SGI® InfiniteStorage Shared Filesystem CXFS™

Features

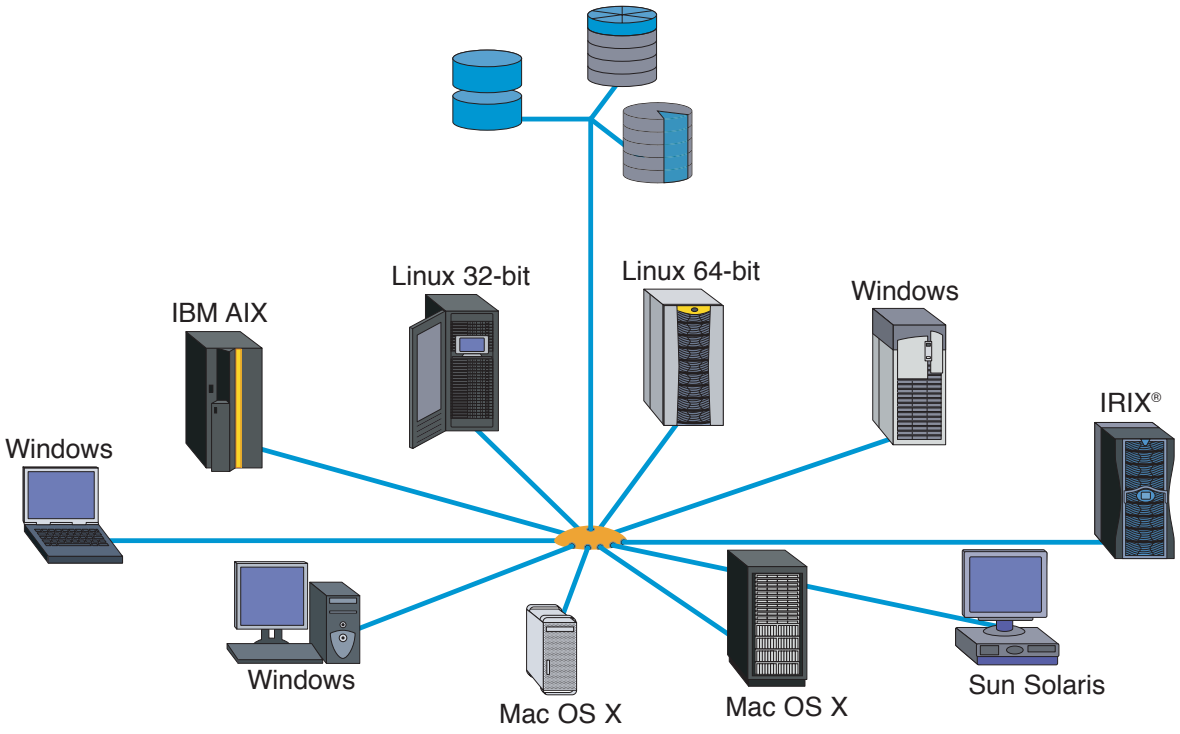
- Delivers on the promise of SANs
- Instant, multi-OS, no copy data sharing
- Time-tested, proven solution
- Scalability and performance ensure investment protection
- Solid, standard data integrity
- Architected to scale up to 18 million terabytes
- Guaranteed application bandwidth priorities with GRIO v2
- High availability with data access failover
- Most robust storage resource management available
- True LAN-Free backup and restore
- Complete SAN/NAS gateway

Delivering on the promise of Storage Area Networks (SANs)

In a typical data intensive environment, 20 to 80% of the compute, network and storage infrastructure and the time it takes to produce results is wasted by slow, resource-intensive data access and sharing methods like network mounts and data copies. SGI® InfiniteStorage Shared Filesystem CXFS™ provides instant data sharing with no network mounts and no data copies. The result is all the total cost of ownership (TCO) benefits of SAN consolidation plus a significant increase in overall productivity for real Return on Investment (ROI).

A SAN provides direct, high-speed physical connections between multiple hosts and disk storage. CXFS provides the software infrastructure to allow simultaneous shared access to that storage—all systems have direct access to all data and are no longer bottlenecked by slow, congested networks or overloaded file servers. This means that in an environment where data is passed on from one task to the next, or needs to be accessed simultaneously by multiple systems, there is a huge reduction in the time needed to do so, and the added benefit of no extra data copies to store and manage.

CXFS combines unparalleled 64-bit scalability and performance with SAN technology, enabling computer systems to efficiently share data and do more work in less time. Currently, IRIX®, Sun℠ Solaris™, Windows®, 32 and 64-bit Linux®, IBM® AIX®, Mac OS® X and 64-bit Linux for SGI® Altix® are supported, with additional operating systems to be added in the near future.



SGI InfiniteStorage Shared Filesystem CXFS

Instant, no copy data sharing among all major operating systems

CXFS is a revolutionary shared SAN filesystem for high-performance computing environments. The ability to share data between computer systems is often mission critical, but until now the technology for no compromise data sharing has been lacking. Data sharing is still most often accomplished using a network filesystem such as NFS or by manually copying files using FTP, a cumbersome and unacceptably slow process.

CXFS allows multiple computers direct access to shared files, while delivering the same performance as XFS®, the industry leading SGI® filesystem that has run reliably on hundreds of thousands of SGI systems for many years. This means that all systems in a SAN enabled by CXFS have access to the same file at the same time, at local or near-local filesystem speeds.

Time-tested, proven solution

CXFS was designed as an extension to the award-winning 64-bit XFS filesystem, widely recognized as the most scalable, highest performance filesystem available. CXFS technology is field proven, running reliably at hundreds of customer sites. CXFS provides:

- Industry-leading performance
- Journaling for reliability and fast recovery
- 64-bit scalability, to support extremely large files and filesystems
- Dynamically allocated metadata space

Scalability and performance ensure investment protection

CXFS is architected to address single files as large as 9 million terabytes and filesystems as large as 18 million terabytes, guaranteeing that CXFS can exceed user data requirements far into the future. Dynamic allocation algorithms ensure that a filesystem can store millions of files without wasting disk space. A single directory can contain millions of files without degrading performance. With the advanced XVM volume manager, disk volumes can be configured across thousands of disks, ensuring that CXFS will grow to meet future storage needs.

Data sharing using CXFS is more efficient than traditional methods, reducing the total cost of ownership for storage. CXFS is designed to reduce costs by centralizing and consolidating storage, reducing data duplication, lowering administration costs, removing LAN bottlenecks, and reducing time wasted waiting for data. CXFS configurations scale easily through the addition of disks for more storage capacity or the addition of Fibre Channels for more bandwidth.

CXFS performance far exceeds that of most filesystems, for typical workloads such as:

- Reading and writing to a file opened by a single process
- Reading and writing to a file where all processes with that file open reside on the same host
- Multiple processes on multiple hosts reading the same file
- Multiple processes on multiple hosts reading and writing the same file, using direct I/O

These capabilities make CXFS ideal for large file-based applications such as film and video post-production, weather forecasting, and geospatial imaging. CXFS performance is superior to that of network filesystems such as NFS. The speed of the network, the size of the server, and the protocol itself can bottleneck NFS performance. NFS protocol is synchronous—each block of data requires two network I/Os. CXFS requires a few metadata I/Os, after which the data I/O is direct to disk. CXFS can use multiple Host Bus Adapters to scale a single system's I/O path. CXFS and NFS work well together. NFS servers can run on any system enabled by CXFS that is exporting the same filesystem. This compatibility allows tremendous scaling of NFS servers and greater failover capabilities.

Solid, standard data integrity

Streamlined metadata management is the key to CXFS performance. Metadata includes information about files (size, ownership, timestamps, location, etc.) as well as information about the filesystem itself. While all file data in CXFS flows directly between servers and disk, transactions that alter metadata are brokered by a metadata server for each CXFS filesystem, which coordinates access and ensures data integrity.

Metadata transactions are routed over a TCP/IP network to the metadata server. Because metadata transactions are typically small and infrequent relative to data transactions, a fast Ethernet connection is typically adequate; however, faster connections such as Gigabit Ethernet are also supported and are recommended for high-availability environments.

Several underlying XFS features enhance metadata performance:

- Fast metadata algorithms with excellent buffering
- Sophisticated structures and algorithms for fast lookups
- Ability to allocate large extents, minimizing metadata transactions for space allocation
- Asynchronous transaction log for even faster metadata operations

Extreme care has been taken to ensure the speed of metadata transactions. CXFS utilizes:

- Specialized thin RPCs
- Excellent buffering of both data and metadata on clients
- Multiple metadata servers (one per filesystem)
- Ability to bypass the CXFS layers on the metadata server
- Ability to dedicate small hosts as metadata servers

CXFS uses a fast and efficient token mechanism to control file access, guaranteeing transaction accuracy that far exceeds that of typical network filesystems. Because file locking is critical, CXFS takes great care in the implementation of POSIX®, BSD, and SVR4 file locks. No application changes are required to use file locking with CXFS.

Guaranteed Rate I/O (GRIO) ensures application bandwidth priorities

Applications like satellite stream acquisition, film post production and media broadcast frequently have the highest quality of service requirements. GRIO V2 is a unique solution which guarantees that a required level of bandwidth is available for a specific user or application within the CXFS SAN.

With GRIO V2, any system or application in the SAN can be guaranteed a dynamic or static stream for a specified period of time. Other systems and applications can be given priorities for their claim on the remaining available bandwidth. GRIO V2 central service throttles non-guaranteed system or application available bandwidth based on the total available SAN bandwidth. This capability scales with the SAN environment and does not restrict configuration flexibility.

GRIO V2 is an option available today for IRIX metadata servers and all clients platforms running CXFS.

High availability with complete data access failover

CXFS is designed to detect and automatically recover from single-system failure, failure of the active metadata server, failure of SAN or TCP/IP network components, and failure of the network or SAN partition.

In a CXFS configuration, one system is the metadata server, and all other IRIX systems may be designated as backup metadata servers. A backup metadata server becomes the metadata server on detection of the failure of the initial metadata server; this failover is designed to take place in seconds, ensuring continual access to data regardless of the state of any single system.

Integrated with the most robust storage resource management available

CXFS is configured, monitored, and managed using an intuitive user interface based on Java™. CXFS can be managed from any Web browser compliant with Java, for the utmost in flexibility. A command-line interface is also available with functionality that can easily be incorporated in automated scripts.

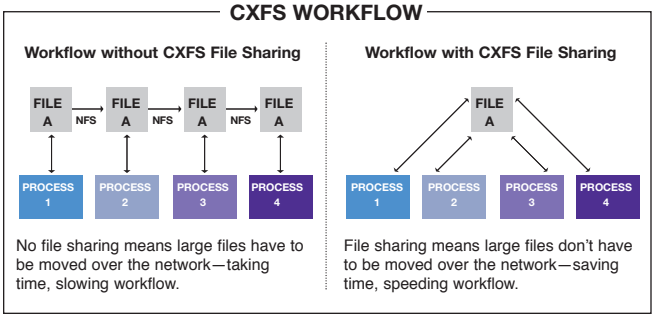
True LAN-Free backup and restore

LAN-based backups are becoming increasingly time-consuming and often put an excessive burden on the LAN. CXFS allows existing backup applications to be easily migrated from the LAN to the SAN. A SAN-based backup server simply mounts the CXFS filesystem(s) and backs up data as if it were local. All data travels across the SAN, the LAN is free for user traffic, and no new software is required.

Complete SAN / NAS Gateway

Systems not participating in the CXFS membership still have access to all data. Nonparticipating systems simply remain in their current configurations and access data in the CXFS volumes through a member system via traditional data sharing methods such as NFS, FTP, and CIFS.

CXFS WORKFLOW



SGI InfiniteStorage Shared Filesystem CXFS

Technology
- Journaled, shared, 64-bit with guaranteed filesystem consistency

Supported Platforms
CXFS is currently shipping on IRIX® 6.5 and is supported on Silicon Graphics Systems Silicon Graphics® Octane®, Silicon Graphics® Octane2™, Silicon Graphics Fuel®, Silicon Graphics® Tezro™, SGI® Origin® 200, SGI® Origin® 2000, Silicon Graphics® Onyx2®, SGI® Origin® 300, SGI® Origin® 350, SGI® Origin® 3000, SGI® Onyx® 300, and SGI® Onyx® 3000 series systems
- CXFS supports the SGI Altix and Prism families of servers and superclusters running 64-bit Linux with SGI ProPack™
- Solaris is currently supported with SunSM Solaris™ 9 and Solaris™ 10 on most SunSM SPARC® hardware platforms
- CXFS supports 32-bit Microsoft® Windows® 2000, Windows® XP and Windows® 2003 on any standard Intel® Pentium® III (minimum) or compatible PC
- CXFS supports 32-bit Linux® SUSE® SLES 9 and Red Hat® EL 4 on any standard Intel Pentium III (minimum) or compatible PC
- CXFS supports 64-bit Linux® SUSE® SLES 9 and Red Hat® EL 4 on any standard Intel XEON EM64T, Itanium 2 or AMD Opteron PC
- CXFS supports Apple® Mac OS® 10.4 on any Power based Mac® and XServe® hardware
- CXFS supports IBM® AIX® 5.2 and 5.3 on any IBM® server or SP cluster based on POWER2™, POWER3™, or POWER4™ architecture
- CXFS works with all storage devices and SAN environments supported by SGI, including 1, 2, and 4Gb, multiswitch fabrics and SGI 9700 with native InfiniBand support

Scalability
Maximum File Size
- 9 million terabytes (or the system drive limits)

Maximum Filesystem Size
- 18 million terabytes (or the system drive limits)

Filesystem Block Size
- Selectable at filesystem creation time
- 512 bytes to 64KB for normal data and up to 1MB for real-time data
- Filesystem extents (contiguous data) are configurable at file creation time using fcntl and are multiples of the filesystem block size; single extents can be up to 4GB in size

Partitioning
- Up to 64,000 partitions, 64,000 wide stripes and dynamic configurations

Physical Disk Sector Size Supported
- 512 bytes

Backup/Restore
- Dump/restore, bru, cpio, tar; Legato NetWorker®, Atempo, and many popular commercial backup packages
- Filesystem freeze and thaw allow RAID devices to do consistent snapshots of the filesystem by flushing all buffers to disk, freezing the filesystems, and then giving access back to the applications
- Restore can restore EFS dumps to EFS, XFS, or CXFS filesystems
- xfs_restore can restore XFS dumps to EFS, XFS, or CXFS filesystems
- Dumps of active CXFS filesystems are supported

Support for Data Lifecycle Management policy automation
- The Data Management API (DMIG-DMAPI) allows implementation of data lifecycle management policy automation software (such as SGI® InfiniteStorage Data Migration Facility - DMF) without any kernel modifications as well as high-performance dump programs without requiring raw access to the disk and knowledge of filesystem structures

Corporate Office
1200 Crittenden Lane
Mountain View, CA 94043
(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

2816 [06.15.2006]

J15165