San Diego Supercomputer Center

Sun Success Story.

Sun Helps SDSC Manage, Move, and Analyze Huge Data Sets

A revolution is going on technical computing and how it will be applied in the future, according to visionaries at the San Diego Supercomputer Center (SDSC). Where this industry used to be focused primarily on compute intensive applications, today, use of High Performance Computing (HPC) disciplines are growing exponentially and becoming much more data intensive. In addition to the growing reliance on data intensive applications, SDSC foresees that the ability to manage, move, and analyze large amounts of data will also continue to grow in importance. Even those applications that have traditionally been floating point intensive, such as computational fluid dynamics, involve more and more data in their calculations. Given this evolution, SDSC needed an architecture that could address both the scientific and data intensive applications, which is why they turned to Sun Microsystems.

SDSC is a national facility that acquires, deploys, and manages the highest performance computer systems, storage systems, and network environments for a diverse community of users. SDSC is one of the leading research and supercomputing facilities in the United States. Approximately half of its resources are devoted to the National Science Foundation (NSF), supporting academic researchers who need access to supercomputing resources. The other half of SDSC resources are focused on independent research funded by government grants. These independent projects fall into several categories:

- Bio-informatics/Computational Biology: data management, Internet services, online databases, and digital libraries
- Environmental systems: data analysis and data mining, enabling technologies
- Computer science research: development platforms for development of data management systems

"Over the last couple of years, we've witnessed a very rapid growth in the size of scientific data sets," says Phil Andrews, Program Director for High End Computing at SDSC. "In addition to growth in size, these data sets must also be network accessible. As a result, there is a torrent of data now available. For SDSC, it has become crucial for us to develop an infrastructure and applications that can actually manage these large volumes of data and changing requirements."

Company

San Diego Supercomputer Center
www.sdsc.edu

Industry/Market

Scientific Research and Supercomputing Services

Products/Services

- Sun Fire™ 15K Server (Starcat)
- Sun Enterprise™ 10000 Server (Starfire‘)
- HPC SAN High Performance Storage Solution
- QFS High Performance Shared File System
- SamFS Advanced Storage Management
- Sun StorEdge™ T3 disk arrays
- Sun HPC ClusterTools‘ software
- Solaris™ 8 Operating Environment
- UltraSPARC® Processors

Key Business Solutions

- Sun's Data Center Readiness Services include testing, configuring, and designing a complete platform architecture
- Sun Professional Services design and implement iForce‘ High Performance Computing Storage Area Network (HPC SAN) environments with network connections
- SunSpectrum Platinumsm support agreements, coupled with skills assessments and training from Sun Educational Services sustain high-performance and low cost of ownership

Key Business Results

- Enterprise-level reliability (more redundancies) forms the cornerstone of rock-solid, central data repository
- Large system memory allows highest levels of I/O throughput
- System bandwidth maximizes scalability, increases speed of calculations, and allows the system to analyze/move/manage large volumes of data more efficiently
- Solaris Operating Environment moves data effortlessly across multiple operating environments with increased flexibility and simplified administration
- More IT infrastructure can be managed with less
- SAN technology scales from 10 terabytes to 250 terabytes and sustains data movement of 2.2 gigabytes per second

The Sun systems are some of the most stable and reliable hardware at SDSC.

"Sun is not only providing the platform architecture, but the vision and the road map for our server and storage infrastructure, thereby allowing SDSC to manage the exponential increases in data growth."

Phil Andrews,
Program Director for High End Computing, SDSC

SDSC and Sun Work Together to Develop Best of Breed Solutions for Data Intensive Environments

Always on the leading edge of technology deployment, the center was one of the first HPC customers to acquire Sun's original 64-processor high-end system, the Sun Enterprise™ 10000 server (popularly known as Starfire™) for traditional floating point computing. SDSC quickly realized its users were changing their methodologies and approach in order to manage larger data sets, as data became an integral part of their overall operations. "With the Starfire system, it struck us clearly that Sun's high-end servers were a natural super-computer for data intensive environments, and both our users and SDSC found new uses for the architecture that we had not originally anticipated," explains Andrews. "Another advantage is Sun's active partnership with SDSC and our exploration into the use of these systems for data intensive applications. In a sense, we were partnering with Sun to explore new applications and new uses of these environments, predominantly involving large volumes of data."

Given the strong alliance between the two technology leaders, when it came time to validate Sun's new high-end server, SDSC was the logical choice to validate the "alpha" system of Sun's next high-end generation, the Sun Fire™ 15K server. The Sun Fire 15K server deployed at SDSC administers and manages the data management aspects of a highly distributed infrastructure.

"We ran the alpha version of the Sun Fire 15K server for over ten months prior to its general availability and became quite fond of its abilities," adds Andrews. "Given our experiences with the alpha system, we've deployed the recently released Sun Fire 15K server as our central data management environment for the quarter petabyte of production data on rotating storage that will be housed at SDSC. It also will be the data management environment used for a national project we're affiliated with – a distributed terascale facility that involves large storage resources disseminated across the nation," continues Andrews.

Sun Fire 15K Server Offers Increased Scalability, Half a Terabyte Memory and More

"Sun was the only player capable of delivering an entire solution stack that connected the end user through the net to the data storage environment and computational platform," says Andrews. Working in conjunction with Sun Support Services, Sun Professional Services

helped to ensure that the platform infrastructure met the quality-of-service requirements such as availability, security, scalability, and reliability. “Sun is not only providing the platform architecture, but the vision and the road map for our server and storage infrastructure, thereby allowing SDSC to manage the exponential increases in data growth.”

Several factors were considered by SDSC before choosing the Sun Fire 15K server as its data management server. One of the more important was that the very large-scale server is designed for the demands of a technical computing data center. For instance, the new design of the system backplane allows concurrent applications to be run on the same processor. In addition, the Sun Fire 15K server can host up to 18 Dynamic System Domains (DSD) under a single Solaris™ Operating Environment. A key advantage is the DSDs are fully configurable while applications are running, thereby enabling more data to be analyzed without any disruption to availability or scalability. The server also offers over half a terabyte of memory, allowing users to cache data moving between the disk storage and applications to and from all systems. This very large shared memory dramatically increases the amount of data the user can keep on the machine in memory and not stored. The result of these new features is much higher I/O throughput, enabling faster calculations while running memory intensive applications concurrently. And the Solaris 8 Operating Environment scales linearly within a single machine, easily handling the huge data sets and compute-intensive calculations that SDSC requires.

Other system advantages include the easy administration and manageability of the Sun Fire 15K server. “One very strong attribute is that the Sun Fire 15K server is administered in exactly the same way as all our other Sun machines,” adds Andrews. “Since our system administrators do not need to learn a new operating environment or technology, we can simply leverage upon the investments we’ve already made in terms of skills assessments and training received from Sun Educational Services.” This means SDSC can manage more IT infrastructure with less management cost. In addition to highly evolved feature sets and performance, reliability, availability, and security were critical quality-of-service requirements as well. SDSC supports portals and data repositories 365x24x7 so availability was key. “We have to ensure that we do not lose data, and that the data we provide is error-free,” explains Andrews. “The Sun systems are some of the most stable and reliable hardware that we have at SDSC.”

Storage Area Networks Set Up SDSC for the Future

In addition to processing large amounts of data, SDSC also needed to manage its storage environment requiring a single point of contact and moreover to reduce their operational total cost of ownership (TCO). SDSC looked to Sun Professional Services for help in designing and implementing a storage area network (SAN) based on Sun’s iForce™ High Performance Computing Storage Area Network (HPC SAN) solution.

“We currently have over forty terabytes of storage data on the Sun HPC SAN environment,” notes Andrews. “We chose the HPC SAN environment because it gives us the ability to manage our storage data as a network versus a direct attached storage environment.

Get the details.

The Sun Fire 15K server's large memory, high number of processors, and excellent engineering, help SDSC analyze/move/manage large data sets effectively.

www.sun.com/servers/highend/Sunfire15K/

This results in enhanced productivity and lower costs." According to Andrews, an additional advantage of the SAN environment is that of flexibility, enabling SDSC to easily move storage resources from one computer system to another or from one dynamic domain to another seamlessly. "SDSC has well over 50 Sun servers that will each be accessing the SAN environment, and it's very easy for us to reassign storage or capacity to one host over another," adds Andrews. "We can also add new servers at very little administrative cost."

"The iForce HPC SAN solution is at the center of TeraGrid," notes Andrews. The Sun Fire 15K server will combine the iForce HPC SAN with its Storage Resource Broker software to manage metadata for over one petabyte of nationally distributed data. The iForce HPC SAN environment consists of approximately 250 terabytes of data, and is expected to operate at efficiency levels over 90 percent and sustain data movement at 2.2 gigabytes per second.

"We have a collaborative working relationship with Sun that has paid real dividends in helping to ensure that our systems sustain the high availability, reliability, and scalability demanded by our end users."

Phil Andrews,
Program Director for High End Computing, SDSC

The SAN infrastructure is increasingly becoming a critical element in SDSC's data center operations. "We're preparing to deploy a 250 terabyte SAN environment, and doing that with direct attached storage would not be feasible," says Andrews. "More importantly, the ability of a storage environment that is geographically distributed is critical in terms of managing storage in a grid environment and operating the IT infrastructure as a network." In particular, the iForce HPC SAN will serve as a gateway to grid-distributed data, providing support for databases, data management, and data mining for TeraGrid project. This is critical when scientific data is shared and accessed by researchers around the world.

Sustaining High-Performance

SDSC values its partnership with Sun and relies on Sun Educational Services for skills assessments, and training related to the Sun server and storage platform. "We value the insight that the training specialists from Sun Educational Services passed on to our IT staff," comments Andrews. "We are able to manage our systems more efficiently and at a lower cost because of the training we received from Sun Educational Services." SDSC has a SunSpectrum Platinum[sm] support agreement in place to help sustain high availability and reliability. Engineers from Sun work on-site alongside SDSC IT staff, to proactively identify any potential problems before they occur. "We have a highly collaborative working relationship with Sun that has paid real dividends in helping to ensure that our systems sustain the high availability, reliability, and scalability demanded by our end users." Indeed, SDSC is sustaining almost 100 percent availability with the help of Sun.

"I've told my colleagues that Sun is a great choice because they won't let you down. It is not any single solution feature that makes their technology stand out. Sun's systems have a good design balance – good memory, good network performance, Solaris Operating Environment, a huge number of applications for Solaris and performance," notes Andrews. "Sun provides the whole stack – products, service, support, and maintenance."

Sun Microsystems, Inc. 901 San Antonio Road, Palo Alto, CA 94303-4900 USA 1-650-960-1300 or 1-800-555-9sun www.sun.com

Africa (North, West and Central): +33-13-067-4680 • **Argentina:** +5411-4317-5600 • **Australia:** +61-2-9844-5000 • **Austria:** +43-1-60563-0 • **Belgium:** +32-2-704-8000 • **Brazil:** +55-11-5187-2100 • **Canada:** +905-477-6745 **Chile:** +56-2-3724500 • **Colombia:** +571-629-2323 • **Commonwealth of Independent States:** +7-502-935-8411 • **Czech Republic:** +420-2-3300-9311 • **Denmark:** +45 4556 5000 • **Egypt:** +202-570-9442 • **Estonia:** +372-6-308-900 **Finland:** +358-9-525-561 • **France:** +33-134-03-00-00 • **Germany:** +49-89-46008-0 • **Greece:** +30-1-618-8111 • **Hungary:** +36-1-489-8900 • **Iceland:** +354-563-3010 • **India: Bangalore:** +91-80-2298989/2295454 **New Delhi:** +91-11-6106000 **Mumbai :** +91-22-2018141 • **Ireland:** +353-1-8055-666 • **Israel:** +972-9-9710500 • **Italy:** +39-02-641511 • **Japan:** +81-3-5717-5000 • **Kazakhstan:** +7-3272-466774 • **Korea:** +822-2193-5114 **Latvia:** +371-750-3700 • **Lithuania:** +370-729-8468 • **Luxembourg:** +352-49 11 33 1 • **Malaysia:** +603-21161888 • **Mexico:** +52-5-258-6100 • **The Netherlands:** +00-31-33-45-15-000 • **New Zealand: Auckland:** +64-9-976-6800 **Wellington:** +64-4-462-0780 • **Norway:** +47 23 36 96 00 • **People's Republic of China: Beijing:** +86-10-6803-5588 **Chengdu:** +86-28-619-9333 **Guangzhou:** +86-20-8755-5900 **Shanghai:** +86-21-6466-1228 **Hong Kong:** +852-2202-6688 • **Poland:** +48-22-8747800 • **Portugal:** +351-21-4134000 • **Russia:** +7-502-935-8411 • **Singapore:** +65-438-1888 • **Slovak Republic:** +421-2-4342-94-85 • **South Africa:** +27 11 256-6300 **Spain:** +34-91-596-9900 • **Sweden:** +46-8-631-10-00 • **Switzerland: German:** 41-1-908-90-00 **French:** 41-22-999-0444 • **Taiwan:** +886-2-8732-9933 • **Thailand:** +662-344-6888 • **Turkey:** +90-212-335-22-00 **United Arab Emirates:** +9714-3366333 • **United Kingdom:** +44 0 1252 420000 • **United States:** +1-800-555-9SUN OR +1-650-960-1300 • **Venezuela:** +58-2-905-3800