WHITE PAPER
May 2005

The role of tape automation in information lifecycle management

ABSTRACT

Information lifecycle management (ILM) is a strategy that maximizes attainment of service levels at the lowest possible cost. Tape automation is an important technology in an ILM implementation. The purpose of this paper is to clearly define the role of tape automation in information lifecycle management and to help IT decision makers make informed decisions regarding their tape infrastructure.

1 Executive summary

Information lifecycle management (ILM) is a sustainable storage strategy that balances the cost of storing and managing information with its business value. To achieve the lowest possible cost while maintaining service levels, IT organizations need to take full advantage of multitiered storage infrastructures.

Tape automation is one of the key elements of the information lifecycle management infrastructure today, and it will be for the foreseeable future. At the high end, tape automation's financial advantage over capacity disk solutions has actually improved since the late 1990s. Advances in tape management have kept high-end tape management and administration costs on a par with, or lower than, disk management costs for comparable capacities. Innovations in tape technology (density and performance) have revitalized the role of tape automation in the information lifecycle management hierarchy. Looking forward, the industry will likely deliver a high-end 1-terabyte (TB) cartridge in the very near future, and it is projected to deliver a 10-TB cartridge near the end of the decade. It is clear that substantive innovation will continue.

The analyst community has recently reported large-scale movement of data from tape solutions to capacity disk. Initially this was perceived as a net increase in capacity, with disk drives being used to buffer and stage data to and from tape. Recently, this appears to have changed, and there have been reports of significant tape replacement activity. The drivers appear to be new requirements for faster restores driven by user pressure and business continuity needs, the repurposing of archived data for customer-facing applications, and the discovery requirements associated with compliance. Organizations are finding all sorts of reasons to access data they thought they would use seldom, if ever.

We believe this tape replacement activity is a short-term phenomenon created by the lack of effective automated solutions for data management and movement between tiers in the storage hierarchy. The situation is aggravated by the long-established optimization of tape applications for *writing* to tape rather than for *retrieving* data from tape. Most of today's solutions for automated data movement are point solutions and have relatively low levels of acceptance across large-scale information technology (IT). New solutions that focus on effective data *use* rather than data *storage* are required. Once these solutions are available, customers will not be forced to make suboptimized either/or decisions regarding devices for data placement; instead, they will have robust automated data movement solutions that will make infrastructure transparent.

Rather than facing a collection of one-off decisions regarding tape versus disk and suboptimizing overall infrastructure, organizations would be wise to consider investing in information lifecycle management wherever possible to meet the full requirements of their applications in the most cost-effective manner. This will enhance IT's ability to contribute directly to the business, serve current needs and prepare for future needs. Taking a systemic view of your organization's information management requirements and mapping out a long-term plan to achieve advanced information lifecycle management will yield significant benefits and make IT a strategic contributor to competitive advantage.

High-end tape systems are more cost effective than disk, and most users are satisfied with them:

- *Tape versus disk total cost of ownership comparisons result in 1:4 to 1:10 cost ratios. Ratios can increase to 1:20 for very large tape automation systems.*
- *Tape administrative costs per unit of capacity are generally better than for low-end open systems disk and roughly equivalent to costs for top-of-the-line zOS disk systems.*
- *Users of high-end tape automation are generally satisfied with their systems.*
- *A significant amount of data is moving from high-end tape automation systems to disk systems.*

2 Introduction

There is considerable evidence that tape automation has a lower total cost of ownership (TCO) per unit of capacity than even low-end disk devices.[1] Further, including all relevant variables in such comparisons generally increases tape's winning edge.

For the purposes of this paper, we need to make a few things clear; our comments and recommendations are directed at a specific market space:

- We are addressing tape automation, including libraries, drives and media.
- We generally refer to high-end technologies in our comments, specifically meaning half-inch high-end and small-form-factor drives and their associated library automation solutions.
- Generally, such devices are in high-end enterprises, although recently they have begun to extend downward into the middle market.

Some will assert that the cost of managing tape will swing the balance in favor of disk devices. Certainly, we can perceive that weakly automated, low-capacity solutions might put this assertion within the realm of possibility. However, estimates of administrative headcount and capacity for even a moderately dense library are lower than such estimates for a similar quantity of open systems disk, and these estimates are roughly equivalent to those for mainframe disk.[2]

Adding to the debate are those who declare tape dead, in spite of considerable market evidence of growth, substantial R&D investment and extraordinary advances in technology.

Most importantly, research sources indicate that high-end tape automation customers are satisfied with their devices.[3] Our experience with customers, although anecdotal, confirms this.

However, a significant amount of data *is* moving from tape to disk. Further, this appears to be a trend led by large enterprises. Some have forecasted it as being the beginning of a major shift in storage. The trend is attributed to compliance, access speed and media management issues. We believe it is slightly more complex than that. We present the case that this is a temporary situation — that in the future, storage management systems will redefine the value of tape automation, and its use will continue to grow with disk.

[1] *Tape Council (www.tapecouncil.org), Horison Information Strategies and NSIS provide a range of data points ranging from 1:10 to 1:4 for the total cost of ownership of tape automation versus disk.*

[2] *Horison Information Strategies, "Tape and Disk Cost Guidelines" (Louisville, Colo.: Horison Information Strategies, September 2004).*

[3] *Enterprise Strategy Group, "Tape Replacement Realities" (Milford, Mass.: Enterprise Strategy Group, March 2005).*

Tape usage:

- *Data protection*
  - *Restore*
  - *Recover*
- *Reference information management*
  - *Discover*
  - *Repurpose*

Tape technology:

- *Continual improvements in density*
- *Continual improvements in automation capabilities and functionality*
- *Keeping pace with or exceeding the requirements of tape's critical role in ILM*

3 Tape trends

The role of tape in most enterprise shops has been evolving for many years. For a truly historical perspective, one must go back to the late 1940s, when magnetic tape was introduced as a replacement for paper tape. Magnetic tape absolutely eclipsed its predecessor. In the late 1950s, disk was introduced, and it evolved through the 1960s and 1970s into the primary method for data storage. Initially, vendors attempted to make disk drives directly competitive with tape and included removable media capabilities. By the early 1970s, it was generally recognized that using more than one storage device might make sense. IBM brought hierarchical storage management (HSM, a conceptual ancestor to information lifecycle management) to the market in the 1970s, supporting tiered storage in multiple virtual storage. The notion that more than one device makes sense continues to be a wise thought today. Tape automation, introduced in the late 1980s and commercialized in the early 1990s, significantly changed tape's role once again, moving it back toward a more primary storage role. Today, the demands of online processing as the primary application model have limited the primary role for tape, leaving tape and tape automation to once again occupy predominantly the secondary or tertiary storage layers.

For the purposes of this paper, we have defined two predominant roles for tape automation today, based on the type of information involved and the objective of the role. Further, we have defined applications within the two primary roles, based on the reason for bringing the information back from tape to production applications.

- **Data protection**, the process of safeguarding dynamic information across the enterprise, has two elements:
  - **Restore** — return of small files, generally saved recently, to correct the loss of production data.
  - **Recover** — return of full-volume data sets and rebuilding of data structures after any disaster that significantly interrupts a primary application or after a full-volume loss.
- **Reference information management**, the process of capturing, organizing and retrieving static (or fixed-content) information across the enterprise, has two elements:
  - **Discover** — access to reference information, via content, across all reference information storage, for the purposes of responding to a legal or business need.
  - **Repurpose** — access to reference information in a business application that is generally different from the creating application. Frequently such applications are customer- or supplier-facing (e.g., customer service, customer relationship management, and so on).

Today, high-performance tape includes demonstrable 1-TB cartridges; according to industry analysts, we can expect a 10-TB cartridge in 2011. This represents an extraordinary growth in density that will continue to align tape as an effective target device for volume backups. Data rates (throughput) have increased at least five times in the last three years, and the pace of change shows little sign of slowing.

Tape market:

- *Thriving after a slowdown (with the rest of IT) in the first half of the decade*

New generations of tape and tape automation technology offer continual improvements that address real customer issues. Issues that plagued high-end tape operations in the past (e.g., mixed-media libraries) are addressed in solutions available today. Advances in tape technology provide the capability to cost effectively manage larger and more critical volumes of cartridges.

There is every indication that tape drive and tape automation technologies will continue to align with the appropriate information lifecycle management requirements for the foreseeable future.

The market for tape drives and automation is large and growing. Although overall drive and automation market revenue declined during the first half of the decade (along with most of the rest of IT), forecasts for the second half of the decade are moderate to strong across most high-end automation and high-end drives.[4]

Tape and disk pricing guidelines:

The Horison Information Strategies study advises: "For typical working automated tape configurations, the price per gigabyte of automated tape ranges from one-fourth to less than one-twentieth the price of an equivalent amount of disk storage."

4 Total cost of ownership for tape solutions

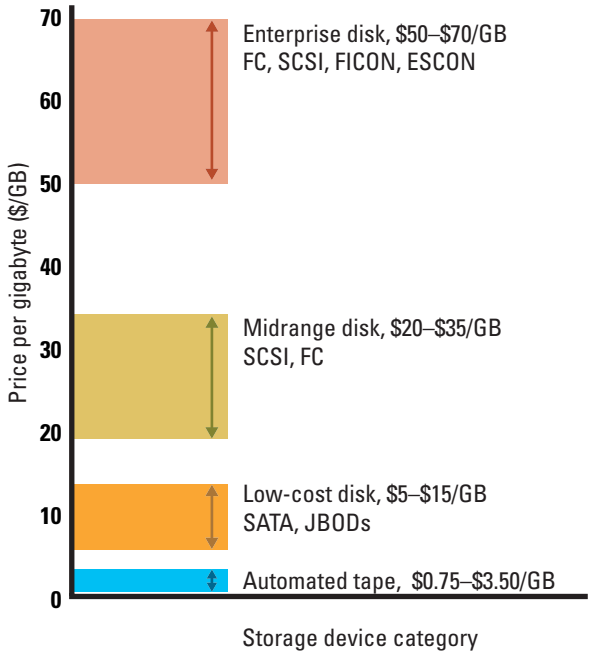
Gross tape and disk prices per unit of capacity have favored tape for most of these technologies' history. During the late 1990s, tape and disk prices were probably closer to each other (on a percentage basis) than ever before, but recent advances in tape technology have actually widened the gap in favor of tape.

Horison Information Strategies recently completed a study comparing the relative pricing of working subsystems of enterprise-class and midrange disk, lower-cost advanced technology attachment (ATA) or subsystems, and tape subsystems (Figure 1). For the working tape subsystem, point-in-time average selling prices for automated tape libraries, ultra-SCSI tape drives and 100-gigabyte (GB) native cartridges were used. Although Horison's comparisons do not include storage management software or facility costs for tape or disk, the results indicate that the pricing gap is actually widening.[5]

[4] *Gartner Group, "Forecast: Tape Automation Systems, Worldwide, 2004–2008" and "Tape Drives Forecast, Worldwide, 2004–2008" (Stamford, Conn.: Gartner Group, 2004).*

[5] *Moore, Fred, Storage: New Game New Rules (Boulder, Colo.: Horison Information Strategies, June 2003), www.horison.com.*

Figure 1: Disk and tape pricing guidelines



- The price per gigabyte decreases as the ratio of cartridges to drives increases, diverging from disk costs.
- Disk prices are for working subsystems.
- Automated tape prices include drives, media and library.
- Tape cartridge capacity is growing faster than disk drive capacity.
- Price for automated tape is nominally about 1/15 to 1/20 the price of disk for UNIX, Linux, Win2K; 1/25 or less for mainframes.

Source: Horison Information Strategies

Tape TCO:

- *With all relevant costs included, the TCO ratio of tape to low-end disk falls between 1:4 and 1:20.*

The Horison study also quantifies the impact of cartridge-to-drive ratio, indicating that high ratios further reduce the pricing of tape automation solutions well below the bottom of the range indicated in Figure 1. This is particularly important in the context of reference information applications, where high cartridge-to-drive ratios should be commonplace.

More complete analysis of tape automation's total cost of ownership versus low-end disk — including software, environmental factors and human resources — indicates a broad range of results, still within the 1:4 to 1:20 ranges indicated. Tape is favored for its environmental and floor-space costs per unit of capacity, and (depending on the study) is affected by human resources costs. At the high end of tape solutions, the cost per unit of capacity in terms of human resources is likely to be significantly lower than for disk:

Table 1. Estimated range of capacity managed per administrator

Storage type	Managed capacity per administrator
Open systems disk	500 GB–2.5 TB
Mainframe systems disk	40 TB+
Tape automation (all platforms)	40 TB–1 EB+ (wide variation is due to library size)

Notes: GB = gigabyte, TB = terabyte, EB = exabyte

Source: Horison Information Strategies, "Storage Facts and Figures" (Boulder, Colo.: Horison Information Strategies, September 2004).

Administration cost comparison:

- *Per unit of capacity, tape automation < open systems disk*
- *Per unit of capacity, tape automation » mainframe systems disk*
- *Large high-end tape automation < any disk*

Different application requirements make TCO comparisons useful in determining the balance of tape and disk to fit a particular set of needs.

After researching several sources of total cost of ownership comparisons (from providers representing a broad set of motivations), we find no credible evidence that tape costs more than disk per unit of capacity with all relevant costs included, and significant evidence that for the right applications, tape remains a lower-cost storage medium than low-end disk.

Unfortunately, most of this is missing the point. Tape's forte is the provision of a low-cost, data-streaming storage medium with portability, a physical break in data flow and extremely stable media. Low-end disk (serial ATA or parallel ATA) is favored in applications that require streaming access to small files, occasional random access and multiple concurrent accesses. Thus, for any specific set of application needs, the discussion of tape versus disk is ideally a matter of the balance of tape and disk and is not an either/or question.

Herein lies the problem: The management costs of a hierarchy that lacks effective policy-based management systems to do automated data movement and management are not simply additive. More likely they are exponential. The inadequacy of today's storage management tools is forcing organizations to choose between technologies. Clearly, a more rational approach would be to establish an intelligent management layer that controls an extremely granular back-end set of technologies and handles data management and movement based on business needs.

Restore:

- *Return small files to users rapidly.*
- *Avoid disruption by having very recent copies that are protected from "copied" corruption.*

Recover:

- *Provide for rapid rebuild of databases and user data lost as the result of a major event.*
- *Support tough new business continuance goals.*

Discover:

- *Provide cross-enterprise search capabilities to support compliance or governance needs.*

Repurpose:

- *Provide access to reference information (archived) for new revenue-generating business applications.*

5 Changing requirements

The requirements for the application areas (and roles) previously defined in this paper have been evolving rapidly over the past few years.

Data protection (for dynamic data) initially was built around the concept of backup and was optimized for write operations. Discussions of pain points like shrinking backup windows and the focus on the design points of media servers/backup applications reinforce the thought that we have been building a write-oriented architecture for some time. Let's face it — it has always been our hope never to have to restore or recover.

The cost of downtime, aggregation of end-user desktop data and user expectations have led to an environment in which we frequently have to restore small files, and we have to do it quickly.

Continuous-operations technologies like mirroring and replication have provided a partial solution to rapid restore requirements, but they only serve this purpose if they can avoid mirroring or replicating corrupted data. Database logging features inherent in today's database solutions also provide the capability to restore transactions or to back out corrupt data changes. For the last few years, IT organizations have been adding copies into their data protection schemes at an unprecedented rate, exceeding the rate of price per unit of capacity improvement for storage. Some think that data protection has simply become too complex to manage.

Recovery from a broader event or from a disaster has received more attention in recent years due to actual business losses after disasters and regulations requiring business sustainability.

On an annual basis, creation of fixed-content data now exceeds creation of dynamic data, and it is forecasted to grow extremely rapidly. Two primary areas of requirements have emerged. One of the underlying drivers of fixed-content growth has been compliance with regulations and internal governance. Discovery (a content-based search capability across all fixed-content archives) is driving a new set of (still incomplete) solutions. Another, perhaps more positive, driver of fixed-content growth has been a focus by major organizations on customer, partner and supplier relationships as a means of achieving competitive advantage. This has driven organizations to repurpose archived (reference) information in online service applications, driving new opportunities for revenue generation.

Today most of the data that is needed to meet these changing requirements is on tape. To meet some of the requirements, organizations are faced with the decision to use tape for an application for which it is not truly suited, or to move the data to disk solutions. Thus, some large organizations are moving data from tape to disk at higher costs. Apparently the value of these applications is sufficient to justify the additional cost.

Information lifecycle management is a sustainable storage strategy that balances the cost of storing and managing information with its business value. ILM provides a practical methodology for aligning storage costs with business priorities.

Underlying capabilities:

- *Business requirement–driven, policy-based management*
- *Automated data placement and movement*

6 Information lifecycle management solutions

Information lifecycle management is enabled by a set of capabilities that make it a sustainable strategy, able to provide cost/business value balance. These capabilities are not completely deployed in today's solutions. While real benefit from working toward an ILM vision can be achieved today, there is more to come.

At the core of managing storage based on a cost-effective business value model are these elements:

6.1 Business-aware policy management

Service level agreements are defined that provide clear direction to policy management layers serving data protection and reference information management applications. Policy capabilities for data retention and reference information management include policy "verbs" appropriate to requirements. For example, data protection policies allow the user to specify restore or recovery time objectives and then to place data in appropriate locations and forms necessary to meet those objectives. Reference information management policies are available for retention, nonrepudiation (e.g., write once, read many [WORM], serialization and audit trails) and access time. Metadata is maintained for all business data objects to facilitate effective management.

6.2 Automated data placement/movement with a hierarchy of devices

Data is automatically placed in a location (device or layer in the hierarchy) appropriate to its policy rules and use. When appropriate, data is also moved across the hierarchy, based on conditions or changes to value. As requirements change, data can be reclassified to promote its placement (and thus performance or access).

Today's solutions are inadequate to completely meet today's challenges.

Today's users are forced to make tough (and in the future, unnecessary) infrastructure decisions that ultimately will be made via storage system intelligence.

7 Today's solutions

Today's information lifecycle management solutions involving tape and disk are relatively rudimentary when viewed in the context just described. From a data protection (recovery) perspective, the closest thing to a mature ILM solution is a virtual tape management system like the Virtual Storage Manager® (VSM®) system. From a restore perspective, the solutions are copy-based and are extremely costly. Current disk-to-disk-to-tape and disk-to-disk approaches also address only part of the data protection requirements, needing extensive disk mirroring and replication to augment their functionality.

Reference information management solutions are generally in an even lower state of development. Very few (if any) combine effective storage hierarchies, content-based searching, business value–driven policy management, retention management and compliance functionality. Partial solutions involving only disk will eventually prove to be cost prohibitive. Partial solutions involving only tape automation will not effectively serve discovery needs and will require a substantial amount of human intervention to serve up data for reuse.

Continuous data protection solutions automating data protection and using a combination of back-end devices will revolutionize the way restore and recovery requirements are met.

- *Tape automation will be critical to effectively managing the cost of massive CDP, while meeting business requirements for safeguarding data.*

8 Tomorrow's solutions

8.1 Data protection

The ultimate data protection solution we can envision today serves both restore and recovery needs. It will be able to serve information back to an application in its optimum form, for either large file or file system data recovery needs, and will be able to restore a small file immediately. It will provide protection from data corruption and will meet requirements for cost. Generally such a solution will be based on the needs of getting data back to applications rather than on writing it.

Enterprise-class continuous data protection (CDP) solutions with integrated (and transparent) back-end architectures involving disk and virtualized tape automation will resolve most restore- and recovery-related needs. Policy management based on recovery time/point objectives defined in service level agreements will drive the solution. Tape automation will have a critical role in containing the cost of CDP solutions.

The fully developed CDP solution will be a "set and forget" implementation. Today's massive administrative costs for data protection will be reduced to setup and change management activities in response to business changes.

Intelligent reference information management solutions will provide discovery capabilities across all archived fixed content and will automate the repurposing of data, driving competitive advantage.

- *Tape automation will likely become more critical to cost-effective reference information management as data volumes grow. By its very nature, reference information is purely additive.*

8.2 Reference information management

Reference information management (RIM) solutions supporting retention management, including support for compliance requirements, will ensure that reference information is managed to best meet repurposing and discovery requirements. In an intelligent RIM solution, a tiered storage infrastructure will be used to contain costs.

Massive tape archives, either contained today in offline libraries or managed via file access in nearline libraries, will be ingested into RIM solutions to create content-based indexing and to allow for transparent placement in the back-end hierarchy. Access to all data in the RIM solution will be independent of the storage media. Policy based on user-driven metadata and policy rules will ensure proper data placement based on application and business requirements.

The ability to repurpose information (use it for a different application than the one for which it was created) will turn years of collected fixed content into a true business information asset. Customer-, partner- and supplier-facing applications will drive revenue opportunities. Today much of this data is in proprietary independent software vendor formats that will require services to translate it and facilitate its ingestion into RIM systems.

Tape automation solutions integrated into the back end of RIM solutions will be crucial. Reference information will be the vast majority of an organization's data and may have unknown future value; thus, low-cost retention will be required. Organizations today are finding innovative uses for information that has been in archives for years, having had no conceived future use, further reinforcing IT's inherent "save everything" behavior.

High-end tape systems are more cost-effective than disk, and most users are satisfied with them:

- *Tape automation is an extremely cost-effective way to store and retrieve data, and it contributes to significantly lowering the overall cost of IT services when used appropriately.*
- *Migrating data from tape to disk is a short-term phenomenon and is a potentially unwise decision driven by the lack of solutions for automated data movement.*
- *Tape automation is an important layer in an ILM hierarchy today and will remain so for the foresee able future.*

9 Summary

9.1 Tape automation is alive and well

Tape automation remains a viable, growing and high-innovation market with a substantially lower total cost of ownership than capacity disk. Tape automation's administrative cost per unit of capacity is on a par with or lower than disk's administrative cost per unit of capacity.

Tape automation's forte is providing a low-cost data-streaming storage medium with portability, a physical break in data flow and extremely stable media.

No single storage layer (solid state, performance disk, capacity disk, performance tape or capacity tape) will meet service requirements at the lowest possible cost. It is only through the thoughtful balancing of an information lifecycle management hierarchy that the cost of service can be optimized.

9.2 Tape migration to disk is a short-term phenomenon

Migration of data from tape solutions to disk solutions is currently a trend in many large organizations.

We believe this is more about the evolution of requirements than about any inherent difficulties or dissatisfaction with high-end tape automation. Further, we believe that the apparent lack of broadly effective automated data management and movement solutions is driving this movement.

Rather than making either/or decisions regarding data placement and incurring higher cost to serve the occasional use of a specific piece of data, a longer-term and more business-effective approach to data management should be investigated.

9.3 Tape automation has a long-term role in ILM

A truly intelligent layer involved in writing data with the goal of reading (and using) it is a key design point, and providing policy-based management and movement will eventually obviate the either/or decisions that today's large IT organizations face.

The development of effective data management, combined with the increasing volumes of data, will require a strategic reinvestment in tape automation to effectively manage volume and demand.

A complete assessment of information management needs addressing business requirements and planning, and a careful — but aggressive — ILM strategy, will address current and future needs …

… and will move IT to a position of great contribution to business value.

10 Recommendations

Information lifecycle management is not a technology-based decision model. It requires significant investment in business/IT alignment, data classification, metadata development and data movement technology. Rather than facing a collection of one-off decisions regarding tape versus disk and suboptimizing overall infrastructure, organizations would be wise to consider an investment in ILM wherever possible to meet the full requirements of applications in the most cost-effective manner.

Today, most ILM solutions are point solutions, but on an application basis, they can provide the benefits of information lifecycle management. Open systems HSM and intelligent virtual tape automation products are available today to provide some of the benefits. Ensuring that interim solutions have reusable components also makes good sense.

A complete assessment of your current information/data management approach may be warranted. After completing a careful inventory of requirements focused on data protection and reference information management and defining a long-term strategy, several of our customers have achieved real — and significant — savings. Taking a systemic view of your organization's information management requirements and mapping out a long-term plan to achieve advanced information lifecycle management will yield significant benefits and will help drive IT/business alignment, making IT a strategic contributor to competitive advantage.

ABOUT STORAGETEK®

Storage Technology Corporation (NYSE: STK) is a $2 billion global company that enables businesses, through its information lifecycle management strategy, to align the cost of storage with the value of information. The company's innovative storage solutions manage the complexity and growth of information, lower costs, improve efficiency and protect investments. For more information, visit www.storagetek.com, or call 1.800.275.4785 or 01.303.673.2800.

WORLD HEADQUARTERS

Storage Technology Corporation
One StorageTek Drive
Louisville, Colorado 80028 USA
1.800.877.9220 or 01.303.673.5151

TT 0011 A 05/05