

Interplay of I/O, Compute and Storage

Convener Graham Heyes (Interplay of I/O, Compute and Storage)

Editors

Questions

Questions to guide Discussion Rounds and Panel Discussions:

- ☐ Access to data and computing resources are vital to advancing the physics program. From your experience of the current state of the art:
 - ☐ What are the greatest obstacles that you face?
 - ☐ What do you like to see improved?
 - ☐ In your ideal world, what system would you design?
- ☐ In allocating resources to move towards your ideal:
 - ☐ What is the first thing that should be fixed?
 - ☐ What compromises would you make in the real world?
- ☐ The interface between the user and the data or computing resources molds both how we think about them and how useful they are.
 - ☐ What constraints does the current user experience impose?
 - ☐ How would you like to interact with data and computing?
- ☐ What do you see as the ideal mix of involvement from NP and groups outside the NP community? How will physicist-developers be best positioned to contribute?

Notes

- ☐ (Author) Your note for the report

Report

Draft for Workshop Report:

Group 1 discussion

Group 2 discussion

Group 3 discussion

Access to data, obstacles

Wes:

- Examples others are doing
 - Google Earth Engine: multi-PB collection of satellite and remote sensed imagery, users write code on GEE that processes data, makes use of Google cloud infrastructure. Interesting dev/coding/execution environment. Move analysis to data.
 - Earth System Grid Federation: collection of climate model output, staged/replicas across the world.
- Wand time:
 - Authentication/authorization: program-wide visibility/solution to enable location transparent access to resources (compute/data)
- What are common use cases, and what are the knobs/parameters to solutions for those common use cases? E.g.:
 - An experiment will run and generate a lot of data in a short amount of time. It is important to have storage for that data. The storage could be local, or could be remote but would require a significant amount of WAN bandwidth.
 - A collection of data needs to be accessible to a potentially worldwide community, but in a controlled/authenticated way.