costs); it takes up less floor space than an equivalent-capacity disk system; tape has infinite capacity: if you need more capacity you don't need to add more drives, simply add more cartridges. Therefore, its cost of ownership is—and will continue to be for the foreseeable future—significantly lower than disk alternatives.

Disk-to-Disk Backup

Declining disk costs have sparked the development of a variety of disk-to-disk (D2D) systems. Not all of these products are alike; in fact, in some respects they are very different. Let's look at the different options.

Disk as Tape (a.k.a. "Virtual Tape")

The idea of disk emulating tape is not a new concept. IBM first introduced the concept with its Virtual Tape Server (VTS), which integrated disk with tape libraries in 1997. Backups were first written to disk in tape image format and then in the background written to physical tape cartridges connected to the library. This can help improve job productivity as well as reduce the physical number of tape systems and cartridges required.

Similarly, the new breed of virtual tape products on the market use disk as tape devices—that is, data is backed up to a disk device, which appears as a tape drive or library to the host and backup application. Data is written in tape, not disk, image format.

Later, management software can be used to copy the backup to physical tape based on pre-set policies (e.g., an administrator can migrate all disk-based backups to tape

immediately, wait until the disk system fills up before migrating data to tape, etc.). When a restore is necessary, the software locates the backed up data (on disk or tape) and the restore process begins. The retrieval process is transparent to the user.

Two basic types of virtual tape products are available today: those that are integrated with tape drives and libraries and those that are not. Implementation of these systems is fairly simple. Administrators simply redirect the target device for the backup to the new disk system. Since the disk device appears to the backup application as a tape drive, no changes to backup procedures are generally required.

While standalone virtual tape products are available, integrated virtual tape products (i.e., those that are integrated with tape drives or tape libraries) provide a more complete solution. Backups are written to the disk system and then automatically migrated to tape at a later time requiring no additional processes to be established by the customer.

Disk as Disk

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

Approach #1: The user makes a copy (sometimes referred to as a 'snapshot') of a primary disk volume and then transmits a copy of that volume to a secondary disk system.

Approach #2: Secondary disk (usually higher-capacity ATA/SATA disk) serves as the target device for backups.

Approach #3: Specialized software/hardware appliances serve as backup targets and provide other options such as compression, continuous point-in-time copies of backups, etc.

Which One Should You Choose?

Disk as tape or disk as disk? Which choice is the right one? Consider how well the solutions will integrate with your current environment. Disk as tape, or virtual tape, is easy to implement, and requires little or no changes to current backup processes. Virtual tape solutions that are integrated with tape drives and libraries automatically migrate disk-resident tape images to tape cartridges.

Disk as disk solutions requires changing the target device from tape to disk and requires additional steps to clone the disk images to physical tape.

To determine which virtual tape approach is the best fit within an existing environment, storage administrators should ask vendors the following questions:

- Which tape libraries and drives (and how many) does the software emulate? Are these drives similar to what is already installed?
- What backup software is supported? Is that software currently installed, or will you be required to bring in additional software—and at what cost?
- For integrated hardware/software solutions, what types of disks are used? What level of RAID protection is available? What is the maximum capacity of the disks? What server connections are supported (SCSI, Fibre Channel, ESCON, FICON, iSCSI)?
- What is the throughput of the system?
- How long does it take to install the product?
- How well does it integrate with existing tape drives and libraries?
- How many different backups can be stored on virtual tape? How many versions of the same backup can be stored?

- Should the product be used for all backups or only for selective backups?
- How is the product managed? Is there a remote management facility? Is the management software included in the price of the product? How well does the management product interface with existing installed backup products?
- How easily are backups migrated from virtual tape to physical tape? Does this require manual intervention or can it be easily automated?

Choosing the most appropriate disk-as-disk solution is a more difficult task than choosing the best virtual tape solution for the simple reason that there are more options available. As a guide, users should ask vendors the following:

- Which operating systems does the system support?
- Is data backed up on a volume or file level, or both?
- What databases are supported, if any? Are specific agents available for applications such as Microsoft Exchange? How much do they cost?
- What is the minimum and maximum capacity of the system? Can you easily upgrade to larger capacities? What RAID levels are supported? What server connections are supported (SCSI, Fibre Channel, ESCON, FICON, iSCSI)? How many connections are supported (minimum and maximum number)?
- How long does it take to install the product?
- What is the throughput of the system?
- Does the solution use its own backup software or does it support existing backup software?
- How many backups, and how many versions of backups, can the configuration support?
- Will the product be used for all backups or just selected backups?
- When restore is required, is data restored to the original primary disk or can it be accessed on the secondary disk system? How granular is the restore process? Can data be restored to a specific point-in-time or is a full restore required?
- Is compression supported to reduce capacity requirements?
- What are the projected restore times?
- How is it managed? Can it be remotely managed?
- How easily can disk backups be migrated to tape? Is this a manual process or can it be automated?

The Problem with D2D

While disk-to-disk (D2D) backup has several benefits, it clearly doesn't replace tape completely. Tape still plays a very important role in the data center and complements a D2D backup solution. Consider the following:

Tape for DR

Since many D2D backup products reside in the same storage area network (SAN) as primary disk, in the event of a data center outage (e.g., a power failure), they too are inaccessible. Therefore, while D2D products can be an effective tool to restore a volume or file that has been corrupted or to assist in the recovery of a primary disk failure, they are often ineffective for disaster recovery. Disaster recovery best practices dictates that a copy of data be stored at an offsite location. Tape is both removable for offsite storage and it is inexpensive.

Tape for Retaining Many Versions of Backup

Database corruption can remain undetected for days; therefore, recovery may require restoring data from a backup that is several days old. Since some D2D products do not have sufficient capacity—or the cost would be unreasonable—to hold several days' worth of backups, a better solution might be to write the current backup to disk and then copy all older backups from disk to tape.

Tape for Scalability and Reduced TCO

Data volumes are growing exponentially, and the capacity needed to back up these volumes is increasing at the same rate. Can D2D products scale to support future backup requirements? 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. Therefore, tape offers what might be considered, "infinite capacity" and is important for user storage consolidation strategies. And though the per-GB cost of disk continues to decrease, innovation in tape technology continues to reduce the per-GB cost of tape and helps it remain the least expensive solution for storing large quantities of data by about an order of magnitude.

Tape for Long-term Retention and Compliance

New regulations are forcing IT organizations to retain data for longer periods of time. 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. Tape is the most cost-effective media for storing infrequently-accessed data for long periods of time.

Also, other new regulatory legislation, such as Sarbanes-Oxley, 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 it is also cost-effective for long-term retention.

Calculating the Costs—Tape vs. Disk

Disk-to-disk backup solutions are often viewed as cost-effective alternatives to tape for backup. Is this true? How much does it cost to backup to disk? To tape? Or to a combination of disk and tape?

To see how the three options compare cost-wise, consider the following: An IT organization needs to store 50 TB of backup data and wants to compare the costs of backing up to tape, disk, and disk/tape over time. Let's assume that data volumes—hence backup data volumes—for this particular IT organization are growing at 20% a year.

Remember, too, that even in the best scenario, disk and tape storage utilization is never 100%; in fact, in open-systems environments, a good rule-of-thumb for IT organizations is 85% utilization for tape media and 70% utilization for disk storage.

In this example, the IT organization would require 179.2 TB of raw storage capacity by the end of the seventh year, which translates into 210.8 TB of usable tape capacity or 256 TB of usable disk storage when utilization rates are factored in for both technologies.

Determine Hardware Configurations

The cost of implementing an LTO 2-based tape configuration was used for the tape configuration.[1]

For the purpose this example, the library is configured with four LTO 2 tape drives and 527 LTO 2 tape cartridges for a total capacity of about 211TB. An LTO 2 cartridge can store up to 400 GB of capacity with 2:1 compression. The list pricing of this configuration is $193,237.[2]

A readily available Serial ATA disk storage system for the disk system.[3] The disk configuration consists of 19 disk controllers. Eighteen of the controllers have three expansion units (18 x 14TB) and one has a single expansion unit (1 x 7TB) for a total disk capacity of about 259 TB. The disk acquisition costs, which include eight racks to house the units, total $1,681,490.

Calculate Environmental Costs

The next step is to calculate the environmental costs for the tape and disk configurations.

Since floor space and electricity prices can vary widely across geographic regions of the country, for this particular example, the floor space cost was estimated to be $20/ft^2 per month and the electricity cost to be $0.10/kwh.

[1] For this example, the HP Storage Works ESL 712e tape library was used.
[2] A more detailed explanation of these costs is outlined in the Data Mobility Group paper entitled, "Is Tape *Really* Cheaper Than Disk (March 2005)?"
[3] The IBM DS4100 SATA disk controller with the DS4000 EXP100 expansion unit was used.

The disk configuration requires eight racks of disk space, or 58 ft^2 of computer room space, bringing the total footprint cost to $13,920 per year. The tape library requires 10.4 ft^2 of floor space, or about 20% less space than the disk system, bringing total footprint costs to $2,496 per year.

Electrical costs for the tape library and drives come to $2,922 per year, while power and cooling costs add another $46,674 per year to the total cost of the disk system.

Therefore, the total yearly environmental costs for the tape configuration is $5,418 per year, compared to $60,554 for the disk implementation, making the disk implementation a whopping 11 times more expensive than the tape example!

Seven-year Cost for Disk and Tape Solution

Extended hardware costs are calculated by adding the initial purchase price to yearly environmental costs for the seven-year period.

The extended cost of the disk solution comes to more than $2M ($2,105,368 to be exact), while the extended cost of the tape configuration is just $231,163, or 11 % of total disk costs.

Some Caveats

The purpose of this exercise is to highlight the effects of environmental costs on long-term storage solutions. The cost calculations are a simplified view of the total cost of a seven-year disk or tape solution; a more complete calculation would also include additional items such as:

- The cost of maintenance after the warranty period expires.
- The cost to replace disk and tape after several years.
- The cost of phasing in equipment purchases. It is not necessary to purchase all equipment in the first year.
- The additional cost of RAID protection, such as RAID-1 or RAID-5, for disk.

Backing Up Data—Which Media to Choose?

In the above example, the tape solution turns out to be about 11% cheaper than the comparably-configured disk solution, factoring in the initial purchase of the systems and environmental costs.

Another option would be to implement a hybrid disk/tape solutions, which combines the cost-effectiveness of tape with the speed advantages of disk.

In this type of scenario, users would back up 25% of their data to higher-performing Serial ATA disk and 75% to LTO 2 tape over the seven-year period, based on data access requirements, etc. Using the 50 TB example above, users would require 159 TB of tape storage and 64.3 TB of disk storage at the end of the seven-year period (assuming average utilization rates).

This type of configuration would require 5 disk controllers and 14 expansion units, which would be housed in 2 racks (versus 8 racks for the all-disk configuration), dropping the purchase price to $440,515. Floor space and power costs for this new configuration would be $3,480 and $11,984, respectively, for a yearly total of $15,464. The 7-year cost of the disk system (floor space + environmental factors) would be $548,763.

On the tape side, this type of configuration would require fewer tape drives and tape cartridges. The tape library would now be configured with 2 tape drives and 398 cartridges costs for a total hardware cost of $155,825. Footprint costs would remain the same, but power costs would drop slightly to $2,863 per year. The 7-year cost of the tape system (floor space + environmental factors) would be $193,338.

Thus, the total 7-year cost of the hybrid solution (hardware + environmental costs) would be $742,101, which is 65% less than the all-disk option with room to grow. The tape library could accommodate an additional 314 cartridges as users' capacity requirements increase.

The Bottom Line

When all is said and done, which configuration you choose—all tape, all disk (D2D), or a disk/tape hybrid (D2D2T)—really depends on access requirements (i.e., how quickly you want to be able to retrieve data), the amount of data backed up annually, and budgetary restrictions (how much you're willing to spend on backup media).

If business service level agreements (SLAs) state that all data must be restorable within seconds, then disk may be your only option (doing replication between two disk systems). However, if the majority of your data has more lenient recovery requirements, then implementing a costly pure-disk system probably doesn't make sense.

A better option might be a disk/tape solution. It allows for fast retrieval of some data while keeping overall storage costs down. In a sense, it offers the best of both worlds: fast retrieval from disk and cost-effective tape storage!