



The government seeks individual input; attendees/participants may provide individual advice only.

Middleware and Grid Interagency Coordination (MAGIC) Meeting Minutes

June 6, 2018, 12-2 pm
NCO, 490 L'Enfant Plaza, Ste. 8001
Washington, D.C. 20024

Participants (*In-Person Participants)

Richard Carlson*	DOE/SC
Devarshi Ghoshal	LBNL
Dan Gunter	LBNL
Joyce Lee*	NCO
Don Riley	UMD
Jack Wells	ORNL

Proceedings

This meeting was chaired by Richard Carlson (DOE/SC). May minutes are approved and will be posted to the website.

DevOps Discussion

Goal: Flesh out ideas and interest on DevOps speaker series.

DevOps: Involved in unifying software development and software operations for continuous improvement in automation. Brainstorming session included the following discussion points:

- DOE HPC centers are very interested. There is some movement starting in the Exascale Computing Project (ECP) (e.g., focusing on deployment of continuous integration capabilities at NERSC, OLCF, ALCF). Fund ECP and/or facilities to procure and deliver new capabilities and continuous integration.
- Perhaps MAGIC could host presentations, etc. We would like information from NSF and DoD.
- LBNL has been doing DevOps for a while. When crossing distributed computing and institutional lab boundaries on networks, we end up doing deployment and configuration activities. We could look for folks who have been involved in workflows, distributed computing in this space for a while and talk about the DevOps side of their activities, i.e., how it is evolving, given the ECP and Amazon activities.
- How do you interface and build a sustainable software environment in a distributed manner?
- Perhaps bring in a discussion of something like JupyterHub (part of work is setting up an environment for others to use on the web to spawn their own analysis tool; i.e., you do the DevOps so they don't have to)
- It is difficult to maintain open source software over time. It would be good to know how people in our community are addressing software, DevOps-related activity and problems? What is required for HPC users and HPC software developers (differs from cloud community)?
- Have projects that host/run websites (e.g., Kbase). Run services on a super computer and identify their challenges in a DevOps realm.

- Deployment of the OpenShift platform on OLCF allows the sustainable provisioning of new services at OLCF (running Kubernetes through OpenShift).
 - Do NSF centers have same kind of issues? NSF has more diverse communities.
 - Kades is facing same issue because they are trying to drive more integration between the ORNL Kades capability and DOE OLCF capability.

Structure of series:

- Run through the summer.
- Kick off series with the research community who will set the level for the rest of the series.
 - How have they been migrating and changing from static development systems to more agile systems? Hear from those who are involved in sustainability (e.g., Globus which is using a model for charging to maintain sustainable products); how have they have been accomplishing sustainability (e.g., Linux RedHat, Open source)?
- Facilities: Operation staff (running machines and software product maintenance) and application developers (sustainability and agile integration)
- Application developers: look at sustainability and agile integration
- Science community: develop or take advantage of what others are developing?
- Wrap up with discussion
- Meeting participants will email suggestions for speakers and topics for the next meeting to Rich Carlson, Rajiv Ramnath and Joyce Lee. We will send out email outlining direction and soliciting speakers from MAGIC members.

CY19 Tasking

Every October, MAGIC presents what we have accomplished and future accomplishments for the following year. This year, MAGIC ran a speaker series on containerization/virtualization and is planning a series on DevOps. While MAGIC decided that a containerization workshop was not needed; it may decide to put on a DevOps workshop.

Possible topics for CY19 tasking:

- DOE is looking into a pilot project across Oak Ridge National Laboratory, Argonne National Laboratory, Brookhaven National Laboratory, Lawrence Berkeley National Laboratory, and the Environmental Molecular Sciences Laboratory to interconnect lab computing infrastructures into a federated environment. We could obtain a briefing from them next summer. This topic includes authorization.
- Are there other ways to address a large, federated distributed computing environment? (See last year's tasking)
- Coming from the network world, when starting to debug network problem, the infrastructure needed to gather data belongs to someone else. So, it is difficult to obtain data. These same problems exist as we move into more distributed computing environments with many workflows; there are no good tools for locating the problem and reporting it to support staff who will act upon it. Perhaps we could shed some light on this scenario and find ongoing work in this area.
- Looking at both federated cloud environments and supercomputing facilities (see Task 2, CY18). Scientific instruments are growing in their ability to push out data; trying to nail an instrument like LCLF to a specific computer isn't how we will get the most productive science out of machines or computers. If we have enough computing power, we can use machines at various locations and distribute the load from science instruments in an intelligent manner.
- At future meetings, we will continue discussing tasking and will finalize it at MAGIC's September meeting, in time for LSN's annual planning meeting. Discussion will include the structure of

meetings; multi-meeting structure may not work for all topics. The group may want to revise the structure to incorporate more discussion during presentations (spread across more months or reduce the number of speakers).

- As we are using different machines? Is there something more foundational, beyond the current, ad hoc solutions, for tracking data? Ties into the provenance issue and verification.

Roundtable

No new updates.

General Discussion:

University Curriculum development and Academic community involvement

At last month's meeting, the need for different skill sets in development teams was raised. Ideally, these skills could be introduced in an academic environment, with additional training conducted at the workplace.

New technology is being incorporated in university curricula (e.g., At UMD, the Big Data course required for Master of Science (MS) information systems and business analytics, is now called Big Data and AI).

- While the main source for technical talent is computer science and electrical engineering, some of these folks come out of Master of Science (MS) programs as well as business schools.
- DOE point of view: many scientists without computer science or electrical engineering degrees are involved in generating code and running experiments; computing is an essential element of their scientific process. There is more growth in all areas.
- The growing need for analytics is blurring the line between domain science and non-technical fields.
- MAGIC will keep an eye out for issues that are related to what we should think about in university programs. We would like to involve the academic community as well as the domain science communities.

Next MAGIC meeting

July 11 (12 noon EDT)