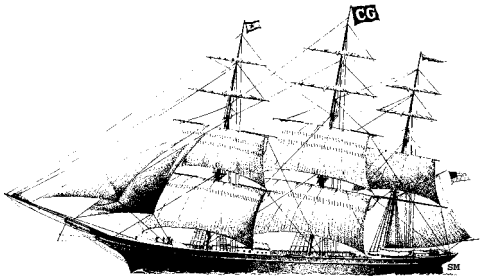
THE CLIPPER GROUP
Explorer™



Navigating Information Technology Horizons℠

Published Since 1996 Report #TCG2004008 February 6, 2004

The Role of Tape in Multi-Tiered Storage — Alive, Kicking, and Rolling Along

Analyst: David Reine

Management Summary

Tape, as a medium, is old, but still very important to most enterprises. The first recorded use of tape in a data processing environment was in 1952 when IBM introduced the Model 726 Tape Drive. That was over 50 years ago. This device stored 1.4MB of data on a movie reel 12" in diameter, the same capacity as a floppy diskette, only significantly larger. Over the next five decades, IT departments have used magnetic tape for scratch space, long-term storage, back-up, and disaster recovery because of its intrinsic qualities of being: removable, transportable, and cost-effective, with infinite capacity. **Despite the proclamations of industry pundits for the last 30 years, magnetic tape not only survives, but also remains an inexpensive, viable storage solution, which has also gained exceptional popularity around the world in support of mission-critical applications.** To quote Mark Twain: *Reports of my death have been greatly exaggerated.*

Low-cost, high-performance magnetic tape solutions are needed more today than ever before. The financial industry generates millions of auditable transactions every hour. The medical industry creates digital x-rays, MRIs and CAT Scans. The government and industry have agreed to mandates for the long-term preservation of digital information through Acts from Sarbanes-Oxley to HIPAA. In the May 2003 ruling, the SEC said that **any electronic media** was acceptable which could preserve records in a non-rewriteable, non-erasable format, and is able to verify the quality and accuracy of the recording process. **The need for an economical, efficient method for the long term storing, archiving and accessing of this content-rich, permanent digital information has never been in greater demand.** Indeed, we have not even touched upon the requirements of preserving enterprise email for multi-year periods or the needs of the entertainment industry, which produces thousands of new DVDs and CDs in digital format every year.

We see a need today for a multi-tiered storage environment, where enterprises can place a value on every database, every file, and every record. An environment where IT departments can control the movement of that data from one storage medium to another under an Information Lifecycle Management (ILM) process. The need to match the value of the information to the value of the storage will govern the process, with mission-critical data on high-performance, high-cost primary storage, archival data on lower-performance, high-capacity, low-cost media. With new tape technology enabling the storing of 1TB of compressed information, and more, on a single cartridge, please read on to see how the new tape architectures fit into the ILM environment.

IN THIS ISSUE

- Information Lifecycle Management..... 2
- Introduction of Virtual Tape Appliances..... 4
- The Future of Tape 5

Information Lifecycle Management

In order to address the role of tape as an information storage medium in today's storage architectures, we need to first step back, widen our focus, and view all of the electronic data stored within the IT walls. However, **it is important to understand that we need to view this data over time**. The value of electronic data changes over time, but not necessarily always in one direction, either up or down. Information introduced into an enterprise database today may have an extremely high value associated to it. The need to process it could have a direct bearing on the profitability of that enterprise. Take an order for example. The faster that Sales can process it and ship product to a customer, the faster that Accounting will be able to invoice that customer and, eventually, receive payment. There is a need to have that data sitting on a storage device that is highly performant and highly reliable. We accept the premium price that it carries.

After Accounting has received payment for the transaction, however, the value of that data may change dramatically. Its primary use may move from one of a billable accounting urgency to one for historical reference. There may not even be a need to look at it again until tax time. Alternatively, you may need to access it when quarterly reports are due. IT may even want to move the data from an accounting file to a marketing function for data mining. In any case, each of these requirements may put a different value on the data in terms of how urgently you need to access it. There may be no need to look at it again for months. **You clearly do not want to waste the most expensive electronic storage resources on information whose value to the enterprise has declined.**

There is a need to establish a multi-tiered array of storage classes[1], with information capable of transitioning from one class, with corresponding costs, to another class with less expensive costs. How many classes does your enterprise need? That depends on the value of its data over time and the characteristics of your storage devices. (See Exhibit 1.) The data center needs to be concerned with the value of every specific piece of information and have a storage receptacle for it.

Exhibit 1 – Device Characteristics

- **Performance** – Speed at which data is delivered to the user; a direct result of throughput and cache size.
- **Reliability** – How much redundancy is available in terms of RAID, hot-swap power, fans, and hot standby drives.
- **Availability** – Access of data through multiple paths to external switches; Ease of use.
- **Serviceability** – Ease of replacement for failed components, access to Customer Service Engineers.
- **Scalability** – Capability for dynamic expansion of various arrays.

The typical enterprise Storage Area Network (SAN) will always contain a high-performance, or premium, tier. This high-end array will probably be from one of the leading storage manufacturers: EMC, Hitachi, or IBM, with *Symmetrix*, *Lightning*[2], and *Enterprise Storage Server (Shark)*. The data center will usually have a high-speed attachment through Fibre Channel in an open systems environment and/or ESCON/ FICON, if a mainframe is involved. There will also be a second tier in this typical SAN. The second tier will usually consist of a mid-range storage array from one of these same vendors or others, such as EMC *CLARiiON*, an HP *EVA*, an IBM *FAStT*, or a StorageTek *D-Series*. This second tier will typically use Fibre Channel throughout, delivering a lower level of performance than is expected from the premium storage, but at a significantly reduced cost. IT administration can also sub-divide this tier into a second and third tier if the various businesses within the enterprise can live with different levels of service. They can install these arrays with either Fibre Channel drives or ATA, or in some cases, both. This could result in three

[1] See **The Clipper Group Explorer** dated August 29, 2002, entitled *Tiered Storage Classes Save Money – Getting the Most Out Of Your Storage Infrastructure* at http://www.clipper.com/research/TCG2002030.pdf.

[2] HDS Lightning is resold by HP and Sun under their own brands.

tiers of storage to the enterprise. By including drives with differing capacities (18, 36, 73, 146GB) and differing speeds (7200, 10K, and 15K RPM), IT can create even more tiers with storage devices of varying cost and performance. Costs of storage to the enterprise, however, extend beyond simple acquisition costs. What about the environmental costs to keep the disks spinning, such as electricity, air conditioning, and floor space? Moreover, what about the maintenance costs to keep legacy disk storage operational? **As the enterprise amasses more and more information, the incremental costs to keep it all on line become prohibitive.**

In some cases, however, the business requirements permit the nearline performance and functionality levels of tape. Furthermore, there are instances where **the business needs demand the low cost of tape.** In these cases, the data center may add, or retain, this additional tier for data protection, data collection or long-term data retention. While we measure the cost of even the least expensive ATA disk media in dollars/GB, we continue to measure the cost of tape in pennies/GB, with that cost continuing to fall.

What are the requirements to which we continue to refer? No matter what your industry, no matter what the size of your data center, management concerns involving continuous access to data and the preservation of that data remain the same. As shown in Exhibit 2, we have a constant need to:

- Improve data center productivity;
- Enhance disaster tolerance;
- Enable a self-managing solution to reduce human interaction; and
- Reduce costs.

Whether the measurement of needs is in gigabytes or in terabytes, you must adhere to corporate information management policies in protecting the enterprise assets to ensure business continuity. There is also the requirement to ensure the conformance to new government regulations such as the Sarbanes-Oxley Act and the Health Insurance Portability and Accountability Act (HIPAA) in preserving an audit trail of information.

Whether it is to comply with regulations or simply to improve relationships with customers and vendors, enterprises are amassing an ever-growing amount of data in order to gain a competitive edge and improve profitability. This could result in significantly higher operational costs if the enterprise is trying to manage and maintain an on-line legacy database. One way to improve profitability is to reduce costs. Archiving data is a solution to the problem of long-term database growth, which leads to reduction in operational, data retention, and storage costs.

The on-line storage of historical data degrades the overall system performance and restricts the availability of mission-critical applications. You can patch the problem by throwing money and resources at it, installling larger drives, faster processors, etc., but eventually the size of the database will become so large that loading, searching, and backing it up begins to impact system availability. **The real fix has to do with changing your storage paradigm.** *Archiving* is the enabling strategy that allows enterprises to access and manage rarely used data from the least

Exhibit 2 –
Customer Requirements

Improve Productivity

Install latest technology in the data center with high-capacity cartridges, high-throughput drives in order to shrink the back-up/recovery window and implement active archiving.

Enhance Disaster Tolerance

Improve data availability through data and control path failover for local solutions; ensure a recoverable remote mirror for long-distance solutions.

Enable Self-Management

Implement an autonomic computing capability with self-monitoring, self-healing policies in a fully automatic process.

Reduce Costs

Lower the Total Cost of Ownership (TCO) by reducing the physical library specifications to manage space, power, and operator intervention.

expensive storage medium available. Once the data has been archived, it can be removed from the application database, improving response times and reducing operational overhead. IT staff can remove additional data from the application database over time, as the currency of the information evolves, adding to the archived data, while at the same time, maintaining the critical database at a manageable level.

IT often views data in four categories:

- *High-performance Online* for immediate access,
- *Lower-cost Online* for fast access (via ATA class disk devices),
- *Nearline* for backup and retrieval,
- *Offline* for archived removable media (e.g., historical email).[3]

A multi-tiered business continuation policy will involve all three of these categories. This policy will automatically migrate the data from one tier to another based upon a specific set of rules. The CIO could establish a requirement to back-up data to a disk-cache-based, or virtual, tape system as well as archiving to a tape library in a Hierarchical Storage Management (HSM) architecture. IT could also implement a mid-level, server-less SAN library solution, or an enterprise-level multi-server attached configuration. It could be a local implementation or a remote one depending upon the level of mission-critical information, and the costs, involved. The decision should be based on a number of factors such as minimizing cost, maximizing performance, and risk avoidance.

By looking at not only the back-up window, but also the recovery window, the IT department can determine the performance level required to complete both a full and incremental backup in the available time slot. In addition, more importantly, they can establish the recovery criteria in order to keep the enterprise operational in case of a man-made or natural disaster. It is here that the value of a virtual tape system may provide the greatest return to the enterprise for business continuity. The number of hours available, the size of the recoverable files, and the number of iterations of information required to be preserved will all contribute to the determination of the configuration of the enterprise's total storage architecture. The technology, the time available, the number of drives, the type of drives, the cartridge capacity, and the number of cartridges required will determine the solution.

Introduction of Virtual Tape Appliances

Many data centers have already or intend to introduce disk into their backup/recovery environment. CIOs will do this in order to cope with the increased volume of backup data and the need to recover the information quickly. **There does not appear to be an intention to replace tape, simply to complement it and improve the backup-recovery architecture.** With a third of all backups taking over 6 hours to complete, and another third between 3 – 6 hours, speed is essential. With perceived cost the only drawback to the acquisition of a virtual tape appliance, the potential savings for each on demand recovery could outweigh that concern. **A reduction in the recovery window is urgent with downtime potentially costing hundreds of thousands of dollars per hour.** The first outage could satisfy the Return on Investment (ROI) by itself.

The implementation of virtual tape appliances or virtual tape libraries, which emulate physical tape libraries, directly addresses the issues of backup and recovery speed and reliability. By introducing an intermediate disk device, or disk cache, such as a low-cost ATA disk, or RAID array for reliability, the data center can significantly reduce backup and recovery time. A broad array of ATA products is available, such as EMC's *CLARiiON*, IBM's *FAStT*, or StorageTek's *BladeStore*. The data center can also improve reliability by removing mechanical failures due to library robots from the recovery scenario. At the same time, IT can improve the mission-critical application response time, by reducing the network traffic from a concurrent backup. This action,

[3] Beware of the risk of a disconnect between the logical and physical ID of a cartridge from the source application to the virtual library when write errors or multi-cartridge conditions occur. This must be managed.

however, only defers the eventual movement of the backup data from disk to tape; it does not eliminate it. The disk is a complement to tape. **The virtual tape architecture requires users to designate an interim destination for the backup data, which, by policy, will be transitioned to tape after a set period.** This allows you to recover the data from an online medium and to rotate the disk space for additional iterations. This reduces concerns over unlimited growth, while protecting mission-critical data in the long term from viruses and other potential corruption. Your data will be safe on tape.

There are many companies delivering virtual solutions to the storage industry today, each with a unique approach to the preservation of your enterprise's most valuable assets. Examples of this are ADIC's *Pathlight VX*, IBM's *Virtual Tape Server*, and StorageTek's *Virtual Storage Manager*, among many others. No matter how large or small the data center, you can find a solution tailored to your environment.

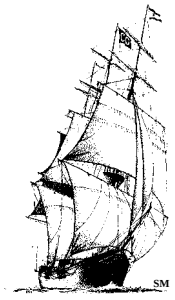
The Future of Tape

Because of the catastrophic events of September 11, 2001, renewed attention to the importance of disaster recovery has returned tape technology to the forefront of enterprise thought. Tape backup and recovery strategies, again, are being viewed as the most cost-effective means to protect the enterprise's most valuable electronic assets, data, and to reduce the risk of the enterprise closing as a result of a similar disaster. The disastrous economic fall of companies such as Enron has also contributed to the need for the retention of increasingly more corporate history, both financial data and email records. **The requirement for low-cost, high capacity, high-performance tape drives has never been greater.**

A gradual shift to ILM has led some enterprises to implement a lower-cost disk strategy for the quick recovery of lost data because of a system outage. ATA disk, however, while cheaper than Fibre Channel, does not support the long-term requirements of the balance sheet concerning reducing the costs of long term data storage. The cost of a tape cartridge, any tape media, appears to be about $.40/GB in a native, or uncompressed, format. The cost of that cartridge in a production mode, however, is significantly less. How much less is a factor of the specific compression ratio for the tape architecture and the content of the data that is being saved. SDLT promotes a 2:1 ratio while IBM advertises a 3:1 capability for the 3592 tape cartridge. In fact, some digital content will not lend itself to compression at all. *Caveat Emptor* - your results may vary!

We will be publishing follow-on reports covering additional virtual tape server topics that you will need to be aware of. These will include an examination of the relative attributes of such midrange tape systems as SDLT, LTO, and SAIT as well as the high-end, traditional data center formats for the IBM 3592 and STK T9940B drives.

In the meantime, if your enterprise encounters any of the problems presented here, then you need to ensure that your solution provider knows the capabilities of Information Lifecycle Management and Virtual Tape. Be sure that they understand the special characteristics of your environment and can offer a variety of backup and recovery solutions for the data center. **When you face a data recovery, the last thing you need is another disaster.**



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

David Reine **is Director, Enterprise Systems for The Clipper Group.** Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. He joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

- ***Reach David Reine via e-mail at dave.reine@clipper.com or at 781-235-0085 Ext. 23. (Please dial "1-23" 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*, 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.