

MINERvA Computing Infrastructure

Heidi Schellman

Northwestern University

MINERvA specifics

- Purpose

Measure neutrino cross sections in the NUMI beamline

- Number of users – 50-100
 - Remote 10-20 growing to ~40
 - Using Fermilab facilities 20-30 growing to 40-60

Experiment schedule

	Pre-2009	2009	2010	2011	2012	2013	2014
Planning	X						
Construction	X	X					
Commissioning		X	X				
Data taking		Test	X	X	X	X	X
Data analysis		X	X	X	X	X	X

Data

- How many events (snarls)/year?
 - Test beam ~5,000,000 (2009/2010)
 - Pedestal and calibration - ~15,000,000
 - Normal data ~15,000,000
 - Normal data after quality filtering? same
- How large is each event (snarl)?
 - Zero suppressed - 200 kB
 - Non zero-suppressed - 1000 kB in 2010, 200kB in 2009
 - Reconstructed – 1000 kB
 - Simulated - 1000 kB
 - Data summary? 10 kB

Central FNAL systems

- CPU used (see table)
- Storage used (see table)
- Uses:
 - Code development
 - Reconstruction and data filtering
 - Calibration and alignment
 - MC Generation
 - User data analysis

Data flow

	Pre-2009	2009	2010	2011	2012	2013	2014
Raw Data, TB		6 TB	30 TB	3 TB	3 TB	3TB	3TB
Processed Data, TB		10 TB	10 TB	10 TB	10 TB	10 TB	10TB
User data, TB		1 TB	1 TB	1 TB	1 TB	1TB	1TB
Simulated data, TB		1 TB	2 TB	4 TB	8 TB	8 TB	8 TB

Please enter incremental quantities

CPU needs in CPU years

	Pre-2009	2009	2010	2011	2012	2013	2014
Running		x	x	x	x	x	x
Reconstruction		1	4	4	4	4	4
Calibration		.2	1	1	1	1	1
Skimming		.1	.1	.1	.1	.1	.1
Analysis		10	20	20	20	20	20
Simulation		10	40	40	40	40	40

Please use CPU-years on a current core.

e.g. # events * time per event in sec * 3×10^7 * reprocessing factor

Simulation will largely be done offsite.

Operating systems

- What OS is used? FSL 4.6
- Do all collaborators have to use the same one?
 - Strongly encouraged due to linkage to LCG and LHCb libraries.
- Move to FSL 5.2 anticipated

Data storage and tracking

- How do you catalog data?
 - SAM?
- How do you provide remote access to data?
 - AFS in past
 - SAM or other synching system in future

Remote systems

- How many remote institutions provide resources for your users/collaboration
 - Currently 4, Rochester, W&M, Hampton, CBPF + Guanajuato (planned)
- Do they have special systems for you or shared?
Shared
- What is done at remote institutions?
 - Code development (yes)
 - Reconstruction (not much)
 - MC generation (most of it will be done offsite)
 - User analysis

Data distribution to remote sites

- Where are data distributed
 - Most remote institutions by 2010
- What kind of data
 - Data summaries and subsamples
- How much data
 - ~ 4-5 TB/year
- How fast does it need to move
 - Not that fast given data rates
- What method is used? SAM or rsynch or copy...

Grid

- Do you use the Grid?
 - Testing phase now, anticipated for reconstruction, large scale MC generation.
- Do you use Grid tools such as Gridftp?
 - testing
- Do you use Gliden or some other tool?
 - Not yet
- Do you use the FNAL Grid exclusively or do you use more general grid resources?
 - Hope to use more general resources.

Databases

- Technology used
 - Currently MYSQL , exploring ORACLE, PostGres
- Size ~ 60 GB anticipated
- Access rate - low
- Are they replicated remotely? likely
- What is stored?
 - Data catalog (ORACLE)
 - Conditions
 - Short term monitoring
 - Hardware tracking
 - Calibration and Alignment constants

Conditions

- How are conditions and calibrations stored?
 - LCG COOL/CORAL packages with DB backend
- How are they accessed?
 - C++ code, python

Code management

- Code repository
 - CVS
- Build system
 - CMT

Standard packages

- What standard packages are used:
 - GEANT4
 - ROOT
 - GENIE
 - NEUGEN
 - LCG
 - CLHEP
 - PYTHIA