Open source XFS™ for Linux™

Providing the World's Most Scalable Journaling-Filesystem Technology to the Open Source™ Community

Open Sourcing Key Technologies

SGI is contributing the XFS filesystem to the Open Source community. XFS is one of SGI's core competencies in high-performance computing. It is the most scalable, high-performance journaling filesystem available today.

SGI's Open Source Intentions

SGI will work with the Open Source and Linux communities in the context of Open Source cooperation. SGI is partnering with universities and government laboratories to provide enhancements to its Open Source technologies.

SGI plans to help scale Linux to run enterprise-class applications by delivering technologies from its core competencies in visualization and high-performance computing. XFS is one of several of these key Open Source contributions. SGI plans to help Linux mature into a world-class, robust, reliable, feature-rich, widely used, and enterprise-ready operating system.

XFS: A World-Class Filesystem

XFS combines advanced journaling technology with full 64-bit addressing and scalable structures and algorithms. This combination delivers the most scalable and high-performance filesystem in the world.

Journaling: Quick Recovery

The XFS journaling technology allows it to restart in less than a second after an unexpected interruption, regardless of the number of files it is managing. Traditional filesystems must do special filesystem checks after an interruption, which can take many hours to complete. The XFS journaling avoids these lengthy filesystem checks.

Fast Transactions

XFS journaling also speeds the read and write data transactions. Its journaling structures and algorithms are tuned to log the transactions more rapidly with general UNIX® filesystem technology.

XFS uses efficient table structures for fast searches and rapid space allocation. XFS continues to deliver rapid response times, even for directories with tens of thousands of entries. The performance of most UNIX filesystems significantly degrades as the number of entries per directory grows—not so with XFS.

21st Century Scalability

XFS is a full 64-bit filesystem, capable of handling files as large as a million terabytes.

$$2^{63}-1 = 9 \times 10^{18} = 9 \text{ exabytes}$$

A million terabytes is about a million times larger than most large filesystems in use today. This may seem to be an extremely large address space, but it is actually needed to plan for the exponential disk-density improvements observed in the disk industry in recent years. Disk densities are now doubling every year, and this rate is expected to continue for years. Filesystems with hundreds or thousands of terabytes are therefore predicted for the next decade.

As the disk sizes grow, not only does the address space need to be sufficiently large, but the structures and algorithms need to scale. XFS is ready today with the technologies needed for this scalability.

XFS Bandwidth

XFS delivers near-raw I/O performance. This was demonstrated on the largest disk configuration SGI was able to assemble: over 4 gigabytes per second (read and write) on 704 disks to one multithreaded process on an Origin™ 2000 IRIX® system.

XFS, the most scalable and reliable file system, will soon be available to the Open Source community as one of many enhancements SGI plans for Linux.

Datasheet

Technical Specifications

Disclaimer:
These technical specifications are for XFS in the IRIX operating system from SGI. SGI is in the process of porting XFS from IRIX to Linux. Some features listed here might not be ported to the Linux version of XFS, due to the differences in the IRIX and Linux kernel features.

Technology
Journaled 64-bit filesystem for IRIX (and soon Linux) with guaranteed filesystem consistency.

Product Span
Available on all systems that run IRIX 5.3 or later. Soon to be available as an open source technology in Linux.

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 using mkfs_512.
512 bytes to 64 KB for normal data and up to 1 MB 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 4 GB in size.

Physical Disk Sector Size Supported
512 bytes

NFS™ Compatibility
With NFS version 3, 64-bit filesystems can be exported to other systems that support the NFS V3 protocol.
Systems that use NFS V2 protocol may Access XFS filesystems within the 32-bit limit imposed by the protocol.

Windows NT® Compatibility
SGI uses the Open Source Samba server to export XFS filesystems to Windows® and Windows NT systems.
Samba speaks the SMB (Server Message Block) and CIFS (Common Internet File System) protocols.

Backup/Restore
Dump/restore, bru, cpio, tar; IRIS NetWorker™ for IRIX and many popular commercial backup packages.

Dump Interchangeability
Restore can restore EFS dumps to either EFS or XFS filesystems.
xfs_restore can restore XFS dumps to either XFS or EFS filesystems.
Dumps of active XFS filesystems are supported.

Support for Hierarchical Storage
The Data Management API (DMIG-DMAPI) allows implementation of hierarchical storage management software with no kernel modifications as well as high-performance dump programs without requiring "raw" access to the disk and knowledge of filesystem structures.

Swap to Files
Swap to files is supported.

Guaranteed Rate I/O (GRIO)
Hard and soft guarantees are supported. Hard guarantees require turning off disk-drive self diagnostics. Guarantees are expressed as a file descriptor, data rate, duration, and start time. GRIO for more than four streams is optional.

Performance
Throughput in excess of 7 GB per second has been demonstrated on a single filesystem using a 32-processor Origin2000 server.
Single file reads and writes exceeded 4 GB per second.

Memory
32 MB recommended.

SGI corporate office:
1600 Amphitheater Parkway
Mountain View, CA 94403
(650) 960-1980
www.sgi.com