

Data Management Workshop - Winter 2012

With Data Management as part of the strategic plan for Lab-wide computing, a workshop would be a good stepping stone to bringing the data management community together and identifying common tools and directions to move in. Data management is here loosely defined as moving and storing files/data, cataloguing and accessing them.

In broad brush:

- get more detail of existing solutions at SLAC: Fermi, LCLS, BABAR...
- survey solutions in the community
- interview projects in need of DM solutions - eg cosmology computing; maybe someone like Todd Martinez; SSRL, LCLS users etc.
- collect the people from the first bullet to discuss common solutions that could be recommended to pursue.
- think about evolution of current DM solutions to a more common base over some timeframe - and what can be done jointly with other projects

Time frame: Feb 2012

Duration: 2 days

Tony Johnson has listed these items (with the odd edit) as potentially of interest:

- encompass disk and tape storage technologies and file-systems (lustre, xrootd etc).
- The aspects of DAQ concerned with pushing those bits out
- Transfer from DAQ to long term storage
- The data access patterns, analysis and reconstruction bandwidth
- Data storage formats (root, fits, hdf5, databases, ...). Optimization of data storage formats for SSD
- Need for scaling storage to support massive parallelization (multiple simultaneous write streams, massively parallel read streams)
- Data Catalogs, eg iRODS etc
- Offsite access to data, e.g. for Fermi download manager, or Globus Online
- Skimming tools, e.g. for Fermi skimmer, astro server
- Long-term data archive, "Cloud" storage