ELBNF Meeting at FNAL January 22, 2015

Review of LBNE Software Effort in 2013-14

Maxim Potekhin Brookhaven National Laboratory

potekhin@bnl.gov

The purpose of this presentation

- Software and its infrastructure is a critical tool and asset for this experiment, and it's important to maintain proper awareness of its status, needs, rate of progress, problem areas etc. With that in mind, it's helpful to present the updated status of what was accomplished in LBNE in recent months.

- The new Collaboration brings together a large number of researchers and engineering personnel with varying fields of interest and expertise in software and computing. Creating a productive software ecosystem and computing infrastructure in this environment will present opportunities to leverage talent and resources, while failure to do so could mean lost productivity and time.

Overview

- Organization and structure of the LBNE computing effort in 2013-14
- References to a few relevant documents
- Events and Milestones of the past 12 months
- Application software and frameworks
- Work areas:
  - Software infrastructure
  - Geometry description
  - Distributed computing
  - 35t prototype readiness, coordination with the DAQ/Online team.

Software and Computing in 2013-2014

- Mission statement:
  - *The goal of the Software and Computing Organization is to make all aspects of computing technology, software tools and resources easily accessible to the largest number of Collaboration members possible, in a most efficient manner.*
- Deliverables - physics applications, simulations, reconstruction algorithms, beamline and detector R&D and characterization - **Physics Tools Group.**
- Infrastructure - build and release procedures, continuous integration, data management, distributed computing etc - **Software and Computing Organization** - essentially services for the Physics Tools Group - √
- Data acquisition, online for the 35t prototype - **DAQ/Online Group.**

Events and Milestones 2013-14

- Initial Computing Plan created in 2013 (DocDB 7818) and presented to DOE
- To guide technology choices and policies, we developed Sofware and Computing Requirements in 2014 (DocDB 8546).
- **Software and Computing Review** by the US Department of Energy in May 2014. Feedback was largely positive, main recommendation is to quantify and develop the Computing Model.
- The P5 Report and events triggered by it - including formation of ELBNF and, of course, this meeting.
- Formation of the US Forum for Computational Excellence (DOE initiative).
- Formation of the international HEP Software Foundation - 2014 meeting at CERN and another at SLAC - took place earlier this week.

Application software and frameworks

- Sharing of effort and expertise at FNAL among the IF experiments lead to adoption of "LArSoft" suite of software tools to model the Liquid Argon TPC.
- LArSoft is based on the *art* framework developed at FNAL which was derived from CMS software.
- *art-daq* used for current online/DAQ work (e.g. the 35t prototype).
- Variety of Monte Carlo tools (cf. GENIE), GEANT4 etc.
- Beam and target simulations, and the LAr TPC present two major branches of work which feed current detector R&D.
- Initial stages of the Near Neutrino Detector simulations.
- Release management is currently done at the package/project level, there are no overarching releases to cover all LBNE software.

Summary of progress "on the ground" in 2014

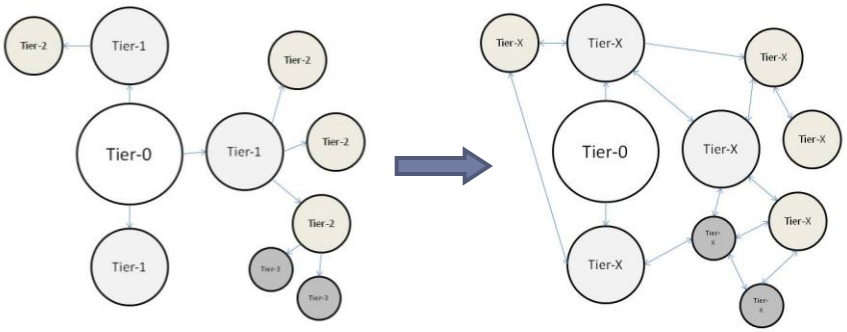
- The “short list” of priority work items remained fairly constant over time
  - Core Software Infrastructure (build/testing/validation/distribution).
  - Data Storage and Management
  - Geometry Description and Model
  - Distributed Computing
  - Collaborative Tools
- Progress has been achieved in all of these.

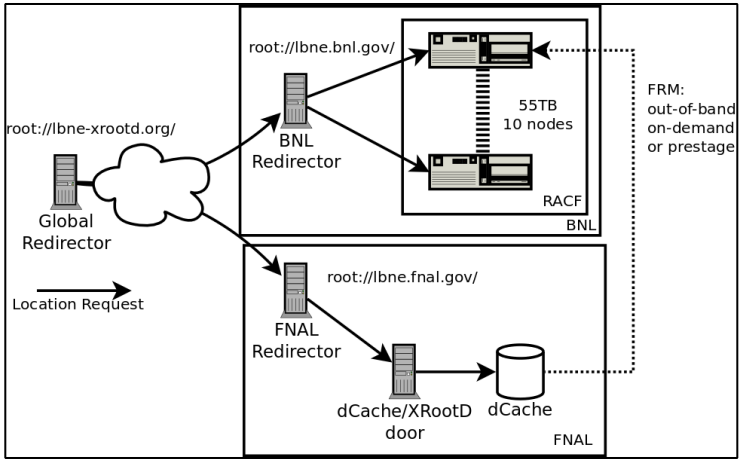
Software Infrastructure

- Build/validation/continuous integration
  - This has been top priority consistently over a year and a half - we found that build procedures can and should be improved to make it easier to build the LArSoft stack (with underlying *art*) from source, without relying on specifics of software configuration practices at FNAL (as explained in the “white paper”, DocDb 10096).
  - They key is to lower the barrier to join the development effort.
  - The “*worch*” build orchestration tool (based on the standard configuration parser and *waf*) has been created to manage the build process (including CMake-based and other systems) - https://github.com/brettviren/worch
  - Recent positive experience with "green field" build from source on a different Linux platform, outside of FNAL and without much help.
- Environment management and packaging software stacks:
  - Recent experiments with Docker - https://github.com/LBNE/lbne-docker
- Release Management
  - Established the role of the LBNE release manager, the scope, policies and procedure are being worked out.

Distributed Data

- Evolution of the LHC Computing Models: a trend to move from the MONARC architecture, flattening the hierarchy of data and workload distribution. We started configuring and testing xrootd (FNAL-BNL data link).





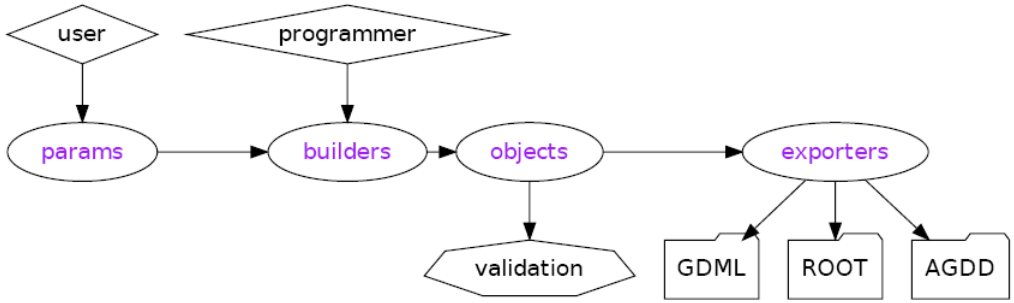
- Data replication, retention and other policies were proposed and documented in the Requirements.
- Currently leveraging storage and Metadata facilities at FNAL.

Distributed Computing

- By and large, until now Fermigrid and other “central” resources at FNAL were sufficient to support MC and other workflows, with some loads also running elsewhere...
- ...but reliable and stable Grid and Workload Management capability is a must have in a large international Collaboration, due to increased scope of the Physics Tools and more computationals tasks to handle.
- We received **excellent** help from the Open Science Grid Consortium
  - Software provisioning via CVMFS
  - As a pilot project, we ran LBNE payloads on ~20 Grid sites
- No choice has been made yet with respect to Workload Management System, something the new Collaboration will probably want to consider.

Geometry

- Brief history
  - Identified as a priority item a while ago, came to the fore due to issues encountered in ongoing Liquid Argon TPC simulation work. Current solution (a mix of Perl and XML) is not optimally structured and can benefit from factorization.
  - While some of us had experience with a more “holistic” approach (pure XML being the only source) it appears that from expediency point of view, combination of “builders” and “configuration” could deliver most benefits relatively quickly.
- Python-based GGD: general geo description
  - Goals: authoring, persistence, provenance, visualization, conversion
  - Components: parameters, builders, objects, exporters
  - See https://github.com/brettviren/gegede



Dynamics of the effort profile

- The long time period before we start taking data presents a problem with attracting graduate students and postdocs. Higher turnover at Universities (compared to Labs) makes it necessary to ensure continuity, documentation and support.
- Until now, we oftentimes had to operate as a skeleton crew with few contributors to both physics tools and infrastructure.
- University personnel seems to gravitate towards physics tools, National Laboratories – to S&C infrastructure. While the latter is good to ensure continuity of support, it may be a good idea to have a more balanced mix.

Summary

- In any scenario the new Collaboration will face, software and computing will remain critically important assets.
- Evaluation of the current software has always been in the plan for the medium/long term. The new international experiment provides even more motivation and opportunity for this. It appears that it is in everyone's interest to survey and assess frameworks, platforms and tools used and/or developed by collaborating institutions, from the standpoint of benefit to the new Collaboration, and doors are open for new ideas and contributions.
- We need to start the process of communication, establishing connections and in general advancing collaborative process in the area of software and computing. We presented this review of prior work done in LBNE with the purpose to increase awareness of progress and issues in this area, and to invite everyone to this discussion.