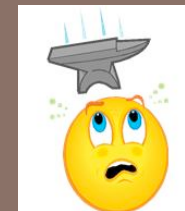


MINERVA @ FIFE



6/4/13

Heidi Schellman for the Collaboration

# History

2

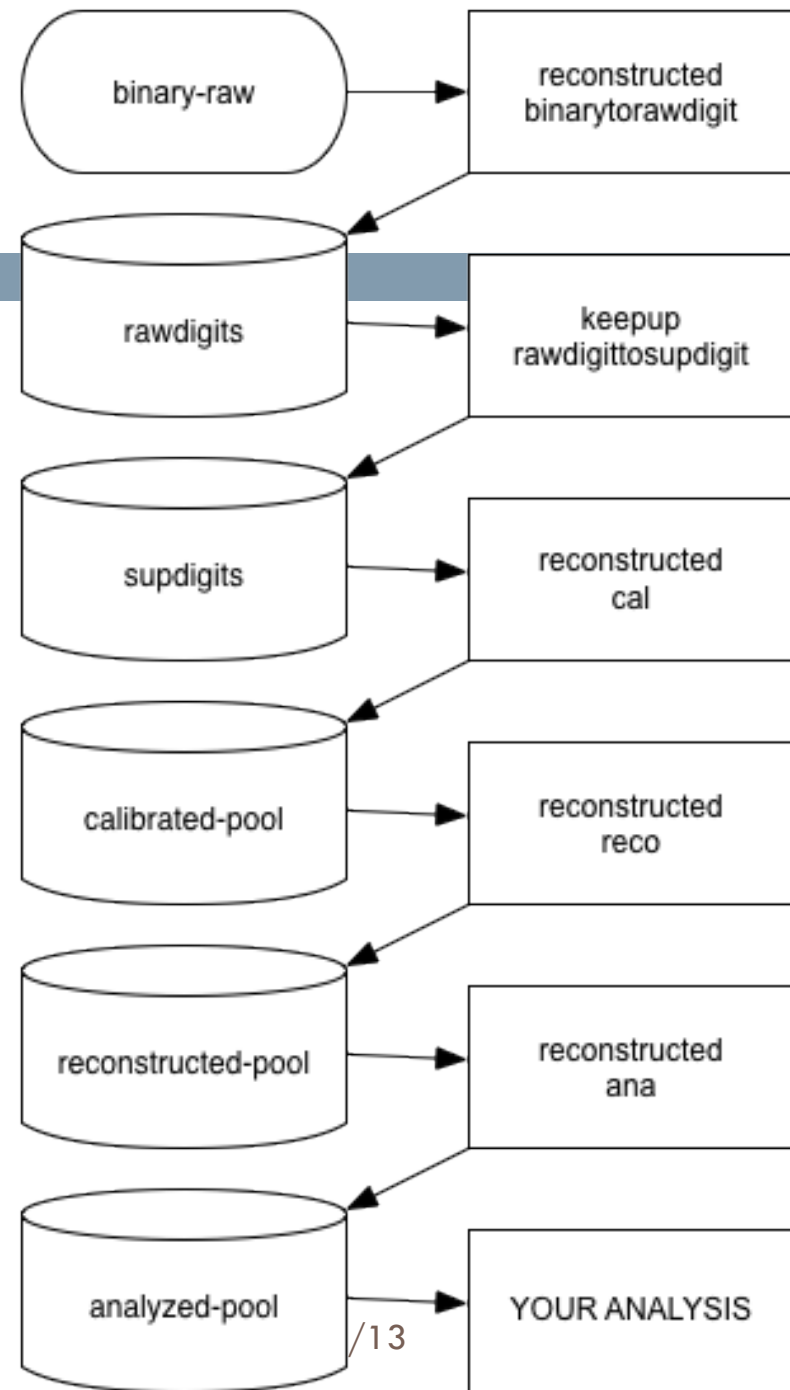
- First big IF experiment approved in 2004
- 2006ish - No Art – decide to use Gaudi like T2K/LHCb
  - ▣ POOL file format, COOL DB
- Windows DAQ → Linux in 2009-2010
- Get 5 if01-05 machines for analysis
- Minerva started taking data in Fall 2009
  - ▣ COOL DB doesn't work
  - ▣ SAM for file catalog
- Big data run 2010-2012 → data ->bluearc
- First major paper in Spring 2013

**We are often the first BIG user of new products.**

# Data size

3

- ❑ 15,000,000 spills
- ❑ 27K **150 MB** files/  
datatier/version
- ❑ 4 TB/datatier/version
- ❑ MC is up to 10 times  
larger
- ❑ All on bluearc



# Databases



4

- Conditions and Calibrations
  - ▣ COOL → unCOOL!!! (Thanks!)
- Beam database (Thanks!)
- MINOS database
  - ▣ thanks to our MINOS friends and replication
- SAM for file catalog (Thanks!)
  - ▣ Run/subrun makes life more exciting!
  
- Note – we know our databases by name – not machine. Would be nice to have a standard mapping schema so we can communicate with sysadmins, dbas, users.

# SAM

5

- We have been using old-sam since we started data taking in 2009.
- Had to tar up files, merge metadata, before writing to enstore.
- Automatic checks of the tar files run daily
- Not useful for analysis as the merged files are not what you really want
- Only used by experts
- Move to SFA/samweb in progress

6/3/13

# Samweb

6

- We are migrating our sam scripts to samweb
- Generally finding samweb to be easier to use
  - ▣ Occasionally find missing functions, Robert adds to the API
  - ▣ More documentation for API would be useful
  - ▣ One issue – duplication of fields that correspond to the full datatier chain – better access methods?

# Small File Archive

7

- Our files are small - needed to merge to store until SFA came along. Makes catalog very hard to use.
- Getting FTS set up to use samweb and SFA to store our files
- Preliminary tests worked
- Need to do a big migration into SFA of all data from the first run -- ready for 300 TB?
- MC needs better metadata << our problem

# Grid

8

- Use fermigrid
  - ▣ Almost too luxurious – hard to migrate offsite.

Fermigrid



Not Fermigrid





# jobsub

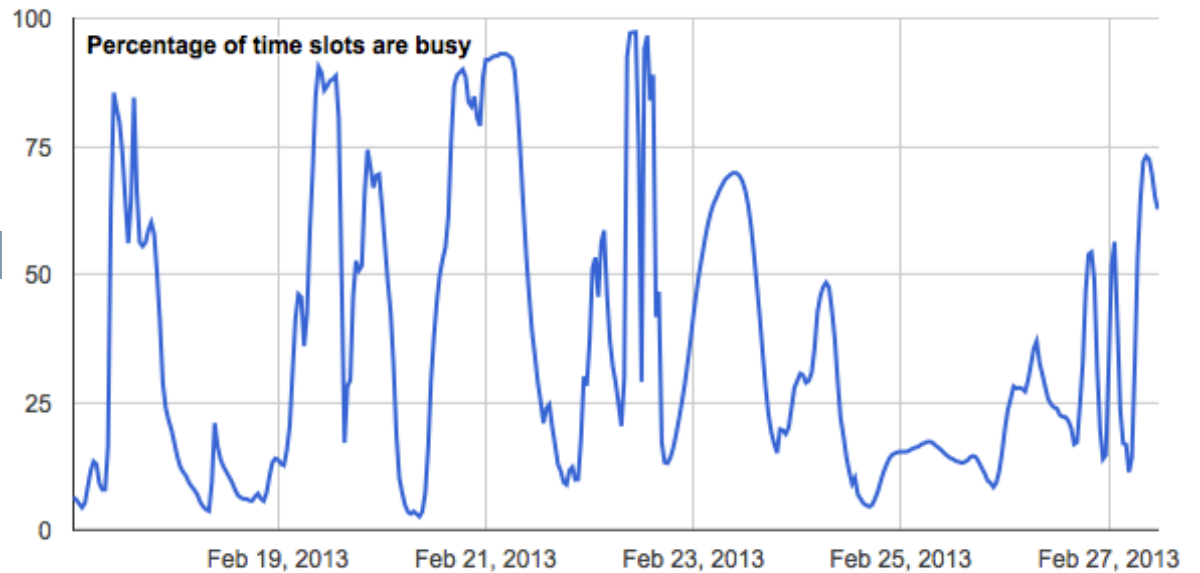
9

```
cat minerva_jobsub
#!/bin/sh
jobsub $@
```

- ❑ Wrapped in minerva scripts
- ❑ Lots of joint development
- ❑ It does a lot! (Thanks!)
- ❑ Working on more samweb integration

# ifdh

10



- ❑ In Feb-April
- ❑ Move to ifdh/gridftp to get correct file ownerships
- ❑ Substantial problems with overload of bluearc
- ❑ cpn locks lock up
- ❑ Lots of work on transfer blocksize, lock strategy, rogue users/files (5TB file anyone?)
- ❑ Now running a lot more smoothly (Thanks!)

6/3/13

# Analysis

11

- 20-50 active users
- ROOT based “anatypes”
- Currently IO bound
- Need to work on solutions for fast data analysis.
  - ▣ Grid is often not the best solution as CPU is wasted waiting for IO
  - ▣ But interactive VM’s with bluearc also not ideal.

# Other things we use

12

## □ Other things we use

- Docdb
- cvs
- ROOT
- GEANT
- GENIE
- Networks
- Lcg/Gaudi
- ECL

## □ Want to use

- cvmfs
- git
- Surprise us!

# Summary

13

## □ Lesson learned

- Need to engage experiment/CD early in planning
- Legacy decisions can be very awkward



## □ Major needs

- Get the storage/caching solutions for files on tape into production
- Work to run simulations and analysis offsite
- Better analysis solutions