WHITE PAPER

Optimizing storage capacity

Reduce storage costs by reserving your most expensive storage for your most critical data

IT organizations face a rapidly growing volume of data. Adding more primary disk capacity to manage data growth is a costly and non-sustainable strategy. Before investing in new capacity, IT managers should rationalize their existing storage infrastructure to maximize use of existing capacity. The storage infrastructure can be made more efficient by assessing data uses, eliminating unnecessary copies, moving less critical data to less expensive disk and repurposing allocated but unused capacity. Optimizing existing storage assets can reduce storage costs by delaying or eliminating the need for new primary capacity to manage information growth.

APRIL 2004

StorageTek®

1 Executive summary

- Adding more of the same disk capacity to manage data growth is an unsustainable strategy.
- By first rationalizing the existing storage infrastructure to maximize its use, IT managers can delay or eliminate the need for new primary capacity to manage information growth.
- Information lifecycle management concepts can be applied to assess data uses and storage assets, identify storage inefficiencies and adjust the storage infrastructure to maximize storage utilization.
- Efficiencies are gained through eliminating unnecessary and inefficient copies, moving less critical data to less expensive disk and recovering over-reserved capacity.
- Balancing data value and storage requirements can result in less dependence on expensive high-performance disk and a lower cost per online gigabyte.

2 The challenge

2.1 Managing data growth

Across multiple industries, the growth rate for data greatly outpaces growth in IT budgets and IT staffing levels. In 2003 alone, Giga Research estimates that data volumes grew at a rate of 40 percent to 75 percent.[1]

- According to Enterprise Storage Group, fixed-content data such as e-mail, presentations, reports and video files is now growing at an annual rate of over 90 percent.[2]
- E-mail alone is becoming a huge management problem at most companies. Storage Magazine estimates that a 3,000-person organization generates one terabyte of e-mail in a year.
- Customer transaction data has mushroomed, as consumer product companies collect and analyze terabytes of data concerning in-store or online customer behavior.
- Recent U.S. government regulations such as the Sarbanes-Oxley Act and the Health Insurance Portability and Accountability Act mandate multiple-year retention periods for certain types of data. Some form of regulations regarding data archival affects virtually all businesses.

The mismatch between data growth and budgets means IT departments can no longer throw money at the problem by adding more of the same expensive disk capacity to store their new data. They need an approach that optimizes their valuable high-performance primary capacity and maximizes the use of lower-cost storage classes — Serial Advanced Technology Architecture (SATA) disk and tape.

3 Storing data according to use frees capacity

IT managers can take practical steps today to manage the growing volume of data. These steps are based on the principles of information lifecycle management (ILM). An ILM-based approach can uncover additional capacity from within existing high-performance primary disk.

Information lifecycle management assumes that all data is not equal. Data is used differently at different points in its lifecycle. For example, monthly financial reports may combine sales orders, shipments, inventory, expense and other data for the month. During the processing cycle the finance department needs real-time access to verify and analyze this data. After the reports are generated, the previous month's data is referenced less frequently as the focus changes to data for the current month.

Storing data according to its use reduces the need to add additional primary capacity, and can free unused capacity for application growth.

[1] "IT Trends 2003, Midyear Update: Storage Hardware," June 11, 2003, Giga Research.
[2] "Reference Information: The Next Wave," 2002, Enterprise Storage Group.

4 Optimizing storage capacity

4.1 A three-step process

IT organizations can optimize primary storage allocation through a three-step process.

1. Review the current environment.
2. Align data performance and availability requirements with storage class.
3. Refine and automate storage processes to maximize capacity use.

4.2 Step one: review the current environment

Reviewing the storage environment has two areas of focus. First, IT must evaluate the data they manage. They need to understand what data they have, rank it in terms of business value, and compare the value ranking to the type of storage system they are currently using.

The second area to look at is the storage environment itself. An audit can reveal the current allocation of high performance disk. This process may identify unused capacity that can be recovered and reallocated.

The review will result in a list of sub-optimal uses of capacity, as shown in the following table. Once these are identified and prioritized, IT managers can develop action plans to free primary capacity.

Data valuation starts with listing data types and ranking them based on how often the data is accessed, who uses the data, or other measure of business value. IT managers can map the current location of the data across a hierarchy of storage systems, from high-performance disk to low-cost tape systems. The resulting map must include all data that IT is accountable for. They can then compare the data's value and performance and availability requirements of the data with where it is stored.

Causes of underutilized storage	Action
·· Non-critical data on high-performance disk	·· Assess use of data and map data to appropriate storage class
·· Data not referenced in past six months	·· Transfer to less expensive storage
·· Protection copies of less critical data	·· Transfer to less expensive storage
·· Storage allocated but unused	·· Review and re-justify allocations
·· Stranded capacity in decentralized disk systems	·· Consolidate into SANs where appropriate
·· Redundant copies (test copies, unnecessary duplication, no defined purpose)	·· Delete

In Figure 1, each dot represents a collection of data related to a specific application. The groups of data are stored in a range of storage classes, but not aligned with the performance and availability offered within each class. For example, disk mirrors, shown in Figure 1 as two data points connected by dotted lines, are often stored on the same class of storage. Yet the protection copy — the lower data point — may have lower performance requirements.

In a lifecycle view of data, age can be a key factor affecting data value. For example, e-mail storage threatens to overwhelm many organizations. The IT staff may find a rapid fall-off in access to e-mail messages after even a few days, and virtually no access to messages more than a month old. If the organization stores all e-mail on primary disk, an opportunity exists to move the older messages to secondary disk or tape. This is just one example of cost recovery that a rigorous data valuation exercise can uncover.

Age drives the usage frequency of some data — but not all. There are many variables that can influence data performance requirements. The classification process provides a reality check for business operations by identifying the variables and their relative weight. Objective measures such as the age of data and how often data is accessed are easy to list. These must be combined with a more subjective weighting of the data's value. Some data may be old and rarely accessed, such as last year's presentation by the COO to a security analyst group. Nevertheless, storing it in the executive assistant's e-mail folder may be justified when the CEO asks for a copy.

4.2.1 Assess disk utilization

The other part of this step is to determine current disk utilization, which can vary widely. Utilization of typical direct-attached storage environments rarely exceeds 50 percent.[3] One factor contributing to this inefficiency is capacity allocation for future growth. Reserved capacity that is never used is rarely identified and recovered.

Another factor that contributes to stranded capacity is decentralized storage management. A large number of non-networked or local network disks may collectively have reserved more extra capacity than necessary. Identifying and recovering this capacity is complicated by the lack of a single management view of all disk resources.

Data protection and recovery strategies are an intrinsic part of a primary storage assessment. For example, physical rotation of disk mirrors multiplies disk capacity. For every change on the primary disk volume, an identical change is made to a redundant backup volume. To get 24 hours worth of protection, a current copy could be rotated out every three hours and replaced by the oldest copy from 24 hours ago. This implies duplicating the primary disk eight times for instantaneous recovery from hardware failure and limited corruption protection. The same level of protection may be attainable by mirroring the data to lower cost disk. Taking such a step must align with the overall data protection strategy.

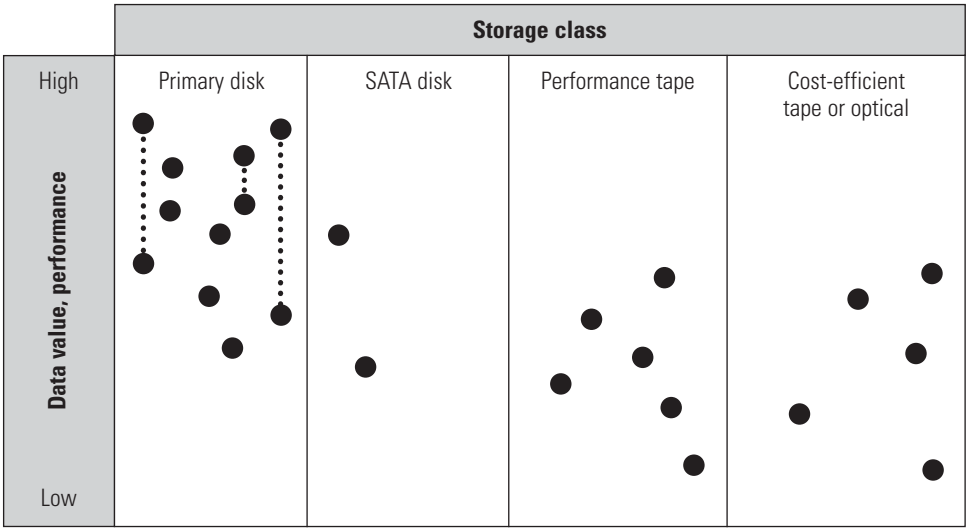

Figure 1. Mapping the value of data by storage system.

[3] "Improving Storage Utilization: The Benefits Are Often A Myth," January 23, 2003, Gartner Inc.

4.3 Step two: align data performance requirements with storage class

The assessment phase should yield a clear view of the requirements of different data types and how well these requirements are met in the current environment. IT can then move data where appropriate to lower-cost and lower-performance storage classes while still delivering adequate performance for mission-critical applications.

4.3.1 Adjust the protection strategy

A key aspect of optimizing storage is to use available technology to accomplish the same level of performance and protection at a lower cost. For example, mirrored physical copies may sometimes be replaced with mirrored logical copies. This technology stores only the data that has changed from one copy to the next, rather than creating a new complete physical copy. Primary disk capacity that had been used for physical mirror copies is now available for application growth.

In a similar way, Serial Advanced Technology Architecture (SATA) disk arrays can store data at a fraction of the cost of high performance disk. In a similar way, SATA disk arrays frees capacity on expensive primary disk.

Point-in-time copies also often reside on expensive primary storage. Point-in-time copies are used for both time shifting backup and for quick restore. When time shifting backup, moving copies to SATA can make a lot of sense. When quick restore is the rationale for the copies, the question becomes how quick is "quick?" SATA disk can often satisfy this need.

Journaling technology can reduce the cost and complexity of creating point-in-time copies and further reduce dependence on primary storage for data protection. This replication method continuously records all discrete data events for instant recovery to any point in time using SATA disk.

4.3.2 Migrate data

Fixed-content data growth is accelerating, approaching 100 percent annually.[4] Fixed content includes all types of non-changing electronic documents, including memos, presentations, streaming data and medical records. Usage frequency can vary widely from near-constant use to practically no reference after the original creation. Yet much of this content tends to be stored and protected for long periods on primary disk.

A usage assessment of data could enable halving the demand for expensive high-performance disk, cutting expense for both hardware and management. Research has shown that after three days the probability is less than 50 percent that a given piece of fixed-content data will ever be referenced again.[5] An analysis can determine the fall-off point for the probability of re-reference for any given application.

Identify and migrate unused content

For example, in a storage environment with 16 terabytes of primary storage and disk mirroring, eight terabytes of disk might be used for primary application and eight terabytes for data protection. Assuming analysis shows a 50 percent drop in usage, four terabytes of the application data not being referenced could be moved from expensive performance disk to SATA disk or tape.

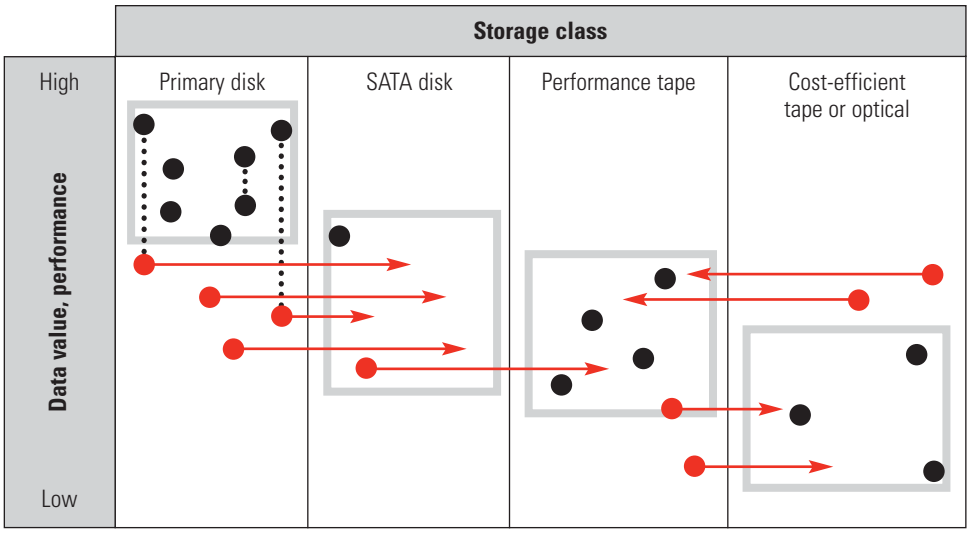

Figure 2. Migrating data to align performance requirements with storage.

[4] "Reference Information: The Next Wave," 2002, Enterprise Storage Group.
[5] Fred Moore, Horison Information Strategies, 2003.

Migrate unnecessary protection copies

A business impact analysis might justify moving protection copies off of expensive primary disk as well. That means the eight terabytes used for data protection could also move to SATA. The end result: four terabytes of primary data on high performance disk out of the original 16 terabytes. Assuming the purchase cost of SATA disk arrarys is one-sixth the cost of high-performance disk, this process recovers a substantial value in primary capacity and lowers the overall cost per online gigabyte.

Twelve additional terabytes of primary disk capacity are now available. With online data growth of 40 percent to 75 percent annually, this process recovers enough reserve capacity to handle the growth of high-performance business applications for several years.

4.3.3 Consolidate storage and recover stranded capacity

Consolidating storage by class can simplify management of the storage environment as it expands to meet business needs. Once data is stored to align with performance and availability needs, consolidate storage systems in like storage classes with Storage Area Networks as much as possible within each class. When feasible, include automated data migration methods to transfer data between storage classes based on age, usage or other appropriate trigger criteria.

Consolidating storage in fewer systems can reduce disk capacity waste due to underutilization. In a decentralized approach, capacity typically gets allocated by application. If the allocated capacity is not used as planned, the extra capacity is stranded within the network serving that application. Creating a pooled storage system enables a smaller amount of total disk capacity to be reserved for application growth. This keeps the reserved capacity to a minimum, and keeps it visible. A pooled approach also simplifies management tasks by reducing the complexity of the storage infrastructure. Highly efficient networked storage environments can operate at 70 percent to 90 percent capacity utilization, two to three times higher than typical utilization for direct-attached storage environments.[6]

4.4 Step three: refine and automate storage processes

Maintaining a balance between data performance needs and storage options is an ongoing process. IT departments can begin this process immediately. As described above, this involves classifying data by value, migrating older unused data from primary storage and recovering stranded capacity from existing disk systems. Automated processes need to be established for repeating these activities. This will ensure the most efficient use of primary capacity as a standard business practice.

Policy-based data management tools can help by automating the migration of data as its value changes. IT should also consider data replication requirements in this equation. New replication processes should be reviewed to ensure business entities can justify the need for additional copies of a class of data, and can quantify the costs.

5 The opportunity

Balancing data value and storage performance can help IT free primary capacity within the current storage environment and gain the following benefits:

- Free wasted primary storage by identifying and moving data with lower performance requirements.
- Accommodate growing data volume.
- Recover stranded capacity due to growing data volumes.
- Improve storage management through consolidation of networked storage systems.
- Reduce dependence on expensive high-end disk for lower total cost per online gigabyte.

5.1 Transitioning data to lower priority storage

One large corporation found a way to balance availability and cost by automating payroll data management and migration. Payroll processing was a mission-critical application, so it made sense to store the data on high-performance disk or virtual array systems during the processing cycle, and to replicate the data every two hours.

[6] "Justifying Disk Storage Networking," January 20, 2003, Gartner Inc.

Once the pay cycle is complete, the automated management system now moves the payroll data to mid-range SATA disk arrays. 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, the data is replicated to a remote location, where it is stored on a backup tape library.

5.2 Finding large amounts of unused data

In another example, a company found that 38 percent of the data it stored on primary disk had not been accessed over the past year. Moving this data to lower-cost disk and tape opened up a terabyte of primary disk capacity without impacting application performance. The IT staff implemented a rules-based process for archiving the data to secondary disk and then to tape. They also eliminated backup tasks associated with the data, which further simplified their management load.

6 Recommended actions

- Consider trusted outside experts to simplify and accelerate a capacity optimization process.
- Start with a data assessment project. Find out what data is there, how it is used and where it is stored. Create a value hierarchy and map data to storage performance.
- Conduct an audit of primary disk utilization and identify unused capacity.
- Prioritize the opportunities for freeing primary disk capacity and create an implementation plan for those with the highest returns.
- Evaluate storage management tools to globally monitor and manage your primary storage infrastructure and processes.
- Measure success against cost per added gigabyte of primary data.

7 Planning considerations

- Does your vendor have an established process for applying the principles of information lifecycle management to data use and primary storage?
- Does the assessment process address your entire primary storage infrastructure and management processes or just a particular application?
- Does the capacity optimization plan make optimum use of the most cost-effective storage systems that meet your data performance needs?

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 1045 A 04/04