White Paper

SGI® InfiniteStorage™ 10000: Replacing Tape with Large-Scale Online Archives

Table of Contents

1.0 Introduction

Many current business requirements—availability, business continuity, regulatory compliance, privacy rights, and corporate governance—have placed demands on IT for continuous access to information assets. The confluences of business, technical, and legal requirements are forcing corporations to re-evaluate their data storage and protection technologies and practices.

Consequently, the data storage industry is in the throws of reinventing itself. Tape storage has long been the standard for mid-to-long-term preservation and archival of corporate data. However, rapidly declining costs for disk-based storage solutions coupled with demands for continuous data access are driving storage vendors to offer new archiving platforms, in a more readily accessible format than tape affords.

Users are quickly recognizing the potential benefits of disk-based backup technologies. Business-critical data can be backed up and restored faster, more reliably, and with less operational cost compared to tape technologies. Many companies have already augmented tape-based backup/recovery systems with disk-based solutions that enable more rapid recovery and reduce operational expenses (OPEX). IT organizations that struggled with insufficient backup windows, recovery problems, or stolen tapes also acknowledge the benefits of disk-based backups.

2.0 A New Storage Paradigm: Large-Scale, High-Density Storage

SGI has recently brought to market a large-scale, high-density (LSHD) disk subsystem. The SGI® InfiniteStorage™ 10000 delivers massive disk storage capacity with a minimal footprint. This solution provides far greater I/O bandwidth than tape for an approximately equivalent cost and with lower associated on-going maintenance costs. The users within traditional high performance computing (HPC) markets can deploy the SGI InfiniteStorage 10000 solutions to increase operational efficiencies and mainstream commercial IT operations can achieve value in the overall combination of performance, administrative efficiency, and lower maintenance costs within the commercial sector.

The SGI InfiniteStorage 10000 represents the future of mainstream IT infrastructures where LSHD disk subsystems can be affordably applied for primary, secondary, and backup storage. Universal disk storage allows IT administrators to consolidate and manage larger amounts of data far more efficiently and to meet the increasing demands for continuous data access.

2.1 Shortcomings of Traditional Tape Storage

Tape media has been considered to be the most affordable long-term archival option since the 1950s. However, managing tapes is more labor-intensive due to the manual requirements for loading and unloading of the media. Additional handling is required for transferring tapes to and from offsite storage facilities. Many organizations outsource the transport and offsite storage of tape media to third-party service providers, which introduces risks of theft or inadvertent losses of valuable corporate information assets. Whenever archived data is required, the tape(s) that contain the desired data must be located, verified, and loaded. Even with today's large-scale automated tape libraries, there is no avoiding the costs and risks associated with tape.

Some organizations consume tens-of-thousands of new tape cartridges every year, many of which are filled with redundant or low-value data simply because IT departments lack adequate knowledge of backup coverage. It is ironic that highly skilled professional resources are consumed to manage a resource (tapes) when a significant percentage of the content has questionable value to the organization. All of the time spent managing tape technology is time not spent managing other applications, such as customer-facing/revenue-generating business processes.

The inherent inefficiencies, costs, and risks for traditional tape-based backup, which also include slow I/O rates and lengthy times to access data, have resulted in a growing demand from IT professionals for an affordable and suitable disk-based approach.

2.2 The Changing Face of Business Risk

IT departments today need to retain more data in a secure and accessible format for longer periods of time due to changes in laws and regulations. Ten years ago, problems restoring data were likely to be met with managers' admonitions directed at IT. Five years ago, with Internet commerce in full swing, the same problems typically resulted in more serious reprimands or even someone losing their job. Today, with pressures stemming from corporate governance and oversight regulations, the inability to quickly and accurately produce an authentic version of data can escalate into a public inquiry involving customers, investors, partners, and even the government. The failure to produce electronic evidence in a timely fashion can result in significant financial penalties as well as creating a negative bias towards the company and unfavorable court judgments. If willful negligence or wrongdoing is determined, there may even be incarceration penalties for those found culpable.

There is no standard case law to inform IT professionals or business executives as to how technology problems will be interpreted by the courts. The court may excuse failures of equipment and media as normal unfortunate events or they may be viewed suspiciously as evidence of failed responsibility.

Considering the legal ramifications and risks of managing and preserving corporate data today, the historical pitfalls associated with tape-based backup/recovery performance and the challenges of restoring data from tape in a timely manner now pose a very real and potent risk to the livelihood of businesses.

3.0 The SGI InfiniteStorage 10000 LSHD Subsystem

The new SGI LSHD storage subsystem addresses the historical liabilities associated with tape. The guiding design priority for the SGI InfiniteStorage 10000 is to integrate the maximum amount of serial ATA (SATA) disks into a compact footprint, meeting the need to retain more data for disk-to-disk backups, active archiving, and rapid retrievals. The SGI InfiniteStorage 10000 is smaller, faster, and more cost effective than many of the disk-based solutions on the market today, and an outstanding alternative to a purely tape-based system.

3.1 Compact Footprint and Operational Efficiency

A key metric for today's increasingly dense storage systems is capacity per square foot. For many IT organizations, managing floor space is an ongoing struggle. The density of the SGI InfiniteStorage 10000—supporting up to 1000 SATA drives—offers a compelling alternative to large-scale tape silos. The basic unit of capacity for the SGI InfiniteStorage 10000 is an ultra-dense 4U disk drawer (tray) that can each accommodate up to 48 drives. Using 500GB SATA drives, this provides 24TB of raw capacity per drawer and up to 200 TB in a single rack.

Tape is often considered to be infinitely scalable since users can simply add more tape cartridges to the system. However, from an operations perspective, this is certainly a poor assumption because adding tapes creates additional administrative burden. Each tape is a discrete physical entity that has to be managed independently. Algebraically stated:

Increased Tapes = Increased administration burden
= Decreased operational efficiency
= Increased CAPEX and OPEX

3.2 System Performance

LSHD disk subsystems can be designed to increase raw bandwidth several times compared to the tape-based counterparts. For example, the SGI InfiniteStorage 10000 can stripe data across disk drives using multiple I/O channels, attaining I/O performance that is orders of magnitude faster than tape subsystems. These performance benefits allow IT teams to carryout backups in less time with less effort and while eliminating manual handling requirements for the media. Eight full-duplex Fiber Channel ports provide excellent aggregate throughput to connect to multiple host systems.

3.3 Reliability, Availability, and Serviceability

The SGI InfiniteStorage 10000 offers significant reliability, availability, and serviceability (RAS) advantages, by virtue of being a disk-based system and through its innovative design. It delivers enterprise-class reliability with dual-channel multiplexing controllers, hot spares, and parity RAID including an optional dual-parity RAID configuration for organizations requiring the highest redundancy levels. Each drawer is powered by an N+1 triple power supply design. All power supplies, fans, and controllers are hot swappable to maintain uptime during maintenance operations. The SGI InfiniteStorage 10000 also supports dynamic multiple paths for highly available connections to host systems.

With so many drives in the system, spares are aggregated much more efficiently than with smaller disk subsystems. Any spare drive can substitute for any failed drive in the subsystem, which means that a lower number of spares can protect storage. The SGI InfiniteStorage 10000 is completely serviceable in the field; users can easily replace individual drives or groups of drives, without requiring a professional services call. The front panel of each drawer has an LCD display indicating drive status and error conditions, and the mechanical design includes slide-out drawers that facilitate access to all the drives in the system.

4.0 Manageability Advantages of Massive Storage

Disk storage can be amassed and managed as a single physical resource with many logical partitions, making it much more efficient to manage—especially when compared to tape. Today's highly scalable file systems enable hundreds of TBs of disk capacity to be addressed by a single administrator and used as a single storage pool. SGI Users familiar with the SGI XFS and CXFS file systems appreciate the power of these advanced 64-bit journaling file systems for storing and managing enormous data sets and extremely large individual files.

4.1 Enabling Efficient Hierarchical Storage Management

The concept of hierarchical storage management (HSM) is a systematic approach for migrating less-frequently accessed data from expensive storage to lower-cost storage. Many HSM implementations are constructed of two layers: primary disk storage and tape. The addition of an intermediate, less-expensive disk storage layer makes HSM much more broadly viable. Considering the enormous capacity in the SGI InfiniteStorage 10000. Built into an HSM system, the SGI InfiniteStorage 10000 system supports an enormous amount of data that would otherwise have to be accessed from tape. Combined with the SGI Data Migration Facility software, this adapted HSM model creates a robust data migration system that automates the administration of data and allows data to be stored on the optimally suited tier.

With an LSHD middle tier in an HSM system, data can be much more aggressively migrated from primary to secondary storage because of the performance and operational simplicity of disk storage. If the average disk access takes less than 10 milliseconds and the average tape access takes ten seconds (which are very conservative estimates), the access time is three orders of magnitude faster for disk storage. The impact on migration and access queues is enormous and translates into much crisper capacity management. Consistent, fast HSM performance is a great benefit for administrators in data-intensive environments.

SGI Data Migration Facility software enables less-frequently accessed, out-of-date, or low-value data to be automatically migrated from production systems to more cost-effective second-tier storage such as the SGI InfiniteStorage 10000 platform. Data Migration Facility is a sophisticated tool that allows administrators to develop actionable strategies for data lifecycle management. When used with this tool, the SGI InfiniteStorage 10000 is a very potent solution for near-term
data management as well as serving as the foundation for future data management and lifecycle processes.

4.2 Enabling the Active Archive

The definition of archiving continues to evolve, but one thing is certain—organizations are feeling pressure to comply with regulations mandating retention periods for certain classes of data. The operational and financial burden of the legal discovery process for digital data falls squarely on the organization that owns and retains these assets. Courts and case law have expectations for how quickly organizations are expected to retrieve and present electronic evidence once a discovery notice has been served. For the majority of organizations today who rely on tape as their primary means of backup and archival, the challenge of recovering even a dozen emails can be a logistical challenge and consume considerable administrative resources.

4.3 Large-Scale Backup and Recovery

Many organizations continue to struggle with backup and recover processes under the pressure of continual data growth. SGI InfiniteStorage 10000 solutions can significantly improve the speed of backup and recovery by virtue of raw performance and scalability. Industries such as media/entertainment, life sciences, energy, and scientific research often need to process massive data sets associated with key applications. The sheer volume of data created in these environments makes it very difficult to adequately protect information. The performance and capacity of the SGI InfiniteStorage 10000 significantly reduce the burden and costs associated with data protection.

5.0 Conclusions

Information is often an organization's most valuable and volatile asset. The ability to access, share, protect, retain and retrieve information is a direct indication of the quality of the organization's information governance practices. Corporate governance is not limited to sound business practices and ethics, but now also spans Information governance.

The growth in data creation, retention, and long-term preservation will continue unabated and encouraged by the continual decline in raw storage costs. The inherent advantages of disk over tape become more important as capacity scales. As a random access media, disk storage provides faster access times and throughput compared to tape and can boast lower administration costs. As a result, IT professionals are turning to disk-based products to address processes formerly associated with tape.

The new SGI LSHD subsystem, the SGI InfiniteStorage 10000, leverages the advantages inherent in disk technology and adds many of its own unique advantages to provide users with innovative, cost-effective, high-performance storage. SGI's track record for developing and delivering storage and data management solutions for the most demanding environments places the company at the forefront for addressing the storage challenges faced by corporate organizations.

For more information about the SGI InfiniteStorage 10000 solutions, please visit www.sgi.com/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

3983 [01.02.2007]
J15218