White Paper

SGI® Storage: Solving the Data Access Bottleneck

1.0 SGI Storage: Solving the Data Access Bottleneck

SGI storage solutions are designed to solve the data access and management problems unique to engineering, scientific, and creative customers engaged in large-scale computing and visualization. As high-performance computing environments incorporated networked storage solutions in the late 90s including networked-attached storage [NAS] and storage area networks [SANs], their limitations became apparent in that neither provided the sufficient access to data that high-performance computational and visualization systems require. We describe this limitation as "the data access bottleneck."

In response to this problem, SGI has developed the revolutionary SGI SAN Server™ 1000 system with the CXFS™ shared filesystem for SANs that delivers near-instantaneous shared access to data from multiple computers and multiple processes within a workflow. CXFS combines the shared data access of NAS with the scalability and performance of a SAN. With SGI storage, projects get done faster and product gets to market faster and at a lower cost than it does with other storage solutions. SGI storage includes the most advanced hierarchical storage management [HSM] solution on the market, integrated backup, and file services to workstations on the network—providing the best end-to-end storage solution available for high-performance computing and visualization.

2.0 The Data Access Bottleneck in Conventional Storage Architectures

Bottlenecks in data access and data storage increasingly impede breakthroughs in technical and creative fields. Many high-performance computing and advanced visualization users find that they could do more, if only they had more storage or faster access to stored data. The increasing deployment of storage area networks is helping solve the data storage problem, but today's conventional SANs only improve data access for unshared data. In situations where data must be shared between different computer systems—and different computing platforms—most users still rely on slow network methods of data access, such as NFS, CIFS, or FTP, because data stored in a SAN is part of the filesystem of each host computer and the only way one computer can share files with another computer in a SAN is by copying from one machine to the other.

Figure 1 shows the three storage architectures available today—direct-attached storage, network-attached storage, and storage area networks—and their respective data access bottlenecks. DAS provides adequate bandwidth to the host server; however, if another system in the network requires access to the data stored on that server, it must be copied over the local area network [LAN] using protocols like NFS or TCP/IP. For example, a 1TB file—not unheard of in high-performance environments—would take four hours to copy from one server to another over a 100Base-T Ethernet LAN, and that's assuming 100% LAN utilization. In many cases, it takes longer to move data over the network than it does to process that data.

NAS has seen wide adoption and has performed reasonably well in many environments. But NAS is showing signs of strain as the complexity of the storage management problem increases. In NAS, a file server sits in front of the physical storage devices and manages the requests from client applications. But as these requests become more frequent and the amount of data requested increases, the file server can quickly bog down. The quick solution to this problem is to add more file servers and then replicate the data across

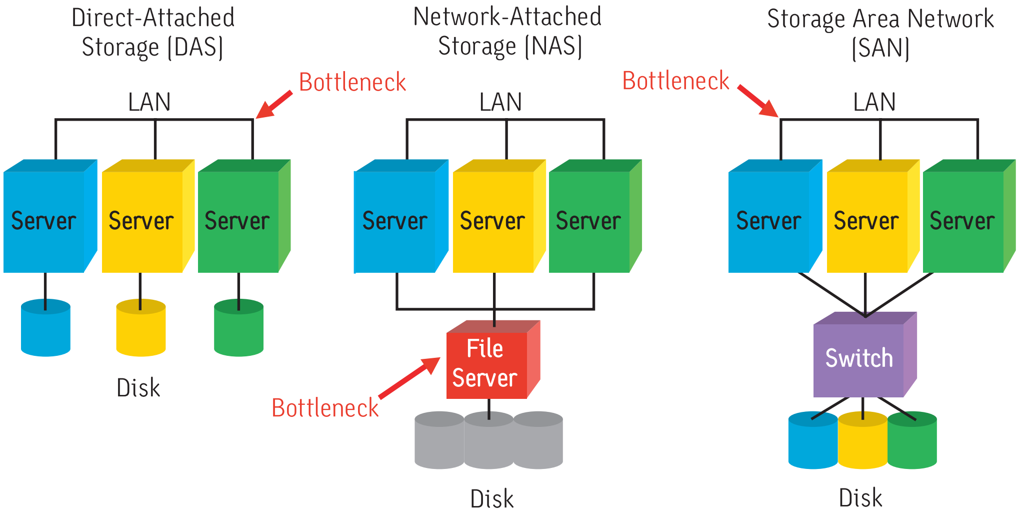


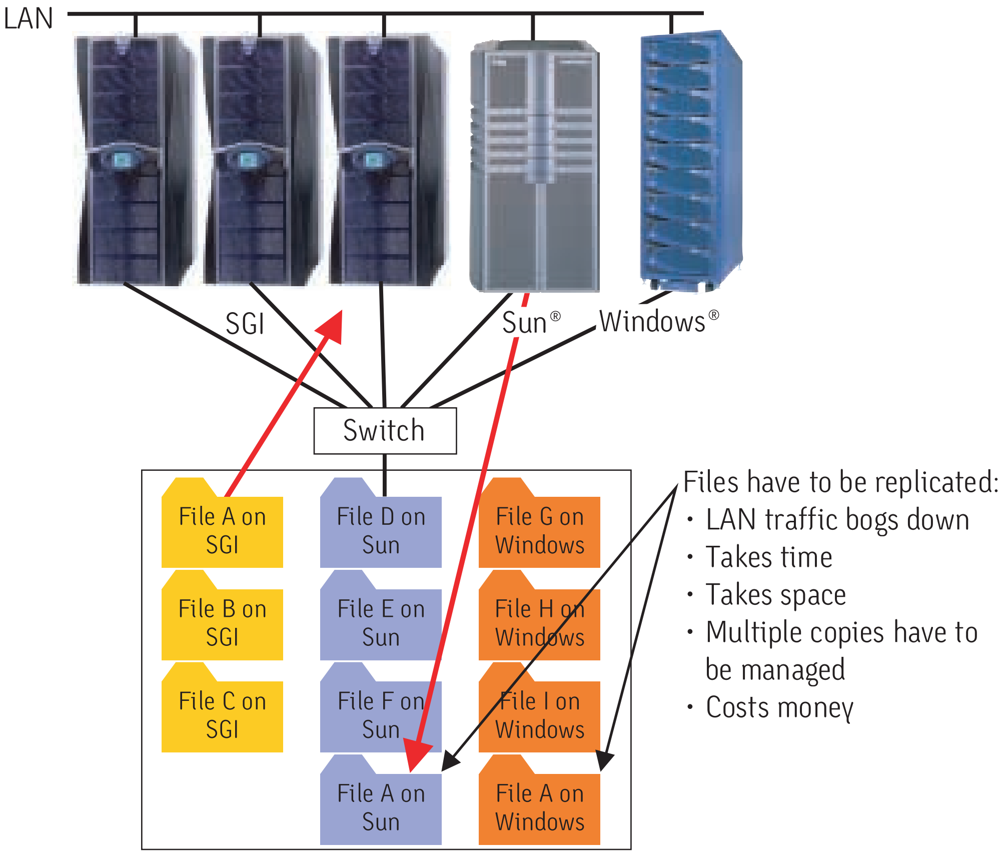
Fig. 1. The Three Main Storage Architectures

the file severs. This creates inefficient use of the physical storage, and keeping the data on the various file servers synchronized and current is complex and expensive.

The first generation of SANs promised to solve many of these problems. With SANs, the storage sits on a dedicated high-performance network along with the computers that access the stored data. Any computer on the network can access any piece of data on the network. But the architecture of ordinary SANs requires that each logical volume of storage be assigned to a single server. (This volume becomes the local volume for that computer.) This server manages requests from other servers for files contained within its assigned local volume. So, for a SAN connected to several computers, only one computer can ever have local access to a file; all other computers in the SAN must undertake the time-consuming process of copying files over the LAN in order to have access to the data in those files. This process is shown in figure 2 and results in slow shared data access; needless replication of files, which takes more disk space; and the need to manage replicated files—all costing time and money.

3.0 The CXFS SAN Filesystem: The Shared Data Access of NAS at SAN Speeds and Scalability

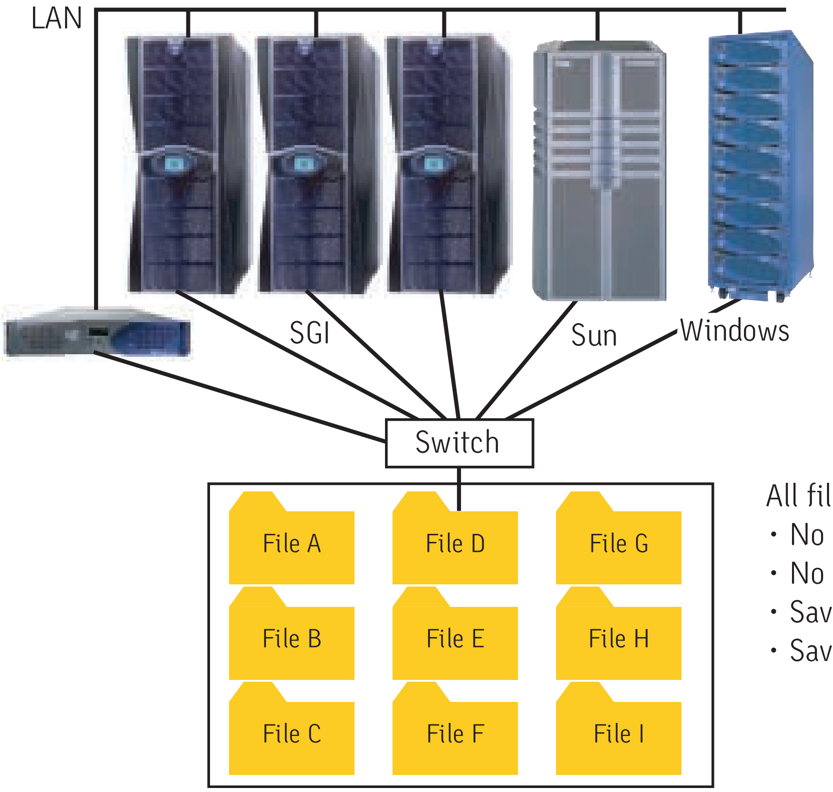
With the introduction of the CXFS SAN filesystem, SGI has laid the groundwork for a new storage paradigm. CXFS combines the shared data access of NAS with the scalability and performance of a SAN. SGI designed the CXFS SAN filesystem specifically for environments where shared data access is critical and local area networks simply cannot provide the necessary bandwidth. CXFS allows all systems in a SAN simultaneous high-speed access to the same filesystem and files. A single system can have multiple connections, making it possible to achieve data rates of multiple gigabytes per second.



Files have to be replicated:

- LAN traffic bogs down
- Takes time
- Takes space
- Multiple copies have to be managed
- Costs money

Fig. 2. File Copying in a Storage Area Network



All files are shared:
- No copying
- No wasted space
- Saves time
- Saves money

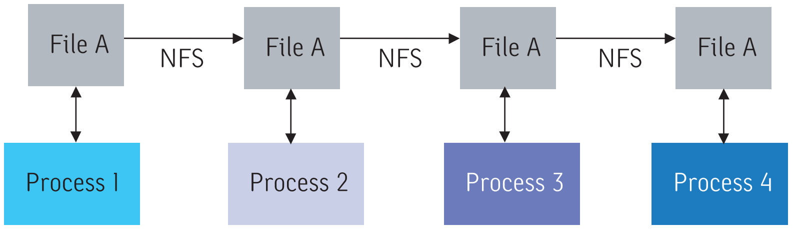
Fig. 3. SAN with SGI CXFS Shared Filesystem

Figure 3 illustrates a SAN with the CXFS shared filesystem. All files can be read and written to by any server in the SAN as if they were their own local files. One system on the SAN acts as a metadata server, controlling file permissions and mediating shared access. Unlike network file sharing, where all data goes through the file server (which often becomes a bottleneck), once the metadata server grants access, systems enabled by CXFS read and write data directly over the SAN to and from disk. CXFS eliminates the copying of files over the network—and the need for the additional time and additional disk space to do so. Should a metadata server fail, a designated backup metadata server automatically takes over management of the CXFS filesystem. This feature—in combination with fully redundant SAN configurations and RAID storage—delivers extremely high availability along with exceptional performance. Even if failures occur, CXFS ensures that a path to access data is always available.

4.0 Positive Effects on Workflow: Getting Projects Done Faster

Creative and technical applications depend on workflows in which shared data access is a necessity. For example, figure 4 illustrates the workflow in the making of a modern movie, where an army of post-production artists uses multiple applications on different systems for digitizing, color correcting, editing, adding special effects, compositing, and other functions. A manufacturing workflow is also illustrated to show the wide range of applicability across many workflows where sharing (or copying) large files is common.

No file sharing means large files have to be moved over the network—taking time, slowing workflow.

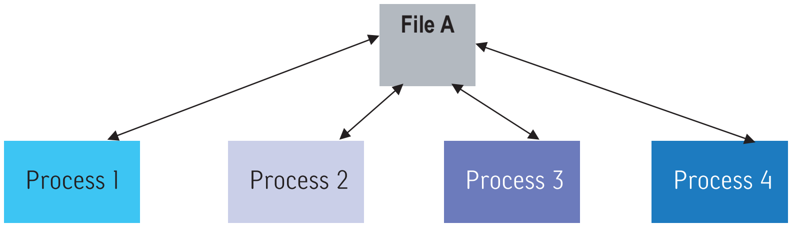


Media	Digitization	Color Correcting	Effects	Compositing
Mfg.	Design	Visualization	Structural Analysis	Crash Analysis

Fig. 4. Workflow without CXFS File Sharing: File A is being copied from one process to the next

These large data files are cumbersome for conventional network filesystems to handle, so users waste valuable time either copying the data over the network from servers to local disk or transferring data manually from one system to another by tape. As shown in figure 5, CXFS significantly boosts productivity where large files are shared by multiple processes in a workflow.

File sharing means large files don't have to be moved over the network—saving time, speeding workflow.



Media	Digitization	Color Correcting	Effects	Compositing
Mfg.	Design	Visualization	Structural Analysis	Crash Analysis

Fig. 5. Workflow with CXFS File Sharing: CXFS provides near-instantaneous access for data-intensive workflows

With CXFS, data-intensive projects take less time to complete at less cost and are easier to manage. With SGI storage, projects get done faster and product gets to market sooner and at lower cost, affecting an organization's top-line revenue-generating capability like no other storage solution available.

5.0 Established Technology

The CXFS SAN filesystem is already in use at over 200 sites, powering some of the most data-intensive applications in the world. Customers in markets such as media, manufacturing, sciences, energy, government research, defense, and education have already embraced CXFS for performance-sensitive data-sharing needs. With the release of CXFS™ 2.1, the proven performance and advanced capabilities of CXFS are now available for the Solaris™ and Windows operating systems in addition to the SGI® IRIX® operating system. By supporting heterogeneous SAN environments, CXFS provides the foundation for a comprehensive data management solution, freeing customers to focus their energy on creative and technical insights without concern for the complexities of data storage.

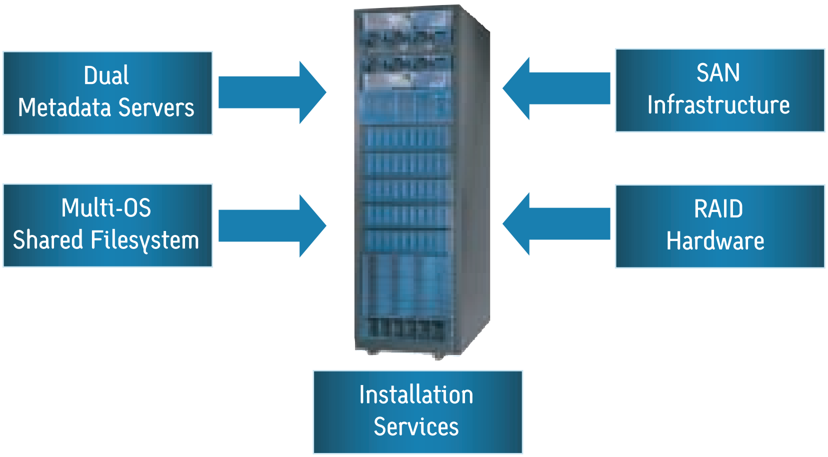


Fig. 6. SGI SAN Server 1000 is a unique solution that simplifies the deployment of a SAN with a shared filesystem

6.0 SGI SAN Server 1000: Turnkey SAN Solutions from SGI

Choosing and integrating the various hardware and software products that make up a modern SAN can be a difficult task. SGI makes it simple for customers to reap the benefits of SAN and CXFS with the introduction of SGI SAN Server 1000.

Illustrated in figure 6, SGI SAN Server 1000 is a turnkey 2Gb-per-second SAN solution with scalable storage and the CXFS shared filesystem. SGI SAN Server 1000 includes a CXFS metadata server and Fibre Channel switch as part of the base configuration. It scales to accommodate up to 29TB of storage with exceptional I/O bandwidth, eliminating many of the deployment complexities associated with SANs. SGI SAN Server 1000 makes a high-performance SAN with CXFS an easy choice for organizations of all sizes, allowing them to increase end-user productivity and decrease the time necessary to solve important problems.

7.0 SGI SAN Server 1000: At the Center of a Complete End-to-End Storage Solution for High-Performance Computing and Visualization

As figure 7 illustrates, SGI San Server 1000 is at the center of a complete data management solution for complex high-performance technical and creative environments in government and defense, manufacturing, energy, media, and the sciences. The SGI end-to-end solution includes backup with Legato NetWorker® and seamless integration of low-cost-per-megabyte “near-line” storage with SGI® Data Migration Facility (DMF). DMF is the industry’s leading hierarchical storage management solution. It automatically migrates less-frequently used online data to less-expensive near-line storage and is integrated with CXFS to provide a single view of all local and migrated files. File serving to multiplatform clients on the network with SGI file servers is also part of the SGI data management offering, along with centralized management of the single shared filesystem with the SGIconsole™ remote multiserver management system. No other vendor in the industry can solve the complex data management problems of high-performance technical and creative environments like SGI can.

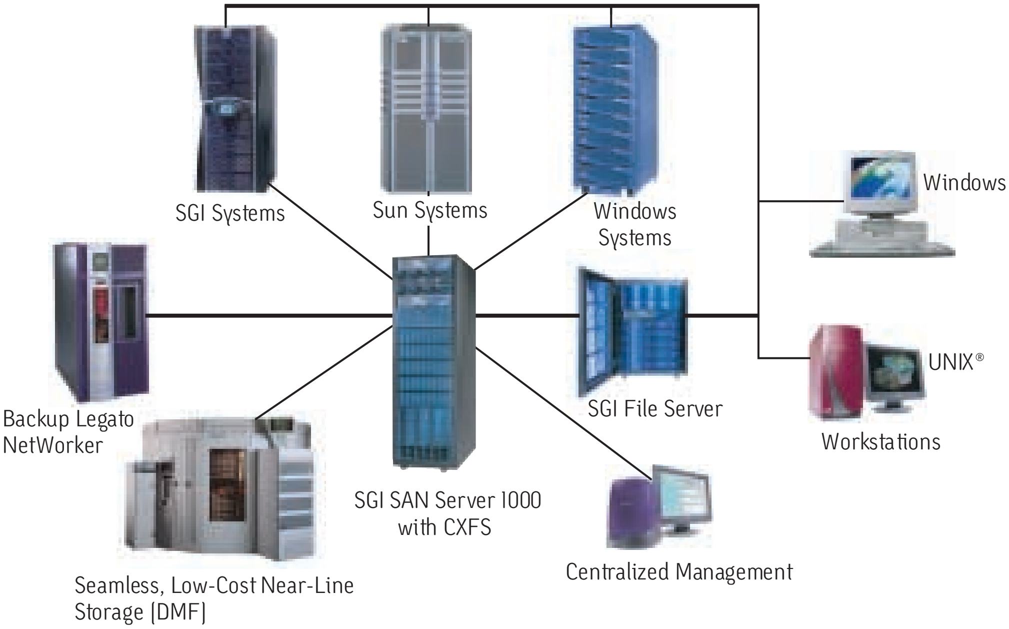


Fig. 7. SGI End-to-End Storage Solution for High-Performance Environments

8.0 Competitive Comparison

Competitive comparisons to well-known, commercially available data-sharing storage architectures help illustrate the superiority of the SGI storage solution. Figure 8 illustrates a sample solution from Network Appliance. The most scalable Network Appliance™ Filer system can hold up to 12TB[1]. In a large-scale data environment, Network Appliance will offer multiple filers to store the required data. This results in an environment with multiple filers, each with its own unique filesystem to manage. As data needs increase, the only solution is to add another filer, and another, and another. The result is increased complexity and management cost.

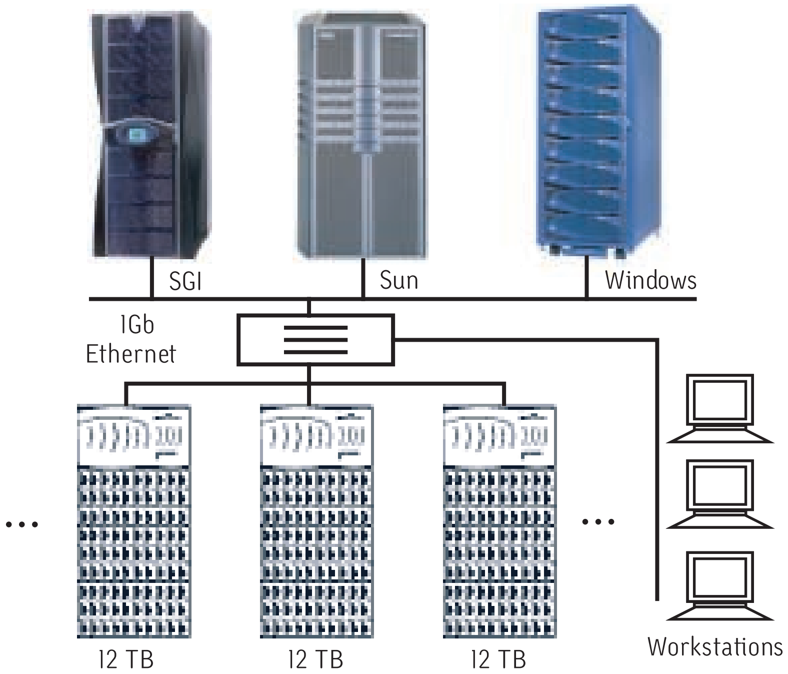
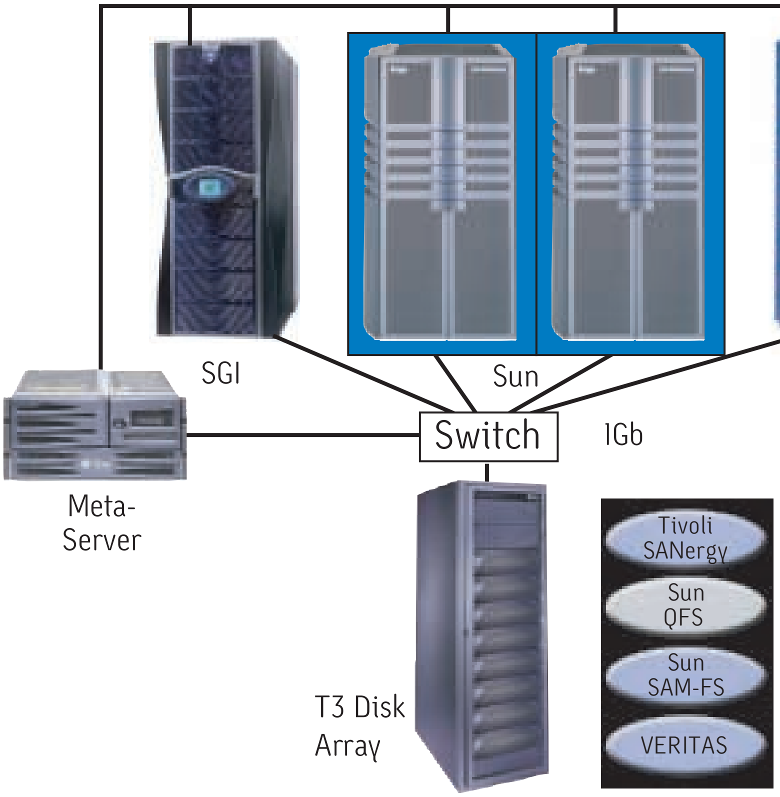


Fig. 8. Network Appliance Solution for High-Performance Environments

[1] Source: *www.netapp.com*; September 2002

Network Appliance data access performance has also been an issue in high-performance environments. Network Appliance's most expensive filer can serve 132MB[2] per second, and this cannot be exceeded by adding multiple NICs or Gigabit Ethernet connections; it is limited by the file-serving processors within the filer head. Compared to SGI, a Network Appliance solution provides much slower data access, limited scalability, and multiple filers and filesystems to manage. This all comes at a very expensive price with painful and even more expensive scalability issues down the road.

The Sun HPC SAN solution is illustrated in figure 9. To provide shared data access in a high-performance computing environment, Sun offers the T3 array with the Sun™ QFS shared filesystem, which only operates on Solaris; the Tivoli® SANergy™ shared filesystem to handle other platforms; VERITAS Cluster Server™ to provide metadata high availability; and the Sun StorEdge™ SAM-FS storage archive for HSM[3]. Sun StorEdge QFS shared filesystem works only on Sun platforms and is a "single-writer, multireader" architecture. That is, only one server in the SAN can write to the filesystem, and that's for all time. Another filesystem is required to handle platforms other than Sun—the NFS-based Tivoli SANergy filesystem—and if high availability is required, then VERITAS Cluster Server must be added. Compared to the SGI solution, the resulting Sun solution has limited functionality (single writer) and slower performance (1Gb SAN), with a collection of unrelated software packages from different vendors and multiple filesystems to manage.



* Competitive information available from vendors' Web site.

Fig. 9. Sun HPC SAN Solution

[2] Source: *www.netapp.com*; September 2002

[3] Source *www.sun.com*; September 2002

9.0 Optimizing Workflow Examples

SGI storage excels in environments where workflow requires large amounts of data to move or be shared among multiple systems in a workflow. The benefits of SGI storage in example workflows are outlined below.

9.1 Optimizing Workflow in Film and Video Post-Production

Working with high-quality digital video assets requires applications to manage huge amounts of data. A single HDTV frame consumes a minimum of 8MB, and a movie displays 24 frames per second (192MB per second). Post-production involves many complex tasks, such as digitizing analog content (35 mm movies), nonlinear editing, digital effects, and compositing. These tasks are usually performed in a workflow process in which data moves from one computer to the next in a sequence.

To date, post-production houses have been relying on cumbersome methods to manage digital assets and move data from one host to another. In some cases, online storage is moved manually from system to system: a direct-attached RAID array on one machine is disconnected after processing on that system completes and then connected to another machine to carry out additional steps in the process. This is an efficient process because it avoids slow copying and offline media, but data availability could be impacted by RAID failures while moving the array back and forth. Also, only one host at a time can access the latest asset.

An asset can also be transferred between systems using tape. Once a processing step is completed, the output of that step is manually copied to tape, carried to the next machine in the processing sequence, and copied back to disk on that system. The efficiency of this process is limited by the bandwidth of the tape device.

Finally, an asset can be copied over a network. Traditional network-based file sharing, like NFS and/or CIFS, or proprietary point-to-point networks are used, but due to limited network bandwidth and protocol overhead, it can take hours or overnight to transfer files between machines, seriously impeding work.

A leader in post-production responsible for more than 30 productions a year (analog and digital) was recently confronted with these challenges in its workflow. With more than 15 machines involved in the process, sharing data had become too complex and too slow, even using Gigabit Ethernet. A SAN enabled by CXFS was created, including SGI® Origin® 200, SGI® Origin® 2000 series, Silicon Graphics® Onyx2®, SGI® Onyx® 3000 series, and Silicon Graphics® Octane® systems. Each machine or set of machines is responsible for a different aspect of processing using specialized applications.

By instantaneously sharing files, it now takes 10 minutes to process an asset that previously required all night to copy from one system to the next. Project completion times are cut, and the production house can complete more projects in the same amount of time, making full use of its post-production assets and gaining dramatic returns on investment.

9.2 Optimizing Workflow in Oil and Gas Exploration

Oil and gas exploration and digital media applications depend on a workflow-processing model in which multiple machines work serially to process a data and information stream. Output from a system is passed to another system for postprocessing and so on. This processing model requires sharing large quantities of data. Many sites are still using inefficient methods such as FTP or NFS to copy data between machines and limit the size of the data sets they process, because of inadequate bandwidth for transferring data. By allowing applications on different machines to share data at high speeds without copying, a SAN enabled by CXFS can save a tremendous amount of time and money.

A large oil and gas company is using a SAN enabled by CXFS in its seismic data analysis operation to help discover new petroleum fields. Specialized applications have been developed in-house to process data from field studies and image geological features below the earth's surface. Compute-intensive applications like this one typically generate so much data that the data set has to be segmented into smaller pieces to keep data transfer times between systems manageable.

The main application begins processing a data set and the output is directed to a file that resides on a CXFS shared filesystem. Once a set amount of output has been created, a second application running on a separate system begins processing the output without waiting for the first application to complete. The second application synchronizes with the first to ensure that it does not read past the current end of file. The second application directs its output to a new output file in the shared filesystem, and the process repeats through several additional processing steps until completion.

The use of SGI storage with CXFS has decreased the time required for start-to-finish processing of a data set by as much as 35%. The customer has also been able to process data sets up to three times larger than were previously possible.

10.0 Summary

SGI has developed a unique storage architecture that combines the file-sharing capabilities of NAS with the bandwidth and scalability of SAN. This new architecture streamlines workflow in data-intensive environments, allowing projects to be completed faster than any other storage architecture available and positively impacting an organization’s top line like no other storage solution can. SGI storage also provides unique lower total cost of ownership (TCO) advantages because it reduces overall storage requirements (fewer disks) by eliminating needless copying and providing a single filesystem view that is easier to manage, back up, archive, and restore. SGI delivers these capabilities in a turnkey SAN solution with integrated hardware, software, and installation services that get customers up and running quickly and easily. Furthering TCO, the SGI storage solution provides seamless integration with SGI Data Migration Facility, the industry’s leading hierarchical HSM solution for near-line storage.

An SGI SAN with CXFS is configurable for no single point of failure and has the highest availability SAN shared filesystem on the market—removing dependence on a single server for data access. With the SGI solution, if any server fails, data is still accessible by all others on the network. Highest bandwidth and scalability are provided by state-of-the-art 2Gb Fibre Channel SAN infrastructure and storage technology and an 18 exabyte (18 million terabyte) filesystem capacity. No other storage vendor can provide such compelling benefits in a high-performance, data-intensive environment as SGI can.

Corporate Office
1600 Amphitheatre Pkwy.
Mountain View, CA 94043
(650) 960-1980
www.sgi.com

North America 1(800) 800-7441
Latin America (52) 5267-1387
Europe (44) 118.925.75.00
Japan (81) 3.5488.1811
Asia Pacific (65) 771.0290