

NERSC creates an ultra-efficient supercomputer

With water-cooled IBM System x iDataPlex servers

Overview

The need

NERSC required a powerful and cost-effective new supercomputing cluster to support ongoing research projects. Maximizing utilization of space, power and cooling capacity in the data center was a priority.

The solution

The new cluster comprises 400 IBM® System x® iDataPlex™ dx360 M3 compute nodes with IBM Rear Door Heat eXchanger water cooling and Voltaire® Grid Director™ 4700 QDR InfiniBand switches in a HyperScale configuration.

The Benefit

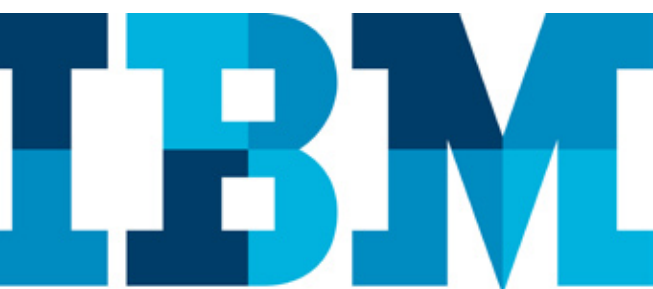
Energy savings make the new cluster a lower-cost, higher-performance option than continuing to run two old clusters. NERSC was able to run all three during the decommissioning phase.

The National Energy Research Scientific Computing Center (NERSC) is the flagship high-performance scientific computing facility for research sponsored by the U.S. Department of Energy Office of Science. NERSC, a national facility located at Lawrence Berkeley National Laboratory, is a world leader in providing resources and services that accelerate scientific discovery through computation.

NERSC needed to replace two existing supercomputing clusters. The organization wanted to create a scalable and modular cluster that would enable expansion at known price-points, and it planned to use the same architecture to support Magellan, an HPC-in-the-cloud project run jointly with the Argonne National Laboratories.

An important objective for NERSC was to enable an orderly switch-over to the new solution, which meant being able to start up the new cluster before the old clusters were decommissioned. This created extreme constraints on physical floorspace and cooling capacity, and NERSC wanted to demonstrate the highest possible density for the new solution. This would, in itself, increase the cooling challenge by packing more hot components into a smaller area.

Following an open procurement exercise, IBM was awarded the contract based on NERSC's best-value source selection process. By deploying IBM System x iDataPlex dx360 M3 servers with an innovative water-cooling solution from IBM and Vette Corp, NERSC was able to achieve its computational goals within the strict power and cooling constraints of its data center. The energy savings enabled by the new iDataPlex cluster make this new—and far more powerful—super-computer less costly than simply maintaining the two out-dated clusters it replaced.



“With IBM’s help, we were able to install the cluster on a five-foot pitch, leaving just 30 inches between the rows—this is the most dense iDataPlex install to date.”

—Brent Draney, group lead for networking and security, NERSC

“With IBM’s help, we were able to install the cluster on a five-foot pitch, leaving just 30 inches between the rows—this is the most dense iDataPlex install to date,” says Brent Draney, group lead for networking and security, NERSC. “What made this possible was the combination of water-cooled doors, our work with IBM engineering to validate the approach, and our work with Vette Corp to create a specialized cooling distribution unit (CDU) that uses return water from another system. By creating such a compact solution, we were able to have the new cluster running before we decommissioned the two old ones.”

High-performance interconnects

The new cluster, named Carver in honor of George Washington Carver, an American scientist, botanist, educator and inventor, comprises 400 iDataPlex servers. Each iDataPlex has two quad-core Intel Xeon X5600 processors running Linux, for a total of 3,200 cores, and the theoretical peak performance is 34.2 TFlop/s. The Voltaire Grid Director 4700 QDR InfiniBand switches are arranged in a HyperScale configuration, which enables high scalability with linear performance, and allowed NERSC to deploy a 1,200 node InfiniBand interconnect using only two racks. The switches provide 40 Gb/s of point-to-point bandwidth for high-performance message passing and access to the global file system.

“We specified our requirement for a high-speed, low-latency interconnect, and IBM included InfiniBand technology in its successful bid,” says Draney. “Voltaire was able to propose a compact, but expandable solution using four of its InfiniBand switches in a HyperScale configuration, which gave us the high port-count we needed for the Carver and Magellan clusters.”

NERSC also installed a second iDataPlex cluster, with 720 iDataPlex nodes, to support its joint Magellan project with the Argonne National Laboratory. These computational nodes are run with a variety of system software models from a traditional cluster setup to a fully virtualized cloud model. The 3000 NERSC users have access to Magellan, and NERSC captures and analyzes workload information as part of its cloud study. Magellan won the prestigious HPCwire Readers’ Choice Award 2010 for “best use of HPC in the cloud”. NERSC also has a strategic architecture policy of keeping shared storage separate from the computational facilities, to make it easy and non-disruptive to upgrade the latter. The organization’s global file system is built on IBM Global Parallel File System (GPFS™), which offers high performance and stability.

The Carver cluster powers a diverse range of important scientific research, from understanding molecular structures to validation of weather forecasting models.

Solution components:

Hardware

- IBM® System x® iDataPlex™ dx360 M3
- Voltaire® Grid Director™ 4700 QDR InfiniBand switches

Software

- IBM General Parallel File System (GPFS™)
- Red Hat® Enterprise Linux®

Business Partner

- Vette Corp
-

Innovative cooling

NERSC is using IBM Rear Door Heat eXchanger for its iDataPlex racks. The organization worked with Vette Corp, which licenses and supplies the doors, to create a customized solution that employs a multi-pass coil. The cold water runs first through the outside of the door to chill the air as it leaves the iDataPlex, then passes back through the inside of the door. Says Draney, “By the time the water is returned to the chiller, it’s actually warmer than the air leaving the back of the door. We control the CDU pumping units to maintain 72 degrees as the returning water temperature, and the higher the water temperature, the more efficient are our chillers.”

The sheer size of the rear doors—which measure four feet by seven feet—means that the air can move more slowly through them and has more time to exchange heat with the water. This in turn enables slower fans, dramatically reducing the energy wasted as noise, and making the Carver cluster the quietest place in the machine room.

The innovative, half-depth form-factor of the iDataPlex dx360 M3 reduces the airflow required across the components, lowering the power needed for cooling. High-efficiency power supplies, larger, better-optimized fans in the 2U chassis, and power management capabilities provide further efficiencies that minimize the dx360 M3’s power requirements.

Says Draney, “Rather than having the alternating hot/cold aisles that many conventional data centers have, we have a linear flow in which our racks are set up so that the exhaust air from one row is the input air for the next. The solution is so efficient that the air coming out of the back of each rack is always below the 75-degree required input temperature for the next iDataPlex rack. This makes the solution very simple and highly scalable: we can add rows of servers without needing to add more air handlers. The iDataPlex cluster is actually a net cooler of the room!”

The power usage effectiveness (PUE) rating of the data center was previously 1.35, meaning that for every watt consumed by the computing resources, a further 0.35 watts are required for cooling. For the iDataPlex cluster, the PUE figure is just 1.15.

“IBM iDataPlex has given us an efficient, reliable and cost-effective building block for supercomputing,” concludes Draney. “The high performance and low thermal envelope of the solution made it the perfect fit for NERSC.”

“The solution is so efficient that the air coming out of the back of each rack is always below the 75-degree required input temperature for the next iDataPlex rack.”

—Brent Draney, group lead for networking and security, NERSC

For more information

Contact your IBM sales representative or IBM Business Partner, or visit the following website(s): ibm.com/systems/info/x/idataplex

For more information about Voltaire visit:
voltaire.com/Products/InfiniBand/Grid_Director_Switches

For more information about Vette Corp visit: www.vettecorp.com

For more information about NERSC visit: nersc.gov



© Copyright IBM Corporation 2010

IBM Systems and Technology Group
Route 100
Somers, New York 10589
U.S.A.

Produced in the United States of America
November 2010
All Rights Reserved

IBM, the IBM logo, ibm.com, GPFS, iDataPlex and System x are trademarks of International Business Machines Corporation in the United States, other countries or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml.

Intel, the Intel logo, and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product and service names may be trademarks or service marks of others.

IBM and Vette Corp are separate companies and each is responsible for its own products. Neither IBM nor Vette Corp makes any warranties, express or implied, concerning the other’s products.

IBM and Voltaire are separate companies and each is responsible for its own products. Neither IBM nor Voltaire makes any warranties, express or implied, concerning the other’s products.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. Offerings are subject to change, extension or withdrawal without notice.

The information in this document is provided “as-is” without any warranty, either expressed or implied.



Please Recycle