



### SAMGrid

SAMGrid is a large scale distributed system to deliver petabyte scale datasets for processing at the CDF and DØ experiments.

It does this by providing the following services in a single unified framework:

- **Managing File Storage**
  - Files are housed on tape and cached on disks around the world
- **Managing File Delivery**
  - Get files from tape or cache
  - Provide location transparency
  - Manage your local cache
  - Use a variety of file transfer mechanisms
- **Managing File Metadata**
  - The SAM database allows metadata based file retrieval
  - User does not need to know a filename
- **Providing Analysis Bookkeeping**
  - What files you ran over, with which application.
- **Managing Jobs**
  - Choose an execution site, deliver job and data to it and store output

**What is a Dataset?**  
**What will theorists & experimentalists want to know 5 years from now?**

- Where is the dataset?
- How much real data does it correspond to?
- Where can I find more information?

- What tools were used?
- What physics parameters were chosen?
- What kinematic cuts were applied?

- What is the parton shower?
- What underlying event model?
- Are photons radiated?

**The Choices Make a Difference In:**  
**Determining Theory Errors**  
**Combining Different Data**  
**Finding New Physics**

More information on  
 Tevatron experience with  
 generating datasets,  
 improving Monte Carlo tools,  
 & applying them to  
**REAL data**

**mc4**  
**run2** <http://cepa.fnal.gov/patriot/mc4run2>