

MAGIC Meeting
NOVEMBER 3, 2010, 2:00-4:00

Attendance:

Linda Akli	SURA
Rich Carlson	DOE
John Chambliss	IRS
Gary Crane	SURA
Christine Feldt	SRA
Shantenau Jha	LSU
Miron Livny	OSG
Paul Love	NCO
Mark Luker	NCO
David Martin	Argonne
Grant Miller	NCO
Larry Peterson	Princeton
Ruth Pordes	OSG
Mike Seablom	NASA
Alan Sill	OGF
Martin Swany	UDeI
Bill Turnbull	DOE
John Volmer	ANL
Wendy Wiggen	NCO
Dean Williams	ANL

Action Items

None

Proceedings

This meeting of MAGIC was coordinated by Rich Carlson of DOE and Grant Miller of the NCO.

Status of OGF: Alan Sill

Alan Sill, OGF Vice President of Standards gave a presentation on Creating a Standards Basis for Science Software Infrastructure. Science computation infrastructure has to provide ease of use to the user community. This requires attention to performance, cost efficiency and optimization of resources for science productivity. Making things easy to use may require software development and result in lower performance and efficiency. Optimizing cost efficiency by using commodity hardware and tools may shortchange science utility and potential. The overall goal is to provide science with tools with optimal performance, that are efficient to develop and deploy and that are easy to use. This goal is furthered by the development of standards.

OGF is an open community of large-scale grid and cloud providers and their user communities. Over the last decade OGF has developed standards for:

- compute-intensive
- data-intensive

- infrastructure-related
- job management

These standards are summarized at: <http://www.ogf.org/standards/>

Major accomplishments of OGF include fostering the development of Grid computing. OGF standards are used across a wide variety of boundaries, including internationally. The OGF standards are developed by the participants in cross-domain projects. The OGF standards are now being extended to cloud computing.

OGF standards have been developed for:

- Applications
- Architecture
- Computation
- Data
- Infrastructure
- Liaison
- Management
- Security

In cloud computing OGF has produced standards including:

- DFDL: Data Format Definition Language
- OCCI: Open Cloud Computing Interface
- CDMI: Cooperatively pursued with SNIA

OGF participated in the NIST workshop (Nov 2010) on Developing Grid and Cloud Standards on a Global Basis and in a workshop on Standards Acceleration to Jumpstart Adoption of Cloud Computing.

For the full details of the briefing, please see:

<http://www.hpcc.ttu.edu/asill/ogf/nitr/MAGIC-Dec2010-OGF-Standards-Intro-Web.pdf>

Cloud Computing and Virtual Infrastructures to Enable Research: Larry Peterson

Larry Peterson discussed his plans under an NSF MRI grant, to develop cloud computing and virtual infrastructures.

PlanetLab has been ongoing for several years. It has many sites and points of presence but only a small calculation capability at each point. PlanetLab has about 500 sites globally but only 2-3 nodes per site. Over the last year, the largest users of PlanetLab have asked for increased compute power and storage at a small number of PlanetLab sites to provide a Virtual Cloud Infrastructure (ViCI). PlanetLab has identified resources to install 1000 core clusters at 5 U.S. and two European sites. They are now acquiring the resources. PlanetLab will contribute technology to both define the cloud and use the cloud. They are primarily working with longer-term users with consistent requirements for cloud services rather than providing dynamic services for short-term users. The cloud sites have good supporting bandwidth. PlanetLab will be putting incentives in place for users to contribute resources in proportion to the resources they use. Capabilities include:

- content oriented file system

- Robust-scalable systems
- Object location service

These facilities will support researchers in their development of new programming models, virtual worlds, and personal clouds. Subsequent research will focus on performance, detecting faults, load-balancing, and security and trust.

The ViCI clusters will be homogeneous. This cloud facility will balance building the facilities with using the facilities. A flexible barrier between users and developers will assure the interests of both communities are addressed. Users will also have developer privileges on the cloud platforms. Identity resources will be standardized so the sites can maintain one identity for users.

Community Roundtable

ESG: Dean Williams

ESG has eight gateways now including 5 U.S. laboratories. ESG is collecting CMIP data from the Australian National University. Over the next year ESG is expected to collect over 10 Petabytes of data. Security is a continual concern. Computation is generally done at the data nodes where the data resides. They will use some of the Magellan capability once it has become available. ESG is currently focused on analyses using the IPCC data that has been developed. They are adding in observational data sets and European data sets to the modeling results that have been completed. NCDC and NOAA data sets are being used.

Open Science Grid; Ruth Pordes

With the release of the Structural Biology Grid portal at Harvard Medical School just before SC10, the usage has sharply increased and there are two specific new groups running production at about ~5% of the total OSG daily usage (includes the LHC!) across the local HMS and more than 15 remote OSG sites. Two application groups have been recently on-boarded: ~24 individual biological researchers attempting to use molecular replacement techniques to solve the structure of difficult x-ray crystallography protein data sets; a few (hospital based, non-clinical/anonymized data users) life sciences researchers who run 100s-1000s of jobs for model simulation (parameter sweep or MC style execution and statistics), data analysis (ECG data), or image processing (cryo-EM).

With the ramp up of support for High Throughput Parallel Computing an increased number of chemistry simulation applications are running on multiple OSG sites submitted by the group led by Professor Juan J. de Pablo's group at the University of Wisconsin Madison. As part of the joint project between OSG and TeraGrid – ExTenci – chemistry simulations from the University of Chicago are also running across more than 5 sites, with a peak daily throughput this week of more than 100,000 CPU hours.

In other work at SC10 we demonstrated visualization of the OSG usage and of disk access patterns for the data intensive LHC CMS analysis jobs on the University of Nebraska Hadoop system. We would point you to the Linux Today article about this installation that has generated broader interest at

<http://www.linuxjournal.com/content/the-large-hadron-collider>.

DOE: Rich Carlson

SciDAC3 was announced at the SC10 PI meeting. The SciDAC institutes will be competed in FY11 and science applications will be competed in FY12.

Next MAGIC Meetings

January 5, 2011, 2:00-4:00, NSF

February 2, 2:00-4:00, NSF