

NOvA Experiment Computing Infrastructure

Peter Shanahan – Fermilab

March 6, 2009

Experiment specifics

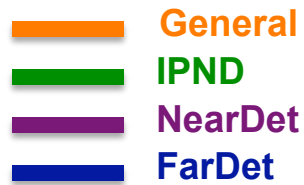
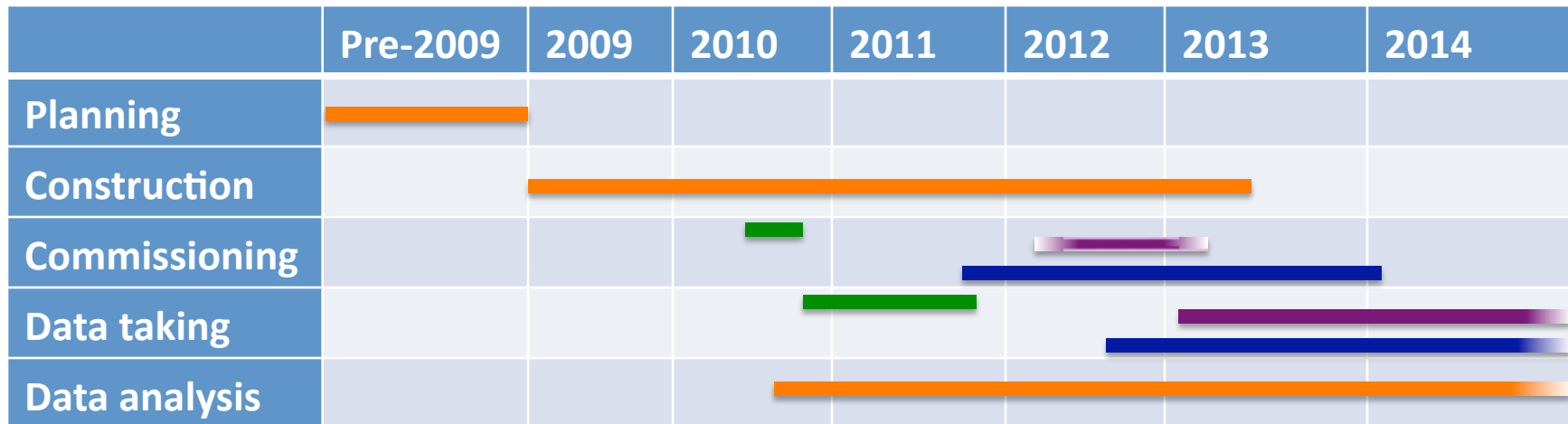
- Purpose

- Principle Physics goal is to search for $\nu_{\mu} \rightarrow \nu_e$ oscillation, and investigate corresponding physics
- Secondary goals include ν_{μ} disappearance, ν interaction physics, and supernovae sensitivity
- Two Detectors:
 - Far 810 km from FNAL with shallow overburden– beam events are negligible fraction of total rate. ~385 kChannels
 - Near at FNAL in NuMI tunnel, roughly equal beam and cosmic rate. ~16 kChannels
 - IPND – Integration Prototype Near Detector in MINOS Service Building, core of Near Detector

Experiment specifics

- Number of users
 - Remote
 - currently, about 10 active in offline software
 - Future: one expects ~50-80?
 - Using Fermilab facilities:
 - Currently: ~10
 - Future: ~50-80
 - Based on assumption that most users end up using FNAL facilities

NOvA Schedule



Note: "Commissioning" includes assembly

Data

- How many events/year?
 - Test beam: 800e6 30us gates at IPND (goal)
 - Pedestal and calibration: 28e6 (1/channel/day)
 - Normal data:
 - 3e9 (100 cosmic triggers/sec)
 - 20e6 (1 beam spill 2 detectors every 2 seconds 2/3 of the time)
- How large is each event?
 - Zero suppressed: ~20 kB (Far)
 - Calibration (threshold): ~8kb
 - Reconstructed: $n \times \text{Raw}$
 - Simulated: $O(2 \times \text{raw})$
 - Data summary?

Central FNAL systems

- CPU used to date
 - We have used $O(1000)$ hours of fnalu batch system 2-3 times per year for MC generation
 - Lightweight flxi04/05 for user work.
- Storage used to date
 - AFS: we've used $\sim 40 \times 8\text{GB}$ partitions
 - Pnfs: 4 TB
- Uses (currently):
 - MC Generation
 - Code development, reconstruction, User data analysis

Data flow

	Pre-2009	2009	2010	2011	2012	2013	2014
Raw Data, TB	0	0	1	1	1.5	20	20
Processed Data, TB	0	0	??				
User data, TB	0.1	0.1	??				
Simulated data, TB	4	4	10	??			

Work in progress.

How much FarDet Cosmic simulations will we need?

Note that unlike MINOS, FarDet spills are as busier than NearDet, due to cosmic rate

CPU needs

	Pre-2009	2009	2010	2011	2012	2013	2014
Running							
Reconstruction							
Calibration							
Skimming							
Analysis							
Simulation							

Work in progress...

Operating systems

- What OS is used?
 - FSL 4.5
 - some collaborators (Indiana) are venturing to FSL 5.2
- Do all collaborators have to use the same one?
 - We haven't had an issue yet that would force a policy.

Data storage and tracking

- How do you catalog data?
 - To date our volume has been low enough to use meetings and manually edited web pages to propagate information about simulation files.
 - Haven't decided yet for large-scale system.
- How do you provide remote access to data?
 - So far scp and afs.

Remote systems

- How many remote institutions provide resources for your users/collaboration
 - Non for general collaborators yet (IU has been producing MC)
 - Cluster at SMU in the works
- Do they have special systems for you or shared?
 - Shared, but expected that NOvA will be heaviest user
- What is done at remote institutions? (Expected)
 - Reconstruction
 - MC generation

Data distribution to remote sites

- Where are data distributed
 - Nowhere yet, but we'll need a facility for using SMU cluster, and eventually others
- What kind of data
 - SMU: Expect Raw data in, MC and Reco data out
- How much data
- How fast does it need to move
- What method is used
 - All of the above to be defined

Grid

- Do you use the Grid
 - Not yet, but we most likely will
- Do you use Grid tools such as Gridftp?
 - unknown
- Do you use Glidein or some other tool?
 - Most likely will use Glidein at least for FNAL grid
- Do you use the FNAL Grid exclusively or do you use more general grid resources?
 - SMU cluster would use OpenScience Grid

Databases

- Technology used
 - postgresql
- Size
 - Requirement of scalability to 500 GB
- Access rate
 - Nominal online requirement, no offline requirement yet

Databases, cont.

- Are they replicated remotely?
 - Not currently the model, but could be implemented if needed
 - However, there will be separate FarDet database for Detector Control System
- What is stored
 - Construction parameters, thresholds, detector conditions (temperatures, etc.)

Conditions

- How are conditions and calibrations stored?
 - In database
- How are they accessed?
 - Online: Used by Epics
 - Offline: via DatabaseBase package in fmwk

Code management

- Code repository
 - CVS
- Build system
 - SoftRelTools

Standard packages

- What standard packages are used:
 - PostgreSQL
 - CLHEP
 - GEANT4
 - Xerces-c
 - GDML
 - ROOT
 - GENIE
 - NEUGEN
 - FMWK

What worked really well?

- Personal opinion: FMWK is very easy to learn and use.

What would you not do again?