Building High Performance Clouds

Most enterprises are aware of how cloud computing helps to cut IT costs, deploy systems more rapidly, and increase agility. But what about scientific efforts? Can clouds – built on widely available platforms and commodity hardware – help researchers to do serious science?

The Magellan system, funded by the U.S. Department of Energy (DOE), is a system designed to run technical computing workloads. The system began life as a part of an evaluation effort, to determine if cloud computing was useful for technical computing workloads. This evaluation proved successful, in a variety of interesting ways. Today, the system is being used to push the limits of technical cloud computing.

The initial phase, the Magellan research project, was an ARRA funded joint project between the Argonne National Laboratory and Lawrence Berkeley National Laboratory, where the National Energy Research Scientific Computing Center (NERSC) is located. At Argonne, a large scale system was constructed to be operated as a private cloud, to assess the usefulness of this approach to scientists. At NERSC, a smaller system was built to assess a variety of scientific workloads and analyze their suitability to a cloud environment. The combined Magellan resources were made available to about 3,000 scientists and researchers.

"Our goal at Argonne was to assess the maturity of the system software stack, its applicability to scientific applications, and the effects that the cloud model had on scientific users," says Narayan Desai, technical lead of the efforts at Argonne. "We expected to see a performance impact because of virtualization, however it was considerably smaller than we expected for many scientific apps. The first major pain point that we felt was the quality and scalability of cloud software stack."

The Argonne team had begun the project running an early open source implementation of the EC2 APIs, but quickly ran into scalability and stability problems. "Things were fine at small scale, but once we hit about 100 nodes the system had substantial issues. Considering our goal scale was about 700 nodes, this approach just wasn't going to work."

They had to decide their next move. "We looked at a number of other cloud stacks, and that's when we took a serious look at OpenStack," says Desai. OpenStack is a large-scale open source cloud computing initiative, founded to drive community established industry standards, end cloud lock-in, and accelerate the adoption of cloud technologies by service providers and enterprises. As a cloud operating system, OpenStack automatically manages pools of compute, storage, and networking resources at scale and is supported by a vibrant ecosystem of technology providers.

OpenStack projects are built through a global collaboration of developers and cloud computing technologists who are producing the open standard cloud operating system for both public and private clouds. Cloud service providers, enterprises, and government organizations around the world are taking advantage of the freely available, Apache-licensed software to build massively scalable cloud environments.

"We found OpenStack very straightforward to build and deploy," Desai says. Today, the Magellan cloud consists of about 750 nodes, including 500 compute nodes, 200 storage nodes, a number of big memory (1 terabyte) nodes and 12 management nodes. "OpenStack worked and scaled much better than our previous platform," he says.

Building a Flexible Environment for Computational Science

The team at Argonne also found that the OpenStack cloud proved ideal at creating an environment for prototyping, software development, and testing for large-scale scientific applications that have not been tailored to the traditional HPC environment.

"We have a lot of users who are doing large-scale computational biology on this cloud. The flexibility that OpenStack has given us has made a class of users – those who do a lot more prototyping and development on a regular basis – extremely productive," Desai says. "This was one of the most surprising findings during the evaluation project: scientific users really benefit from direct access to computational resources, with the flexibility to design a full software environment for their applications. Moreover, this benefit vastly outweighs the performance penalties for many application types, particularly in loosely coupled applications. Both of these conclusions were unexpected."

Magellan Going Forward

As a result of this project, Argonne decided to continue running the system as a cloud platform after the completion of the evaluation project. "This evaluation has been a big success," Desai says. And the lab is continually training more researchers on how to utilize the OpenStack cloud. "There are a lot of researchers who have gotten lots of science done on the OpenStack cloud, and we're going to keep learning how we can push this cloud further," he says.

For now, the focus on the system is twofold. The last year was spent transitioning the system from a testbed into a production grade system. The other major focus is closing the performance gap between traditional HPC platforms and the Openstack platform. "While we expect that virtualization will never be free, the costs are getting low in a lot of areas; we need to gain a better understanding of the performance tradeoffs, as well as techniques for tuning performance in virtual machines."

To this end, Desai's team has been working on building a performance-optimized cloud. Network and storage performance are often cited as key challenges for cloud systems. Scientific workloads often depend on movement of large data sets from site to site for analysis and visualization, so the bottlenecks here would have a substantial impact on daily use of the system.

Their first priority was to assess network performance off system. Using a development deployment of OpenStack, they demonstrated near saturation of a wide-area 100 gigabit Ethernet link. "We were able to demonstrate 99 gigabits of traffic flowing from 10 VM instances at Argonne to LBL across ESNet, the DOE research network," says Desai, "We had expected to need many more instances running across 20-30 nodes, but the fact that our network interfaces were the limiting factor was excellent; that demonstrates the low overhead virtualization can have as well as leaves room for improved node performance. Even more important, all of our network tuning could be accommodated without any modifications to OpenStack."

The team has also been working on improving storage performance. The have built a custom storage solution for ISCSI storage that delivers more than 2GB/s per server. "In aggregate, our storage servers should be able to provide enough bandwidth (12.5 GB/s) to stream data across the 100 gigabit Ethernet link at line rate. Soon, it will be feasible for researchers to dynamically provision cloud resources to move data cross country at tens of gigabytes a second."

Major Applications and Future Plans

Magellan is driven by its applications; its future architecture will be driven by their needs of their workloads. Many early users are bioinformatics applications. For example, one major user of the system is the DOE Systems Biology Knowledge Base, a collaborative effort to build predictive models of microbes, microbial communities, plants, and their interactions. The KBASE project uses Magellan to build data intensive computational methods and services. Another project using the system is the MG-RAST metagenomic annotation system, which assesses microbial communities' composition and metabolic function. Desai's team plans to generalize the system to other application domains over the next year, including cosmology and materials science. Moving forward, the team will continue to work on network and storage performance to power additional compute-heavy and data intensive DOE applications that are advancing scientific understanding in a variety of disciplines.

Argonne National Laboratory

http://www.anl.gov/
Industry: Academic / Research
Headquarters: Chicago
Size: 5,001 to 10,000 employees

OpenStack technologies Argonne National Laboratory, US Department of Energy uses:

#0
#0
#0
#0