



# Lustre™ File System

## High-performance storage architecture and scalable cluster file system



The Lustre™ file system redefines I/O performance and scalability standards for the world's largest and most complex computing environments. Ideally suited for data-intensive applications that require the highest possible I/O performance, Lustre is an object-based cluster file system that scales to tens of thousands of nodes and petabytes of storage with groundbreaking I/O and metadata throughput.



### Highlights

- **Unparalleled scalability:** Employs an object-based storage architecture that scales to tens of thousands of clients and petabytes of data—a file system virtually without limits
- **Reliable:** Deployed in production on many large and small clusters, meeting the uptime requirements of business and national security applications
- **Proven performance:** Delivers dramatic increases in throughput and I/O by intelligent serialization and separation of metadata operations from data manipulation
- **Cost effective:** Significantly reduces deployment and support costs through support for industry-standard platforms and heterogeneous networking environments
- **Open source, open standards:** Developed and maintained as open source software with an open networking protocol and POSIX file system semantics, ensuring broad support for industry-standard platforms and heterogeneous networking environments

### Unique storage architecture

The highly scalable and distributed Lustre file system combines open standards, the Linux operating system, an open networking API, and innovative protocols. Together, these elements create the world's largest “network-neutral” data storage and retrieval system.

Applying intelligence throughout its architecture, the Lustre file system turns commodity hardware into smart storage devices that manage data objects. The objects are dynamically distributed horizontally across the servers, shattering the performance limitations of traditional storage systems.

Building a Lustre cluster requires a Lustre MetaData Server and Lustre Object Storage Servers, each with disk storage. A pool of client systems can access these servers through one of many supported networks. The Lustre file operations bypass the MetaData Server completely and fully utilize the parallel data paths to all Object Storage Servers in the cluster. This unique approach of separating metadata operations from data operations results in significantly enhanced performance.

Like other UNIX® and Linux file systems, Lustre files are represented by inodes. However, a key difference in the Lustre file system is that its inodes simply contain references to the objects storing the file data.

### Open source

Lustre technology has been developed and maintained as open source software under the GNU General Public License (GPL), enabling broad support for industry-standard platforms.

### Heterogeneous networking

The Lustre network architecture provides flexible support for a wide variety of networks and high-performance features. The Lustre file system interoperates with network vendor-supplied libraries through Lustre Network Drivers, which utilize advanced features such as Remote Direct Memory Access, OS-bypass for parallel I/O, and vector I/O for efficient bulk data movement. Lustre Network Drivers exist for many networks, including TCP/IP, Quadrics Elan, many flavors of InfiniBand, and Myrinet GM—each with performance exceptionally close to the raw device throughput.

### Rugged high availability

The Lustre file system organizes all servers in active-active failover pairs. Together with protocol interoperability between versions, live cluster upgrades are now routine.

### POSIX compliance

The Lustre file system provides a tested and fully compliant file system interface in accordance with the POSIX standard.

### Simple configuration

With version 1.6, the Lustre file system revolutionizes configuration simplicity. Routine formatting and mounting of server devices aggregates them into a global high-availability cluster file system.

### Innovative protocols

The Lustre file system employs a distributed lock manager to handle access to files and directories and synchronize updates. This feature improves the metadata journaling approach used by most modern file systems.

### Extreme parallel computing

The Lustre lock manager automatically adapts its policies to minimize overhead for the current application. Files being used by a single node are covered by a single lock, eliminating additional lock overhead. Nodes sharing files get the largest possible locks, which still allows all nodes to write at full speed.

### Intent-based locking

To dramatically reduce bottlenecks and to increase overall data throughput, the Lustre file system uses an intent-based locking mechanism, where file and directory lock requests also provide information about the reason the lock is being requested. For example, if a directory lock is being requested to create a new unique file, the Lustre file system handles this as a single request. In other file systems, this action requires multiple network requests for lookup, creation, opening, and locking.

## Lustre File System Specifications

The Lustre file system version 1.6 is now available to the general public under the GNU GPL.

### Operating systems

- Red Hat Enterprise Linux 3 and later
- SUSE Linux Enterprise Server 9 and 10
- Linux 2.4 and 2.6
- Solaris™ Operating System (coming in 2008)

### Hardware platforms

- Intel Architecture 32-bit/64-bit
- x86/x64
- PowerPC

### Networking

- Quadrics Elan-3 and Elan-4
- TCP/IP
- InfiniBand
- Myrinet

### Proven performance

Lustre technology powers most of the world's largest Linux supercomputers and is the first production-tested, object-based Linux cluster file system.

### Recent results

- File I/O % of raw bandwidth: >90%
- Achieved single OSS I/O: >2.5 GB/s
- Achieved single client I/O: >2.0 GB/s
- Single GigE end-to-end throughput: 118 MB/s
- Achieved aggregate I/O: 130 GB/s
- Metadata transaction rate: 15,000 ops/s
- Maximum clients supported: 25,000
- Maximum file/file system size: 1.25 PB/>32 PB

### Learn More

For more information on the Lustre File System, visit [sun.com/lustre](http://sun.com/lustre) or [www.lustre.org](http://www.lustre.org)

### SunSpectrum support and services

Sun provides comprehensive software support for your Lustre application with the SunSpectrum Software Service Plan at the Standard level. Having access to Sun's expert technicians, in-depth knowledge base, software updates and releases helps ensure the success of your deployment and the availability of your solution.

For more information see: [sun.com/service/serviceplans/software](http://sun.com/service/serviceplans/software)

And, to further ensure the success of your project, take advantage of Lustre administration training. In a packed 3-day course, Sun Learning Services covers best practices in deployment and administration as well as configurations from very simple to very complex.

### Featured partners

- |                        |             |
|------------------------|-------------|
| • Bull                 | • Quadrics  |
| • Cray                 | • Scali     |
| • DataDirect Networks  | • Dell      |
| • HP                   | • Red Hat   |
| • Hitachi Data Systems | • LSI       |
| • Linux Networx        | • EMC       |
| • Novell               | • SGI       |
|                        | • SiCortex  |
|                        | • Terascale |