DOE to explore scientific cloud computing at Argonne, Lawrence Berkeley National Laboratories

October 14 2009

Cloud computing is gaining traction in the commercial world, but can such an approach also meet the computing and data storage demands of the nation's scientific community? A new program funded by the American Recovery and Reinvestment Act through the U.S. Department of Energy will examine cloud computing as a cost-effective and energy-efficient computing paradigm for scientists to accelerate discoveries in a variety of disciplines, including analysis of scientific data sets in biology, climate change and physics.

Cloud computing refers to a flexible model for on-demand access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services, and software) that can be easily provisioned as needed. While shared resources are not new to high-end scientific computing, smaller computational problems are often run on departmental Linux clusters with software customized for the science application. Cloud computing centralizes the resources to gain efficiency of scale and permit scientists to scale up to solve larger science problems while still allowing the system software to be configured as needed for individual application requirements.

To test cloud computing for scientific capability, DOE centers at the Argonne Leadership Computing Facility (ALCF) in Illinois and the National Energy Research Scientific Computing Center (NERSC) in California will install similar mid-range computing hardware, but will

offer different computing environments. The combined set of systems will create a cloud testbed that scientists can use for their computations while also testing the effectiveness of cloud computing for their particular research problems. Since the project is exploratory, it's been named Magellan in honor of the Portuguese explorer who led the first effort to sail around the globe and for whom the "clouds of Magellan" - two small galaxies in the southern sky - were named.

One of the goals of the Magellan project is to explore whether cloud computing can help meet the overwhelming demand for scientific computing. Although computation is an increasingly important tool for scientific discovery, and DOE operates some of the world's most powerful supercomputers, not all research applications require such massive computing power. The number of scientists who would benefit from mid-range computing far exceeds the amount of available resources.

"As one of the world's leading providers of computing resources to advance science, the Department of Energy has a vested interest in exploring new options for meeting the overwhelming demand for computing time," said Michael Strayer, associate director of DOE's Office of Advanced Scientific Computing Research. "Both NERSC and ALCF have proven track records in deploying innovative new systems and providing essential support services to the scientists who use those systems, so we think the results of this project will be quite valuable as we chart future courses."

DOE is funding the project at $32 million, with the money divided equally between Argonne National Laboratory and Lawrence Berkeley National Laboratory, where NERSC is located.

"Cloud computing has the potential to accelerate discoveries and enhance collaborations in everything from optimizing energy storage to

analyzing data from climate research, while conserving energy and lowering operational costs," said Pete Beckman, director of Argonne's Leadership Computing Facility and project lead. "We know that the model works well for business applications, and we are working to make it equally effective for science."

At NERSC, the Magellan system will be used to measure a broad spectrum of the DOE science workload and analyze its suitability for a cloud model by making Magellan available to NERSC's 3,000 science users. NERSC staff will use performance-monitoring software to analyze what kinds of science applications are being run on the system and how well they perform on a cloud.

"Our goal is to get a global picture of Magellan's workload so we can determine how much of DOE's mid-range computing needs could and should run in a cloud environment and what hardware and software features are needed for science clouds," said NERSC Director Kathy Yelick. "NERSC's users will play a key role in this evaluation as they will bring a very broad scientific workload into the equation and help us learn which features are important to the scientific community."

Looking at a spectrum of DOE scientific applications, including protein structure analysis, power grid simulations, image processing for materials structure analysis and nanophotonics and nanoparticle analysis, the Magellan research team will deploy a large cloud test bed with thousands of Intel Nehalem CPU cores. The project will also explore commercial offerings from Amazon, Microsoft and Google.

In addition, Magellan will provide data storage resources that will be used to address the challenge of analyzing the massive amounts of data being produced by scientific instruments ranging from powerful telescopes photographing the universe to gene sequencers unraveling the genetic code of life. NERSC will make the Magellan storage available to

science communities using a set of servers and software called "Science Gateways," as well as experiment with Flash memory technology to provide fast random access storage for some of the more data-intensive problems.

The NERSC and ALCF facilities will be linked by a groundbreaking 100 gigabit-per-second network, developed by DOE's ESnet (another DOE initiative funded by the Recovery Act). Such high bandwidth will facilitate rapid transfer of data between geographically dispersed clouds and enable scientists to use available computing resources regardless of location.

"It is clear that cloud computing will have a leading role in future scientific discovery," added Beckman. "In the end, we will know which scientific application domains demonstrate the best performance and what software and processes are necessary for those applications to take advantage of cloud services."

Source: Lawrence Berkeley National Laboratory (news : web)