NERSC NEWS
NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER
BERKELEY LAB
Office of Science
U.S. DEPARTMENT OF ENERGY
MAY 2007

Up to Speed

SCIENTISTS, VENDORS MEET TO DEFINE OBSTACLES, SOLUTIONS IN DESIGNING, DEPLOYING PETASCALE SYSTEMS

A DOE workshop hosted by NERSC brought together the international supercomputing community in May to identify challenges for deploying petascale systems, a collaboration that resulted in a series of recommendations.

The two-day meeting in San Francisco attracted about 70 participants from roughly 30 supercomputer centers, vendors, other research institutions and DOE program managers from Advanced Scientific Computing Research (ASCR), which funded the workshop. The discussions covered a wide-range of topics, including facility requirements, integration technologies, performance assessment and problem detection and management.

While the attendees were experienced in developing and managing supercomputers,

continued on page 2

Wanted: Killer Apps

NERSC IS POISED TO TAKE ON MORE SOFTWARE DEVELOPMENT PROJECTS

NERSC staff members are playing an increasingly active role in high-performance computing software integration and development, as they face a greater demand for marrying software from a variety of supercomputer makers with applications from a wide range of scientific fields.

That's the message from David Skinner, head of NERSC's Open Software and Programming Group, which carries out in-house software development and deployment projects to help researchers run their codes on NERSC systems. Skinner, who also is leading the new SciDAC Outreach Center, gave a talk on software development strategies recently as part of the lunchtime speakers series at NERSC.

Aside from the growing gap between what's available from vendors and what's needed for running a computer center, two other trends are contributing to a greater demand for in-house software development, Skinner said. One is the emergence of multidisciplinary computational science teams who layer software to build custom application frameworks. The other trend deals with the challenges presented by exploding concurrencies and data volumes.

Aside from NERSC staff, the audience also included representatives from high-performance computing vendors. Skinner emphasized the great need to "model good software practices for our own benefit and in order to help the HPC community achieve higher standards of software quality, both in terms of reliability and performance."

Skinner encouraged the audience to develop software that goes beyond fixing problems in a one-off way and toward contributing directly to software projects. He cited several software development assignments he and others have under-

continued on page 3

When Atoms Collide

RESEARCH CREATES BETTER MODELS FOR SIMULATING SUPERHEATED PLASMAS FOR FUSION ENERGY

Harnessing fusion power as an endless source of low-cost energy has been an intriguing challenge for scientists. Just ask George Vahala, who has relied on the computing power at NERSC to help other scientists understand the dynamics of generating fusion energy.

Vahala, a physics professor at the College of William and Mary, has been one of the most active users of NERSC resources this year. He has spent most of the 1.18 million computing hours on tackling tricky problems for modeling the flow of hot plasma inside a magnetic chamber as the atoms collide and fuse, a leading approach for generating megawatts of power in a few seconds.

continued on page 5

The streamlines in the image illustrate the velocity distribution of the particles as they move along the lattice links, demonstrating the effectiveness of the lattice-Boltzmann method to model fluid dynamics.

Open for Science

SCIENTISTS CAN NOW MANAGE THEIR RESEARCH AT NERSC AND ELSEWHERE VIA ONE ACCESS POINT

More NERSC users can now launch and manage their work at multiple computing sites by going through a centralized grid, thanks to efforts to connect NERSC's systems to the Open Science Grid (OSG).

The SGI Altix 350 visualization server, named DaVinci, joined the OSG last month, the latest in a long-range plan to connect all NERSC supercomputers to the grid, said Shreyas Cholia, a member of the Open Software and Programming Group at NERSC.

Making NERSC part of the more than 50 sites in the United States, South America and Asia that make up the OSG will save valuable time and reduce headaches for scientists who carry out

continued on page 4

Petascale

continued from page 1

they recognized that the leap to petascale computing will require more creative approaches.

"We are going through a learning curve. There is fundamental research to be done because of the change in technology and scale," said Bill Kramer, NERSC's General Manager who led the workshop.

In his welcoming remarks, Dan Hitchcock, Acting Director of the Facilities Division within ASCR, urged more collaboration, noting that various stakeholders in the high performance computing community have historically worked independently to solve thorny integration problems.

Mark Seager from Lawrence Livermore National Laboratory and Tom Bettge from the National Center for Atmospheric Research (NCAR) helped jumpstart the workshop by sharing their experiences with designing state-of-the-art computer rooms and deploying their most powerful systems.

Both Seager and Bettge said having enough power to supply massive supercomputers will become a bigger headache. A search for more space and reliable power supply led NCAR to Wyoming, where it will partner with the state of Wyoming and the University of Wyoming to build a $60-million computer center in Cheyenne.

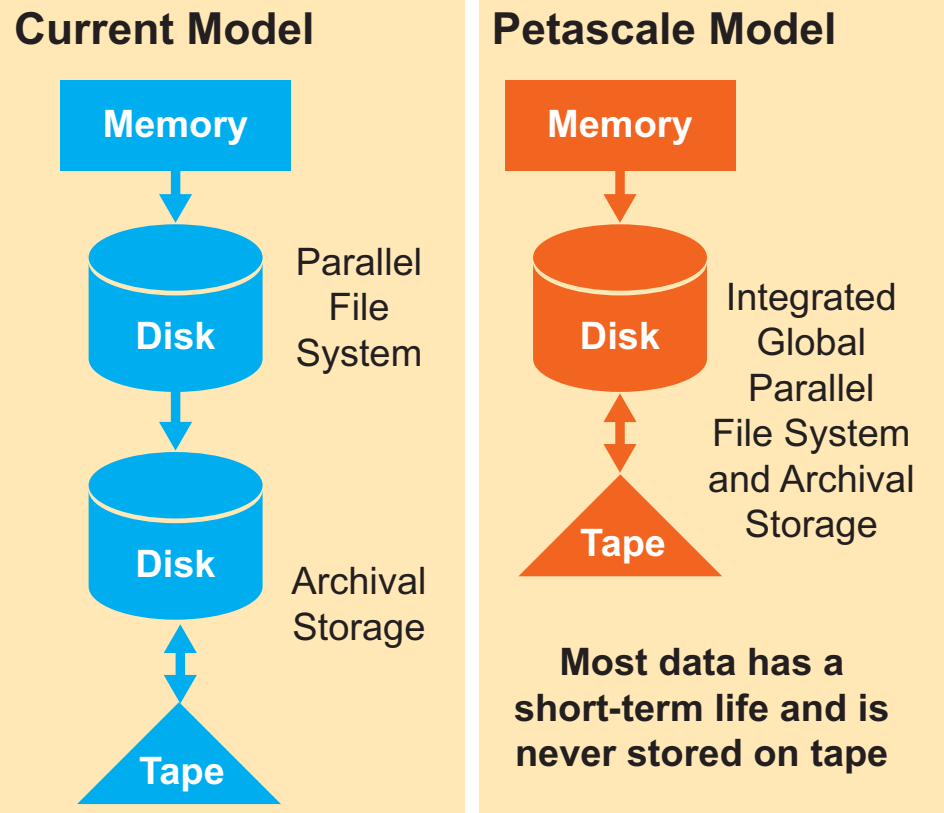
Seager also advocated the creation of a risk management plan to anticipate the worst-case scenario.

"I would argue that the mantra is 'maximizing your flexibility,'" Seager said. "Integration is all about making lemonade out of lemons. You need a highly specialized customer support organization, especially during integration."

Six breakout sessions took place over the two days to hone in on specific issues, such as the best methods for performance testing and the roles of vendors, supercomputer centers and users in ensuring the systems continue to run well after deployment.

The workshop program also included a panel of vendors offering their views on deployment challenges. The speakers, representing IBM, Sun Microsystems, Cray and Linux Networx, discussed constraints they face, such as balancing the need to invest heavily in research and development with the pressure to make profit.

A second panel of supercomputer center managers proffered their perspectives on major hurdles to overcome. For example, Patricia Kovatch from the San Diego Supercomputer Center hypothesized that the exponential growth in data could cost her center $100 million for petabyte tapes. Currently the center spends about $1 million a year on tapes.



The move to petascale computing will also drive changes in data flow, as Patricia Kovatch of the San Diego Supercomputer outlined at the recent workshop on deploying next-generation systems.

"We feel that computing will be driven by memory, not CPU. The tape cost is a bigger problem than even power," Kovatch told the audience.

Memory is clearly one of many challenges. Leaders from each breakout sessions presented slides detailing the daunting tasks ahead, including software development, acceptance testing and risk management.

At the end of the workshop, Kramer led the discussion to prioritize major challenges that emerged from the breakout sessions and enlisted the audience's input to define strategies for tackling those problems. A report detailing the discussions and recommendations will be issued.

The workshop had moments of levity. During dinner, attendees shared humorous tales about blunders they made or witnessed in their line work. The stories included a machine room design that left only a 3-inch headroom for computer cables; a cheapskate landlord who shut off power to the cooling system at night and weekends to save money; and a comment from a former U.S. president: "Oh Seymour Cray, I've heard of him. Doesn't he work for IBM?"

Learn more about the discussions and see the slides from the workshop at http://www.nersc.gov/projects/HPC-Integration.

WHAT IS NERSC NEWS?

NERSC News publishes every other month and highlights the cutting-edge research performed using the National Energy Research Scientific Computing Center, the flagship supercomputer facility for DOE's Office of Science. NERSC News editor Ucilia Wang can be reached at 510 945-2402 or Uwang@lbl.gov. Find previous NERSC News articles at http://www.nersc.gov/ news/nerscnews.

SPOTLIGHT

CONTENT MASTER

Harvey Wasserman

Harvey Wasserman from the Science-Driven System Architecture Team has been chosen as chair of the technical program for SC07, which will take place in Reno in November. He oversees the committee that reviews and selects papers, posters, workshops, panels, and tutorials. Wasserman also leads the effort in securing plenary speakers and in organizing the wide variety of awards presented at the conference, including the Gordon Bell Prize, Seymour Cray Award and Sidney Fernbach Award. Find out more about the technical program for this international conference on high-performance computing at http://sc07.supercomputing.org/?pg=techprogram.html.

COGNITIVE COMPUTING, BETTER SEARCHING

NERSC and CITRIS, the Center for Information Technology Research in the Interest of Society, co-hosted two meetings at UC Berkeley this month that focused on cognitive computing and online search technologies. NERSC Director Horst Simon gave the opening remark in both meetings, which attracted a total of 400 people from around the world.

The first meeting, "Cognitive Computing, a Multi-Disciplinary Synthesis of Neuroscience, Computer Science, Mathematics, Cognitive Neuroscience, and Information Theory" examined how scientists and engineers can build computer systems to become more flexible and adaptable, much like how the human brain and nervous system work.

The second meeting, "The Future of Search," brought together UC Berkeley scientists and industry leaders, such as Yahoo, Microsoft and Google, to set search goals for building next-generation search engines. During the meeting, attendees discussed how to reach the ultimate goal of creating search engines with cognitive and natural language capabilities.

More information about the cognitive computing meeting can be found at http://www.citris-uc.org/CognitiveComputing07. Learn more about the future of search at http://www.citris-uc.org/FutureSearch.

VISIT FROM STUTTGART

Michael Resch, the head of the High Performance Computing Center Stuttgart (HLRS) visited NERSC last month, a stop-over to discuss issues facing both supercomputer centers.

Resch spoke with NERSC staff about plans for upgrading systems and supporting scientists. They also chatted about operating energy-saving, large-scale supercomputers.

HLRS is part of the University of Stuttgart in Germany. Resch, who serves as the director at the university's Department of High Performance Computing, also discussed a plan by the president of the university, Wolfram Ressel, to visit NERSC in October.

"I've known NERSC and Horst Simon for a long time, and I was especially impressed that NERSC is able to keep a leading position in HPC over such a long time," Resch said.

continued on page 4

Killer Apps

continued from page 1

taken during his talk, such as the authentication program, MyProxy, that became part of the Open Science Grid's software distribution.

Other examples included Integrated Performance Monitoring (http://ipm-hpc.sf.net), which provides performance profiles on the execution of parallel jobs at computing centers run by DOE, the Department of Defense and the National Science Foundation.

"Some of the projects started out based solely on the needs that the center NERSC has, but those software solutions turned out to be useful to larger communities," Skinner said.

He presented a list of suggestions for staff interested in starting or joining a software development project. The list included the following tips:

- Scope the goals, needs and requirements.
- Scope the license and intellectual property constraints.
- Assemble a team and have frequent discussions.
- Establish a home for your code.
- Have someone outside the group of developers evaluate progress.
- Fan out. Find others who need similar functions and leverage their resources.
- Hand off. Feed your work back into someone else's development efforts if possible and prudent.

Crafting a smart project proposal is equally important. NERSC's web site contains a template that describes clearly the scope of the work, including costs, benefits and goals.

Toward the end of his presentation, Skinner invited fellow NERSC colleagues to participate in his group meetings and learn more about software development opportunities. His group is responsible for providing a host of services, including source code control, bug tracking and coding assistance.

A similar set of services also are offered to SciDAC researchers through the SciDAC Outreach Center, which serves as a clearinghouse for SciDAC activities and resources and a liaison within the high performance computing community.

For information about SciDAC outreach, check out http://outreach.scidac.gov.

Science Grid *continued from page 1*

their research at several computing facilities. Instead of dealing with different authentication processes and software at each site, the scientists can go through the OSG to manage their computing jobs and data files.

"Being part of OSG gives scientists a one-stop interface where they can run and access all of the computing jobs," Cholia said. Nearly four dozen NERSC users already have registered to use the OSG.

Although the PDSF system at NERSC has been part of the OSG for several years, it's primarily for projects in high energy physics and nuclear science. Making other NERSC supercomputers available over the OSG has been a priority over the past year. NERSC managers tapped Cholia to work on integrating OSG services and hired Jeff Porter to be part of the OSG team that validates and tests software tools. Bill Kramer, NERSC's General Manager, chairs the OSG Council.

Connecting NERSC systems to the OSG requires bridging different pieces of software and building a new infrastructure, particularly when some of the NERSC computer architectures are new to the OSG. Cholia is responsible for examining the OSG software stack and figuring out ways to support it, either through using solutions already available or by developing new ones.

Porter, on the other hand, works on making sure the software released by OSG several times a year can be deployed without glitches. He runs validation tests and solicits feedback from selected users who also get to run the new tools over a testbed, or a small grid.

The map shows the locations of more than 50 computing and storage sites that form the Open Science Grid.

In the next few months, Jacquard, the Opteron cluster at NERSC, will join the OSG. Adding Jacquard will bring parallel computing resources to the OSG and make such resources available for scientists who need to run MPI codes and larger parallel jobs. Cholia and Porter also are evaluating other NERSC systems such as Bassi and Franklin for the grid.

Having NERSC be part of the OSG helps the OSG to fulfill one of its goals—attracting a greater variety of users from different disciplines. NERSC hosts research from a wide-range of scientific communities, including life sciences, material sciences, climate research and chemistry.

NERSC also has set aside computing time specifically for projects carried out over the OSG, part of an effort to attract new research and users.

"We are keeping up with the cutting edge methods of how high-performance computing is done," Porter said.

Learn more about the Open Science Grid at http://www.opensciencegrid.org.

SPOTLIGHT *continued from page 2*

NEW TO THE TEAM

Several people recently joined NERSC as full-time staff members, serving as experts in areas such as storage, system infrastructure, network and security. Here is a short introduction of the new team members:

Michael Welcome

Michael Welcome, who is now a member of the Mass Storage Group, has over 23 years experience in the field of computing and computational science. For the past seven years, Welcome has worked in the Future Technology Group within the Computational Research Division of Berkeley Lab, doing both application performance analysis and low-level network and interconnect programming.

His recent projects include porting GASNet, the Berkeley UPC runtime system, to the Cray XT3/XT4, as well as porting and optimizing Miranda, a Lawrence Livermore National Laboratory (LLNL) hydrodynamics code, to the LLNL BlueGene system. Welcome also analyzed the performance of a CCSE adaptive mesh refinement code on a collection of high-end systems, including the Cray X1 at Oak Ridge National Laboratory, IBM Power

continued on page 6

Fusion Energy

continued from page 1

For years, Vahala has worked on improving the lattice-Boltzmann method, a relatively new approach to simulating fluid flow that is ideally parallelized and scales almost perfectly with the number of processors available. By embedding the fluid dynamics into a higher dimensional kinetic space, the solution trajectories are easier and more efficiently solved. The trick is to reduce the required memory and to ensure numerical stability at very low transport coefficients.

Over the years, Vahala and his team have made tremendous progress in solving one of the toughest problems: how to ensure the stability of a lattice-Boltzmann code when a high Reynolds number is used to better quantify turbulence? His work also has made it possible for the leap from two-dimensional to three-dimensional modeling using lattice-Boltzmann methods.

While the initial runs were performed on the latest Department of Defense machines (an IBM-P5), it was absolutely necessary to turn to the old work-horse, NERSC's Seaborg machine (an IBM-P3), to complete the simulation runs in 48-hours bursts on 4096 processors.

"Seaborg may be viewed as slow nowadays – but its many processors and large memory made it utterly indispensable for our research," Vahala said.

In many instances, the method has proven to work better than other computational fluid dynamic codes in modeling systems with complex physical geometries. In fact, researchers have used lattice-Boltzmann to study a wide range of phenomena, from star formation to the dynamics of cell proteins. In the commercial sector, oil companies have turned to the lattice-Boltzmann method for figuring out how to extract more black gold out of the field.

For the past seven years, lattice-Boltzmann has been used to characterize magnetohydrodynamics, the effects of a strong magnetic field on the turbulent flow of electrically conducting fluids such as plasmas. Moreover, the algorithm automatically permits the divergence of the magnetic field to be zero to machine accuracy. Understanding magnetohydrodynamics, in turn, will help researchers design better tokamaks, the machines that create the magnetic field.

In 2005, he co-authored a paper showing how well his three-dimensional lattice-Boltzmann code scaled on the Earth Simulator. The code achieved a sustained 26.25 teraflop/s, registering the best performance by a scientific code on that supercomputer. Jonathan Carter from NERSC was the paper's lead author while Lenny Oliker from Berkeley Lab's Computational Research Division was one of the co-authors. The research was published in the Proceedings of SC05.

Vahala's other recent publications include "Entropic Lattice Boltzmann Representation Required to Recover Navier-Stokes Flows", Physical Review, E75, 036712::1-11 (2007), "The Lattice Boltzmann Representataion for Plasma Physics" in Physica A362, 48-56 (2006) and "Quantum Lattice Representation for Vector Solutions in an External Potential" in Physica A362, 215-221 (2006).

Aside from Carter, Vahala's current research team includes Jeffrey Yepez (Hanscom Air Force Base), Linda Vahala (Old Dominican University) and Min Soe (Rogers State University).

Swiss Connection

NERSC General Manager Bill Kramer (left) led a tour of NERSC's computer room for visitors from the Swiss National Supercomputing Centre. The man from right is Richard Alexander (Senior System Engineer), Hussein Harake (System Engineer), Domink Ulmer (COO, Head of Administration) and Ladina Gily (Event Manager, HR Manager, Head of Facility Management).

NERSC hosted four visitors from CSCS (Swiss National Supercomputing Centre) last Friday, May 4. After NERSC staff members gave talks about the center's supercomputing and storage capabilities, they heard from Dominik Ulmer, the COO and Head of Administration at CSCS. Ulmer spoke about the upgrades to its flagship system, a Cray X3 with 3,328 dual-core processors, and plans to build a new computing facility by 2010.

He also talked about a recent initiative called Swiss ALPS Programme, which allocates millions of computing hours to each project and aims to boost the CSCS's ability to run large-scale projects. Historically, a typical allocation at the center ranges from 40,000 to 200,000 computing hours. Many of the projects being carried out at the CSCS are in the field of molecular dynamics, although the center is striving to broaden the scientific scope in the next four years.

CSCS is an autonomous unit of the Federal Institute of Technology in Zurich (ETH Zurich), but the center is located in Mano, in the Italian speaking part of Switzerland.

"NERSC is always impressive," said Ulmer, who previously had visited NERSC twice. "And it's always a good opportunity to exchange ideas and learn. We have many things in common."

SPOTLIGHT *continued from page 4*

5 at NERSC and the NASA Columbia system.

Previously, Welcome had worked in the NERSC Computational Systems Group, managing the Cray T3E and the J90 cluster as well as writing and maintaining the client-side accounting and account management software. Prior to this, Welcome worked at LLNL from 1983-1996 developing adaptive mesh refinement codes for both compressible and incompressible fluid dynamics.

Jason Lee

Jason Lee, who just joined the Networking, Security, Servers and Workstations Group, holds a master's degree in computer science from San Francisco State University. Lee has worked at Berkeley Lab for the last 15 years, working on projects that range from developing a parallel storage system to grid architecture in the Global Grid Forum.

More recently, he worked on Bro, a network intrusion detection system developed by the Lab and the International Computer Science Institute. Lee was involved in the early quality-of-service (QoS) network tests with Van Jacobson back in 1994, when Jacobson was the head of the Network Research Group at the Lab and was known for his work on improving IP network performance and scaling.

Lee worked on carrying out some of the first gigabit test beds, including the Multidimensional Applications and Gigabit Internetwork Consortium (MAGIC) project in 1993, Bay Area Gigabit Network (BAGNet) in 1995 and National Transparent Optical Network (NTON) in 1997. He also worked with the Advanced Internet Research group at the University of Amsterdam on 10-Gbps transatlantic links in 2002.

Tina Declerck

Tina Declerck recently retruned to the NERSC's Computational Systems Group after working there from 1997 through 2001. In her previous tenure at NERSC she managed the PDSF and Cray J90 SV1 clusters. During her hiatus, she worked at Sistina Software, the company that developed the Global File System (GFS) and was later bought by Red Hat. At Sistina, Declerck was in charge of growing and managing the customer support team.

At 3Ware, which manufactures SATA RAID controllers, she defined new product features, acted as the primary interface between the field support and engineering teams, provided technical training to resellers and worked on I/O performance and tuning.

Now back at NERSC, Declerck is the system lead for the SGI Altix named DaVinci, and she provides secondary support on Franklin, the Cray XT4, that was recently installed.

Zhengji Zhao

Zhengji Zhao joined the NERSC's User Services Group after working for three years as a postdoc in the Computational Research Division's Scientific Computing Group (SCG) at Berkeley Lab. In SCG, she worked on developing new methodologies for computational nanoscience: the linear scaling 3D fragment (LS3DF) method for large-scale electronic structure calculations and other methods.

Zhao received her Ph.D. in computational physics from New York University for developing the reduced density matrix (RDM) method, a highly accurate alternative to wavefunction-based computational chemistry methods. She received a master's degree in computer science from the Courant Institute of Mathematical Sciences, New York University; a master's degree in theoretical physics from Peking University in Beijing, China; and a bachelor's degree in theoretical physics from Jilin University in Changchun, China.

Akbar Mokhtarani

Akbar Mokhtarani came to NERSC from Berkeley Lab's High Energy and Nuclear Physics Computing Group, where he supported the BaBar detector at the Stanford Linear Accelerator Center (SLAC) and the IceCube and Majorana experiments. His responsibilities included migrating a database to new software, working on a Java/C++ monitoring system and building and packaging simulation software.

Mokhtarani, who is in the Mass Storage Group, is now applying his database and software development skills to the SciDAC Petascale Data Storage Institute (PDSI), for which he is gathering data and reporting on storage and file system reliability, and working on I/O benchmarking of selected petascale applications. He also works closely with NERSC's Science-Driven System Architecture Group and the team in charge of the Global File System.