

MINERvA data management review

E. Snider
Fermilab

April 8, 2014

The problem

- MINERvA has been working to change the data handling model
 - Currently using project disk to host production input and output data
 - Mounted via nfs on Fermigrid
 - I/O contention is a serious bottleneck
 - Limited capacity
 - Installed volume will not scale with data
 - Adding disk does not generally increase total throughput
 - Will have an increasingly negative impact on physics output if not changed
 - Have been working to integrate SAM data handling system + IFDH tools
 - Data stored on tape, delivered via caching system
 - Scalable I/O
 - Tape volume more easily scales with data
 - Grid compliant transport
 - Avoids mounting project disk
 - Also need high-throughput staging areas for data not going to tape
 - Scratch dCache

The problem

- ...but, the transition has been difficult
 - Have had difficulty establishing reliable, robust operation in testing
 - Encountered performance issues due to sub-optimal usage patterns
 - Required functionality sometimes missing
 - Often in the position of debugging a newly deployed services
 - Most of the scripts to use the new services have yet to go into production
- The backdrop
 - Increasing demand for storage to complete short term physics goals
 - Driven by large number of associated high-throughput computing tasks
- The purpose of this review is find a remedy to this situation

The charge

The MINERvA spokes and SCD senior management would like to understand the resources needed to upgrade the MINERvA data management and workflow systems to effectively use the SCD supported services with the intent to provide more robust and reliable operations. The focus should be on production and user analysis jobs for both data and simulation. In particular the committee should review:

- Requirements for data handling within the experiment.
- Current MINERvA data management and workflow systems. This includes file characteristics, account management, services and technologies currently deployed.
- Current operations model including manpower availability from MINERvA and SCD.

The charge

The MINERvA spokes and SCD senior management would like to understand the resources needed to upgrade the MINERvA data management and workflow systems to effectively use the SCD supported services with the intent to provide more robust and reliable operations. The focus should be on production and user analysis jobs for both data and simulation. In particular the committee should review:

- Requirements for data handling within the experiment.
- Current MINERvA data management and workflow systems. This includes file characteristics, account management, services and technologies currently deployed.
- Current operations model including manpower availability from MINERvA and SCD.

The charge

The overall system needs dramatic improvements to meet the production processing that MINERvA needs to meet its milestones, e.g., results for August neutrino conferences. The committee is charged with producing a report that describes the feasibility of a MINERvA upgrade given the resources available from both MINERvA and SCD. The committee should take into account MINERvA imposed constraints. If a major upgrade is deemed unfeasible, then alternatives for more incremental improvements should be considered.

We would like the report delivered by April 14th.

The charge

The overall system needs dramatic improvements to meet the production processing that MINERvA needs to meet its milestones, e.g., results for August neutrino conferences. The committee is charged with producing a report that describes the feasibility of a MINERvA upgrade given the resources available from both MINERvA and SCD. The committee should take into account MINERvA imposed constraints. If a major upgrade is deemed unfeasible, then alternatives for more incremental improvements should be considered.

We would like the report delivered by April 14th.

We should think broadly about how to address issues, including changes to scripts, workflows, architectures, documentation

Structure of the review

- This morning, early afternoon
 - Introduce physics goals of the experiment, the targets, deadlines, etc
 - Provide an overview of the system, requirements, issues
 - Presentations on the major workflows, storage and data transfers
- Later this afternoon
 - Go into details for some use cases to see how things are done at low level
 - Discuss the specific services and products used by the experiment

In all cases, there is a lot of time scheduled for discussion.

- Tomorrow morning
 - Outline a plan of action, make writing assignments.
 - Hope to be done by lunchtime, but have FCC1W for the afternoon if needed

One constraint for today: MIPP seminar at 4:00 pm. MINERvA collaborators need to attend this

Introductions

- “The committee”

- Erica (chair)
- From MINERvA: Mousumi Datta, Laura Fields, Mike Kirby, Jaewon Park
Cheryl Patrick, Gabe Perdue, Phil Rodrigues, Heidi Schellman
- SCD: Mike Diesburg, Stu Fuess, Adam Lyon, Marco Mambelli, Marc Mengel,
Andrew Norman, Saba Sehrish
- Others I missed or who are in attendance

- Presenters

- Debbie Harris
- Heidi Schellman
- Cheryl Patrick
- Emily Maher
- Jaewon Park
- Mousumi Datta
- Phil Rodrigues