http://oss.sgi.com/projects/xfs

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 Linux community using the copyleft GPL license. 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, Government laboratories, and other companies to provide enhancements to its Open Source technologies, such as XFS.

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.

View some XFS source code, white papers, and design documents at **oss.sgi.com**.

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.

XFS Features

- Sub-second filesystem recovery after crashes or power failures (never wait for long 'fscks again)
- 64-bit scalability: millions of terabytes, millions of files, and a million files per directory (no more 2 GB limits)
- High reliability and performance from journaling and other advanced algorithms

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 grows¾ 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.

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.

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.

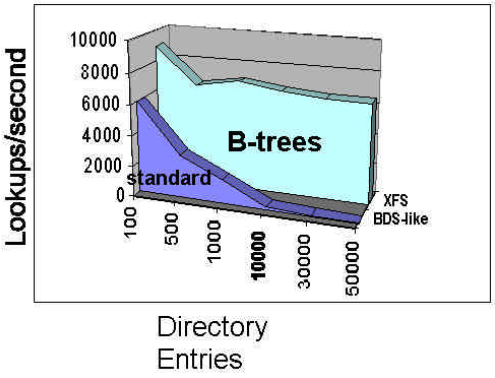
XFS features

- Scalability which will support disk growth for decades to come:
  - Scalable file sizes (9 million terabytes)
  - Scalable filesystems [18 million terabytes]
  - Scalable algorithms for high performance, even with huge file systems
    - Large numbers of files
    - Large files (including sparse files)
    - Large directories
    - Advanced algorithms for fast performance on huge filesystems
    - Rapid recovery
- High performance
  - Extremely fast transaction rates
  - Extremely high bandwidths
  - Extremely fast directory searches
  - Extremely fast space allocation
- Advanced features
  - Hierarchical storage [HSMs, DMF]
    - Off-line storage virtually on-line
- Compatibility
  - **Backup with popular commercial packages** such as Legato Networker® for IRIX and Veritas NetBackup™ or with dump/restore, bru, cpio, or tar.
  - **Support for multiple HSMs**, including SGI DMF™ and Veritas™ HSM through the DMIG-DMAPI interface
- **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.

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

Algorithms and internal structures which scale to support millions of files per filesystem and a million files per directory.

XFS uses efficient table structures [b-trees] 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 grows¾ not so with XFS.



Filesystem Block Size

512 bytes to 64 KB for normal data and up to 1 MB for real-time data.
Filesystem extents (contiguous data) are configurable up to 4 GB in size.
512 byte physical disk sectors supported.

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.

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

kentk (7/99)