WHITE PAPER

Purpose-driven storage

Using information lifecycle management to reduce storage costs, manage information growth and enhance data protection

The amount and importance of data continue to grow while IT budgets remain flat. Against this backdrop, incremental data storage and management strategies fall short. Information lifecycle management responds to these challenges. It reduces overall cost per managed gigabyte by aligning data value with storage infrastructure costs. An information lifecycle management process can be applied to existing infrastructures to accommodate data growth and tighter budgetary controls.

APRIL 2004

StorageTek®

1 Executive summary

- Information lifecycle management is most useful when viewed as a process for aligning information's value with the storage infrastructure, rather than as the infrastructure alone.
- A total system approach to information lifecycle management yields the highest return on investment, even if deployment is distributed over time in limited-scope projects.
- Information lifecycle management is best implemented in three steps: assess current assets, adapt the infrastructure, maintain the information-to-storage balance.
- The thoughtful application of an information lifecycle management process has helped organizations free primary disk space, reclaim lost staff productivity, better comply with data retention regulations and simplify backup and recovery processes.

2 The challenge

2.1 Managing rapid data growth in a cost-effective manner

According to Giga Research, data volumes grew at a rate of 40 to 75 percent in 2003, despite a slow economy. Along with this continuing growth, storage issues are becoming more complex. Enterprises are dealing with an increase in the variety of data types, in the variety of data uses, and in the retention periods for data.

The typical response to data growth is to maintain the status quo. For IT managers, this means: Buy more disk, back up all data in the same manner, and manage the increasing volumes of data through established labor-intensive processes. But this tactical response to data growth has an inherent problem: It isn't sustainable. It works only if IT budgets are increasing and if an organization has or can add the staff resources necessary to manage the increasing data loads.

This isn't the case at most enterprises. A recent Information Week survey of 400 business-technology executives showed that, on average, IT's share of the corporate budget will decline in 2004 — despite improving business conditions. Typical IT organizations can't simply throw money at the problem by buying more primary disk, nor can they throw more staff at the resulting storage management problem. Budget constraints demand new strategies that enable an organization to do more with less — or to store more data at a lower cost per gigabyte.

A process-based, total system approach to information lifecycle management provides this new strategy. This approach balances the costs and performance of storage options with the purpose of data at different points in its lifecycle. It distinguishes the value of a piece of information based on such characteristics as its age, its importance to day-to-day business operations, user access requirements, and legal and regulatory guidelines.

By better matching the purpose of information with the cost of storing it, this approach responds directly to the challenge of rapid data growth in a time of falling IT budgets.

3 Information lifecycle management

3.1 A new perspective

Information lifecycle management is not a new storage device, solution or management technology. It is a process for managing information that can be applied today with readily available resources and technologies to improve IT productivity. Once an information lifecycle management process has been implemented, adapting to information growth — and new opportunities to use data — becomes more manageable.

A total systems approach to information lifecycle management can be applied to improve the way an organization manages data, uses data, protects data and keeps data. This approach can address those needs either individually or collectively.

Specifically, an information lifecycle management process helps reduce the cost and complexity of information-related challenges in four areas: information management, primary storage, data protection, and archiving and compliance.

Enterprise storage options vary widely in performance and cost. Information lifecycle management leverages this diversity to reduce over-dependence on expensive disk systems. With an appropriate information lifecycle management process in place, more data can be managed without proportional growth in budget. The end result: better management of data at a lower overall cost per managed gigabyte.

Disk and tape purchase cost per gigabyte	
Enterprise disk (FC, SCSI, FICON, ESCON)	$50–$70/GB
Mid-range disk (SCSI, FC)	$20–$35/GB
Low-cost disk (S-ATA, JBODs)	$5–$15/GB
Automated tape	$0.75–$3.50/GB
Optical media	< $10/GB

Table 1. Source: Horison Information Strategies and Giga Research, 2003

3.2 Managing data according to its purpose

A total systems approach to information lifecycle management strives to manage information according to its purpose. This means that an organization needs to:

- Automate management tasks to reduce the cost and complexity of data management.
- Match data performance needs with the most appropriate and cost-effective storage options.
- Match data value with the most appropriate and cost-effective data protection options.
- Match data retention requirements with compliant archive and retrieval options.

Ultimately, a total systems approach to information lifecycle management offers a way to optimize storage environments by classifying data into value levels, classifying storage options into tiers and making certain that the value of the data is used to determine its location in the storage hierarchy.

A total systems approach to information lifecycle management considers all aspects of a company's storage infrastructure. Information requirements are increasingly interdependent across systems and applications. Once a holistic implementation plan is developed, changes to processes and the infrastructure can be deployed over time.

3.3 Automating data management for reduced complexity

Information lfecycle management uses automation tools to better manage data throughout its life. Systems and processes are designed to reduce the cost and difficulty of operating increasingly complex storage infrastructures.

Under a typical information lifecycle management storage hierarchy, data moves as its role changes. Data begins in memory on a server. Once the data has been created, it moves to disk. Some data moves directly to secondary disk. Other, more valuable, data moves to primary storage and/or solid-state devices.

Some data is sufficiently large and sequentially accessed to be moved directly to tape. As time goes on, the data/storage balance is maintained through data movement within the tiers of storage (primary disk, secondary disk and tape). Behind it all is an archive. Any data that must be retained for the long term is duplicated to the archive layer.

3.4 Matching performance needs with storage

An information lifecycle management strategy should strive to closely match data performance needs with storage capabilities to improve efficiency. Higher value data that requires faster access times is stored in higher-cost, higher-performance systems. Lower value data is stored in more cost-efficient systems with slower access times. (See Figure 1).

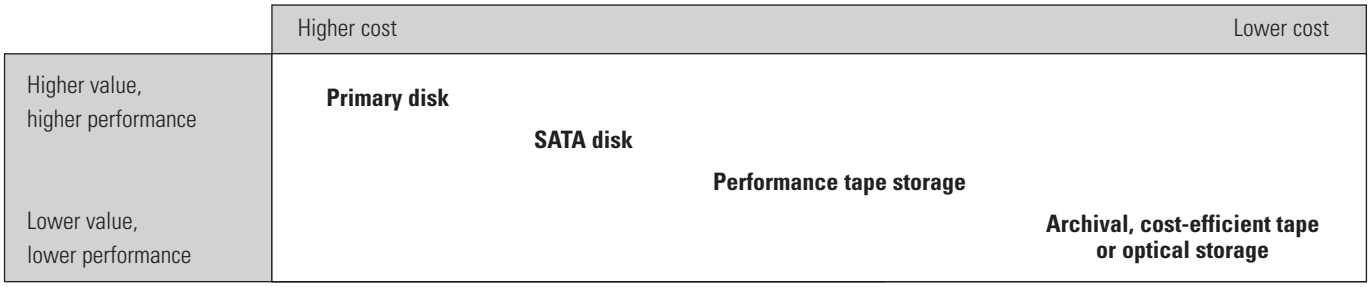

Figure 1. Matching performance needs with storage options.

3.5 Matching data value with protection

With a total systems approach to information lifecycle management, data is protected based on its value, using a tiered protection methodology. This approach strives to lower the total cost per gigabyte of protected data — while making certain that all data is appropriately protected. (See Figure 2).

3.6 Matching data retention requirements with archiving

With a total systems approach to information lifecycle management, data is retained based on its likelihood of recall, using a tiered archive methodology. This approach strives to lower the total cost per terabyte of archived data — while complying with information retention and protection regulations for its recall ability — measured by its integrity, search ability and retention period. (See Figure 3).

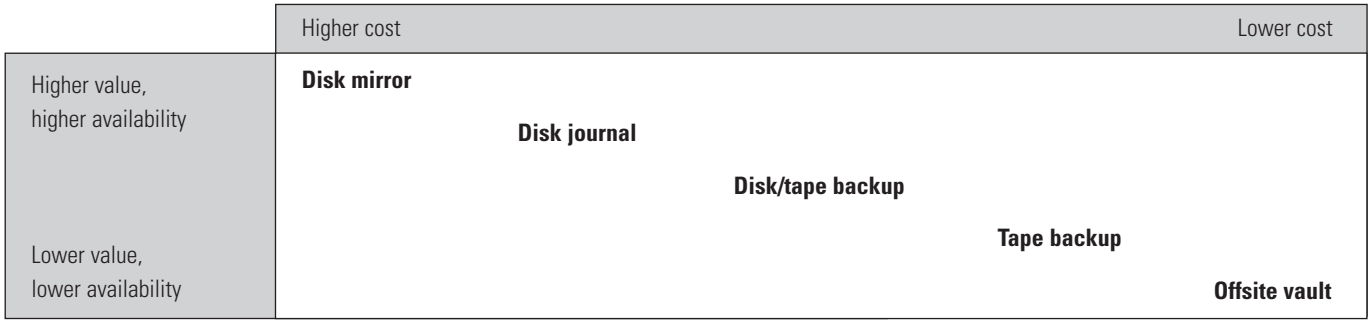

Figure 2. Matching data value with protection options.

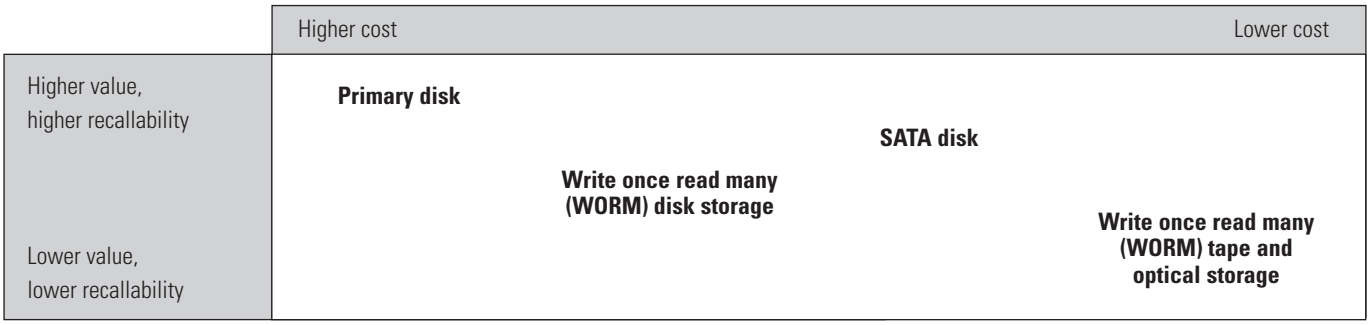

Figure 3. Matching data retention requirements with archiving options.

4 Applying an information lifecycle management process

4.1 A three-step approach

A total systems approach to information lifecycle management can be applied today to improve the productivity of information infrastructures. This approach does not hinge on the availability of new technology, nor does it require the purchase of costly hardware or software solutions. It can be implemented today by making strategic improvements to a current storage infrastructure.

An information lifecycle management implementation encompasses three steps:

- Assess information assets and uses.
- Adapt the storage infrastructure to store information according to its changing purpose.
- Maintain the data balance with integrated information management tools.

Each step in the process includes specific steps that support the broader information lifecycle management strategy.

4.2 Step one: assess information assets and uses

4.2.1 Classifying data by purpose

The information lifecycle management journey begins with classifying data according to its value and purpose. This process values different types of data across one or multiple dimensions to better understand the gaps. Dimensions can include the applications that data is used in, the role of the data in key business operations, and the expected frequency of access to the data. Data can then be segmented into performance, cost, retention and availability requirements — from critical to simply required.

The data classification process might yield a conceptual model similar to Table 2. As this table suggests, the age of data often relates to its value. This, for example, is the case with e-mail — a source of data that threatens to overwhelm many organizations. According to research from the University of Berkeley, the probability that a user will need to re-access an e-mail message falls to less than 50 percent after three days, 25 percent after two weeks and less than five percent after 60 days (source: Horison Information Strategies).

Value data across one or more dimensions	Segment data by performance, cost and availability requirements	Store and protect data classes appropriately
Applications — ERP to e-mail	Critical	Synchronous mirrors; Asynchronous journaling to offsite backup
Modules in applications — financial reporting to employee databases	Vital	Asynchronous mirrors; Journaling or snaps to ATA disk; Offsite backup to tape
Users — senior executives to rank and file	Sensitive	ATA for primary copy; Daily incremental tape backup; Weekly offsite full tape copy
Time of expected next access — days, weeks or months	Required	ATA for primary copy; Daily incremental backup to ATA; Biweekly offsite full copies on secure media technology

Table 2.

Yet in many organizations, virtually all e-mail is stored in online primary disk systems. This suggests the opportunity to cut costs by automatically migrating older e-mail from online systems to lower-cost disk and later to tape-based archiving systems. A detailed data classification system can highlight such opportunities for storing information in a more cost-effective manner.

In practice, the value of any particular data is influenced by multiple variables that must be factored into an information lifecycle management process. The age of data is just one of many variables that must be considered.

4.2.2 Assessing the current environment

Once the value and role of data is known, assessment of the effectiveness of the current storage strategy and tools is required. This phase seeks to compare the performance of the current infrastructure with the purpose of the data being managed. Assessment includes identifying unused capacity and highlighting opportunities to move data from primary disk to lower-cost storage, such as ATA disk or automated tape libraries.

Step one of the information lifecycle management journey is essentially a series of focused assessments. Each assessment should yield a close-up view of current storage issues and opportunities. Specific areas of study should include the following:

Gap analysis — Identify opportunities to improve global processes as they relate to storage. These gaps can appear as manual processes, loss of personnel, or inefficient use of software and hardware.

Backup and recovery assessment — Look specifically at opportunities for gains in backup and recovery processes. Examine such issues as backup staffing, errors in backup processes, resource utilization and backup performance. Recovery deserves special attention, since recovering data typically takes three times as long as backing it up.

Storage profiling — Survey the storage environment, including hardware, applications and file systems. Understand how many copies of replicated data are sitting on your primary disk. Observe data aging patterns and identify opportunities to move cold data off primary disk to free space on your performance disk. Identify unused capacity and table space consumption. Examine network performance. Evaluate backup, e-mail and database application performance and usage.

SAN assessment — Create a map of the existing SAN environment and identify areas that could benefit from optimization. Identify opportunities where the addition of a SAN would fill gaps in processes or the information infrastructure.

4.3 Step two: adapt the storage infrastructure

Step two puts the principles of information lifecycle management to work by aligning the current storage infrastructure and processes to address the gaps and overlaps identified through the earlier assessments. This step addresses information challenges in four areas: information management, primary storage, data protection, and archiving and compliance.

Information management — Simplify storage infrastructures by consolidating management and support systems. More fully utilize capacity by putting systems in place to replicate data consistently to the most appropriate media — and only as needed. This more efficient management of data copies should reduce the cost and difficulty of operating a complex storage infrastructure.

Primary storage — Make more efficient use of existing primary storage by reducing the use of unnecessary and expensive primary mirrors. Open up primary disk storage capacity by moving targeted data to secondary disk and tape-based solutions, and deploying unprovisioned capacity. These efforts help meet growing online storage requirements and reduce the need to purchase more primary disk.

Data protection — Simplify backup and restore processes to reduce data production labor and improve the reliability and speed of backup and recovery. Simplification comes through measuring actual backup and restore performance for backup window length and recovery time, including identifying root causes of failures. Next, a plan is developed to align actual performance capabilities with protection service level agreements (SLAs). This work should result in faster and more reliable backup and recovery with reduced labor costs and administrative burdens — and shorter problem resolution times.

Archiving and compliance — Deploy solutions to improve archiving efficiency and make certain that information is retained and can be recalled to meet compliance and business needs. Compliant systems meet requirements for information integrity (safety and traceability),

accessibility and retention periods. Store and maintain data to match varying requirements with automated management and migration tools. With a total system approach to information lifecycle management, archiving capacity can grow as needed because only a fraction of the archived data is stored on higher-cost disk.

4.4 Step three: maintain the data balance

The third step of information lifecycle management implementation focuses on maintaining the data balance. The work here uses integrated information management tools to automate the adjustments to the storage infrastructure that keep data performance, protection and retention systems in line with data's changing role. Storage resource management tools and well-defined change management processes can keep the information infrastructure optimized through normal business change.

Maintaining the balance also includes ongoing refinements to the data value classification system developed in the Phase I assessment process. The classifications established then should be periodically reviewed and adjusted to account for changes in the business.

What percent of data is currently classified as critical? Has the data of certain users risen in importance to the business? Do access and performance requirements need to change to address shifts in business priorities? These are the kinds of questions that should be reviewed on an ongoing basis to make certain that a data classification system remains current.

This ongoing work enables the realization of the information lifecycle management vision — in which data is stored and protected according to its value to the organization.

5 The opportunity

A total systems approach to information lifecycle management can generate significant, quantifiable gains. These gains stem from the implementation of systems and processes that enable more cost-effective approaches to information management, primary storage, data protection, and archiving and compliance.

The opportunities presented by a total systems approach to information lifecycle management are illustrated in the following case studies.

5.1 Information management: Migrating data automatically

Following information lifecycle management principles, a large corporation cut its storage costs by automating the management and migration of payroll processing data.

The company's payroll application requires the highest availability and performance, so payroll data is stored on primary disk or in a virtual array system. This information is protected continuously via high availability connections and by making instantaneous copies of data every two hours.

At the end of the pay period, payroll data is migrated from the higher-performance disk to mid-range ATA disk. At this stage, users can access payroll data from the company's Web site for a period of three months.

After three months, the data is written to a tape library, which is on the same campus as the data archive. For disaster recovery protection, a storage domain manager mirrors the data to a remote location, where it is stored on a backup tape library.

5.2 Primary storage: opening up valuable disk space

In the course of assessing its information management needs, one organization determined that 38 percent of the data held on primary disk had not been accessed over the past year. This finding suggested the opportunity to cut costs by moving this inactive data to lower-cost storage.

Ultimately, this organization opened up a terabyte of disk space by moving seldom-accessed data to lower-cost disk and tape. This was accomplished via the implementation of processes for rules-based archiving to secondary disk and then to tape. This freed up costly primary disk space for storage of current data that requires the fastest recall times.

At the same time, this change lessened management burdens by eliminating the need to mirror the data to primary disk, back it up weekly and ship it to an offsite disaster recovery location.

5.3 Data protection: automating backup processes

A multi-campus community college used information lifecycle management principles to back up growing volumes of application data in a more cost-effective and nondisruptive manner. Based on management policies, a disk-based data protection appliance determines which

volumes of data should be protected and for what length of time, and then creates an image copy of the protected volume.

Backups are now performed from image copies rather than the protected servers. This effectively eliminates backup windows and the associated server downtime. Additionally, the college can now restore a corrupted database in minutes rather than days.

5.4 Archiving and compliance: reducing e-mail costs, burdens

A global technology company with more than 5,000 employees used information lifecycle management principles to address skyrocketing e-mail growth. E-mail usage was growing faster than IT resources, resulting in restrictive policies that reduced end-user productivity.

An information lifecycle management assessment identified the opportunity to use policy-based data management to create what amounted to bottomless mailboxes for users. The ensuing implementation of a disk, tape and software solution led to dramatic gains. The company recovered 321,000 hours of wasted end-user productivity, reduced total cost per e-mail user by 75 percent and recovered nearly one terabyte of network storage.

5.5 Realizing information lifecycle management benefits — today

These specific cases illustrate the types of gains made possible by a carefully applied information lifecycle management process. While it remains a vision for many organizations, information lifecycle management is already a reality for growing numbers of businesses. The successes of these organizations show that an information lifecycle management strategy can be applied today to reduce the cost and complexity of storing and managing information.

6 Recommended actions

- Prioritize your highest growth environments and highest value applications for improvement with information lifecycle management.
- Consider a thorough storage assessment before undertaking any information lifecycle management-related changes.
- Don't limit the assessment to storage capacity. Create a list of your most pressing information challenges. Include current systems, budgets, staffing and regulatory requirements.
- Gain business management support by briefing executive-level personnel on your organization's information challenges and potential opportunities.

7 Planning considerations

- Does your vendor have an established process for implementing information lifecycle management?
- Does the assessment process address your entire storage infrastructure and all of your storage management processes?
- Does the information lifecycle management plan include the optimum mix of disk and tape storage systems, given the range of your information needs?
- How much of your existing infrastructure can be preserved when implementing the information lifecycle management plan?
- How much existing primary disk capacity will the information lifecycle management plan reclaim?

ABOUT STORAGETEK®

Storage Technology Corporation (NYSE: STK), a $2 billion worldwide company with headquarters in Louisville, CO, has been delivering a broad range of storage management solutions designed for IT professionals for over 30 years. StorageTek offers solutions that are easy to manage, integrate well with existing infrastructures and allow universal access to data across servers, media types and storage networks. StorageTek's practical and safe storage solutions for tape automation, disk storage systems and storage integration, coupled with a global services network, provide IT professionals with confidence and know-how to manage their entire storage management ecosystem today and in the future.

StorageTek products are available through a worldwide network. 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

MC 1040 C 04/04