

Sun's Open Network System for HPC

Breakthrough application performance, density,
and energy efficiency

A broad range of organizations now depend on high-performance computing (HPC) technology. From the largest supercomputing clusters to more mainstream commercial and technical applications, effective HPC infrastructure is more essential than ever. Unfortunately, even as some strive to deploy greater capacity, they are often met with escalating server, storage, and switch sprawl. In addition to adding complexity, cost, and heat, these challenges can significantly limit the compute power that can be deployed to address an organization's greatest challenges and opportunities. Beyond faster processors and systems, a fundamental shift is required in the ways that HPC infrastructure is built.

Sun remains focused on bringing the wealth of technology advances happening in the marketplace to bear on organizations' most compelling challenges. This system includes:

- Blade server technology based in the Sun Blade 6048 Modular System chassis
- A revolutionary system for consolidating and aggregating InfiniBand fabrics
- The most advanced and open storage technologies for cluster caching, data retention, and archiving
- Sun HPC Software, providing pre-built and optimized HPC software in both Linux Edition and a Developer's Edition for the Solaris™ Operating System

Highlights

- Sun Blade™ X6275 server modules offer two compute nodes, each with two sockets for Intel® Xeon® Processor 5500 Series CPUs, and up to 96 GB of memory.
- The Sun Blade 6048 QDR InfiniBand Switched Network Express Module (NEM) connects up to 24 compute nodes to a QDR InfiniBand fabric.
- The Sun Lustre™ Storage System provides leading parallel file system I/O for HPC clusters, built around the Sun Fire™ X4540 and X4250 servers and Sun Storage J4400 arrays.
- Sun™ HPC Software helps Sun provide a complete hardware and software HPC solution that helps scale application performance and speed results.

Sun's Open Network System for HPC dramatically increases performance, density, and scale — with high-performance server modules and tightly-integrated QDR InfiniBand networking for the Sun Blade 6048 Modular System.

Scale, simplicity, and savings

Sun's Open Network Systems approach is focused on innovating around the convergence of open compute, storage, networking, and software technologies. This approach is designed to deliver the best performing, most efficient, and most scalable systems on the market to maximize the economics of computing. In the case of HPC infrastructure, the result is simplified cluster architecture that scales effectively from a single rack to the world's largest supercomputing clusters — all with the same carefully engineered components.

Dense and scalable modular systems

The massively scalable Sun Blade 6048 Modular System provides compute power, configuration, and I/O flexibility as well as high density and efficiency. A variety of server modules are supported, based on x86/x64 and SPARC® architectures. For HPC deployments, the Sun Blade X6275 server module provides two complete compute nodes on a single server module — each with two sockets for Intel® Xeon® Processor 5500 Series CPUs, and up to 96 GB of system memory. This configuration essentially doubles the density of some competitive offerings.

Dense and innovative QDR InfiniBand connectivity

Providing dense connectivity to servers while minimizing the number of cables is one of the issues facing large HPC cluster deployments. Sun offers solutions for both dual data rate (DDR) and quad data rate (QDR) InfiniBand fabrics that apply significant innovation in this area. For QDR InfiniBand interconnects, the Sun Blade 6048 InfiniBand QDR Switched NEM offers the ability to connect up to 24 nodes in a single Sun Blade 6048 shelf (up to four per chassis). Each NEM provides all of the connections necessary from the individual server modules to two 36-port InfiniBand switch chips. Mellanox InfiniScale IV 36-port switches offer very low 100ns latency, and QDR speeds across all ports. Sun Blade 6048 InfiniBand QDR Switched NEMs can be connected together directly to provide mesh or torus fabrics.

QDR Clos fabrics can be created by connecting Sun Project M9 QDR InfiniBand switches, supporting fabrics of up to 5,184 nodes.

DDR InfiniBand Clos fabrics can also be built using the Sun Datacenter Switch 3x24 and 3456 coupled with Sun Blade 6048 InfiniBand Switched NEMs. In contrast to competitive offerings Sun can reduce the overall switch infrastructure within the datacenter, reducing the requirement for cabling through industry-standard 12X connectors and 12x cables that aggregate three 4x InfiniBand connections (Table 1). QDR InfiniBand HCAs integrated onto Sun Blade X6275 server modules provide additional savings.

Cost-effective network attached storage

For fast and cost-effective storage, the Sun Lustre Storage System provides access to the latest technology and open software — powered by an integrated solution of Sun hardware and Lustre software.

Table 1: InfiniBand connector, switch, and cable consolidation

Sun InfiniBand Innovations	Benefits
Industry-standard 12x connectors	Increased switch density
12x InfiniBand Cables	3:1 reduction of 4x InfiniBand cables
Fully integrated switched NEMs	Switch consolidation and reliability

Optimized Metadata Servers (MDS) as well as highly-available Object Storage Servers (OSS) allow Sun to provide scalable and proven architecture for accelerated network storage I/O. Two fundamental design goals of the Sun Lustre Storage System are simplified configuration and implementation, and maintaining a price/performance lead over competitive designs. These goals are primarily achieved by standardizing the hardware architecture and choosing leading hardware components. In addition, a modular approach is used to provide scalability. Organizations can start with the number of OSS modules required for their cluster application and can easily grow capacity and throughput by adding additional OSS modules.

Robust technology choices

For HPC infrastructure, Sun's Open Network Systems provide key technologies that help to fully utilize the resources of these powerful systems, address the most challenging opportunities, and produce the most compelling results. For example, larger clusters can employ the Sun Lustre Storage System while smaller clusters can use Sun Storage 7000 Unified Storage Systems. These unified storage systems provide support for Hybrid Storage Pools that integrate memory, solid state drives (SSDs), and conventional hard disk drives (HDDs) through the scalable 128-bit Solaris ZFS™ file system.

Compelling results

Mechanical Computer Aided Engineering (MCAE) applications represent an important class of HPC applications for manufacturers and other scientific and engineering organizations. Using the InfiniBand technology similar to that in the Sun Blade X6275 server module, Sun has measured impressive results. Using the integrated InfiniBand infrastructure in the Sun Blade 6048 Modular System, Sun projects improvements of up to 3.2 times the performance utilizing 50% of the datacenter space and achieving 3.2 times the performance per watt — relative to that of competing systems¹ using Gigabit Ethernet technology.

For more information, please see world-record results published at <http://www.sun.com/ons/performance>. With these results, Sun's Open Network System for HPC provides:

- Scale — innovative systems optimized for increased throughput and performance
- Simplicity — reducing management overhead of deployment and system management
- Savings — lower acquisition cost and Eco efficiency on challenging HPC workloads

1. Based on in-house testing using Radioss Neon_1M benchmark. Performance results were measured on Sun Fire X2270 server using Infiniband and were approximated for one-node of Sun Blade X6275 server module, based on architectural similarities. Performance results for HP BL460c (2xIntel Xeon E5456 @ 3.0GHz) with Gigabit Ethernet, running Windows HPC Server 2008 RC2 as reported on <http://www.altairhyperworks.com/Benchmark.aspx#/>, as of 04/11/2009. Power consumption was obtained using power sizing tools available from HP.