LBNE Computing Needs for FY11 & FY12

Tom Junk

April 14, 2010

NuComp Steering Meeting

With input from Brian Rebel, Panagiotis Spentzouris, Greg Sullivan,
Bob Svoboda, Bob Tschirhart,
Jon Urheim, Brett Viren, and Sam Zeller

LBNE = Beamline + Near Detector + Water Cherenkov + Liquid Argon

Life Cycle Overview: 2010

- The big task for 2010 – CD-1. We are targeting September 1 2010 for submission of the CD-1 documentation. Many design questions need to be addressed on this timescale with the resources we have.

  Examples from WCD: Size and shape of cavern, do we need a veto layer, and what PMT coverage and quantum efficiency are needed to meet the physics goals. ND, LAr and Beamline have similar needs.

  -- we need preliminary simulation, reconstruction, analysis, validation, and confidence extrapolating to planned designs this fiscal year. Investigation of several alternatives.

- Standalone tools already available – Geant4-based

  WC simulation runs now, SK reconstruction algorithms have been adapted, questions are being addressed with what we have

Life Cycle Overview: 2010

- LAr uses the LArSoft Geant4 simulation framework
- LAr needs additional computing resources to produce CD-1 deliverables.
- BNL has promised necessary CPU and disk to perform WC simulation and reconstruction needed to produce CD-1 deliverables on the RACF

Life Cycle Overview: 2011 and 2012

- CD-2 target is mid-2012
- More detailed simulations of the proposed alternatives
- Develop LBNE-specific reconstruction algorithms (ongoing)

(SK code is proprietary, physicist input for LAr reconstruction needed.

- Incorporation of tools into a common framework
  - Desire to share this framework between ND, Far Detector in order to minimize the support needed, and to reduce the number of learning curves a collaborator needs to climb.
  - Different constituencies have preferences for different frameworks. BNL prefers Gaudi, FNAL has CMS-lite and FMWK proponents.

- Evolve the computing environment towards what we want to have for the experiment.

Justification for Resources

(estimated data types/sizes and CPU uses)

Resource Request is Combined form Beamline, ND, WC, and LAr.

Advantage: We do not yet know how much we will need and sharing the request reduces the chance of resource under-utilization.

Disadvantage: Competition for resources

Currently we are not using all the resources we have available to us, but as our needs grow, we will have to use dedicated resources instead of borrowing them.

Justification for Resources

- Disk – we need to store
  - Monte Carlo samples – raw and processed versions with

    varying amounts of "truth" information. MC samples will become obsolete rapidly, being replaced with newer, better ones, so we prefer disk to tape at this stage.
  - Code
  - Documentation
  - User home directories

- Tape: Some archival storage of CD-1 and CD-2 MC sets for future reference is needed
- The future: Large amount of raw data (mostly noise) produced by the online DAQ – filtered sets available in several places. Near Detector data most convenient to handle at FNAL.

- Backed-up storage for code and documentation required.

  Home directory backup a plus.

Justification for Resources

- Interactive login and batch:

Depends on how collaborators are used to doing their work. MINOS and NOvA have a large, shared interactive cluster with access to BlueArc disk and shared code.

CDF relies more on users stripping subsamples of the data with micro-ntuples on desktops made on the batch farms (infrequent access to the entire data set, while LBNE will have much more frequent access to the entire data set).

-- Large pool of interactive cores connected to disk with the full data is requested.

I like to have a good debugger available. (Totalview is nice, open to others.)

Justification for Resources

- Grid: Need to run simulation and reconstruction for WCD, LAr, ND, and beamline simulation

Brett's and Greg's estimate of WCD CPU needed once we are running (3 100 KTon WC baseline):

Data: 2 passes x263 CPU years per year of collected data for production (in 2010 CPU-years)

Cosmic MC: 109 CPU years – can be re-used for each year of data

Neutrino MC: $2x47xN_{running\text{-}years}$ CPU years.

We don't need all this in FY11 and FY12, but we will need a subset of it to design the detector and evolve our computing environment.

Justification of Resources

- Servers for DB, Web, special purpose

  Already have multiple docdb's, web areas (wiki, BNL trac, svn repositories for code development and CDR drafts)

  May need a special purpose calibration DB server

- Special CD manpower needs:

  Batch tools: submission, monitoring, security

  Data handling

  User support

  Incorporation of tools into framework

Processing Beyond FNAL

- Computing resources available now at RACF at BNL (we just have to use them!) for WCD work.
- Current installed dedicated LBNE resources:

  160 cores (=80 with hyperthreading). 55 TB of disk.
- Condor Batch installed
- Data access across nodes by XRootd
- Parasitic access to a much larger system (currently used by ATLAS and RHIC)
- Estimated data transfer for WC detector: 90 TB/year raw data, 60 TB/year simulation. FY11 and FY12 much smaller
- Request already put in for $53K for FY11 – will double the nodes (both disk and CPU will double).

Summary

totals are cumulative

Need	FY 2011	FY2012	Comments
Disk (TB)	20	50 (30 additional)	
Tape (TB)	5	15 (10 additional)	
Int./batch (cores)	48	96 (48 additional)	
GRID (slots)	100	200 (100 additional)	
Servers	0	1	Maybe earlier? Repurpose interactive equipment but needs support
Personnel	0	1.5	

Need physicist input to simulation, reconstruction, tool development the most at this stage. Give people resources that are easy to use – add structure as we go.