MAGIC Meeting Minutes
February 6, 2013

Attendees

GJ Ahn	Arizona State U.
Jim Basney	NCSA
Bob Bonneau	AFOSR
Rich Carlson	DOE/SC
Michael Dresher	EGLeu
Geoffrey Fox	Indiana U.
Chris Greer	NIST
Shantenu Jha	LSU
Krishnan	
Craig Lee	Aerospace Corp.
Bryan Lyles	NSF
David Martin	Northwestern U.
Andre Mersky	
Grant Miller	NCO
Alan Sill	TTU
John Towns	NCSA/XSEDE
Von Welch	Indiana U.

Action Items

Proceedings

      This MAGIC Meeting was chaired by Rich Carlson of DOE/SC.
Alan Sill organized a session to discuss the NIST Cloud Computing and Big Data Forum &
Workshop and the NIST Standards Acceleration to Jumpstart Adoption of Cloud Computing
(SAJACC) Program.

The briefings are summarized below. For the complete briefings, please see the MAGIC Wiki site
under the February 6, 2013 Meeting at:
http://www.nitrd.gov/nitrdgroups/index.php?title=Middleware_And_Grid_Interagency_Coordi
nation_(MAGIC)#title

NIST Cloud Computing and Big Data Forum and Workshop: Chris Greer

      NIST held a workshop January 15-17 to explore the intersection between cloud
computing and big data. There were 600 participants and 1000 Webcast viewers. The
workshop goal was to report progress on developing standards for cloud services and
capabilities, to identify potential cost savings, and to identify increased opportunities to address
big data using cloud computing. The next NIST workshop, in September or October, will
address the intersection of cloud computing and mobile computing. The workshop discussed
how to promote progress in the big data community by using clouds. It addressed definitions,

FOR OFFICIAL GOVERNMENT USE ONLY
c/o National Coordination Office for Networking and Information Technology Research and Development
Suite II-405 · 4201 Wilson Boulevard · Arlington, Virginia 22230
Phone: (703) 292-4873 · Fax: (703) 292-9097 · Email: nco@nitrd.gov · Web site: www.nitrd.gov

big data, and applications and developed a taxonomy, reference architecture and a roadmap for big data applications.

Big Data and Clouds: Challenges and Opportunities: Geoffrey Fox

Geoffrey Fox reviewed the proceedings of the January NIST workshop. The charge to the presenters was to discuss opportunities and challenges presented by the intersection of clouds and big data., e.g.,

- The ability of clouds to make big data approaches feasible and cost-effective
- Bringing the computation to the data rather than the data to the computation

There will be a huge need for cloud workers soon (Microsoft predicts 14 million cloud jobs by 2015). Curricula are needed to educate people to design/use clouds running data analytics processing big data to solve problems in informatics. Areas needing addressing include data curation/management, analytics, run-time, and applications. A promising approach for education is Massive Open On-line Courses (MOOCs).

Clouds will be needed for scientific data analysis. Many communities have been developing cloud applications to address central problems including the LHC and genomics communities. We need to generate a consensus on data architectures for simulations. Improvements in data analytics improve the capability of clouds to address big data, including:

- Better algorithms and hardware
- Data analytics libraries
- Data mining, image processing, information retrieval, dimension reduction
- Collaboration between academia, business and government to build robust, scalable algorithms

FutureGrid is an international testbed modeled on Grid5000. It offers computing testbed as a service including provisioning, image management, dynamic IAAS and NAAS, and expert management.

Federated Community Cloud Team Report: Alan Sill

The Cloud Computing Standards Report was incorporated into the January workshop. It addressed Requirement 5: Frameworks to support seamless implementation of federated community cloud environments. It discussed how hybrid cloud and cloud broker elements can be leveraged and harmonized in federated community cloud settings. It discussed:

- Current usage patterns and projected trends
- Applicability of Grid and cloud solutions to federated cloud scenarios
- VO use cases address many of the federated community cloud requirements such as:
    o Privacy and security
    o Compliance adherence
    o Trust infrastructure
    o Membership
    o Policies and procedures
    o Common governance
    o Etc.

Scenarios were defined including catastrophic dynamic event response and specialized remote medical care.

The document discussed VOs sharing platforms to federate clouds. It also documents current use patterns and projected near-term trends in architectures and tools. There are several activities in the community that indicate the developing standards including:
- Cloud Plugfest series (OGF, SNIA, ETSI)
- Emergence of open source cloud broker products and tools
- EC-funded projects supporting federated clouds: Contrail, FI-Ware, Helix, Nebula

Federation of resources is enabled by federated identity management. Identity provisioning from the Grid community provide a range of levels in degree of security.

Cloud-Based Disaster Response: Craig Lee

Federal agencies have public trust responsibilities such as Disaster Response that will not be met by the commercial community. The National Geospatial-Intelligence Agency contracted for a pilot project of a Geospatial Intelligence Community Cloud for Disaster Response. That:
- Leverages industry best practices
- Enables data mobility
- Enables each organization to manage the data they provide
- Protects data from unauthorized use

The project is developed in three cycles: Infrastructure development; software development, and enhancements. The on-demand collaboration requires:
- Federated identity management
- Federated access control
- Trust federations

XSEDE: David Lifka

Extreme Science and Engineering Discovery Environment (XSEDE) has the goal to enhance research productivity. Clouds are a critical resource such that XSEDE will integrate cloud services into its portfolio to support use cases that are not currently well served. XSEDE announced a survey in August 2012. Over 75 submissions were received to date and more are coming in. Survey data will be made available by February 2013. The survey includes a wide array of provides including, e.g., Google, SDSC Cloud, Amazon Web services,… A wide range of data is being collected such as cloud providers, bandwidth in/out of the cloud, data moved into/out of the cloud, storage, software,… A wide range of science disciplines are represented. User benefits include:
- Pay as you go
- Lower costs: capital expenditures and administration
- Computational scalability and elasticity
- Data scalability and elasticity
- Software as a service
- Education as a service
- Ease of use
- Fast scientific workflows
- Rapid prototyping
- Expanding data analysis to small labs

User issues include: the learning curve, database stability and scalability, security issues, data costs, storage costs and performance, and funding issues.

SAJACC: Alan Sill

SAJACC will provide a portal for a public internet-accessible repository. It will enable interoperable cloud computing before standards are finalized and faster standards development. Elements of SAJACC will include:

- A NIST cloud standards portal incorporating use cases, validated specifications, reference implementations and proposed specifications and reference implementations
- Government-run validation exercises

SAJACC cloud management use cases will enable users to open, use and close an account, copy data objects into/out of the cloud, erase data objects, VM control and query of cloud provider capabilities and capacities. SAJACC will enable interoperability use cases between cloud providers. It will provide security services: identity management, authorization policy management, e-discovery, and security monitoring. The heart of the SAJACC process is DEMONSTRATIONS. Future use cases include the cloud management broker, transfer of data ownership within a cloud, fault-tolerant cloud group, and multi-provider integration. SAJACC provided recommendations for moving forward:

Next MAGIC Meetings

- March 6, 2:00-4:00, NSF, Room II-415
- February 3, 2:00-4:00, NSF, Room II-415