White Paper

SGI® InfiniteStorage™ 10000: SATA Systems and Nearline Storage Models

Table of Contents

1.0 Executive Summary 1
2.0 Background: Nearline Storage 1
2.1 Disk-based Nearline Storage 1
2.2 Demand for Disk-based Solutions 1
3.0 The SGI InfiniteStorage 10000 2
3.1 Improved Cost Models with Nearline Storage 2
3.2 Fast Reliable, Scalable Storage 2
3.3 Support for a Large Storage Pool Architecture 3
3.4 Other Benefits of the SGI InfiniteStorage 10000 4
4.0 Nearline Storage Applications 4
4.1 Target Markets 4
4.2 Other Suitable Applications of Nearline Storage 4
4.3 Restrictions 4
5.0 A Tiered Storage Model 4
5.1 Enterprise Storage 4
5.2 Archival Storage 4
5.3 Economy Storage 5
6.0 Summary 5
Appendix: Optimizing Storage Capacity 7
Step One: List All Storage and Data 7
Step Two: Align Data Performance Requirements with Storage Tiers 8

1.0 Executive Summary

Across most organizations, rapid data growth outpaces storage budgets. Adding high-performance main enterprise storage to correct this imbalance is a costly strategy while the alternative, offline archival tape storage, does not have the performance to meet most application needs. Serial ATA (SATA) disks represent a new type of storage solution that can fill the need for something in between—disk-based nearline storage.

As a nearline storage solution, SATA disks complement enterprise and archival data and storage capabilities in terms of reliability, accessibility, performance, and costs. This innovative solution allows IT managers to introduce new levels of storage efficiencies, keeping up with data growth by using affordable SATA disk storage.

The SATA disk-based SGI InfiniteStorage 10000 Nearline Storage System enables the creation of the world's largest, most-scalable, nearline disk-based storage solutions. With built-in intelligence, it delivers continuous on-the-fly data protection with no loss in performance. Organizations can deploy the SGI InfiniteStorage 10000 to implement a reliable enterprise-class nearline solution that can scale up to 480 TB non-disruptively.

As organizations consolidate data into large disk pools, additional levels of robustness and resiliency are required. The SGI InfiniteStorage 10000, with the groundbreaking ability to withstand multiple drive failures within and across individual parity groups, maintains data availability, host performance, and quality of service under the most severe failure modes. SGI InfiniteStorage 10000 allows an organization to build cost-effective highly scalable secondary storage in the smallest footprint available. This space efficiency delivers lower cost in terms of storage purchases and operating expenses. With the industry's best price per capacity per square foot, the SGI InfiniteStorage 10000 makes it possible to introduce worry-free disk-based solutions that cost-effectively serve as a platform for nearline storage, archive storage, fixed-content and large-file storage, and backup and recovery storage.

This white paper provides an overview of the SGI InfiniteStorage 10000 and its capabilities. Nearline storage and tiered storage models are also overviewed to show how the SGI InfiniteStorage 10000 can enable more cost-effective storage deployments for today's organizations and businesses.

2.0 Background: Nearline Storage

Nearline storage is used for applications with less frequent read/write activity. These applications require quicker access and higher throughput compared with long-term archival storage, but do not require the continuous, instantaneous access provided by online solutions using enterprise-class disks. Nearline storage can be classified as "infrequent, fast-read/fast-write" storage.

Prior to SATA drives, tape libraries were the only alternatives for nearline storage solutions. With the advent of the affordable SATA disk and its quicker access and higher throughput, new storage solutions based on large capacity, inexpensive disks are now possible.

2.1 Disk-Based Nearline Storage

Adding more primary disk capacity to manage data growth is a costly strategy, and tape does not have the access or throughput speed to meet production data needs. Organizations need storage solutions that allow them to create a more efficient storage infrastructure by providing the right storage for each type of data, moving less-critical data to less-expensive disk, and efficiently reallocating unused capacity.

Disk-based nearline storage solutions meet all of these needs. By filling the gap between main enterprise storage and tape-based archival storage, disk-based nearline storage dramatically reduces overall storage equipment costs as well as operating costs while satisfying the requirements for application performance.

2.2 Demand for Disk-Based Solutions

Many trends and market conditions are increasing the demand for disk-based nearline storage solutions:

- **Data storage requirements are growing** — For most organizations, the growth rate for data outpaces the growth rate for storage budgets. According to a 2003 study done by University of California at Berkeley School of Information Management and Systems, new stored data grew by 30 percent per year from 1999 to 2002, with 92 percent of data stored on hard disks. This trend is expected to continue over the next few years.

- **Regulatory compliance requires businesses to "save everything"** — Recent legislation, such as the Sarbanes-Oxley (SOX) Act and the Health Insurance Portability and Accountability Act (HIPAA), is forcing companies to retain vastly more information for much longer. That information must be more readily accessible, making tape impractical.

- **IT budgets are flat or shrinking** — While stored data increased by 30 percent per year from 1999 to 2002, most IT storage budget increases were in the single digits. This mismatch between data growth and storage budgets means organizations can no longer afford enterprise storage to meet all of their storage requirements.

- **Customers are realizing that much of the data is inactive** — Experience is showing that 70 to 90 percent of data is inactive, with most new data going inactive after about 30 days. While this data may be important or even vital to the organization, it is infrequently accessed.

- **SATA storage promises** — SATA technology enables large capacity drives that are less expensive than Fiber Channel drives, but have improved performance compared to tape. They can effectively serve as “cheap data tubs.” The combination of cost and performance promises to make SATA drives the best storage for less-frequently accessed but essential data.

3.0 SGI InfiniteStorage 10000 Storage

SATA disk-based SGI InfiniteStorage 10000 storage has many characteristics that make it an ideal nearline solution. In terms of performance, the SGI InfiniteStorage 10000 provides industry-leading levels of throughput in a high-density package. Combining high-performance and high-density disk in a small footprint provides a unique nearline solution that is distinguished from tape and MAID. Also, SGI InfiniteStorage 10000 solutions are up to five times more scaleable than other disk-based storage systems.

This scalability is matched with multiple levels of redundancy and continuous on-the-fly data protection, making the SGI InfiniteStorage 10000 a solution with enterprise-class reliability. All this storage and reliability is housed in the industry’s smallest footprint delivering customers the best price per capacity per square foot storage solution.

SGI InfiniteStorage 10000 storage, with built-in intelligence, delivers continuous data protection and true dual-redundancy disk capability. SGI InfiniteStorage 10000 solutions can overcome double disk failures in the same redundancy group without adversely affecting data availability or system bandwidth. This capability allows vital business environments to implement a reliable enterprise-class nearline solution that can scale up to 480 TB.

3.1 Improved Cost Models with Nearline Storage

To understand the unique value of the SGI InfiniteStorage 10000 as a nearline storage solution, consider the situation of an online Internet-based photo gallery. Online photo storage requires high capacity, but these service providers are very cost sensitive.

Online customers expect it to be easy to store, view, and share photos with friends and family. This service also enables online photo companies to provide additional services like editing and creative tools. These are combined with specialty products such as photo books, calendars, cards, mugs, mouse pads, aprons, frames, and other revenue-generating product opportunities. A key success factor is providing online users the freedom to view what they want, when they want—giving them a positive experience that encourages them to take advantage of the other revenue-producing services.

Internet-based photo gallery providers find that there is a rapid fall-off in access of photos after the first few days, and virtually no access to photos more than a month old. However, with the “view what they want, when they want” business philosophy, short access time and increased throughput rates are an important requirement. In the past, this has forced photo galleries to employ more expensive storage solutions.

SGI InfiniteStorage 10000 storage provides an alternative. Photo galleries can scale to 240 TB in a single rack and to 480 TB in two racks. The parallel, continuous, on-the-fly data protection architecture of the SGI InfiniteStorage 10000 platform delivers reliable access to photos in only milliseconds or seconds.

With power-saving features and a volumetric-efficient package, service providers can reduce overall storage operating expenses. This optimal combination of low-cost scalable capacity and improved availability enables a cost-effective storage foundation.

3.2 Fast, Reliable, Scalable Storage

Organization requirements mandate that a block-level storage device be capable of quickly accessing Logical Unit (LUN) storage targets from very large pools of storage. SGI InfiniteStorage 10000 storage embodies this with a broad, parallel backend that includes twenty fully redundant Fiber Channel loops controlled by multi-processing multi-ported block-level storage controllers. This truly parallel access from very large and deep storage pools driven at wire speeds allows for quick, reliable access for infrequently accessed data.

With large pools of vital business data, reliability becomes a critical consideration. SGI InfiniteStorage 10000 controllers support a fully redundant host and disk-side parallel architecture with on-the-fly dual read/write parity protection. This architectural parallelism enables mass scalability with robust reliability. The InfiniteStorage 10000 also offers dual-redundancy disk capability to overcome double disk failures in the same redundancy group without adversely affecting data availability or system bandwidth. This function provides data access protection even in the case of multiple simultaneous disk failures in individual parity groups—an especially important feature for very large disk pools and modern ultra-capacity drives that can take time to rebuild.

Architectural parallelism extends through the entire system allowing drive configurations that support up to 960 SATA disks. Built into this parallelism are advanced data protection features enabled by RAID protected cache and disks. This parallel, fully redundant, dual-pathing front and back end delivers true mission-critical data integrity and with no loss in performance.

Collectively, these features allow users to:

- Scale storage as large as required using less expensive economy storage
- Provide the highest levels of Quality of Service (QoS) for consistent, predictable data delivery
- Reduce storage operating costs
- Deploy, support, and manage storage simply and easily, reducing management costs
- Ensure high data availability and uptime even while servicing full host loads

3.3 Support for a Large Storage Pool Architecture

SGI InfiniteStorage 10000 storage uniquely supports large storage pools with reliable sustained bandwidth from media. The current flagship SGI InfiniteStorage 10000 platform supports up to 960 disks and, due to the family's in-band parallel parity processing engine (P3E), calculates not only write but read parity in real time for sustained bandwidth performance even in degraded mode. There is no performance impact in the case of a dual disk failure or even under a disk rebuild—and rebuilds can be done in a timely manner.

Using 500 GB SATA disk drives, a single SGI InfiniteStorage 10000 can scale to an impressive 480 TB in total capacity. The platform's ability to virtualize the storage through LUN aliasing, WWN masking/filtering, and port zoning enable very easy deployments and simplified ongoing system management. Additionally, a large variety of statistical data is available to enable straight-forward tuning, optimization, and network troubleshooting.

The unique parallel architecture of the SGI InfiniteStorage 10000 allows even a single "PowerLUN" (RAID disk target) to provide the entire aggregate bandwidth of the system. This is accomplished without file system-based software striping that can cause excessive switching latencies and storage-port contention in Fiber Channel fabric switches. Enabled by its shared-memory architecture and advanced cache coherency capabilities, the parallelism of the SGI InfiniteStorage 10000 allows for far greater scalability than any system built with generic RAID systems. Almost all generic RAID systems are designed as general-purpose devices for databases and back-office environments with smaller disk support requirements. With the SGI InfiniteStorage 10000 multi-port architecture, many SAN storage networks may be built without any Fiber Channel switches at all.

Figure 3-1 contrasts the SGI InfiniteStorage 10000 approach on the left with a comparable generic storage area network (SAN) on the right. The generic SAN is characterized by multiple RAID systems (typically dual controllers, with at least two LUNs per controller) with a limited number of disks striped by the file system across and through a Fiber Channel switch. Note that

Figure 3-1: SGI InfiniteStorage 10000 with ten 48-slot trays in a single rack

each and every computer, for each and every I/O it performs, must stripe across all of the eight storage ports (an "eight-way stripe"). Each one of those I/O operations therefore generates eight switching events, encumbers eight switching latencies, and monopolizes each switch port connected to the RAIDs for one-eighth of the time. This does not include wait states introduced when ports are busy with other computer's requests. When all eight computers operate simultaneously, RAID switch ports can become congested and impose further delays—a situation very similar to getting a busy signal on a telephone switch. Port contention and the build up of switching latencies quickly rob access performance from the SAN and severely inhibit scaling.

In contrast, the SGI InfiniteStorage 10000 provides access in a much more efficient and reliable manner. Access to the storage can be compared to having many straws in a glass of water—each of the eight computers in the diagram has private and full-time access to the SGI InfiniteStorage 10000 "liquid" (storage), without any switching events or port contention. All straws can simultaneously access the liquid, for truly parallel sharing.

SGI InfiniteStorage 10000 systems offer simplicity and superior scalability because of this unique parallel architecture.

3.4 Other Benefits of the SGI InfiniteStorage 10000

Host and disk interface modularity is implemented using a well-defined physical hardware and software driver structure that will enable porting of future interface types. Faster versions of Fiber Channel, InfiniBand, SATA/SAS, and other interface options such as Ethernet and PCI Advanced Switching may become common storage interconnects as well. The SGI InfiniteStorage 10000 platform frees users to choose the right connectivity infrastructure and disk types for their particular requirements, and be assured of a rich growth path and investment protection.

4.0 Nearline Storage Applications

Nearline disk-based capacity solutions can be used for applications where data needs to be accessible in milliseconds, seconds, or even minutes, but where the data is less frequently accessed than other enterprise data.

As with the Internet-based photo gallery example already described, nearline storage must enable the rapid ability to call up infrequently accessed data. Nearline storage can be used for data staging for supercomputing processing or disk-to-disk-to-tape backup, bulk storage for email, as well as archiving of SOX data.

4.1 Target Markets

Nearline disk-based storage solutions are gaining popularity within a variety of markets including scientific research, simulation and modeling, bioinformatics, medical imaging, and oil and gas were large amounts of infrequently accessed data needs to be read or written quickly to support R&D efforts.

4.2 Other Suitable Applications of Nearline Storage

SATA storage will work well when used in nearline environments where performance is less of a factor but the data/drives are heavily used. Applications that stream data or require significant amounts of data movement without a high transaction rate are a good fit. Moderate performance requirements but nearly continuous read/write activity supports heavy usage applications such as file serving, business intelligence, and data warehousing and mining.

4.3 Restrictions

Nearline solutions do not fit well in applications involving high transaction rates where continuous disk operation is a must. These environments benefit from high-performance disk drives with higher rotational speeds—10K or 15K RPM—and are designed with rapid access times to satisfy a 7 x 24-hour duty cycle.

5.0 A Tiered Storage Model

For years, vendors and analysts in the storage industry have used a pyramid or triangle to depict a hierarchy of storage products and solutions that span all levels of price, performance, and capacities. Broad ranges exist within all levels, however this discussion will focus on a tiered storage model based on disk and tape solutions (Figure 5-1).

The purpose of this tiered framework is to assist organizations in managing storage based on data and application requirements. Data Life Cycle Management (DLM) building blocks are defined to lower storage cost by migrating data from primary to secondary storage and ultimately to an archive (tape storage). Data movement is based on factors such as age or criticality to the organization, with the value of data typically changing over time.

The tiered model can also assist organizations with the development of technology and storage roadmaps. SGI InfiniteStorage 10000 systems allow for easy integration with best-in-class solutions and partners including cluster solutions, disk-to-disk backup, remote disaster recovery, and long-term archival.

5.1 Enterprise Storage

Enterprise storage falls into Tier 1—online and primary storage where access times are important and continuous disk operation is a must. Maximum disk reliability supports heavy, nearly continuous read/write activity for mission-critical or business-vital applications.

Enterprise storage is traditionally based on high-performance Fiber Channel or SCSI disks, where fast rotational speeds (10 or 15K RPM) and a 100-percent duty cycle is necessary for nearly continuous read/write activity. These drives are designed and best suited to applications that require constant, instantaneous access to data, such as high-performance simulation and analysis for intense supercomputer processing, online transactional processing, decision support, and enterprise messaging.

5.2 Archival Storage

At the other end of the spectrum is *archival storage* such as tape or MAID (Massive Array of Idle Disks), which is used primarily for applications where infrequent serial access is required, such as long-term data storage. Both of these technologies require significant time to access the first byte of data, which may take from as little as 15 seconds to as much as several minutes or even hours if a tape is not housed in the library.

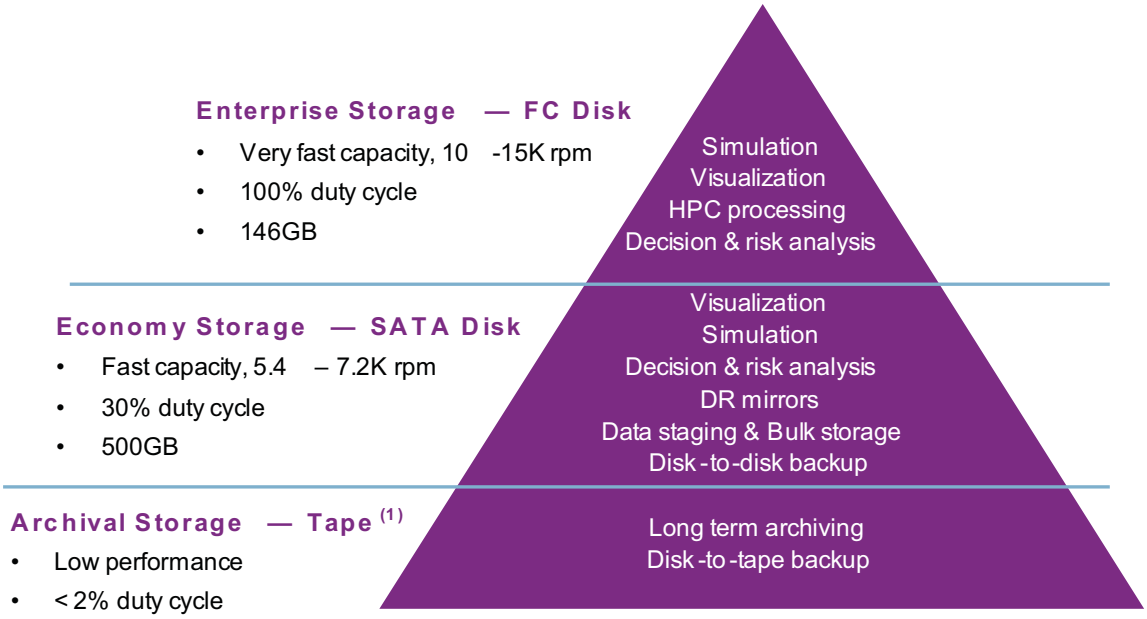

Figure 5-1: Sample SGI Tiered Storage Model[1]

5.3 Economy Storage

In between enterprise and archival storage is economy storage based on Serial ATA (SATA) disk drives. Economy storage—another name for nearline storage—spans a range of applications, including online or active archival, reference data, backup, and remote disaster recovery. Economy storage is for applications that require quicker access and higher throughput of data compared with long-term archival storage, but that do not require the continuous, instantaneous access provided by enterprise storage.

Economy storage—shown as Tier 2 in the previous model diagram—serves as secondary storage and enables organizations to store data more economically than compared to enterprise storage. Utilizing less-expensive SATA disks instead of high-performance SCSI or Fiber Channel disks, these Tier 2 solutions offer capacity at a lower price. SATA disks receive and send data faster than tape drives, shortening access time and increase throughput rates without resorting to expensive enterprise storage.

These SATA disk-based solutions are not intended as a tape replacement but as an intermediate step to accommodate increasingly complex storage demands. In fact, since SATA-based storage solutions can support the most popular tape backup software, it acts as a repository of tape data for nearline recovery. Backing up to a SATA storage solution and then to tape enhances data protection management and improves primary storage and tape library performance. It's also faster and consumes less application-server CPU than direct backup to tape.

The economy storage market is growing quickly with IT departments finding new uses for this low-cost solution. For example, financial services can apply economy storage where they must be able to rapidly call up large amounts of simulation data for analysis. Economy storage is gaining popularity in a variety of applications including scientific research, simulation and modeling, bioinformatics, medical imaging, and oil and gas were large amounts of data need to be read or written quickly to support R&D efforts.

6.0 Summary

The InfiniteStorage 10000 allows organizations to build cost-effective highly scalable secondary storage in the smallest footprint available. This massively scalable storage solution lowers storage purchase costs as well as storage operating expenses. Create worry-free disk-based solutions for nearline storage and archive, fixed content and large file storage, backup and recovery systems, and nearline production storage with the industry's best price per capacity per square foot storage solution.

The primary benefits delivered by the SGI InfiniteStorage 10000 solutions are summarized in the following table.

[1] SGI partners with tape vendors to povide this solution.

Table 6-1. SGI InfiniteStorage 10000 Benefits

Data Accessibility	How quickly can work be done? Disk drives access data in milliseconds to seconds, even when in power saving mode. Tape takes minutes to hours or days depending if it is on site or not. If data accessibility is a key requirement, the SGI InfiniteStorage 10000 is the industry leader for cost-effective accessibility.
Cost	The InfiniteStorage 10000 controller and its high-density drive package deliver a disk-based solution that approaches the price of tape. With *sleep* mode and other power-saving features, the SGI InfiniteStorage 10000 saves on power, cooling, and floor space for the best price per capacity per square foot disk solution in the market.
Performance	How much data needs to be moved in how much time? With a parallel architecture and broad back end, the SGI InfiniteStorage 10000 delivers 10.8 TB-per-hour bandwidth. There is nothing in the nearline market that even comes close.
Data Protection	The SGI InfiniteStorage 10000, with continuous on-the-fly dual-parity protection, allows continually monitoring of media reliability and corrective measures to avoid data loss. The solution also monitors drive health and responds with preventative measures to maintain data availability.
Reliability	With the dual-redundancy disk feature, the SGI InfiniteStorage 10000 nearline solution delivers enterprise-class reliability. This innovative feature protects against a dual-drive failure even in the same redundancy group without adversely affecting data availability. No other nearline solution has this level of reliability.
Scalability	SGI InfiniteStorage 10000 is up to five times more scaleable than other disk-based storage systems, packing up to 240 TB in each cabinet and 480 TB in two. This extremely dense solution delivers a volumetric solution that reduces the need for expanding the data center.

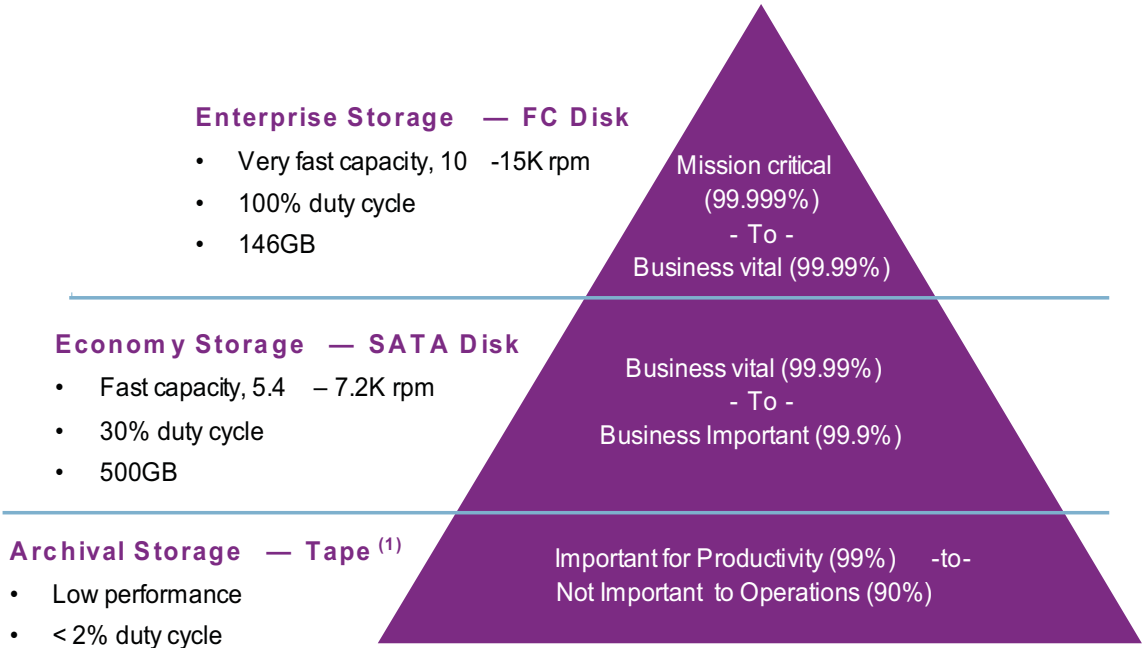

Figure A-1: Data Value by Performance

Appendix: Optimizing Storage Capacity

With a new storage option—nearline storage, or Tier 2 economy storage—IT organizations must evaluate existing storage situations to determine the most appropriate uses. The following two-step methodology is presented to help an IT department optimally match storage to data and take advantage of the cost savings of economy storage.

Step One: Identify All Storage and Data

Review the storage environment to determine what data exists and identify its locations, characteristics, and levels of importance to the organization.

This review takes place in a two parts. First, all data must be identified and defined in terms of its importance. This classification of the data can be done by segmenting data into three, four, or five discrete tiers (Figure A-1). For example, 10 percent of an organization's data might be deemed ***mission critical*** and needs to be accessible in milliseconds; 20 percent is ***business vital*** but can be accessed between milliseconds to seconds; 20 percent is ***business important*** and could be accessed in minutes to hours but not exceeding eight hours; and the remainder ***not important to business operations*** but necessary to keep for regulatory purposes, or else ***important for productivity*** but could be accessed in three days from tape.

The second part of this step requires mapping data to storage (see Figure A-2). Each dot represents a collection of data related to a specific application. These groups of data are mapped to their respective storage tiers. Note, however, these data groups may not be aligned with the performance and availability offered within the storage tier. For example, mirrored data, shown in Figure A-2 as two data points connected by dotted line, is often stored in two tiers. The main, active data might be on a higher-peroformance tier, with the protection copy—the lower data point—on a lower-performance tier.

Age can also be a key factor affecting data needs. For example, Internet-based photo galleries find there is a rapid fall-off in access of photos after a few days, and virtually no access to photos more than a month old. If an organization stores all photos on enterprise storage, an opportunity exists to move the older photos to less-expensive economy storage.

Age does not always drive the usage frequency of data. There are other variables that can influence data performance require-ments. The assessment process provides a reality check for the organization's operations. Objective measures such as 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 and the need for immediate access. Some data may be old and rarely

accessed, such as a call centers customer list from last Christmas. Nevertheless, storing that data on economy storage may be justified to provide an excellent experience when the same customer calls next Christmas.

Profiling data by its access, throughput, and availability characteristics will reveal sub-optimal uses of capacity and opportunities to more efficiently apply the right storage to the right data. Less-critical data can be moved to less-expensive disk storage and regained top-tier storage reallocated.

Step Two: Align Data Performance Needs with Storage Tiers

Step Two aligns data performance requirements with storage tiers (see Figure A-3). Each dot should yield a clear view of the requirements of different data types and how well these requirements are met in the current environment.

This alignment will identify data that is optimally using existing storage resources and data that could be residing on nearline storage or archived to tape.

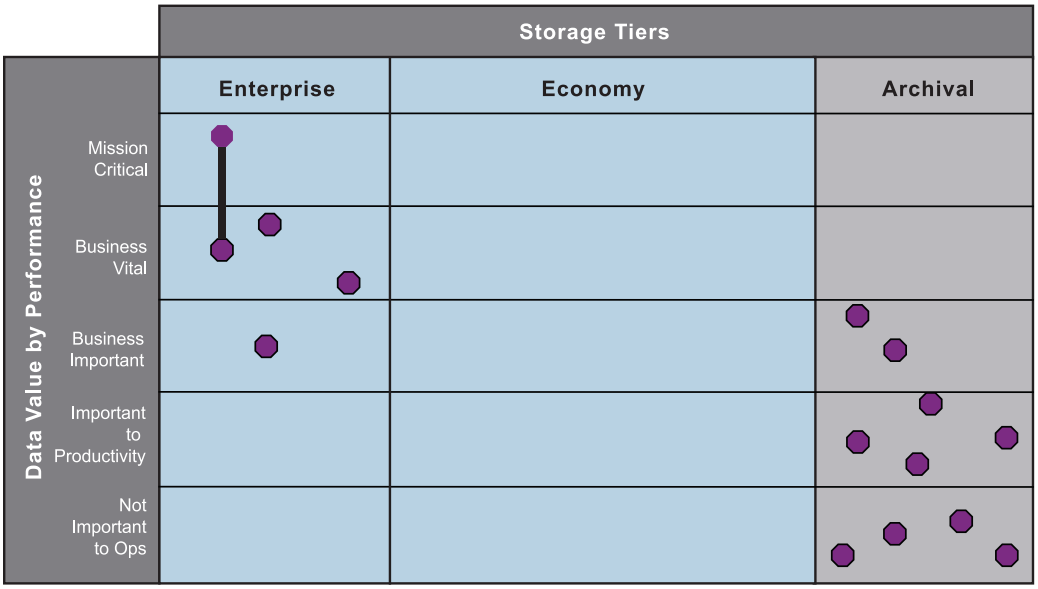

Figure A-2. Mapping the value of data by storage tiers

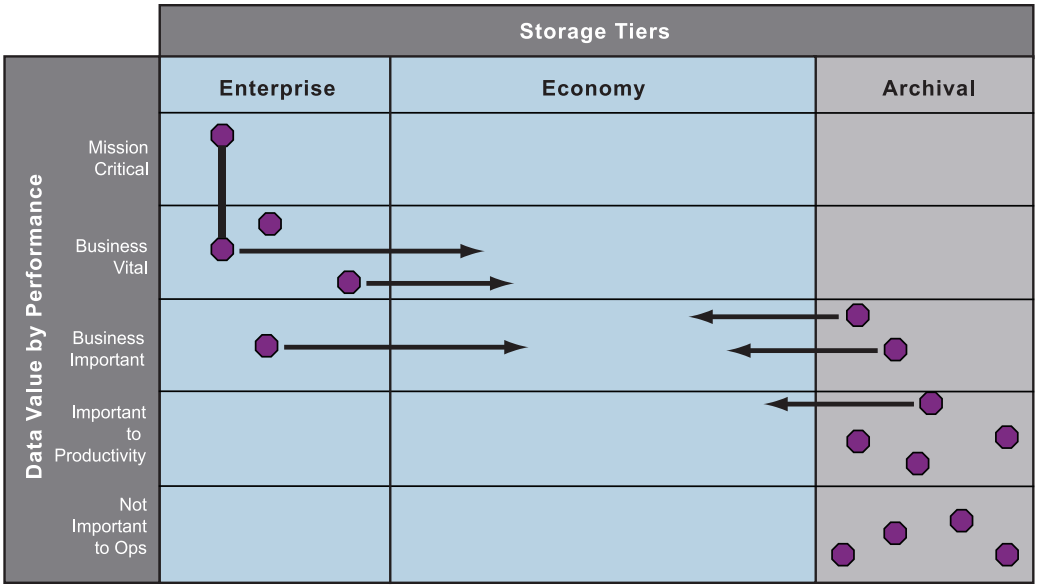

Figure A-3. Migrating data to align performance requirements with storage

Corporate Office
1140 E. Arques Avenue
Sunnyvale, CA 94085
(650) 960-1980
www.sgi.com

North America +1 800.800.7441
Latin America +55 11.5185.2860
Europe +44 118.912.7500
Japan +81 3.5488.1811
Asia Pacific +1 650.933.3000

3985 [01.08.2007]

J15226