Networking and Information Technology
Research and Development

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

Middleware and Grid Interagency Coordination (MAGIC) Meeting Minutes[1]
May 1, 2019, 12-2 pm
NCO, 490 L'Enfant Plaza, Ste. 8001
Washington, D.C. 20024

Participants (*In-Person Participants)

Ilkay Altintas (SDSC)	Joyce Lee (NCO)*
Ben Blaiszik (UChicago/ANL)	Glenn Lockwood (NERSC)
Richard Carlson (DOE/SC)	David Martin (ANL)
Dhruva Chakravorty (TAM)	Gilberto Pastorello (LBL)
Kevin Constantine	Don Petravick (NCSA)
Devarshi Ghoshal (LBNL)	Hakizumwami Birali Runesah (UChicago)
Diego Davila	Alan Sill (TTU)
Jake Fries (NCO)	Derek Simmel (PSC)
Dan Gunter (LBL)	Suhas Somnath (ORNL)
Florence Hudson (NE Big Data Innovation Hub)	Dale Stansberry (ORNL)
Margaret Johnson (NCSA)	Nathan Tallent (PNNL)

Proceedings

This meeting was chaired by Richard Carlson (DOE/SC) and Vipin Chaudhary (NSF). March 2019 meeting minutes were approved.

Speaker Series: Data Life Cycle

- Ilkay Altintas, Chief Data Science Officer, San Diego Supercomputer Center, Division Director, Cyberinfrastructure Research, Education, and Development Director, Workflows for Data Science Center of Excellence, *Storage Challenges and Opportunities for Data Science in the Continuum*
- Ben Blaiszik, University of Chicago, Globus and Argonne National Laboratory, Data Science and Learning Division, *A Data Ecosystem to Support Machine Learning in Materials Science*
- Glenn Lockwood, Storage Architect. Advanced Technologies Group, National Energy Research Scientific Computing Center (NERSC), Lawrence Berkeley National Laboratory, *Simplifying Data Management Through Storage System Design*

[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.

Ilkay Altintas - *Storage Challenges and Opportunities for Data Science in the Continuum*

Overview

SDSC, Data Science Hub at SDSC

- Halicioglu Data Science Institute (couple undergraduate and graduate program/data science and engineering);
- Workflows for Data Science Center of Excellence at SDSC
- Focus on questions spanning computational science, bid data, data interpretation/exploratory data science through more intelligent workflows that sit on intelligent solution architectures)
- Goal: perform workflow problems

Common theme:

- "Big" data, computational science, data science, cyberinfrastructure and their applications
- Problems are increasingly data driven, need collaboration and are very heterogeneous. Typical team that engages in collaboration and communication -> pipeline that needs to be scaled.

Today's Charge:

- Science is collaborative at a continuum; hard to imagine data science without storage, which has different modalities and life cycle of what goes into storage;
- Storage needs to be coupled with data management services; and
- Need dynamic storage optimization through intelligent workflows that can take advantage of underlying intelligent infrastructure

Storage is Data!

How to ensure what we store is valuable; how to turn what we store into knowledge? How to distill data using HPC? How make products, different applications available, queriable, etc, to different communities? Then will find problems that can bring benefits.

Each problem has its own needs and way of interacting with systems. Focus on problems to solve at the application integration level. Data driven problem solving requires:

- Collaboration. As problem solving becomes more complex, collaboration involves expertise on heterogenous systems, data management, data driven methods, scalable tools for dynamic behavior. Need the points from today's charge in a proper storage system coupled with data systems.

SDSC Data Gateway: Built solution architecture; any system, storage, networking, computing systems managed by resource managers.

- Something runs on them; performance profile – can compose services running on those systems in an optimal way;
- Working on workflow management problems is to coordinate and create optimal schedule;
- Product of management operations needs to build interface for the community to use;
- All stackable layers need data life cycle management, reproducibility, security and collaboration to be useful to science overall; and
- 4 parts relating storage: data life cycle management, composable services, resource management and composable systems. Need an abstraction on top of storage to enable intelligent decisions from virtual management.

Using workflow for process integration through dynamic composition

- Learn from exploratory collaboration? Process for the practice of data science.
    - Learn during collaborations and turn into services, based on what we learn.
    - Any workflow system can be coupled with certain environments. So, exploratory learning can be turned into scalable execution across systems.
    - Dynamic composition requires systems to collaborate with it, dynamic network , compute and storage resources and intelligent software to steer applications.

Pacific Research Platform (PRP) network

- Built Cognitive Hardware and Software Ecosystem Community Infrastructure (CHASE-CI)
- Machine Learning (ML) layer built on top of PRP; can couple some storage and simple computing to create capability for doing ML on big data much faster.
- Storage block: Many things coming together (CEPH, block, object storage and other storage from datalake or other storage for further scalability - all coordinated through ROOK, a cloud native storage orchestrator.

Scott Sellers Connect application: Rapid 4D Object Segmentation of NASA water vapor data

- Identify water vapor through imagery and ML to create database of precipitation
- PRP shortened his workflow from 19.2 days to 52 minutes; requires constant download of data and executing ML on the data.
- Note distributed nature of storage and cloud computing, and the needs around that dynamic.

Summary

- Dynamic storage composability matters
- Integrating optimization workflows and solution workflows; integrate as part of optimization scenarios. Next, need to think of data management as part of abstraction because the lifecycle of storage can be optimized together with the efficiency of the application.

Ben Blaiszik - *A Data Ecosystem to Support Machine Learning in Materials Science*

Growing opportunity for ML in the sciences (more publications across domains, for ML and informatics in material science – increase in 2014-15, continuing trend).

Challenges: access to datasets, models and codes; lagging benchmarks; while speaking to material science, most infrastructure building is general to other domains.

Motivation: real-time coherent diffraction inversion using deep generative networks
e.g. beamline scientist and this would be useful. How use it? Where's model, code, data?

Cohesive infrastructure emerging for AI-driven science: (Diagram of data)

- Data – see improved capture, organization, analysis, sharing, moving and delivering
- Much should be automated, especially organization and capture
- Compute should be accessible at many scales (hpc), types (cpu, gpu), local or distributed
- Model and codes – discoverable, well-described and portable (e.g., containerized), so easy to use on various computers
- Workflows – easily discovered, adapted, etc

Materials Data Facility – built to empower data publishing by researchers, heterogeneous or distributed. Operate to automate data and metadata extraction and ingest (critical)

Goal: enable unified search and discovery across different data sets

- Try to extract as much information as possible from files to make it discoverable and reusable
- Use APIs to enable automation
- Flow:
    - Submit data from lab or programmatic interface (can be ingested from many locations)
    - MDF Connect: Extraction and Transformation process
        - Extract material specific things (Diagram 6) and Transforms representations or change vocabularies to support interoperability of many services
    - Dispatch to various sources (e.g., MDF Publish, MDF Discover, NIST MRR, Citrine)

Automating Metadata Extraction – key goal of data triage as large amounts of data is generated; how sifting and making it discoverable. Automate as much as possible, in an intelligent manner to determine which extractors work on particular data sites; (e.g., building set of parsers on best-in-class Python library; making openly available).

Open source: Materials IO, see Github under MDF organization

- Parsers takes input file and parses out information which is registered in our search index, which users can query and perform more in-depth analysis using only a few lines of code.
- Shown by Logan Ward: built force-field potentials from different datasets that he had indexed in MDF; now can cross-index easily through few lines of code.
- After making data open, making data easy to find is critical second step.

MDF – Enabling Flexible Data Publication

How does data get into the ecosystem?

- Web Interface (holds over 30 TB of data and scaling quickly) or Python script (automate processes)

MDF – Curating Datasets

Many capabilities. Asynchronous process temporarily halts indexing flow. Allows users from defined group to approve a submission (e.g., can require metadata fields, prescribed licenses)

Models & Codes: An Emerging Data Infrastructure Ecosystem (Diagram)

DL Hub – Data and Learning Hub for Science – collect, publish and categorize models and processing logic from various disciplines. Also, building capabilities to serve those models via a cloud hosted service to make more user friendly.

DL Hub- A Data and Learning Hub for Science (Examples)

User provides trained model → DLHub creates unique endpoint for user (makes predictions against the model). Examples: Exascale cancer research, x-ray science, energy storage, tomography

Bringing it together – train models to predict bandgap of material, given an input image.
Query MDF for input data, which is sent to DLHub to code images and obtain the predictions.
Same infrastructure at different scales is applicable to some of the automation of metadata extraction.

Select DLHub Use Cases

Model-in-the-Loop Science: Many things headed this way: models (DL hub, etc) will be driving science.
What is the next simulation or experiment to be completed at the beam line?

- Starting to ingest some potential models

- Interest in community and model benchmarking: can see automatically applying models to data set; see which models are best for predicting crystal structures, etc.

DLHub Architecture and Performance
https://arxiv.org/abs/1811.11213: Discussed the Infrastructure built; performed scale testing; investing caching, batching

Glenn Lockwood - ***Simplifying Data Management Through Storage System Design***
Issue and challenges around data storage:
Managing the complexity of storage system through better system hardware and software – need to simplify user interfaces.

NERSC Mission: support broad range of sciences funded by DOE Office of Science; i.e., anyone researching that requires supercomputing services at moderate to large scale; but experimental and observational facilities generating data now need supercomputing resources to process data. Currently supporting large scale data analysis from experiments as well as traditional, computational sciences.

NERSC Mission
Dual mission: Large scale simulation and Large scale data analysis
Internal Dual mission:

- supporting scientists by providing productive, highly customized software programming environments and advanced app/and performance expertise for users
- DOE ASCR Mission: to deploy advanced HPC technologies – equip computational scientific community with hardware that will be commonplace in 5 -10 years.

Storage architecture is getting complicated
Not cost effective to deploy fastest possible storage for all. Results in stratification of storage hierarchy in HPC in past 5-10 years.

Storage Pyramid – in production today (CORI). Top to Bottom:

- Burst buffer Fastest storage (Flash-based storage)
- Scratch Tier: large capacity, disk based storage system
- Project tier/Community file system tier: higher capacity tier with more data resilience; not purged
- Archive Tier: tape-based, cheap and high capacity, slow storage with tiers in between for users to move and store data in most economical location in relation to performance requirements (e.g., longer wait to access data from archive)

Addressing complexity of data management : Levels to address complexity
Outermost layers

- Top layer: change user applications: work with code teams and users to optimize their data patterns for underlying storage system; however, portability challenges
- Develop better middleware and tools: libraries and methods that applications can interact with, so user can write single source code and run at different facilities, using middleware tools/abstractions to hide these differences/characteristics from application
- Optimization: by delivering better/smarter system software- make storage system more agile/responsive to users' intent (e.g., APIs for users to tag their metadata)
- Simpler storage system hardware (e.g., eliminating tiers in storage hierarchy to extent possible)

Simplifying storage system architecture

- 2020: Deploy Perlmutter system that will eliminate disk-based scratch tier and have same capacity and performance of burst buffer(all flash); users not have to choose between 2 when working with hot data.
- 2025: Requires smarter storage system software to manage/hide complexity of deciding data placement for users through abstractions.

Towards intelligent data management systems

- need better understanding how different storage tier is used
- tools to understand hardware is complex because each component is complicated by system mobile software monitoring.

State of the art in understanding I/O today

Old way: storage expert for monitoring, not sustainable as storage systems become more complex.

- Total knowledge of I/O (TOKIO) project: provides abstraction on top of monitoring data
- Communicates with monitoring tools running on storage tier
- Answer questions: Why is I/O slow? What causes unpredictable I/O performance when users want to read or write their data? Can see all components of I/O system and if do in aggregate, can be informed on what to optimize for future systems. (e.g., contention with other users)
- Designed to work at all data centers and across data centers. Started examining how data moves across data centers; See how data is moving into and outside of storage and compute systems per site and amount of data moving between sites that ESnet touches (track Globus transfers, big picture of hottest data paths, and storage system usage – and optimize systems accordingly)
- Vision: Develop intelligent data management system that automatically adapts to workload

Discussion

- Experience with debugging pathological behavior of user activity in storage? When something acting poorly in storage system, TOKIO gives insight of everything else that is going wrong in system. If part of system is dead, key is better resilience.
- Security concerns:
    - SDSC: part of layering of services in storage on top of intelligent infrastructure, detect security issues.
    - MDF: most of data and models are open, building on services (Globus Auf).
    - NERSC: allow users to discover what security and integrity storage system provides. Rely on middleware and applications to use that information accordingly.
- DL Hub – centralized service where calls come in, compute distributed among execution sites Implementing model to pull down container and run locally. Sustainability (e.g., set up other sites)
- Rapid growth in volume of data from simulations and experimental facilities: how aid science community in discarding data from data streams to enable effective storage?
    - NERSC: trying to provide foundational tools to enable communities to make their decisions – not qualified to do it for the communities (data management policy is to delete data after 30 days)
    - MDF: provide automated tools to index data to help inform users; decision
    - SDSC: data triage is an issue. One approach – allocate imagery and leave to users to decide

Data Life Cycle Series Planning
Outcomes/product discussion: Start considering what to do with material presented (e.g., high level summaries; include publishing) or in depth workshops – not there yet.
Workflow is a unifying topic. Informing agencies and community to assist in directing resources.

Topics:
Data Analysis
Tools- Jupyter notebooks; hub; binder hub effort (June)

- Jupyter project (Rich will ask around; possibly Dan Petravick: architect –main implementer of the DES (involves Jupyter Hub, etc)
- Open on Demand (David Hudak, Ohio Supercomputer center)
- Binder Hub project (need POC)
- Dan Gunter (LBL)- Uncertain, can ask around

Reproducibility: Victoria Stodden and Dan Katz- Ben Blaizik working with them to incorporate DL Hub with their tools

Data Use/re-use

Publication of code and data: Lorena Barba GWU and Dan Katz (Journal of Open Source Software)

Data Triage: July 3 meeting

- In situ analysis, energy issues, business processes
- Discard data – data curation speaker (librarians)
- Can ask Fran Berman (RPI) who can't make June meeting– developed working groups, interested in output

Large Scale Networking IWG (LSN) Deliverables
Containerization and DevOps Reports: Brief reports to be derived from last year's speaker series and will be delivered to LSN. Dhruva Chakravorty (TAM) is doing the containerization report to be submitted to MAGIC by June 7. Will develop a road map – either build on it or go from there.

Roundtable/Events
None to report.

Next meeting
June 5 (12 pm ET)