

Technical Case Study: SpinServer® Deployment at GX Technology Enhances Performance and Storage Management while Reducing Cost

Network Appliance | October 2004 | TR-3346

TECHNICAL CASE STUDY

Network Appliance, a pioneer and industry leader in data storage technology, helps organizations understand and meet complex technical challenges with advanced storage solutions and global data management strategies.

Executive Summary

GX Technology (GXT) is a seismic processing services company based in Houston, Texas. Because GXT must complete processing on time to meet customer commitments, and because seismic processing is extremely compute- and data-intensive, GXT began migrating its time-processing service from its legacy Sun™ enterprise servers to a grid infrastructure consisting of clusters of bladed Linux® servers and network storage. A Network Appliance™ SpinServer storage system was selected to provide all storage for the application because of its superior reliability and performance.

SpinServer is a flexible, high-performance grid storage system that can scale to encompass as many as 512 cluster nodes and a total of 11PB (11,000TB) of storage in a single installation. GXT currently utilizes 16 SpinServer nodes with 250TB of storage under a single global namespace. This unique feature of the SpinServer architecture allows network clients to view all storage on any of GXT's 16 nodes as if it were a single large file system and access it all with a single mount. Virtual file systems (VFSs) stored on individual nodes can be transparently migrated between servers to balance workload and optimize performance without disrupting ongoing work. GXT uses this feature regularly and depends on the overall ease of use of SpinServer to help achieve critical processing goals.

As a result of the SpinServer deployment, GXT has been able to expand production and run more jobs in parallel with greater performance. The great flexibility and ease of use of the system have allowed GXT to redeploy one of the two administrators formerly responsible for the Sun servers. A single individual manages the 250TB SpinServer installation plus performs other duties. Because of cost savings due to decreased management costs, decreased server acquisition costs, and increased performance, GXT estimates that the SpinServer deployment has resulted in at least a tenfold gain in overall price/performance over that of the previous environment. This paper discusses the deployment of SpinServer at GXT in some detail, exploring the unique features and benefits of the solution that make it uniquely suited for busy processing environments such as that at GXT.

Introduction

Network Appliance SpinServer is a flexible, high-performance grid storage system that can scale to encompass as many as 512 servers and a total of 11PB (11,000TB) of storage in a single installation. This scalability enables SpinServer to provide a tremendous amount of aggregate bandwidth to support the most data-intensive computing environments. Advanced load-balancing capabilities and superior ease of use make SpinServer ideal for high-performance computing (HPC) installations such as grid environments with thousands of Linux compute servers. This technical case study describes the deployment of SpinServer at GX Technology, a state-of-the-art seismic data-processing company based in Houston, Texas.

Background

GX Technology (GXT, www.gxt.com), a division of Input/Output (I/O), provides seismic processing services for major oil companies and other clients involved in oil and gas exploration and production. Seismic processing is both compute- and data-intensive. A single processing task may consume hundreds or thousands of processors and require weeks or months to complete, while consuming tens of terabytes of storage in the process. Data from GXT studies is often used to make decisions about bids on leases, so it is critical that processing complete prior to the date bids must be submitted. As a result, GXT customers typically contract for completion by a certain date, and ensuring that processing proceeds according to schedule is critical.

GXT offers a variety of services, including acquisition, time processing, prestack depth and time imaging, 4D, and multicomponent. For its time-processing service, GXT originally utilized up to 10 Sun E6800 enterprise-class servers, each with 24 CPUs, 48GB of memory, and 15TB of direct-attached storage. This configuration worked acceptably until projects began to exceed 15TB in size. To continue operations, GXT had to spread the data for a single project across two or three systems. Data stored on remote systems was accessed via NFS. Spreading data across multiple systems significantly increased data management overhead, and the mixed workload of computation and file serving put an excessive burden on the Sun systems, which were no longer able to deliver the required performance.

GXT had previously begun deploying clusters of bladed Linux servers to provide flexible, economical compute capabilities for some of its other services. When the Sun servers used for time processing became a problem, it was natural to begin moving to a grid-like configuration with Linux servers and consolidated network storage.

After a careful evaluation, GXT selected NetApp SpinServer to meet its storage needs. GXT's major priorities were reliability and performance to ensure that important processing jobs continue around the clock without interruptions or bottlenecks. SpinServer was evaluated against competitors and found to offer not only superior reliability and performance, but also a variety of other features that made it the ideal choice.

SpinServer Deployment at GXT

A SpinServer cluster is a group of individual SpinServer nodes connected through a dedicated cluster interconnect (typically Gigabit Ethernet). Each SpinServer node controls one or more

Customer Challenges

- Expand processing capability and improve efficiency
- Consolidate direct-attached storage from 10 Sun servers

SpinServer Value Proposition

- Cost-effective grid storage
- Single global namespace
- Ability to dynamically balance load across storage systems for maximum performance without user disruption

Deployment Results

- Increased production capacity
- Streamlined management, enabling redeployment of one headcount
- Lower storage/maintenance costs
- Tenfold price/performance improvement

storage pools. Every storage pool, in turn, contains one or more virtual file systems (VFSs). VFSs are directory trees, each consisting of a root directory and a set of nested subdirectories and associated files.

Global Namespace

A SpinServer cluster exports a single global file system by arranging the complete set of VFSs from all SpinServer nodes into a single tree. One VFS is identified as the root VFS, which defines the root directory of the global file system. All other VFSs are joined to the global file system in a designated manner. A network client sees a single large file system and can access any part of the global namespace without having to mount the specific SpinServer node that stores the desired data. The aggregate bandwidth available from the single global namespace scales linearly with the number of systems in the cluster, making SpinServer ideal for applications where high bandwidth is critical.

For its time-processing deployment, GXT has a single SpinServer cluster, which currently consists of 16 cluster nodes, each with about 15TB of storage, for a total cluster capacity of about 250TB. GXT allocates storage for each project when the project starts. Each VFS is configured to be about 1TB in size. VFSs are arranged within the global namespace to meet the needs of each project. A low-priority project might draw all its storage from one or two cluster nodes, while a high-priority project might utilize VFSs on all cluster nodes to ensure maximum throughput. If a project unexpectedly requires more storage, it can be flexibly allocated from free space on any of the 16 servers and made available in the correct location in the global namespace.

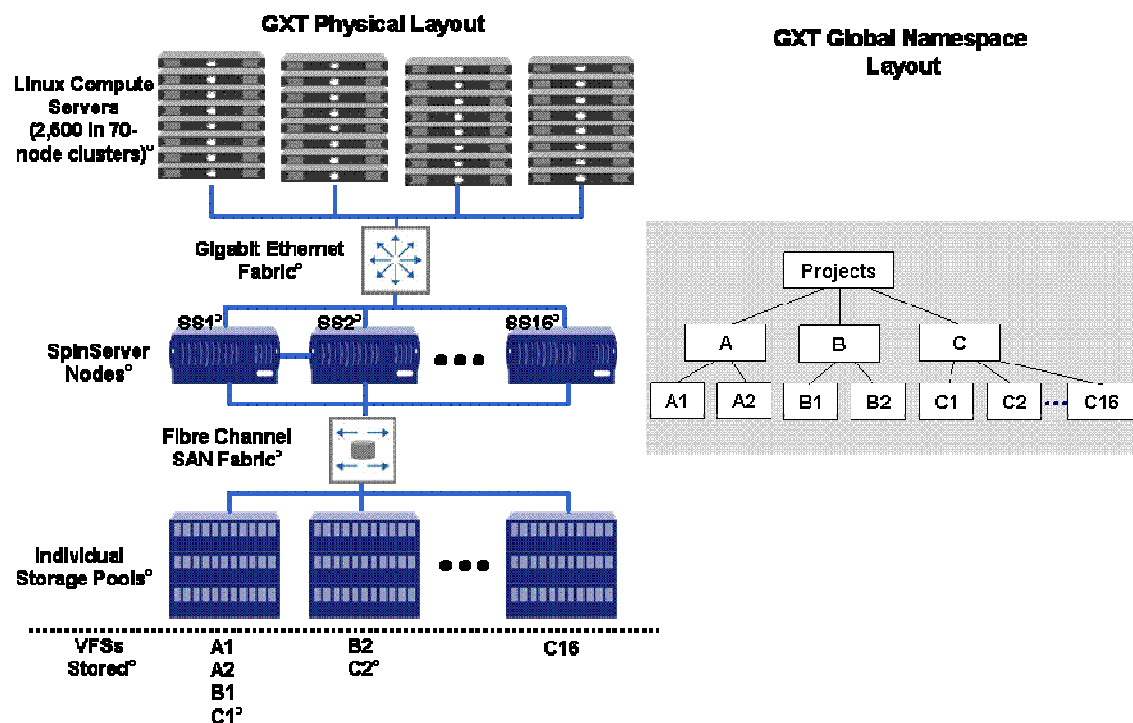


Figure 1) Physical and logical infrastructure at GXT.

Workload Balancing

Because of the workload-balancing features of the SpinServer architecture, it is simple for GXT to adjust the storage configuration for a project to increase performance. Since the location of a VFS in the global namespace is completely independent of the node storing the VFS, a VFS move operation allows the contents of a VFS to be moved from one server to another at any time.

The first stage of a VFS move copies the contents of the VFS over the cluster interconnect. Read and write operations occur with no disruption and are directed to the original location. These updates are propagated to the new location during the second stage. The final stage of the move operation transfers file-locking information between the source and destination servers, at which point the new VFS becomes active, and the old VFS is released. The names and locations of the files in the global namespace do not change as a result of the move, nor do any mount points become invalid.

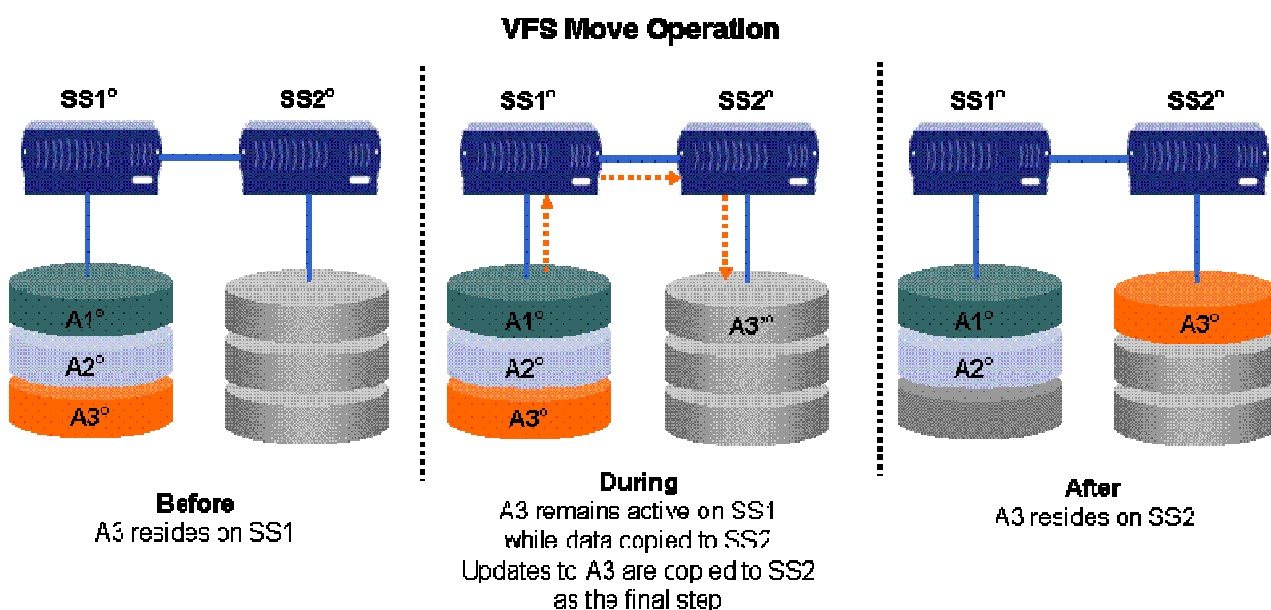


Figure 2) VFS move showing the migration of VFS A3 from SS1 to SS2.

GXT performs this type of operation all the time, to balance workload or increase the bandwidth available to a particular project.

If storage becomes a bottleneck for a particular project, or the priority of a project increases, a GXT administrator can simply use VFS move to spread VFSs across more systems. Likewise, if a particular server in the cluster becomes hot because it contains too many highly active VFSs, these can be spread across more nodes to balance the load.

Ease of Management

A typical time-processing project lasts only a few months, so project setup and teardown occur frequently. This is another area where the SpinServer architecture is a major advantage for GXT. Both physical and logical storage configuration can be done from an administrator's desktop, since everything is done in software. Reconfiguring for a new project takes GXT less than 30 minutes. During daily operation, GXT simply monitors the performance of its SpinServer nodes and makes adjustments with VFS move when necessary, without downtime or disruption to ongoing work.

When more storage bandwidth is required, another SpinServer node can be added to the cluster without bringing the cluster down, and workloads can be transparently migrated to that system, which becomes part of the global namespace. If an individual system requires maintenance, its VFSs can be transferred to the other cluster members (assuming they have enough free storage space) so the system can be serviced without disrupting operations.

Implementation

Migrating several hundred terabytes of storage from Sun servers to SpinServer was a significant undertaking for GXT, made even more complicated by limited physical space in its busy data center plus the need to keep all projects moving forward.

To complete the migration, GXT purchased 25TB of additional storage capacity. Data was migrated from one Sun server to the SpinServer cluster across the network using NFS, and then the storage arrays attached to the Sun server were disconnected, reattached to a SpinServer system, formatted, and readied for another round. This process was repeated until all data was transferred from the Sun systems to SpinServer. All existing storage arrays were recycled as part of the SpinServer configuration for significant cost savings.

SpinServer Impact

The SpinServer implementation at GXT has been extremely successful, delivering not only great reliability and performance but also improving flexibility, simplifying management, and reducing costs across the board.

The combination of bladed Linux servers and SpinServer has resulted in a flexible grid architecture that adapts readily to changing demands. According to Bill Eue, enterprise storage administrator for GXT, "With 10 to 20 active time-processing projects at any given time, GXT typically utilizes four or five 70-node Linux clusters (280 to 350 systems) to meet its processing needs, but we can flexibly allocate up to 10 such clusters from across the company if necessary for big jobs and tight deadlines. These Linux systems only require a single mount of the SpinServer namespace to gain access to all the storage for a project, so both server and storage resources can be quickly allocated or reallocated to meet changing needs.

"For us, being able to physically control data locality is critical to achieving optimal performance to meet our processing deadlines," Eue continues. "SpinServer makes it possible to load balance across all cluster nodes. I periodically move VFSs between systems, and I do it without leaving my office. Physically controlling which systems store which data allows us to achieve maximum performance."

Ease of management has resulted in considerable increases in management efficiency and reduced overall management costs. According to Eue, "We used to have several systems engineers responsible for maintaining the 10 Sun systems. Now one person can manage the SpinServer system and have 50% free time for other duties. This has given us a net gain of at least one headcount, which we have redeployed to accommodate growth instead of making a new hire."

The SpinServer deployment resulted in a significant increase in both overall performance and price/performance. "With SpinServer and our overall processing architecture, we can now complete more projects in parallel and expand our production," says Eue. "A single SpinServer system has the same storage capacity and twice the data throughput of a 24-CPU Sun server but costs significantly less, and naturally a SpinServer system costs much less to maintain. I would guess that our price/performance increased by a factor of at least 10."

Conclusion

The SpinServer implementation at GXT has been extremely successful. The scalability, performance, and ease of management of the solution combined with the single global namespace and workload-balancing capabilities have had far-reaching benefits. According to Eue, "The implementation has met or exceeded our expectations in every area."

Because of this success, GXT is continually evaluating SpinServer for other applications and other services as well as expanding the capabilities of its current system. "We are currently planning to deploy a MetroCluster by the end of the year to improve our disaster recovery capability," concludes Eue. "We are consolidating operations from three data centers in the Houston metropolitan area down to two. Each location is expected to have components of the SpinServer cluster, and we plan to use the built-in mirroring capability to maintain copies of all critical data in both locations. That way, if one location is hit, we would still have access to everything. Even in this distributed configuration, SpinServer appears as a single global namespace and can be managed as a single entity from one location."

Additional Resources

"SpinServer Systems and Linux Compute Farms," Network Appliance TR-3304, February 2004, www.netapp.com/tech_library/ftp/3304.pdf.

"Meeting the Storage Challenges of Today's Linux Grid Computing Enterprise," Network Appliance TR-3325, 6/3/2004, www.netapp.com/tech_library/ftp/3325.pdf.