Minutes
MAGIC Meeting
March 10, 2010, 2:00-4:00
NSF, Room II-585

Action Items
1. Argonne will provide a presentation to MAGIC on the Argonne implementation of Magellan at a future MAGIC meeting, possibly April.

Proceedings

This meeting of MAGIC was chaired by Susan Turnbull of DOE and Jennifer Schopf of the NSF.

Magellan at NERSC: Presented by Jeff Broughton, NERSC

Midrange computing and data management are important capabilities in supporting large-scale science applications. Computational resources are limited but the number of ways for providing compute power and data management is increasing. Midrange computing is particularly appropriate for:

- Serial or scalability-challenged codes
- Science not requiring tight coupling
- Science that can run at low concurrency
- Data-intensive science
- Staging to the large centers

Advantages to using cloud computing include:

- Cycles on-demand: bypass allocations process
- Overflow capacity to supplement existing systems
- Customized and controlled environments
- Parallel programming models for data-intensive science
- Creating science communities around data sets

Magellan serves to answer questions for DOE regarding mid-range computing and cloud environments:

- How much of DOE's workload can be served economically by commercial or private-to-DOE cloud?
- What HW and SW features are needed?
- Do emerging cloud computing models provide viable approaches to computational science?
- Can clouds at different DOE facilities be federated?

Testing at NERSC has identified the slow-down of processing due to cloud computing for several types of science applications. Slowdown ranges from about 1.2 to 17 depending on the type of codes used. Performance was investigate3d for using 32, 64, and 128 processors. For Hadoop and Task Farming. Conclusions of this testing indicated that:

- Only loosely coupled applications are appropriate for cloud computing

- There are practical limits to the size of a cluster
- There are non-uniform execution times
- Costs for data storage and I/O are substantial

The NERSC Magellan Cloud for science applications includes 720 nodes, 5760 cores in 9 Scalable Units (SUs) providing 61.9 Teraflops. The NERSC Magellan infrastructure provides node aggregation into virtual clusters, provisioning of clusters for specific research projects, and dynamic provisioning of multiple software environments. It also provides InfiniBand QDR, a global file system, and bulk storage. The key is providing flexible and dynamic scheduling of resources.

Science communities are enabled around data sets. Custom hardware and software provide remote data and computing resources. The cloud computing enables remote working with large data sets.

Discussion identified that NERSC and Argonne intend to cooperate on joint cloud computing capabilities. Small businesses can apply for resources on Magellan. Magellan is seeking collaborators for testing capabilities. Funding is for two years.

One definition of cloud versus Grid computing states that: Clouds are about infrastructure, hosting, and capabilities; Grids are largely about federation activities.

For the full briefing please see:
http://www.nitrd.gov/Subcommittee/lsn/magic/index.aspx

For further information contact Jeff Broughton at: jbroughton@lbl.gov

AI: Argonne will provide a presentation to MAGIC on the Argonne implementation of Magellan at a future MAGIC meeting, possibly April.

Argonne on Magellan
Argonne implementation of Magellan is similar to the NERSC implementation. Argonne is using a different software model (they are not using MOAB for example). Argonne is exploring tools and software for science support. They are looking at MPI on the Argonne cloud. There is a strong emphasis on data-intensive cloud computing. Argonne is holding a "Welcome to Magellan" day soon, largely internal to Argonne but open to others as well. Argonne has held-back some funding to add active storage that will come on-line at the end of summer 2010. Argonne is working with SBIRs to investigate opportunities with small businesses in using Magellan.

Roundtable

CEDPS: Steve Tuelke
CEDPS held a conference last week with about 100 attendees bringing together many production grids. Slides are available at Globusworld.net

TeraGrid: Dan Katz

TeraGrid has a proposal with OSG at the NSF to look at technologies that can be used in both. They will do testing with independent user communities. They will develop cooperating infrastructures.

TeraGrid has 11 resource providers (11 universities). One university is sensitive to who their users are. Tracking may impute responsibility so other universities are hesitant to specifically identify users. The universities are beginning to discuss this issue. Users currently sign up for 11 policies when they join.

NASA: Hugh LaMaster

NASA is interested in cloud computing, particularly with its nebula project. They are also interested in the networking to support cloud computing.

OGF

OGF is holding a meeting in Munich next week. Attendance expected to be around a few hundred.

FCC: Mike Nelson

The FCC is issuing its Broadband Plan next week. Cloud computing will likely be mentioned. It will provide increased emphasis on 40-100 Gbps bandwidths.

Federated Identity: Don Stewart

Google is considering implementing OAUTH which would enable a Facebook manager to access your accounts on your behalf, providing fine-grained control. It is anticipated to affect Web 2.0 applications and use.

Next MAGIC Meetings

April7, 2:00-4:00, NSF
May 5, 2:00-4:00, NSF