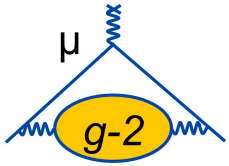


Fermilab

Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

Job Submission and Workflow

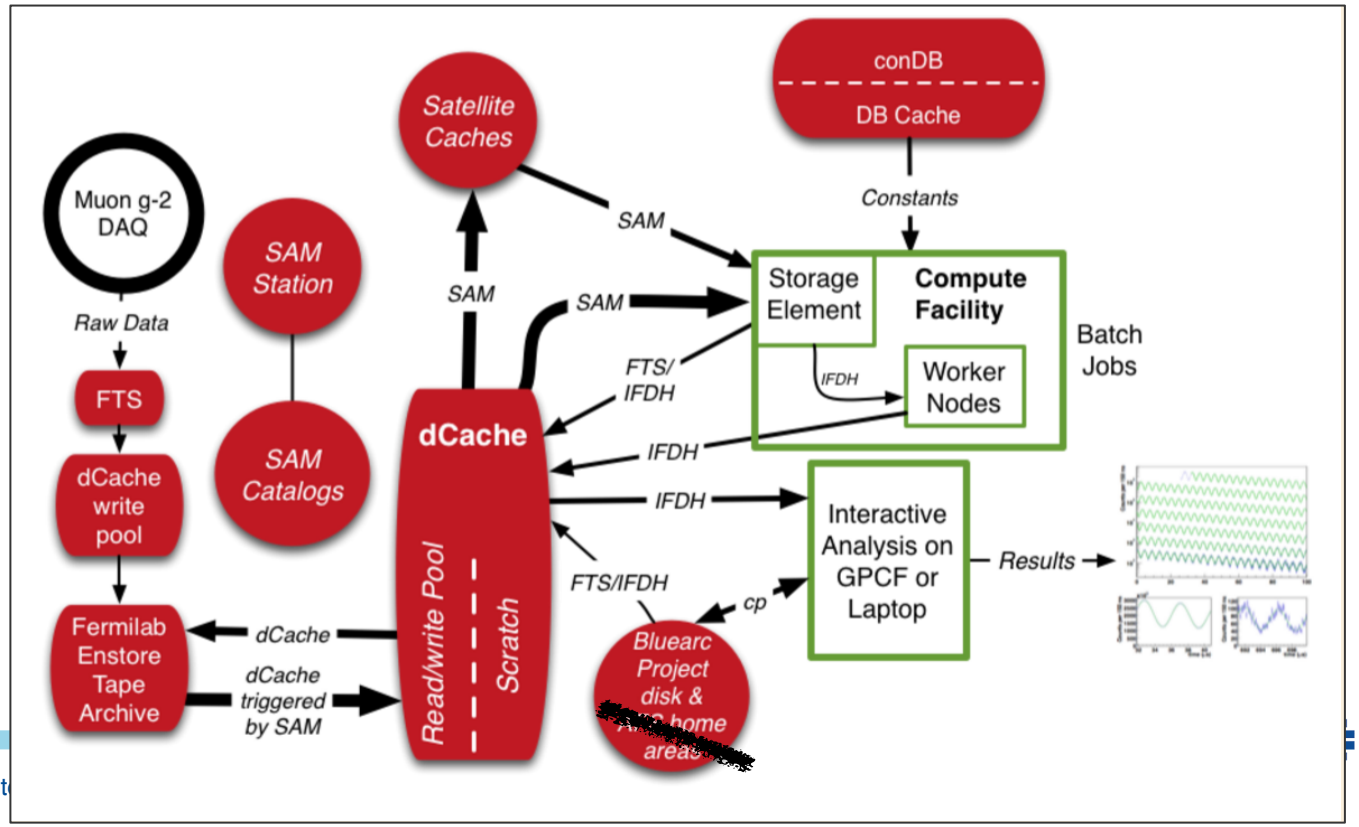
Tammy Walton

Computing Readiness Review

Nov 7-8, 2016

Overview

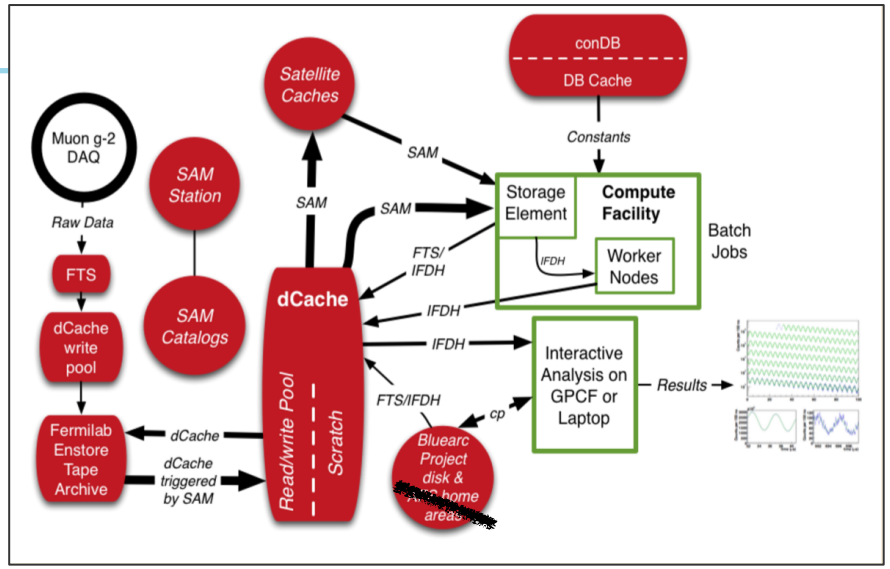
- Adam stressed that we will NOT reinvent the wheel.
- We are using Fermilab resources and services for job submission.
- This talk covers :
  - resources/services in operations
  - quantify how much experience we have using the services
  - directions and plans to incorporate additional resources

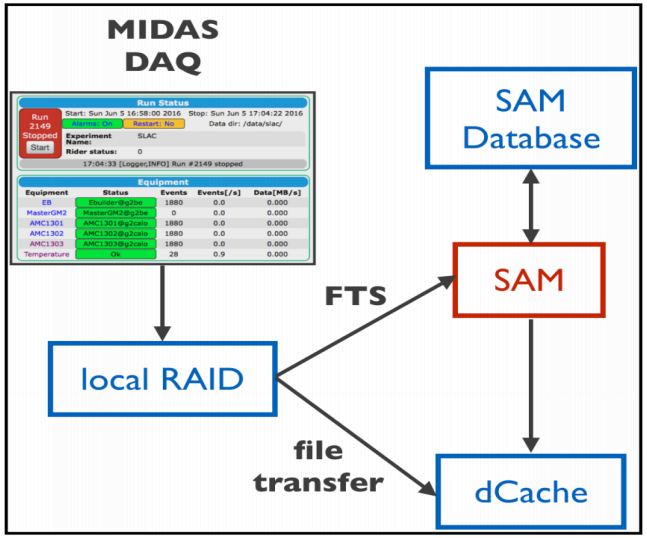


Fermilab

Status

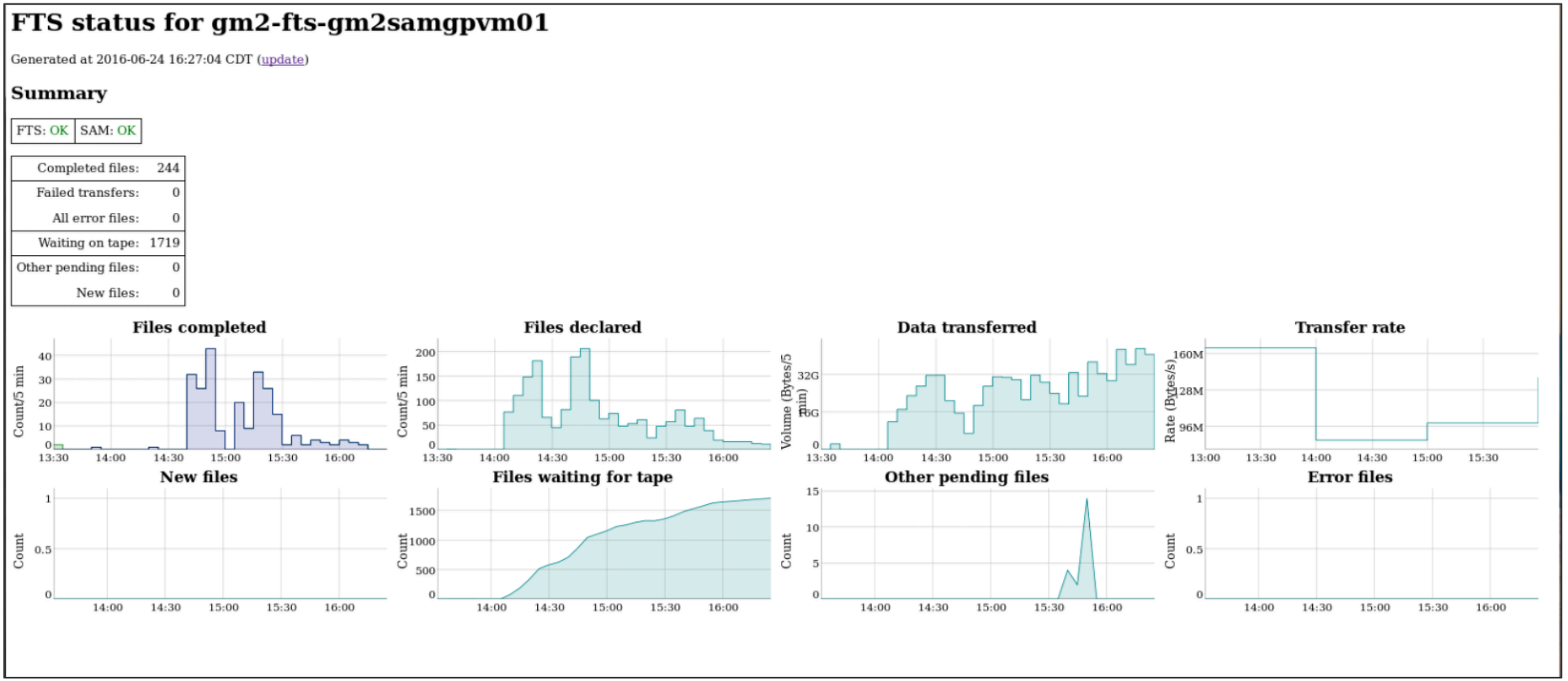
- SLAC test-beam used the FTS for storing data.
- FTS system copies the MIDAS raw files to the Fermilab dCache and are written to tape and cataloged in SAM.
- We need more practice using the services. In the very near future, we will gain more experience using both the calorimeter and tracking QA test-bench data.





Status : Using File Transfer Services and SAM

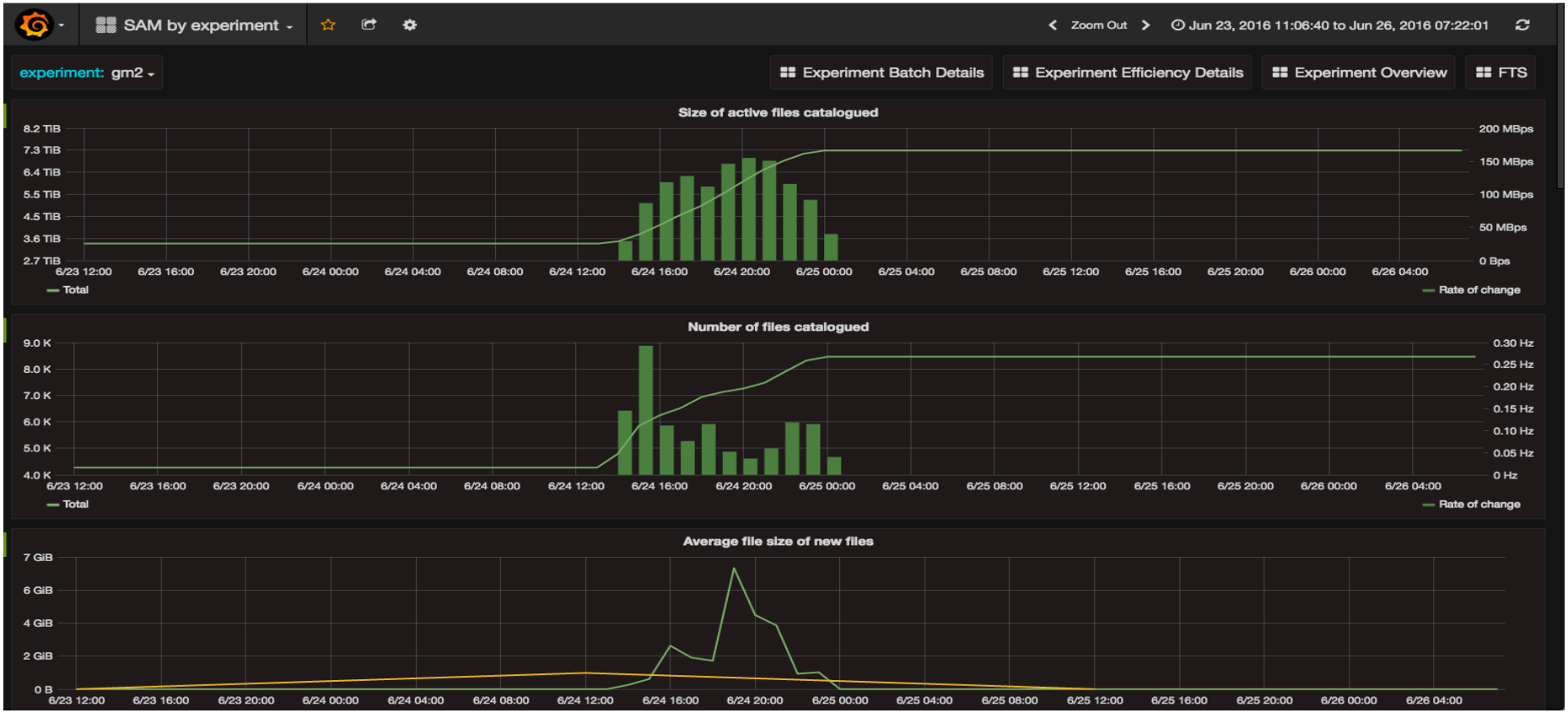
Success of using and setting up the server (SLAC test-beam data)



Fermilab

Status: Using File Transfer Services and SAM

Success of accessing the data from SAM and processing the data using Jobsub (SLAC test-beam data)



Fermilab

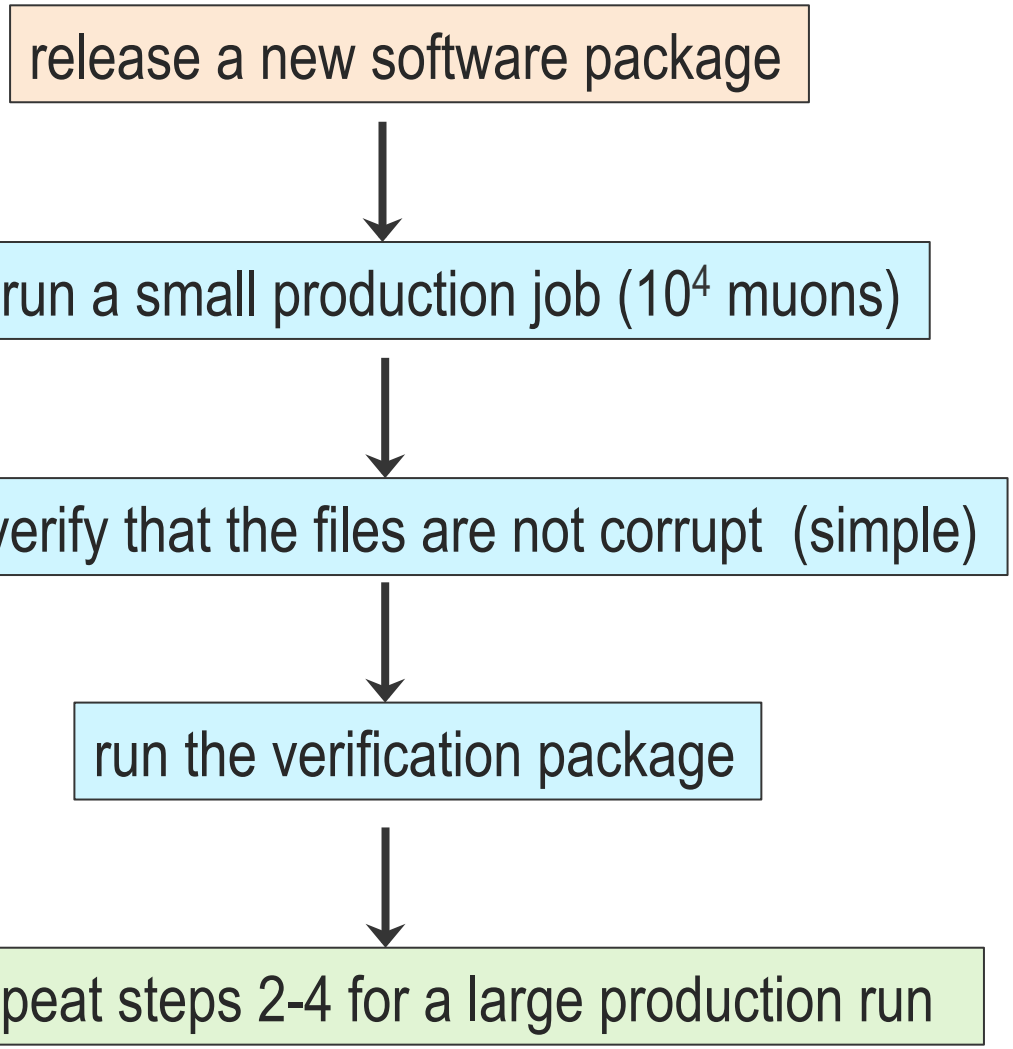
Status: Simulation (More on using Jobsub)

- We have the most experience using jobsub to process the simulated data.
  - Mock Data Challenge
    - Used the data for understanding the simulation and geometry (see Renee's and Leah's talks).
    - Produced around 50 million muon events (spanned 4 software releases)
  - Things to come with the mock data challenge production
    - Goal is 10% of the real data (10^{11} muons) using a gun that tracks the muons around the ring.
      - Study calorimeter and tracking calibration, digitalization, and reconstruction
      - Develop the physics analysis tools
      - Validate the framework structure
      - Practice data handling, management, and storage
      - More realistic (physics) studies of the muons (eg, spin tracking, muon losses, etc)

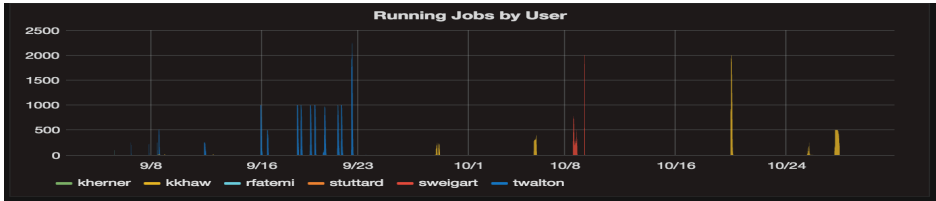
Status: Simulation (More on using Jobsub)

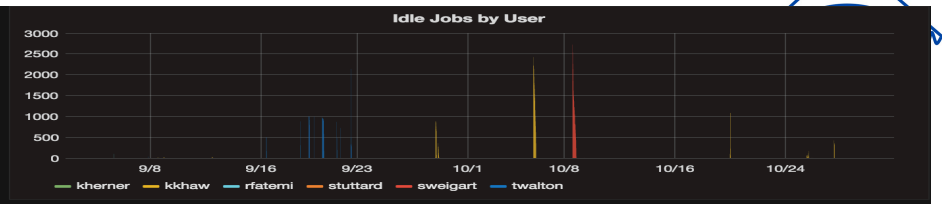
- The scripts are simple and standard for running jobs on the Fermilab grid.
  - Approved by OPOS
- We haven't had the need to use offsite resources.
  - Take advantage of Mu2e expertise.
- We are using bash shell scripts.
  - In the very near future, the scripts will go through a python upgrade phase.
- Running jobs on the Fermilab grid.
  - Use ifdh to copy the simulated and log files to the user /pnfs/GM2/scratch area.
  - We are not using a production user.
  - Not copying files to tape or adding datasets to SAM.

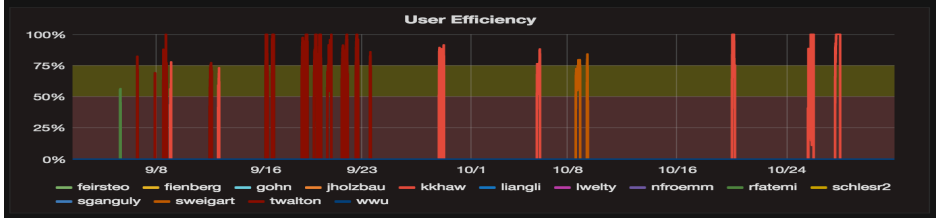
Status: Current Workflow for Processing

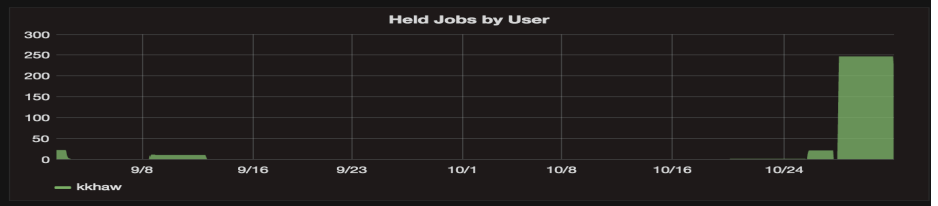


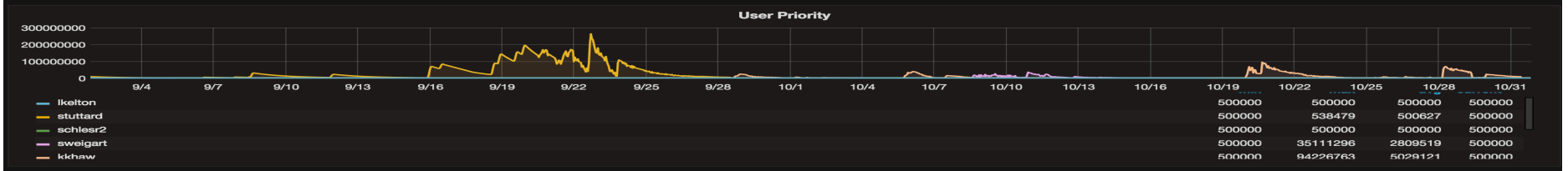
Status : Running Jobs On the Grid



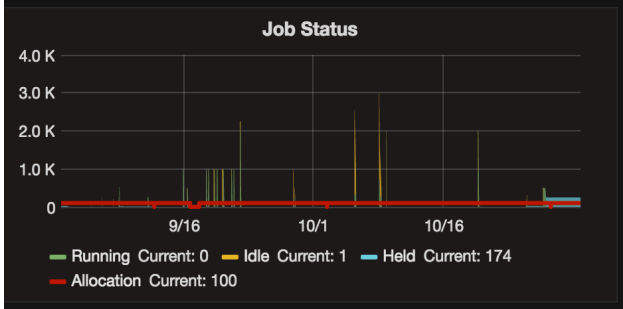


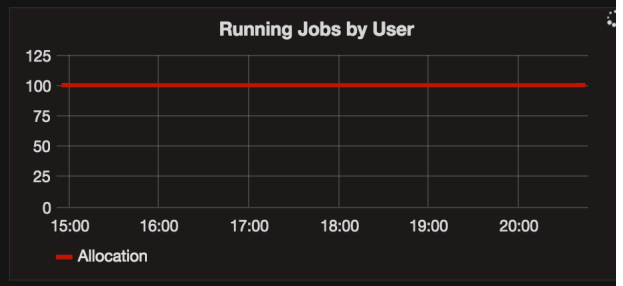




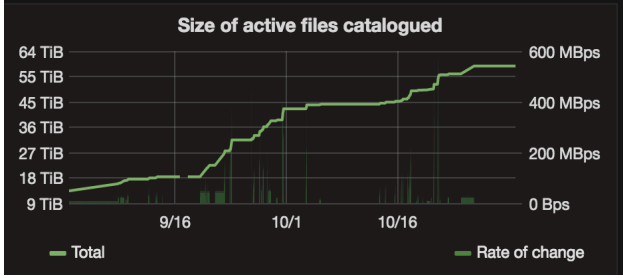


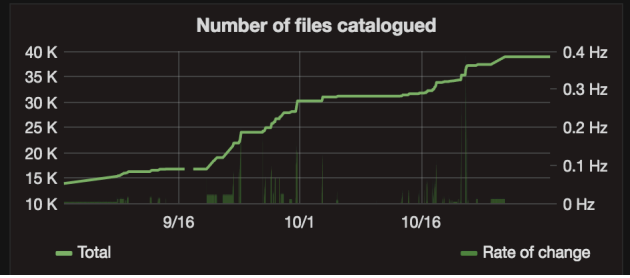
BATCH

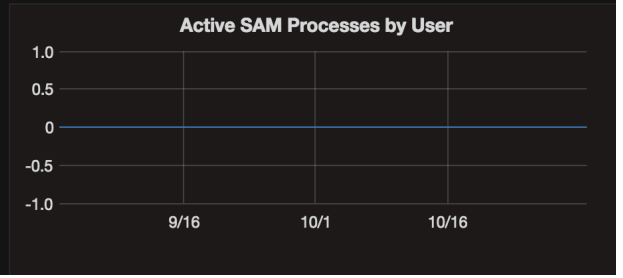




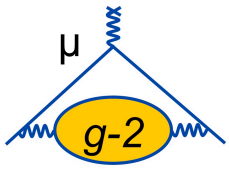
SAM







Status: Checking the Quality of the Files



- A simple python script (findMissingFiles.py) which searches for missing files and checks if the file completed successfully.

- Below is the report from the QA script.

```
Searching for missing data files



Missing files= 0/93
        Number of files that abruptly stopped = 0
Good files= 93/93
Time Summary (minutes) -- Average File CPU= 177.3736 and Real= 163.4864
Average Events per File= 5000
Average Size per File (MBytes)= 235.05
```

Status: Using *art* SAM services in the Production Chain

- We have not taken full advantage of the *art*-SAM services.
  - Do not have an official production team.

- We have started the initial steps.
  - Incorporate the FileCatalogMetadata service in the production fhicl file.
  - Plugin to write experimental metadata.
    - We have had many discussions regarding the experimental metadata.
    - At the moment, ***fhicl file name*** is the only additional data that we need to store.
    - However, we have not decided whether to store the name or the contents of the file.

```
services: {

  FileCatalogMetadata : {
      applicationFamily  : "gm2ringsim_offline"
      applicationVersion : "v7_03_00"
      fileType           : "mc"
  }
```

```
  module_type  : RootOutput
  fileName     : "gm2ringsim_muon_gas_gun.root"
  dataTier     : "truth"

  // use plugin to defined generic metadata paramaters that are not part of sam services
  FCMDPlugins  : [ { plugin_type : Metadata
                     fhicl_name  : "gm2ringsim_mdc0.fcl"
                   } ]
```

Status : Documentation

production summary wiki

Using redmine for documentation.

Current layout is not ideal.

Goal of the production team is to reorganize the web page.

g-2 Production Summary

Mock and Simulation Data and Processing/Access

MDC0 Data

Files are located on the dcache area. Currently we are not moving files into the sam database.

- **v6_02_00**

Files are produced using the software release, where a bash script submitted 10 jobs every thirty minutes.
This production contained 1752 files for a total of 5,256,000 events.

Files Location	Events per Job	Notes
/pnfs/GM2/scratch/users/jholzbau/production_v6_02_00/production3	3000	
/pnfs/GM2/scratch/users/jholzbau/production_v6_02_00/production4	3000	
/pnfs/GM2/scratch/users/jholzbau/production_v6_02_00/production5	3000	
/pnfs/GM2/scratch/users/jholzbau/production_v6_02_00/production6	3000	

- **v6_03_00**

Files are produced using the software release.
Total number of events is 18,114,000

Files Location	Events per Job	Notes
/pnfs/GM2/scratch/users/jholzbau/production_v6_03_00/production0	3000	250 x 3000 = 750,000
/pnfs/GM2/scratch/users/jholzbau/production_v6_03_00/production1	3000	2998 x 3000 = 8,994,000
/pnfs/GM2/scratch/users/jholzbau/production_v6_03_00/production2	3000	missing most of the files, issue with kinit. 20x3000=60,000
/pnfs/GM2/scratch/users/jholzbau/production_v6_03_00/production3	3000	277 x 10 x 3000 = 8,310,000
/pnfs/GM2/scratch/users/jholzbau/production_v6_03_00/production4	3000	2999 x 3000 = 8,997,000

Files are produced using feature branches CoordSystemService (gm2ringsim) and strawTrackReconstruction (gm2geom) (Date 3/01/2016 version)
Total number of events is 7,365,000 (475*1000 + 460*5000 + 459*10000)
WARNING: These jobs have a bug in StrawArtRecord::momentum, which is now fixed (7/4/16, see simulation elog 304). You can use StrawArtRecord::px_local etc instead.

Files Location	Events per Job	Notes
/pnfs/GM2/scratch/users/twalton/v6_03_00_geom_test1/2016-03-01-16-07-32	1000	
/pnfs/GM2/scratch/users/twalton/v6_03_00_geom_test1/2016-03-01-21-57-03	5000	
/pnfs/GM2/scratch/users/twalton/v6_03_00_geom_test1/2016-03-02-00-24-22	10000	

Files are produced using feature branches CoordSystemService (gm2ringsim) and strawTrackReconstruction (gm2geom) (Date 3/22/2016 version)
Total number of events is 7,600,000 (100*1000 + 500*5000 + 500*10000)
WARNING: These jobs have a bug in StrawArtRecord::momentum, which is now fixed (7/4/16, see simulation elog 304). You can use StrawArtRecord::px_local etc instead.

Files Location	Events per Job	Notes
/pnfs/GM2/scratch/users/twalton/v6_03_00_geom_test2/2016-03-22-13-18-04	1000	
/pnfs/GM2/scratch/users/twalton/v6_03_00_geom_test2/2016-03-22-16-00-33	1000	
/pnfs/GM2/scratch/users/twalton/v6_03_00_geom_test2/2016-03-22-16-12-22	5000	

Plans

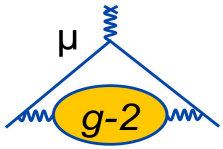
- Include the calorimeter, tracking, and auxiliary detectors into the production.
  - Design a layout for running various detector packages.
  - Our digitalization, calibration, and reconstruction have a straight-forward work flow.

- Learned from NOvA and MicroBooNE
  - Examples of using FTS system
  - Convert production bash shell scripts to python scripts

- Incorporate SAM services into the simulation production flow.
  - Use NOvA model as a template for deciding which information should go into the metadata.
  - Determine the experimental metadata that are needed.
  - Note, Adam has used the full SAM services and we are users of SAM4Users.

- Investigate using POMS (Production Operation Management Service) for submitting, managing, and tracking the production.

Summary



- Due to the status of the simulation, we have NOT need to run large and continuous production jobs.

- Therefore, we have not yet form an official g-2 production team.
  - Now, it's time to bring the production efforts up to par and ensure we are ready for data taking.
  - Will ask for help from the g-2 collaboration at our December Collaboration Meeting.
  - Since, we are using Fermilab resources and services, the data management will be ready long before data taking.