SGI Capability Brief

Wide Area Filesharing Based on CXFS: High-Performance, Multiplatform Data Sharing for WANs

Introduction

As the amount of stored data and the size of individual data sets continues to increase, data sharing—enabling multiple computer systems to access and manipulate the same data—has become increasingly difficult.

Data sharing is rapidly evolving from network filesystems such as NFS and CIFS to high-performance storage area network (SAN) filesystems such as SGI® CXFS™. CXFS meets the data sharing needs of small organizations, campuses, and—by using multiple fibre links up to 10 km—may span even greater distances. However, existing solutions are unable to meet the data sharing needs of global corporations, military and government applications, or research organizations with multiple installations. Data sharing across significant geographical distances is largely dependent on local copies of data. Copies are created either by transferring data over a wide area network (WAN) or by copying the data to tape and sending it by overnight delivery. In either case, the result is duplicated data. Maintaining multiple copies of large data sets is prohibitively expensive and it is often impossible to find enough storage to accommodate copies. Having multiple copies of data results in inevitable problems with data integrity, because changes in one copy of the data are not reflected in the others. The lack of synchronization makes real-time collaboration using the same data at multiple locations impossible. In addition, many valuable staff hours are wasted waiting for data to be copied. In summary, current solutions to the wide area data sharing problem are slow, expensive, and difficult to administer.

At Supercomputing 2002 (SC2002), SGI, LightSand, and Marconi have for the first time demonstrated the ability to enable true data sharing over wide area connections using a geographically distributed Fibre Channel SAN and technology derived from the SGI CXFS SAN filesystem. This solution allows users to share a single filesystem in a multi-OS environment over great distances. This capability holds significant promise for wide area data sharing, data protection, and academic and commercial grid computing, while some are already envisioning new commercial possibilities.

Enabling Wide Area Data Sharing

The wide area filesystem demonstrated at Supercomputing 2002—Wide Area CXFS—utilizes a specially adapted version of the SGI CXFS SAN filesystem and LightSand™ S-series gateways to transmit Fibre Channel and IP traffic generated by CXFS across a wide area connection. Under all conditions tested, data access was faster than with FTP over the same network, demonstrating a performance advantage in addition to the advantages of a single namespace and no duplication of data.

The solution offers all the standard features of SGI CXFS, a standards-based 64-bit filesystem capable of simultaneously supporting SGI® IRIX®, Solaris™, and Windows® platforms. Unlike network file sharing, where all data goes through the file server (which often becomes a bottleneck), systems with CXFS read and write data directly over the SAN to and from disk.

Adapting CXFS for use over a wide area network required that CXFS be internally modified to accommodate greater latencies and that IP metadata traffic and SAN data traffic be bridged together over the WAN. Bridging is accomplished using LightSand S-series gateways. These gateways efficiently transport both Fibre Channel and IP data over a WAN, directly coupling Fibre Channel onto readily available SONET, DWDM, or dedicated fibre while preserving the buffer-to-buffer credit mechanisms that give Fibre Channel its high performance. The LightSand™ S-600 gateway provides data rates of up to 74MB per second at distances in excess of 8000 km.

For proof-of-concept tests, six SGI® Origin® 300 servers were configured in a single Fibre Channel SAN cluster spanning a wide area connection. Each side of the connection was configured with three servers and local storage and connected to a LightSand S-600 series gateway, which in turn connected to a WAN

simulator. The WAN simulator—capable of simulating distances up to 8000 km (5000 miles)—was used in place of an actual WAN to facilitate testing.

This configuration achieved disk I/O in excess of 60MB per second in each direction across the simulated WAN for an aggregate performance of 120MB per second. A similar configuration was demonstrated in the SGI booth at Supercomputing 2002 with an SGI® system on one side of the WAN connection and SGI, Sun, and Windows systems on the other.

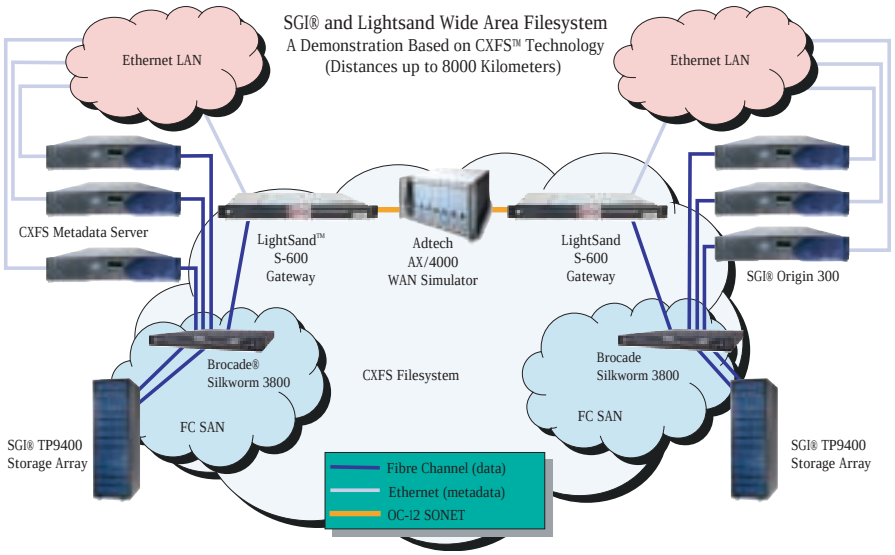

Fig. 1. Proof-of-Concept Configuration for Wide Area CXFS

Demonstrating Wide Area Data Sharing in the Real World

A real-world demonstration capable of achieving significantly greater bandwidth over distance was also presented at Supercomputing 2002 using technology from SGI, LightSand, Marconi plc, Brocade, Bay Microsystems, and the Naval Research Laboratory (NRL). An SGI® Onyx® family system and a Silicon Graphics Fuel™ workstation running a satellite imagery application were configured as CXFS clients and linked via the Advanced Technology Demonstration Network (ATDnet) to an imagery database residing on a CXFS SAN at the NRL facility in Washington, D.C.—a distance of almost 50 miles from the convention site.

At the core of the demonstration was the Marconi® BXRTM-48000 ATM switch. This 480-Gb per-second switch-router was configured with new interface cards that provide OC-192c (10Gb per second) physical interfaces. This OC-192c connection provided the link between the NRL and the SC2002 show floor (in the Marconi booth). Once the data arrived at the SC2002 show, the BXR-48000 broke out two OC-48 ATM streams and delivered those streams to a Bay Microsystems Internetworking Processor (InP). The InP converted the two OC-48 ATM streams back into OC-48 Fibre Channel over SONET for delivery to two LightSand S-2500 gateways. The S-2500 gateways split the data into four 1Gb-per-second Fibre Channel streams that were connected to a pair of Brocade® 3800 Fibre Channel switches, which in turn connected to the SGI systems.

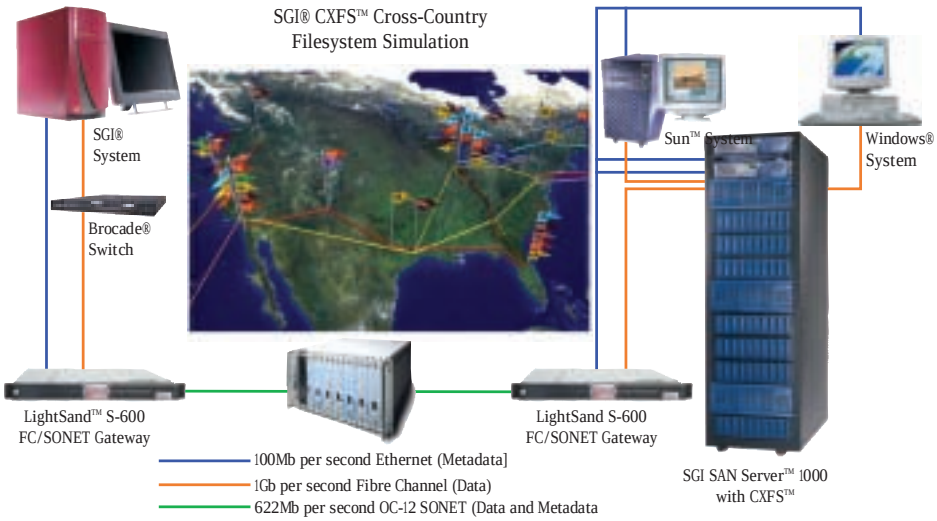

Fig. 2. Demonstration Configuration from Supercomputing 2002

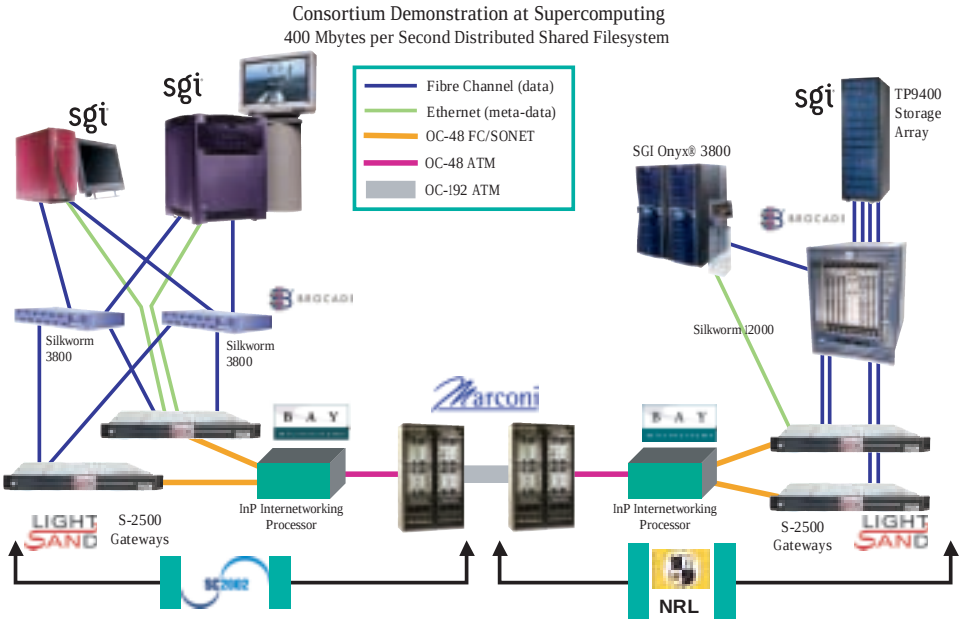

Fig. 3. Demonstration Configuration from Supercomputing 2002

The satellite imagery application is an advanced "space to face" application that utilizes NRL's MOADB (mother of all databases), a 50TB data set. The application includes satellite footprint modeling, ocean temperature simulations, and hyperlinks to other data sources. The application demonstrated sustained bandwidth of over 300MB per second per host between the two sites. NRL was pleased with the overall performance of the application over Wide Area CXFS, especially given that the technology is only at the demonstration phase.

New Opportunities

The technology demonstrations at Supercomputing 2002 have tremendous implications for the future of wide area data sharing, showing for the first time that—with appropriate network technology—it is possible to achieve extremely high data rates on a single filesystem shared between geographically remote sites.

Judging from the intense interest, organizations that have already deployed or are deploying the necessary bandwidth between sites will be quick to adopt this technology when it becomes commercially available. SGI is planning to release a solution based on this demonstration technology within the next six months.

It's true that dedicated OC-12c or higher bandwidth is an expensive resource. However, many large corporations and government and military research installations already have the available bandwidth to begin deployment. In addition, existing and planned computing grids with high-speed networks joining different locations are ideally suited for this technology, since it helps to solve one of the difficult data management problems that grid computing environments are currently facing. Copying large data sets from one location to another to use different grid resources is incredibly cumbersome, but until now no alternative has been available.

Some attendees at the Supercomputing conference were quick to envision new types of network services in which companies could outsource data processing of various types without losing control of their data. They could simply make a filesystem available via CXFS for the duration of the needed processing. No copies of critical data would ever be needed (or allowed) and output results would be written back to the company's own storage.

Whatever the eventual applications of this technology, it clearly represents a next step in the evolution of computing—allowing data storage and computing to be geographically separated.

Additional Information

http://www.sgi.com/products/storage

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) 6771.0290

© 2003 Silicon Graphics, Inc. All rights reserved. Silicon Graphics, SGI, IRIX, Origin, Onyx, and the SGI logo are registered trademarks and CXFS, XFS, and Silicon Graphics Fuel are trademarks of Silicon Graphics, Inc., in the U.S. and/or other countries worldwide. Windows is a registered trademark or trademark of Microsoft Corporation in the United States and/or other countries. All other trademarks mentioned herein are the property of their respective owners.

3440 03/10/2003]

J14203