



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

Middleware and Grid Interagency Coordination (MAGIC) Meeting Minutes¹

August 7, 2019, 12-2 pm
NCO, 490 L'Enfant Plaza, Ste. 8001
Washington, D.C. 20024

Participants (*In-Person Participants)

Wes Bethel (LBL)	David Martin (ANL)
Richard Carlson (DOE/SC)*	Tom Morton (DoD CIO)
Vipin Chaudhary (NSF)	Michael Nelson (Georgetown)
Sharon Broude Geva (UMich)	Don Petravick (NCSA)
Ian Foster (ANL/UChicago)	Hakizumwami Birali Runesha (UChicago)
Dan Gunter (LBNL)	Arjun Shankar (ORNL)
Florence Hudson (NE Big Data Innovation Hub)	Alan Sill (TTU)
Joyce Lee (NCO)*	Suhas Somnath (ORNL)
Miron Livny (UW-Madison)	Sean Wilkinson (ORNL)

Proceedings

This meeting was chaired by Richard Carlson (DOE/SC). June 2019 meeting minutes were approved.

Speaker Series: Data Life Cycle Summary (February – July 2019)

- Rich Carlson (DOE/SC), Vipin Chaudhary (NSF) and Joyce Lee (NCO)

Last fall, MAGIC decided to started series on data life cycle, from inception to long term archival storage. Changes in data field (data analytics ML, AI); different ways of dealing with data. Larger machines and simulations are generating more data. MAGIC thought it important to look at what has changed and their impact on agency perspectives and the programs to examine. Want to identify changes and for programs to incorporate and take advantage of these programs.

- February: Started examining data sources and topics interested in discussing.
- March : Brought in scientists and science community groups from environmental science and smart cities community to discuss is changing in several science communities; environmental and societal issues at large.
- April: Data analytics and processing.
- May: Data management and storage activities. Add importance of understanding IO patterns to extract maximum performance from HPC calculations (Glenn Lockwood talk: [Understanding and Tuning I/O Performance ATPESC 2019 Glenn K. Lockwood National Energy Research Scientific](#)

¹ Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program.

[Computing Center Lawrence Berkeley National Laboratory](#)). NERSC developing tools to improve ability to analyze and diagnose. This talk includes some observations regarding the importance of developing new tools to understand I/O performance to be included in the NITRD MAGIC summary of data lifecycle topics. Exemplar of kind of code that funding agencies. Should be pushing so can get better value out of their computing.

- June: Interactive analysis of tools
 - Thomas Mendoza will be speaker on Jupyter Security (Sept. 23) at Trusted CI Webinar, NSF Cybersecurity Center of Excellence; “Jupyter Security at LLNL with Thomas Mendoza” will be shared through Trusted CI community, an open community <https://trustedci.org/webinars>
 - Rick Wagner working on integration of Globus Auth
- July: Data triage and discovery

Summary of Scientific Data Life Cycle prepared by Rich Carlson, Vipin Chaudhary, Joyce Lee (see slide deck). Speakers, Challenges and Solutions

Speakers

Another set of speakers for next month.

October 10: LSN Annual Planning meeting.

- MAGIC needs to be prepared for next topic.
- Next topic(s): speaker series? Other topics? Will have more detailed discussion in time for APM
- MAGIC presentation on what we have accomplished to engage federal agencies.

Discussion of Topics

How to share MAGIC’s activities?

- Consider proposal presentation for PEARC ’20 (e.g. lightening talks of important discussions) to Increase awareness, discuss challenges and solutions) PEARC 20 (Oregon, in July 2020) <https://pearc.acm.org/pearc20/>
- Dissemination of MAGIC’s activities
 - Possible proposal presentation for PEARC ’20. <https://pearc.acm.org/pearc20/>

Data integrity for scientific endeavor (data provenance, security)

- Data integrity is focus of NSF Cybersecurity Center of Excellence (invite Von Welch or Jim Basney)
 - Data integrity, etc within data life cycle (working on encrypted data (secure nodes); sites setting up data enclaves associated with HPC centers)
- Von has put together group for Supercomputing to discuss identity management; he is unavailable this year. Perhaps can bring in speakers during the year.

National Cybersecurity Center of Excellence (NIST)

- NIST is part of LSN, but haven’t been participating in LSN calls.
 - NIST proposals requesting input: Data confidentiality – identifying and protecting against breaches and How to recover from breaches
 - Interesting topics? Something new in data confidentiality? New data confidentiality projects

Focus on data integrity, etc within data life cycle

- Wide range of topic areas: from working on encrypted data (secure nodes); sites setting up data enclaves associated with HPC centers

Data confidentiality (National Cybersecurity Center of Excellence (NIST); note NIST proposals requesting input on identifying and protecting against breaches and How to recover from breaches)

Implications of new AI to science workloads for HPC ecosystems (what computers, what middleware is needed, network structures, etc)

- DOE running series on AI for science (need summary)
- AI for optimizing HPC – work on IO systems

ROI and cost efficiency for academic and lab based computing

- 2 PEARC papers led by Craig Stewart (Indiana University) – calculating ROI quantitatively and financially and in human terms.
- CASC: discussing topic of optimizing cost efficiency in delivering computing is a sensitive topic, but need to address it. Can't have intelligent conversation about use of cloud computing without thoroughly examining this topic. – invite Craig, others to review results

Direct integration of energy sources and computing facilities

- Multi-disciplinary - integrating energy production and computing more closely. Many startups locating data centers remotely, near sources of renewable energy
- Zero carbon cloud: Andrew Chin working on feasibility of stranded power to power small data centers
- Training is requiring larger and larger machines – using HPC resources do conduct training. Training is power intensive

Re-examine networking infrastructure underlying middleware

- Attempts to put more middleware in network layer (authentication, encryption, security)
- Major network providers could discuss what doing to support high bandwidth distributed computing

Note: Look into Research Data Access & Preservation Association (RDAP)

Deliverables

Containerization Report: pending approval by management. Dhruva Chakravorty (TAM) recognized for organized report. Will circulate when approved.

DevOps Series Report: Seeking volunteer to summarize 2018 DevOps series. Will get credit for summarizing. Joyce Lee will assist as needed.

Roundtable

September 23 - Trusted CI Webinar, NSF Cybersecurity Center of Excellence; “Jupyter Security at LLNL with Thomas Mendoza” (<https://trustedci.org/webinars>)

September 25-27- [CASC meeting](#), The Alexandrian, Alexandria, VA

Nov 1, 2019 -[CSSI](#) deadline

November 14, 2019 [OAC Core](#)

November 19, SC19, Colorado Convention Center, Rm 711 (1:30 – 3:30 p.m. MT)

Next Meeting: September 4 (12 noon ET)