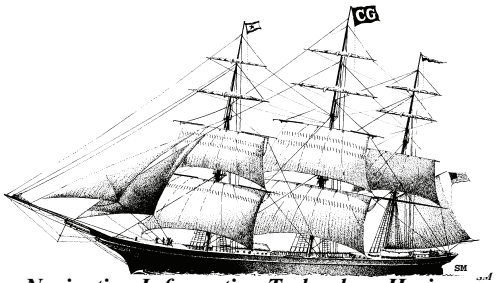
Clipper Notes™
VENDOR-AGNOSTIC EXPLANATIONS AND ADVICE FOR THE INFORMATION TECHNOLOGY BUYER



Navigating Information Technology Horizons℠

New in 2007 Report #TCG2007013 February 1, 2007

The Evolving Role of Tape in the Data Center

Analyst: Dianne McAdam

Management Summary

The disk versus tape debate continues to rage within the trade publications. Disk supporters claim that the emergence of low cost Serial ATA (SATA) drives proves that disk is now a viable, inexpensive media for backup. In fact, some proponents claim that tape is dead…or at least dying.

The very eloquent writer, Mark Twain, once said, "The rumors of my death have been greatly exaggerated." Tape vendors could say the same thing – the rumors about the death of tape have been greatly exaggerated. Tape is neither dead nor dying – but the role of tape is changing within the data center.

It is said that some things don't change very much in IT. It does seem that the disk-versus-tape debate has not changed over the last several years. The prices of disk drives are dropping, making disk an attractive option for some traditional tape-based applications. Nevertheless, tape vendors continue to develop and deliver larger tape cartridges, reducing the cost per gigabyte. Therefore, the gap between the cost of tape and disk remains the same.

But some things do change in IT. Enterprises, both large and small, are now faced with a new challenge – the challenge of storing more and more data every year. Enterprise data centers report data growth rates of 50 to 100 percent or more. Unstructured data, such as digital images, is growing at a faster rate than structured data (such as databases) and these unstructured files can be very large. For example, if we stored the complete works of Shakespeare, we would need five megabytes (MBs) of storage. Five MBs is also the average size of just ***one*** digital photo. Not only are we required to store more and more data, but also we have to keep it longer than we did before. Regulations require financial companies to keep emails for seven years or more. Health care agencies are keeping records for the life of the patient and longer. The result – storage requirements are growing at extremely high rates every year and there appears to be no end in sight.

IT is responsible for storing lots of data for a very long time. But how often do we need to access that data? The majority of this data is very seldom accessed. The old *80/20 rule* still applies today. Twenty percent of the data is accessed eighty percent of the time. The other eighty percent of the data is not accessed very often. The need to store data for a long period of time, coupled with the knowledge that much of data will be infrequently (or never) accessed helps to define one of the new roles of tape in the data center.

Read on to examine the new roles of tape and disk in data centers today.

IN THIS ISSUE

The Evolving Role of Disk and Tape

Disk drives were originally designed to support random I/O operations. They have been and continue to be the right storage device for online applications that require fast access.

The explosion of internet commerce has required some applications to be available 24 hours a day. Other applications that support online customers have very limited backup windows. These new availability requirements leave little, if any time, to bring down applications for backup processing. This has led to a new role for disk in the data center; disk is now a target for backing up high priority application volumes and files.

Unlike disk, tape is a sequential device. It is well suited for reading and writing large streams of data. Large data streams include feeds from satellites, video surveillance, or broadcast news clips, for example. Backup applications, on the other hand, cannot always continuously stream data to tape drives. The performance of backup applications can be limited by the amount of bandwidth that is available between the source and target devices. Tape drives operate at maximum performance when they continuously write (or read) data. Backup applications typically write data to target devices in small blocks. In situations with limited bandwidth, the blocks do not follow each other quickly enough. Tape drives are forced to write the first block of data, then stop and reposition the tape while waiting for the next block to arrive. This write-stop-reposition-write again motion causes tape drives to operate at less than optimal performance, which can elongate the backup process. Applications with constrained backup windows and limited bandwidth can benefit from disk-based backups, which do not suffer from the same degraded performance when data arrives in small bursts. Disk-based backups may not be necessary for all applications. Applications that have larger backup windows can continue to benefit from backing up data to lower cost tape. **Tape continues to maintain its role as a backup target for applications that do not have constrained backup windows or stringent Recovery Time Objectives (RTO).**

Does this mean that all backups should be saved on disk? Absolutely not! Most restores occur within the first few days after a backup has been completed. The likelihood that you will need to recover data is reduced the more time that has elapsed since you did the backup. Therefore, it may not make economical sense to store **all** copies of backups on disk because not only are you adding cost but also because you only need to recover from disk for a short period of time. Older disk-based copies of backups can be migrated to tape, freeing up disk space to store more backups that are current.

Writing backups to disk (disk-to-disk), then migrating older backups to tape (disk-to-disk-to-tape) blend the speed of restoration from disk with the economy of tape.

What Tape Does Well

Tape has always been easily transportable. Tape cartridges can be shipped offsite in boxes, transported by truck or hand carried to remote locations for secure offsite storage[1]. If a disaster occurs in the data center, the offsite tapes can be retrieved to restore applications.

Recently, some vendors have developed removable disk drives. Data is written to a disk device that can later be ejected and transported to other location. Current tape cartridge capacities exceed the capacities of removable disk drives. However, we expect the capacities of these removable devices to increase in the future. For now, we continue to view tape as the most commonly used transportable media. **Tape continues to maintain its role as a low cost media for offsite storage.**

Archival Storage

Tape, by its very nature, works at its optimal performance when reading and writing large streams of data. Some examples of large data streams include the following.

- Images from surveillance cameras that record activities at ATM machines, in public buildings, on city streets, and toll booths;
- Broadcast news events; and
- Digitized medical images, such as MRIs

Broadcast companies must retain news clips for many years. Today's news story can be part of tomorrow's review of a politician's career. Government agencies and security organizations may need to review surveillance tapes days, weeks, or months after an event. Many of these large images must be kept for a long time.

For example, new regulations require hospitals to keep medical records for the life of a patient, if

[1] Tapes containing sensitive data that are transported offsite should always be encrypted to ensure that the data is protected if the tapes are accidentally lost or stolen.

not longer. A single digitized X-ray can be 10 to 20 MBs in size. Ultrasound studies increase sizes dramatically since images are taken thirty times per second. Suppose a child breaks a leg while playing football. X-rays initially are taken to determine the exact location of the break; later, more x-rays are taken to diagnose how well the break is healing. The hospital is required to keep these images and all future records for decades. Multiply those images by thousands of patients and the storage requirements are staggering. The majority of these images may never be accessed again. Storing these images on disk can be an expensive option. **Tape can dramatically reduce the cost of archival storage**. The following example illustrates this point.

Comparing Acquisition Costs of Tape and Disk

To compare the acquisition cost of tape and disk, we need to configure two similarly sized systems. We chose an automated tape system from Quantum and SATA disk system from IBM. Let's assume that a corporation must store about 100 TBs of data the first year and they predict that their data storage requirements will grow at a modest 20% per year. They are evaluating storage solutions that will meet their projected needs for the next five years.

With a 20% data growth rate, they will require the following.

- 100 TBs of storage in the first year,
- 120 TBs of storage in the second year,
- 144 TBs of storage in the third year,
- 173 TBs of storage in the fourth year, and
- 207 TBs of storage in the fifth year.

Tape and disk systems are never allocated at 100%. Storage utilization can vary widely from one data center to another. For this exercise, we choose to use utilization rates of 70% for disks and 85% for tapes. These rates are commonly attained in mainframe environments but may not be achieved easily in open systems environments.

So, for our example, we actually require 244 TB of tape storage to store 207 TBs of data (assuming an 85% utilization rate). Additionally, we would need 297 TBs of disk storage (at 70% utilization) to store 207 TBs of data.

Disk and Tape Configurations

For the tape system, we chose Quantum's PX720 library with four DLT-S4 tape drives. The PX720 library is available in several configurations that can hold up to 726 tape cartridges in one frame. The entry model of the library, which supports 190 cartridges, is sufficient to meet the five-year storage needs in this example. Each DLT-S4 cartridge holds 800 GBs of data or 1.6 TBs of data compressed at the industry standard average of 2:1. Different types of data compress at different rates and your mileage may vary. For the purpose of these calculations, we choose to use the industry average of 2:1 compression. We need 153 cartridges to support this project, leaving 37 slots for expansion. (See Exhibit 1, below.)

For the disk configuration, we choose a SATA disk array from IBM. One dual controller configured with seven expansion units can support up to 44.8 TBs of storage. For this project, we need six dual controllers (each controller is attached to seven expansion units) and one dual controller attached to five expansion units for a total of 302 TBs.

Exhibit 1 –
Purchase Price for Quantum *PX720/DLT-S4* Tape System

1 PX720 Library	$64,940
4 DLT-S4 drives ($16,495 per drive)	$65,980
153 DLT-S4 cartridges ($110 per cartridge)	$16,830
Total	$147,750

Note: MSRP for PX720 base frame with entry slots, 1 drive cluster and a communication module. This also includes Web-based remote management and monitoring software, redundant power supplies, and fans.

Source: Quantum

Exhibit 2 –
Cost for IBM *DS4100* Disk System

7 dual controllers ($26,547 per controller)	$185,829
47 expansion units ($10,999 per unit)	$516,953
7 expansion unit attachments feature ($4,999 per feature)	$34,993
4 Racks ($4,850 per rack)	$19,400
Total	$757,175

Note: The dual controllers (model 1724-100) are configured with DS4000 EXP100 expansion units. Attaching four or more expansion units to a dual controller requires the attachment feature.

Source: IBM

The cost to purchase the disk solution is over five times more expensive than the tape solution. (See Exhibit 2, on previous page.)

Some Caveats About Disk Pricing

The calculated disk acquisition costs are for the minimum requirements. We did not factor in the cost of protecting disk-resident data by RAID schemes. Mirroring, or RAID-1 protection would double the cost of the disk systems. Other RAID schemes, such as RAID-5, require additional disk storage without requiring double the amount of disk. Many RAID-5 implementations require at least 20% more disk capacity.

Also, we did not calculate the cost for spare disk drives. The disk systems were configured without any disk spares. Most disk vendors recommend one spare disk for each expansion frame. This will also increase the cost of the disk systems.

Comparing Electrical Costs of Tape and Disk

While acquisition costs vary by a large margin, the cost to power and cool the systems varies by a greater margin. Electricity costs vary by state, region, and country. For example, electrical costs in California and New England are higher that the Midwest. We choose to use $.145 per KWH, which is the average cost of electricity for commercial customers in New England.

The yearly electrical costs were calculated by:

1. Adding the power requirements and the cooling requirements for each unit.
2. Multiplying that sum by the number of units.
3. Multiplying that product by 8760 (the number of hours in a year).
4. Multiplying that product by .145 (the cost per KWH).

Exhibit 3 – Yearly Electrical Costs for Automated tape Library

	Power Requirements Per Unit	Cooling Requirements Per Unit	Total Yearly Electrical Costs
1 PX720 library	1.6 KWH	1.6 KWH	$4,065
4 DLT-S4 drives	.018 KWH	.018 KWH	$183
Total Yearly Cost			$4,248

Note: Based on 14.5¢ per KWH.

Note that the DLT-S4 drive uses 30 watts when writing data, and 14 watts when idle. We used 18 watts as an average, to represent writing 25 percent of the time and idle the rest of the time. (See Exhibits 3 and 4, below.)

The cost to power and cool the disk solution is over eleven times more expensive than the tape solution.

Why Calculate Electrical Costs?

A yearly electrical bill of $49,386 to power and cool the disk systems may not seem excessive when compared to the initial acquisition cost of $757,175. However, at the end of five years it will cost $246,930 for electricity or about one-third the initial acquisition cost.

The tape systems will only require $21,240 in electricity in five years (or about one-seventh of its initial lower acquisition cost).

Why calculate electrical costs? The previous calculations assume that energy costs will remain unchanged over the next five years. However, most energy consultants fear that electrical costs will continue to rise, adding to the costs to store data on disk.

New England commercial customers are all too familiar with the rise in energy costs. In January 2005, the average cost of electricity was $.113 per kilowatt-hour. In January 2006, that cost rose to $.145, a whooping 28% increase.

We would like to believe that electrical costs would not rise at 28% per year. Let's take a more conservative approach and assume that a New England commercial company sees an increase of 10% per year.

With a ten percent increase in electricity, it will now cost $301,507 to power and cool the disks for five years – or 40% of the original acquisition cost. Tape systems run up less of a bill

Exhibit 4 -Yearly Electrical Costs for SATA Disk System

	Power Requirements per unit	Cooling Requirements per unit	Total Yearly Electrical Costs
7 disk controllers	.33 KWH	.39 KWH	$ 6,402
47 expansion frames	.33 KWH	.39 KWH	$ 42,984
Total Yearly Costs			$49,386

Note: Based on 14.5¢ per KWH.

– only $25,935 in five years (or 18% of the original cost).

Other Factors

These cost calculations are an exercise to highlight the effects of acquisition and electrical costs on long-term storage solutions. A more complete study would include factors such as:

5. **The cost to replace disk and tape after several years**. Many enterprises replace disk systems every two to three years and replace tape systems about every five years**.**
6. **The cost to phase in equipment purchases as needed.** Not all of the equipment in the initial configuration would be required in the first year.
7. **The cost of maintenance when the warranty expires.** Warranty periods differ by vendor and by system. Maintenance pries also vary depending on the vendor and the type of coverage required.
8. **The software and personnel costs to manage the storage devices.**
9. **The cost to backup or replicate the data stored on these systems for disaster recovery.**

What About Data Reduction?

Many vendors have developed or will be soon announcing software that can dramatically reduce the storage requirements for backup. This technology is called by several different names, such as *data reduction* or *data deduplication*. The technology is complex, but the idea is simple. We back up the same data over and over again. Data reduction techniques examine the backup streams and eliminate repeated sequences. Compression eliminates repetitive characters; data reduction software eliminates larger blocks of repeated data sequences.

Many enterprises report significant reductions in storage when data reduction techniques are applied to backup data. Results vary widely, but many report greater than 10-to-1 or 20-to-1 reduction. Since results do vary widely, we chose, in this paper, not to calculate the savings resulting from data reduction.

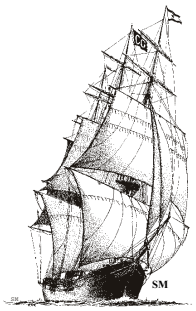
Since data reduction techniques can provide significant storage savings, enterprises that implement such technology discover that they can direct more backups to disk or store backups on disk for longer periods of time without requiring additional disk purchases. In these environments, tape can continue to play a role as a low cost media for offsite storage.

Conclusion

The role of tape has evolved in the data center. It is no longer the primary backup target for mission-critical applications that have stringent RPO and RTO objectives. Yet, it continues to maintain its role as the primary backup target for applications with less stringent requirements.

Tape has a predominant role in disk-to-disk-to-tape strategies. It is a cost effective media for migrating older backups from more expensive disk-based targets. Tape should also be viewed as a low-cost media to store archived data for long periods of time.

When evaluating whether to use tape or disk as storage devices, some customers may focus on the initial acquisition costs, but fail to think about other costs, such as electrical costs, that can add to the cost of storage over several years. Storage hardware costs are decreasing every year. However, customers have few choices when it comes to electricity. They do not have the option to go to another energy provider. The only way they can control energy costs is to choose storage wisely. **Take control of energy costs by choosing the right storage for the job.**



About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Dianne McAdam **is Director of Enterprise Information Assurance for the Clipper Group**. She brings over three decades of experience as a data center director, educator, technical programmer, systems engineer, and manager for industry-leading vendors. Dianne has held the position of senior analyst at Data Mobility Group and at Illuminata. Before that, she was a technical presentation specialist at EMC's Executive Briefing Center. At Hitachi Data Systems, she served as performance and capacity planning systems engineer and as a systems engineering manager. She also worked at StorageTek as a virtual tape and disk specialist; at Sun Microsystems, as an enterprise storage specialist; and at several large corporations as technical services directors. Dianne earned a Bachelor's and Master's degree in mathematics from Hofstra University in New York.

- ***Reach Dianne McAdam via e-mail at dianne.mcadam@clipper.com or at 781-235-0085 Ext. 212. (Please dial "212" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group Captain's Log**, **The Clipper Group Voyager**, **Clipper Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.