Minutes
MAGIC Meeting
January 6, 2010, 2:00-4:00
NSF, Room 1150

Action Items

1. Linda Akli will coordinate developing a briefing on SURAGrid for the February MAGIC meeting.

2. Grant Miller will provide coordination between JET and MAGIC on issues of common interest.

3. Mike Nelson will send the JET information on the January 13 Webinars on the Digital Cloud

Proceedings

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

Developing Data-Intensive Applications for Heterogeneous Distributed Platforms

Presentation by Shantenu Jha of Louisiana State University

Data intensive applications (DIA) generally use distributed data that is supported by heterogeneous infrastructure. It is important to decouple application development from the underlying infrastructure and to provide interoperation across Grid clouds. A flexible infrastructure should support multiple applications on differing infrastructures. The goal is to develop distributed data-intensive applications to use a broad range of distributed systems without vendor lock-in or disruption while retaining the flexibility and performance that scientific applications require.

Frameworks address some of the primary challenges for developing distributed data-intensive applications. The frameworks need to address:

- What are the components and how are they coupled
- Layering, ordering and encapsulation of components

Distributed applications objectives include:

- Interoperability
- Distributed scale-out: using multiple distributed resources concurrently
- Extensibility
- Adaptivity
- Simplicity

The challenge is how to develop distributed applications with these objectives.

SAGA provides:

- General-purpose basic and common Grid functionality for applications.

- Building blocks as a basis for constructing higher levels of functionality and abstraction
- Meets the needs of a wide range of applications

SAGA provides an integrated, stable, uniform and high-level interface and it provides application developers with units needed to compose high-level functionality across distributed systems. SAGA Components include:

- Python API wrapper, CAPI wrapper, Native C++ API
- Functional API packages for: jobs, files, replicas, CPR, and SW
- Adapters for: files, jobs, replicas, CPR
- Middleware services

SAGA C++ is open source and is implemented as a set of libraries that include:

- Core
- Functional packages
- Language wrappers
- Middleware wrappers

These provide horizontal extensibility, vertical extensibility (different adaptors for different middleware), and extensibility for optimization and modular design. SAGA adapters include: Fork, SSH, Condor, Globus, GRAM2, OMII, GridSAM, Amazon EC2, and Platform LSF. File adapters include: Local FS, GlobusFTP, Hadoop Distributed File System, Cloudstore KFS, and OpenCloud Sector-Sphere.

SAGA has been used on four classes of problems:

- SAGA NXM Framework: compute matrix elements where each is a task
- MapReduce Framework: Control the distribution of tasks
- Sphere: Stream-based processing
- Directed Acyclic Graph (DAG) Execution

SAGA provides a basis for performing experiments to understand performance tradeoffs, configurations, and infrastructure, for a broad range of applications. Designing and implementing SAGA-like frameworks is challenging.

For the analysis results of these classes of problems and the full briefing please see: http://www.nitrd.gov/Subcommittee/lsn/magic/index.aspx

Roundtable

Earth Systems Grid: Dean Williams, See http://esg-pcmdi.llnl.gov/

Last month the ESG its alpha version software. This month the beta version will be released with a Beta2 version scheduled for the following month. ESG is gearing up for CMIP5.

CEDPS: Steve Tuelke

Last year CEDPS developed a project plan for the remainder of the CEDPS program. It centered on getting the Globus service up and stable and getting early users onto the system. They are considering how to integrate Condor with data movement.

CDIGS:

CDIGS is working on implementing the Globus toolkit, Version 5 which will be released in a few days. They are moving to Gram5 which is backwards compatible with Gram2. Testing is going well. They are looking at how to implement logging and integrating measurements from the global world, likely using PerfSONAR.

TeraGrid: Dan Katz

TeraGrid is ending and Expedia is starting up. Initial deliveries of equipment will start next Monday. TeraGrid had three times as many requests as they could fulfill. Reviewers evaluated requests for resources and the staff allocated resources. In general all meritorious applications were scaled back on their request to allocate available resources fairly.

TeraGrid is installing several new systems:

- Two remote data and analysis visualization machines are arriving by Fall
- Dash: a prototype for Gordon, coming in 1-2 years, a flash-based system that will store data in flash for faster access than from disk
- A GPGPU system is expected from Georgia Tech in the fall
- FutureGrid, an experimental resource system for research with nodes in Indiana U., Purdue U., U. of Chicago, TACC, U. Florida, and UCSD. Most nodes will have VMs running so the resource can be operated as a cloud environment

The DOE PIs are implementing a Magellan node at Argonne National Laboratory in Chicago.

OSG: Miron Livny

OSG is streaming real data. The data movement is stable. Currently security implementations are impinging on the ability of the researchers to accomplish their work.

SURAGrid: Linda Akli

SURAGrid developed a strategic plan in August 2008 and SURAGrid is currently working to the milestones established in that plan. The plan provides for an environment that is more robust and stable. TTU provided a document on how to make SURAGrid a stable operational environment.

AI: Linda Akli will coordinate developing a briefing on SURAGrid for the February MAGIC meeting.

AI: Grant Miller will provide coordination between JET and MAGIC on issues of common interest.

Meetings:
January 27, NSF: Workshop on the Open Innovation Model

AI: Mike Nelson will send the JET information on the January 13 Webinars on the Digital Cloud

Next MAGIC Meetings

February 3, 2:00-4:00, NSF, Room 1150
March 3, 2:00-4:00, NSF, Room 1150