



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

Middleware and Grid Interagency Coordination (MAGIC) Meeting Minutes
September 7, 2022, 12-2 pm ET

Virtual

Participants

Dan Gunter	Kevin McMorrow
David Martin (ANL)	Marcy Collinson (Oracle)
Hal Finkel (DOE/SC)	Mike Heroux
Jay Park (NSF/OAC)	Rich Carlson (DOE/SC)
Jeff Conklin (NCO)	Sharon Broude Geva (University of Michigan)

Introductions: This meeting was chaired by Rich Carlson (DOE/SC) and Jay Park (NSF)

Planning Discussion:

- Jeff Conklin gave an overview of the topics that were covered over the last year.
 - Cloud efficiency, cloud economy
 - Network capacity and infrastructure issues
 - Best practices for delivering on-prem resources
 - Workforce Development
 - Maintenance challenges for software
 - Leadership Scientific Software
 - AI/ML to advance grid and cloud computing
 - Use of grid and cloud computing to advance AI/ML
 - New AI Architectures
 - Creating a computing continuum
- Potential topics for 2023
 - AI/Deep Learning/ML – Potential series
 - Programming Modules
 - Multitenancy
 - AI for the Edge Instruments
 - Troubleshooting Challenges

AI/Deep Learning/ML

- Marcy Collinson from Oracle said they could find a subject matter expert for a few different types of topics, including:
 - Health care and AI
 - HPC within the research, industry, and cloud bursting
 - Offered to take lead on finding a presenter for one of those
- Rich asked what we want to get out of the AI/Deep Learning/ML? Current state of the art? Future directions? He said there are three different areas here, with healthcare, cloud bursting, weather. Do we want to go with other elements in here? Do we have other things?

- How many months? What is the overall theme?
- Dan Gunter said it would be really nice to get a few talks to touch on some different domains so we can see where the similarities and differences are and how deep learning is being used. Bio vs Materials science vs Physics etc.
 - Could be one or two sessions.
- Rich said: How are the science communities using AI/ML? Would include the health care system, the researchers in health care, etc. Another aspect is How do our facilities find it useful? The light sources? Computer centers? Are the people running the infrastructure using this? Miriam's been doing something in this for the networking world.
- Dan mentioned self-driving labs. That kind of combines AI and the facilities.
- Rich asked Marcy: Is Oracle using AI to do planning for how their infrastructure is used? Is there anything public that they would be able to say in how they are using it internally rather than supporting their user community?
- Marcy said if we want to get more specific about a sub-domain or a specific AI topic, she can take that to their AI/ML team and see what they can disclose.
- Mike Heroux said AI-assisted programming or AI-assisted software development is also an interesting topic.
- Rich asked if Mike thought that was also moving into the debugging aspects? Performance tuning area? Mike said very much so.
- David Martin suggested that the interplay between traditional HPC and the new AI accelerators would be interesting.
- Sharon Broude Geva mentioned the ethics for AI and the equity and diversity issues that go with it. She wondered if looking at it from a research perspective in a way that would tie in more closely to what people have to think about as they do computational research might be interesting.
- Marcy said that Oracle is sponsoring the brain hack effort, which is all about neuroscience and ethical hacking. She shared a link to the blog post: <https://blogs.oracle.com/research/post/oracle-sponsors-brainhack-2022-advance-neuroscience-cloud> She said she thinks it is an excellent topic and she would be happy to coordinate with the researchers they are working with if that's of interest. They are based in Australia.
- Rich said in addition to ethics, there is the bias. If we're looking at science data and we're trying to draw conclusions, what kind of biases can sneak into an AI system that would impact your conclusions?
- Jay said that if we consider AI as a service and a system that provides that kind of service then definitely fair AI service would be an interesting topic for this group.
- Dan Gunter said that he thinks this coincides with the topic of AI interpretability. Can you interpret those results and understand them and therefore know what the biases may be or be able to interpret them differently.
- David Martin said one of the things that we're also struggling with is how you share your models.
- Dan said that we also have this issue in that AI is part of a bigger model than it feeds into, because then it's hard to see how the effects are propagating.
- Sharon said these go with the interplay between HPC and AI accelerators or other aspects of workflows. It might be interesting to see different types of workflows that include AI and ML.

Programming Models

- Rich said that we just did several months on software sustainability. He asked Mike Heroux if there are elements of that or new programming models that we'd like to touch on?

- Mike said we have these community developed portability layers. Things like Kokkos and RAJA. We have open MP, we have a target offload, we have open ACC, we have the C++ language standard itself that supports on-node parallelism, etc. It also takes a really long time for new features to make it into this language standard. Fortran, in particular, has really struggled to provide modern features in a way that's available to commercial compilers or the community compilers. He said there are a lot of opportunities for discussion and around trying to accelerate our ability to efficiently express parallelism in our software.
- Rich said that at one point there was a lot of discussion and exploration of domain-specific languages. Is that activity still going on?
- Mike said he thinks it is. The challenge is that with most DSLs in the scientific community, there's not a big enough community to support them.
- Sharon asked if it would be interesting to look at the interfacing part of these languages.
- Rich asked David and Marcy how we deal with getting these things deployed and user support.
- David says it's a good question. Deploying them isn't horribly hard, but supporting a really heterogeneous environment is.
- Marcy said this is an interesting problem that Oracle is looking to solve through product development. She said they will be making a public statement around a product they've been working on that's research specific. She said she's not sure how much analysis around specific language development has earned with that product as of yet. She said that her team is in favor of working hand in hand with the research community to try to problem solve for these types of issues and make sure that cloud interfaces are taking language and code into consideration.
- Rich asked Marcy if it would be in her public domain to list the primary challenges of programming inside that computer environment you have. Or is it not a big challenge because people specify what kind of resources they want when they ask for an allocation?
- Marcy said a little bit of both. They are doing in-depth market analysis around these types of topics, specifically with the research community as well as federal. She said she would be happy to take this back to her team and if further discussion is warranted, or if we wanted to establish something like a working group around this, there is the possibility to talk about multi-industry collaboration.
- Rich said we don't want to ask for anything that's proprietary or business specific.
- Rich asked Marcy if there's a set of challenges that she might think are unique to programming in the cloud. He asked David what's unique about programming for HPC.
- Marcy said it's multifaceted in the cloud. She said she would have to assess the individual aspects
- David said hearing from a commercial standpoint compared to an LCF would be interesting.
- Marcy said these things are becoming more complicated every single day.
- Sharon said maybe going off in a different direction, have these new programming models changes how we need to look at requirements for reproducibility?

Multitenancy

- Rich described the multitenancy issue.
- David said it depends on what you mean by multitenancy.
- Rich said there are more and more accelerators and other unique devices or multiple cores.
- David said they've already got the problem in Polaris now, where it has GPUs and CPUs and people load up the CPUs and ignore the GPUs.
- Marcy said or they want GPUs and GPUs are not readily available. And there's problems with establishing GPU integration with multitenancy capabilities.
- David said they haven't figured out any way to allocate GPUs separately.

- Sharon said there's the security and compliance aspect.
- Marcy said that Sharon had an excellent point.
- Dan asked if this overlaps at all with the concept of the types of workloads.
- Marcy suggested cloud bursting scheduling or schedulers.
- Rich said that that is a persistence issue. Then the other issue is the real-time aspects. Can a real-time job share resources with a long-running cube job.
- Sharon said going back to the question about workloads, given that there are new audiences for a lot of these resources, and a lot of them are long tail. Their jobs may be much smaller and so that causes even more need to be able to multi-tenant resources.
- Rich said in the HPC world, what he's seen is people using scavenger services. But it's not a multitenancy environment, it's a limited use resource.

Other

- Rich said that he thinks the AI for edge systems can get rolled up into the higher level for the AI systems and that part of the discussion.
- Rich said troubleshooting challenges might fit in with the programming models.
- Rich said typically we have a session dealing with cybersecurity – current status
- He said we have typically tried to use the meeting co-located with SC to bring in our international partners for some things. And we've either done cybersecurity there or just what's going on in other parts of the world and modeling the MAGIC area.
- He said we are planning to have a hybrid meeting in Dallas so we will cancel our normal monthly meeting in November.
- Sharon suggested adding NAIRR.

Roundtable

Next Meeting

October 5 (12 pm ET)