IDC
Analyze the Future

WHITE PAPER

Accelerating File-Level Storage with NAS

Sponsored by: Silicon Graphics Inc. (SGI)

Brad Nisbet

September 2007

IN THIS WHITE PAPER

This white paper examines how organizations of all types are currently addressing storage and data management needs to support the growing amount of file-level activity — not only within industries that traditionally have file-intensive needs but among a growing population of midsize and large enterprise organizations. The paper then identifies the pain points associated with current solutions, discusses the benefits of network-attached storage (NAS), and focuses on SGI's line of InfiniteStorage NEXIS NAS solutions and the benefits to be gained by customers.

SITUATION OVERVIEW

Introduction

The Data Explosion

More than ever before, today's businesses rely on the power of digital information. Across a wide variety of organizations, from small businesses to large enterprises, companies are generating an ever-increasing amount of data — digital information that is either directly the lifeblood of the business or a significant component behind the scenes. For many, simply storing this information is as overwhelming as managing, leveraging, and transforming the digital information into a true business asset.

As a result of this data explosion, IDC expects that the capacities shipped on external disk storage systems will grow about 62% per year, on average, increasing from 2.2 exabytes in 2006 to 24.2 exabytes in 2011. However, it is clear that IT budgets will not grow nearly at this rate and that CIOs, IT managers, and storage administrators will not be willing to pay a significant amount of money to address their data explosion issues. IDC expects that spending on overall disk storage systems will increase at a single-digit rate between 2006 and 2011.

What does this mean for IT and storage administrators who are drowning in a rising sea of data? Educated and savvy end users will seek storage solutions that enable them to consolidate IT resources and improve efficiency in data management, all while keeping costs at bay.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

Growth of File-Level, Unstructured Data

For a very large population of organizations, the increase in digital information and the associated growth in storage capacity will be driven by file-level data, much of which is unstructured and/or "fixed" (unchanging) content. Indeed we have seen a new generation of organizations that almost entirely rely on the storage, management, long-term retention, and timely retrieval of files as the core of their business. Examples include:

- Internet services such as online audio, photo, and video repositories or online email sites
- Industries or application areas that engage in high-performance or technical computing such as oil and gas, life sciences, entertainment (e.g., rendering), government, and other simulation application areas

However, the growth of files is not relegated to Internet services or other industry-specific and high-performance segments. Organizations of all types, from small and medium-sized businesses to larger enterprise datacenters, are experiencing a flood of file-level data as they continue to:

- Digitize business information in the form of files
- Increase the numbers of internal and external users generating and/or accessing the files
- Seek to share business or office files in an efficient manner to improve productivity or help achieve business goals
- Generate larger files associated with the continued advancements in application features, graphics, and processor power
- Generate more files, particularly as a result of increased email activity and collaboration
- Seek to leverage the value of file-based data in the form of business intelligence and analytics via data warehousing
- Address business governance and regulatory compliance, which will increase and complicate the retention of business information, forcing IT managers to develop ways to save more (or all) files generated for longer periods of time

As file-level data continues to garner an increased focus, enterprise organizations have had several options from which to choose in addressing their file-level needs. For most, the deployment of general-purpose servers (GPSs or "servers") to accommodate file sharing remains the most widely adopted method. Organizations that are capable of making more strategic investments to consolidate or reduce management complexity rely on alternative solutions such as advanced file systems or NAS.

One such company that is addressing the file-based needs of a wide variety of customers is Silicon Graphics Inc. (SGI). The company has a long history of serving the file-based needs of the high-performance and technical computing communities, and its ongoing investments to enhance its cost-effective, extremely scalable, and high-performing NAS product line will help a broad base of customers address a wide variety of file-based storage needs.

Data Types and Growth of Files

For companies to address the issues related to the storage of unstructured, file-based data or digital content, it is useful to understand how and where this type of data fits in to the entire realm of digital business information. What types of data are today's businesses looking to keep? In short — almost everything. Closer inspection reveals some interesting characteristics about the data that businesses desire to store and leverage:

- **Traditional transaction-oriented, structured data.** These data sets are typically stored and processed in the form of structured databases such as order entry systems, supply chain management systems, reservation systems, and so forth. While this represents a significant portion of the type of data stored on disk storage systems today, it also will be a shrinking portion of the overall storage problem, as its relevance and visibility within organizations will be overshadowed by secondary and longer-term retention of other information, typically unstructured or file-based data. Although this is still a growing market (terabytes shipped are expected to rise about 37% through 2011), IDC estimates that about 36% of all external disk storage systems shipped in 2006 were to support these transaction-oriented types of data and expects this portion to shrink to about 20% by 2011.

- **Replicated data.** Much of the transaction data is replicated for a variety of reasons, including backup; disk-based disaster recovery; copies for development, testing, and design; copies to meet either compliance or business governance; or replication for business analysis. This is also a growing data type, with terabytes shipped increasing by about 47% each year through 2011. However, its share, much like that of transaction-oriented data, will shrink in overall percentage from 38% of all external disk storage systems capacity shipped in 2006 to about 29% in 2011.

- **File-level data.** This data largely consists of digital content that is unstructured and much of which is "fixed." This file-level data can exist as storage pools within organizations whose entire business is centered around the storage, management, and delivery of files (e.g., Internet/Web services or high-performance computing) or is pervasive among a huge population of small, medium-sized, or large enterprise organizations that have a blend of transaction/database information and a growing set of files such as office documents, emails, Web pages, and digital images/audio/video. File-level or unstructured data represents one of the fastest-growing types of data storage today; it accounted for about 27% of the data in 2006 and is expected to exceed 50% by 2011. The terabytes shipped are expected to grow 84% per year on average through 2011. For many organizations, doubling or tripling the amount of file-related data each year is a very real possibility.

For good reason, the ability to store, manage, and leverage digital information in the form of files is a growing concern for many CIOs, IT managers, and storage administrators. In particular, organizations are looking for ways to satisfy the high-performance needs to deliver file-level services to a growing base of users and applications, all while maintaining control over physical resources and costs.

To meet the requirements necessary to deliver an IT solution capable of supporting current and future file-level needs without stressing IT budgets, companies will need a storage solution that offers extreme scaling efficiency, in terms of both capacity and performance. This advanced scalability for file-level data has first been observed in areas of high-performance and technical computing. These application areas are extremely file-intensive — in terms of the numbers of files, the size of files, or the intense compute power needed to generate, store, and retrieve the massive amounts of file-level data generated.

The scalability characteristics common among these high-performance and technical computing environments are now needed by and useful among many mainstream enterprise business applications whose file-level needs have grown tremendously. These enterprise organizations currently have several options to address their file-level needs.

Available Options to Address File-Level Storage

There are several options that CIOs, IT managers, and storage administrators consider when developing a storage solution to address file-level data:

- Advanced file systems such as SAN file systems and other clustered file systems
- General-purpose servers
- Network-attached storage

This paper discusses each option but focuses on the use of GPSs and NAS. While the use of GPSs is perhaps the most pervasive option deployed today, NAS offers a growing set of benefits when faced with daunting file-level storage requirements.

Advanced File Systems: SAN or Clustered File Systems

Many large organizations that rely on large pools of file-level data have deployed SAN file systems, such as SGI's InfiniteStorage Shared Filesystem CXFS. While these solutions are robust and are certainly helping customers address data management and organization of large pools of file-level data, they are generally reserved for very complex environments. Certain organizations that have invested heavily in SAN technology also find these types of solutions attractive because they allow organizations to leverage their SAN investments.

A typical SGI SAN is used as pooled storage for a large shared memory compute cluster for high and deterministic performance applications. Using CXFS, the SGI SAN Servers enable multiple computers with different operating systems to have direct access to a shared file system. This means all systems connected to the SGI SAN Servers have access to the same file, at the same time, at local or near-local disk speeds.

Customer Pain Points Regarding SAN or Clustered File Systems

SAN or clustered file systems are generally not as popular among mainstream enterprise organizations that are looking for more complete, turnkey solutions. Although these types of advanced file systems can be effective, they are typically reserved for organizations that can engage in more complex deployments. Many times, such deployments occur either within very large enterprise organizations or within the high-performance and technical computing communities, which often prefer and have the resources to be more involved in the creation of the solution.

The SGI InfiniteStorage Shared Filesystem CXFS is appropriate for large shared memory compute clusters with shared SAN storage. While CXFS is a great example of a high-performance file-level solution for SAN, the majority of mainstream enterprise applications today do not represent this architecture.

General-Purpose Servers

General-purpose servers represent the most widely adopted choice among end users to address their file-level storage needs. Servers can be deployed for a variety of workloads, including network file sharing. IDC estimates that approximately $5.4 billion was spent in 2006 on general-purpose servers deployed for the function of network file sharing.

These servers are typically deployed with internal hard disk drives to facilitate file-level storage, and they are usually based on Windows, Unix, or Linux operating systems. Each server typically has a separate file system that uses a single type of protocol for sharing files over the network, commonly CIFS for Windows servers and NFS for Unix or Linux servers. The scalability of the internal DAS is often contained within each individual server.

A general-purpose server with internal DAS remains the most widely deployed file-sharing solution for several reasons:

- Customers are loyal to existing channels of IT infrastructure. In most cases, they will tend to be loyal to one or several of the larger server vendors that traditionally focus on server-based solutions.
- Many organizations might perceive the notion of migrating to a new storage architecture as costly or too complex.
- Many organizations may be unaware of alternative solutions, such as NAS. In some cases, customer loyalty could actually be the result of indifference or ignorance to alternative architectures.

Customer Pain Points Regarding GPS Solutions

The increase in files in organizations is placing stress on administrators and their infrastructure to accommodate growth. For many organizations that have already deployed GPSs for general file sharing, the typical reaction to increasing file-level storage capacity or activity is either to undergo disruptive upgrades to higher-capacity hard drives or to deploy additional file servers to accommodate an increase in capacity or performance in an effort to serve files efficiently to all users and applications.

Many times, the storage capacities or performance capabilities of general-purpose servers are inadequate, especially as organizations grow over time. Upgrading individual servers by manually replacing the drives with larger drives is extremely disruptive and involves considerable time and effort on the part of IT or storage administrators. Along with taking servers offline during the upgrade, administrators are subject to the complexities of manually reconfiguring and rebooting each server.

Eventually, when maximum capacities or performance is reached, additional servers are added to increase file-level services. This proliferation of multiple devices leads to:

- Increased complexity of architecture
- Increased use of resources and higher cost to manage multiple devices and multiple file systems
- Increased backup activity and complexity
- Multiple devices for multiple user environments or operating systems
- Increased licensing costs for each server added

Network-Attached Storage and Benefits over GPS Solutions

NAS is a disk storage system that attaches directly to the user network and contains its own file system, which has been optimized to store and share files. NAS is a mature technology that has found success among users who wish to consolidate their file-level services, typically consolidating from multiple general-purpose servers.

IDC estimates that the NAS market will grow from $2.6 billion in 2006 to about $4.8 billion by 2011, representing a robust 13% CAGR. NAS provides many of the following typical benefits over general-purpose server solutions:

- **Ease of use and dedicated purpose.** NAS is designed to be easier to deploy and manage than server file-sharing solutions. It is a storage system designed and optimized to store, share, and manage files.
- **Greater capacity and scalability than GPS solutions.** Higher initial capacity points can provide consolidation of storage and management into fewer devices. Further, NAS is typically designed to scale capacity external to the initial deployed NAS device, greatly enabling the ability to scale even further as capacity needs amplify.

- **Greater performance.** Because NAS devices are purpose-built for file sharing, they generally provide greater performance than their server counterparts. In addition, depending on design, the ability to scale performance is often enhanced with NAS solutions versus GPS solutions because enhanced performance scalability with some NAS solutions does not necessarily translate into increased management complexities.

- **Consolidation.** NAS typically offers not only increased initial capacities and greater performance than GPSs but also advanced scalability with respect to both of these metrics. This often allows a single NAS solution to replace a group of desperate GPSs, each needing its own management and each constrained by its own scalability limitations. NAS allows for tremendous opportunities to consolidate.

- **Multiprotocol file sharing.** Most NAS solutions support multiprotocol environments simultaneously, unlike general-purpose servers, which are each dedicated to a single operating system. Typical environments supported include Windows (via CIFS protocol) and Unix/Linux (via NFS protocol).

- **Lower licensing costs.** Because NAS typically allows for the consolidation of multiple GPSs, it has the potential to lower the cost associated with fewer server licenses.

SGI's InfiniteStorage NEXIS NAS Product Family

One solution on the market today that is helping a broad spectrum of customers address the growing requirements related to file-level storage comes from SGI and its line of InfiniteStorage NEXIS NAS products. In particular, SGI's NAS products offer enhanced scalability with respect to capacity and performance, providing customers the flexibility to make appropriate initial investments that are aligned with current needs while providing the ability to store, manage, and scale their growing file-level environments over time in an easy and cost-efficient manner.

The InfiniteStorage NEXIS NAS products are based on SGI's XFS file system — an extremely robust and well-tested high-performance computing file system developed by SGI in 1992. XFS has served as the basis for the company's longstanding expertise and success in providing file-level storage solutions to the extreme demands of the high-performance and technical computing communities. The InfiniteStorage NEXIS NAS products are designed to provide all the benefits of XFS' performance, single point of management, and nondisruptive scalability within the building blocks of a simple appliance-based architecture.

In addition, SGI has developed a set of effective management tools, delivered under the InfiniteStorage Appliance Manager, that consolidates the management and growth of vast amounts of file-level data into a single point of management.

Given the wide differences in customer applications and data types, the SGI InfiniteStorage NEXIS NAS product family offers the following entry points and characteristics to match a wide variety of current and future capacity and performance needs:

- **InfiniteStorage NEXIS NAS 500** is an entry-level offering based on SATA drive technology that provides an integrated NAS server and storage in a single 2U chassis. It is appropriate for smaller or medium-sized organizations or smaller workgroups within larger enterprise organizations.
- **InfiniteStorage NEXIS NAS 2000** is available in SATA (capacity-oriented) or SAS (performance-oriented) configurations. The NEXIS NAS 2000 represents the entry point for SGI's enhanced scalability based on XFS file system technology. The SAS configuration is best suited for applications such as storage consolidation and file sharing among midsize enterprises or media, education, and scientific workgroups. The SATA configuration is targeted for small, online archive applications or moderate to large file sequential workloads.
- **InfiniteStorage NEXIS NAS 5000** is a midrange, higher-capacity offering based on SATA drive technology aimed at secondary storage applications such as disk-based backup, corporate or compliance archiving, or audio/video/graphic repositories.
- **InfiniteStorage NEXIS NAS 7000** is an enterprise-class storage array–based Fibre Channel (FC) drive technology. The NEXIS NAS 7000 is aimed at price-performance applications such as large enterprise file-level data consolidation, data mining, media production, and other high-performance or technical computing applications such as oil and gas or simulation modeling. The NEXIS NAS 7000 is also offered in a highly available, active-active configuration for more stringent uptime requirements.
- **InfiniteStorage NEXIS NAS 1000 Gateway** offers the separation of the NAS file system and intelligence from the back-end storage, enabling the flexibility to apply NAS connectivity and file-level services to existing or complementary pools of SGI InfiniteStorage SAN storage. Ideal applications would be similar to those of the NEXIS NAS 2000.
- **InfiniteStorage Appliance Manager** is a simple, GUI-based, single point of management console that provides ease of use for deployment and scalability. The advanced management software is designed to enable installation of base configurations in as few as 20 minutes and offers system and network performance monitoring, including alerts and problem detection tools.

All InfiniteStorage NEXIS NAS solutions leverage SGI Altix server technology and are built on Intel Xeon or Itanium 2 processors. Multiple connection protocols include both CIFS and NFS, as well as iSCSI for block-level connectivity. Snapshot technology is an option to enhance data protection and backup. In addition, all NEXIS NAS products include hot-swappable hard disk drives and redundant CPUs, power, cooling, and mirrored boot drives.

A Long History of Helping Customers Address File-Level Storage

SGI has a long history of providing scalable, high-performing storage solutions that have helped thousands of customers in the high-performance and technical computing space address intense file-level storage requirements. The company is successfully expanding this expertise to a broader set of midsize and enterprise customers who are now experiencing the pain associated with storing and managing increasing pools of file-level data.

While SGI's CXFS shared file system enhances the ability of customers to address file-level storage while leveraging a SAN architecture, the company has introduced its line of NAS solutions to provide a broader choice of options, particularly for customers who are looking for the benefits of scalability delivered in an appliance package. For many customers, SGI's InfiniteStorage NEXIS NAS, based on the XFS file system, provides an easier, cost-effective method to store and manage growing numbers of files.

Customer Benefits of SGI InfiniteStorage NEXIS NAS

File-Level Consolidation

A broad spectrum of CIOs, IT managers, and storage administrators will stand to benefit from the storage consolidation that derives from SGI's cost-effective and scalable NAS solutions. SGI's InfiniteStorage NEXIS NAS solutions are designed to deploy and scale well beyond the typical internal DAS capacities of most general-purpose servers.

The SGI InfiniteStorage NEXIS NAS has the potential to simplify an organization's file environment dramatically by reducing the number of devices necessary to satisfy a given organization's file-level demands. This capability in turn can help reduce capital costs and lower management costs as well.

Scalability

SGI's advanced capability to nondisruptively scale capacity and single file system size well beyond 100TB (theoretically into exabytes) represents a tremendous leap forward for many organizations that are currently using general-purpose servers to address file-level storage. A single InfiniteStorage NEXIS NAS appliance deploys and scales well beyond the typical internal DAS capacities of most general-purpose servers, which are normally under 1TB. The nondisruptive nature of capacity and file system upgrades ensures increased uptime, which for many organizations translates directly into increased revenues or business opportunities.

In addition to the benefits of providing large pools of capacity, the ability to scale to a very large, single file system nondisruptively is highly beneficial to administrators and CIOs. Administrators will appreciate the improved efficiency, ease of management, and simplified backup related to a single file system environment, while CIOs will take note of the reduced costs associated not only with the reduced numbers of physical devices but also with the reduction in time to manage more complex environments resulting from multiple file systems.

Performance and Diversity of Applications

The high performance and advanced scalability of the underlying XFS file system make the InfiniteStorage NEXIS NAS product family an appropriate storage solution for a broad spectrum of file-level applications, ranging from those that are oriented toward large file sequential I/O and focused on throughput to those that are oriented toward small file random I/O and focused on I/Ops. This diversity allows for a wide variety of applications from high-performance and technical computing to more horizontal enterprise or datacenter deployments.

In addition, the high performance of the InfiniteStorage NEXIS NAS is capable of supporting millions of files per directory and thousands of clients and concurrent users, all with the ease of administering a single file system.

Tiering Options Lead to Cost Savings

The ability to align the importance of business data with the appropriate IT infrastructure is of growing importance to many administrators of midsize and large enterprise organizations. With three classes of hard disk drives (SATA, SAS, and Fibre Channel), the InfiniteStorage NEXIS NAS product family collectively offers customers multiple tiers of storage under a single management console. This capability allows IT managers to fine-tune and match the capacity, performance, and cost of the storage with the corresponding business value or importance of the file-level data.

Gateway Can Leverage SAN Storage

SGI's InfiniteStorage NEXIS NAS 1000 Gateway provides the ability to leverage existing pools of SGI SAN storage. This has the potential to greatly enhance the ability of existing SGI SAN customers to deploy file-level services using NAS technology on top of their existing investments. It also allows new customers to deploy complementary systems that combine SAN storage with high-performance NAS capabilities that are simple to deploy and manage.

CHALLENGES/OPPORTUNITIES

SGI has helped many customers build successful compute and storage solutions within the high-performance and technical computing space. The company is now extremely well-positioned to bring its advanced file-level expertise and success to an expanded set of enterprise customers in the form of NAS. SGI should strive to address the following items to expand its market opportunities as well as to increase its value to existing and potential customers:

- Like all storage suppliers, SGI will need to continue to keep pace and stay ahead of technology advancements in terms of capacity, performance, and scalability. The existing NAS solutions are indeed robust, but increased competition will not cease, especially in the area of file-level services. Most disk storage systems suppliers have recognized the importance and opportunity that lie in supporting advanced technologies for file-level data storage.

- SGI will need to continue to add functionality and drive deeper value by internally developing or working with ISVs for even tighter integration between storage, applications, and third-party functionality such as backup, archiving, or data classification. In particular, SGI should focus on partnerships that will foster continued success for storage solutions deployed in midsize/large enterprises or datacenter environments. The company is well-positioned to advance this integration, but it will need to execute successful strategies to do so.

- The InfiniteStorage NEXIS NAS 1000 Gateway is an ideal product to leverage the SGI installed base of SAN storage and enable file-level services via NAS. SGI must work internally and with other storage systems vendors to expand the list of qualified, external storage arrays to embark on new opportunities and gain more NAS footprint in the datacenter.

CONCLUSION

CIOs and IT/storage administrators will continue to increase their focus on the expansion of file-level data as more users and new applications among midsize and large enterprise organizations generate more and larger files. As business information continues to become more file-based, administrators will be looking for new ways to control and support this increase in file-level data without increasing staff, complexity, or costs. For many organizations, this will include the deployment of external or shared file systems in the form of NAS.

SGI has a long history of developing advanced file-level storage solutions for the most demanding customers — the high-performance and technical computing communities. This history of success builds a strong foundation on which the company can offer its high-performance and cost-effective NAS solutions to an expanded base of midsize and large enterprise organizations.

SGI's InfiniteStorage NEXIS NAS solutions address many of the pain points that enterprise storage administrators face today. The advanced nondisruptive scalability of SGI's NAS appliances allows organizations to consolidate file-level services while enabling the ability to scale performance and capacity at a minimum cost.