Microboone Offline Computing

SCD Projects Meeting
Oct. 16, 2013

H. Greenlee

Outline

- Development environment.
  - Art framework.
  - Larsoft.
- Data Handling.
- Grid computing.
- Portability Issues.

Microboone Development Environment Overview

- Interactive computing is supported locally on the uboonegpvmXX (XX=01,02,03) machines.
  - Each machine has four cores and 12 Gb total memory.
- Microboone has 50 Tb of bluearc disk space in /uboone/app and /uboone/data.
- Microboone (along with many other experiments) uses the art framework as its offline software framework.
- Microboone (along with several other LAr TPC experiments) uses the larsoft project as its main offline software.

Art Framework

- In general, we are happy with art. It is sufficiently powerful and flexible.
  - I would not hesitate to recommend other experiments to use art.
- Like any software product, art has its own issues. Some of them are as follows.
  - Poor documentation.
    - I know that the art workbook project is trying to address this.
  - Art project team can be slow in responding to requests.
    - Needs more manpower (art team was hit pretty hard by layoffs).
    - Art has several issues regarding integration with sam that need attention (like not automatically generating required metadata).

Larsoft

- Larsoft currently maintains a common code repository and development environment for all LAr TPC experiments.
  - Until recently, microboone did have its own separate development environment (and still mostly doesn't).
- Earlier this year, management of larsoft was transferred from PPD (Brian Rebel and Eric Church) to SCD (Rick Snider with help from Lynn Garren).
- Along with the management transition, several technical transitions are in progress or planned.
  - Code repository svn → git + git flow.
  - Build system srt → mrb + cmake.
  - Experiment-specific software builds.

Larsoft (cont.)

- Microboone is supportive of the technical changes being made in larsoft, and wants them to be completed asap.
- Since the management transition, there has been a tendency for the different experiments to go their separate ways regarding algorithm development.
  - Some of this separation was inevitable, necessary, and a good thing.
  - The partial separation shouldn't be allowed to progress to progress to a complete separation.
    - There should continue to be a shared repository for common code (while allowing for experiment-specific code).
    - The larsoft general meeting can continue to be a venue for intellectual cross-fertilization among the LAr TPC experiments.

Microboone Data Handling

- Microboone data handling is based on sam + enstore.
- Microboone has been working hard to integrate sam and related products into its offline framework and production system.
  - We now generate sam metadata automatically in art programs.
  - MC production files from two major MC productions (so far) have been stored in sam + enstore.
    - Metadata and bluearc disk locations entered into sam database.
    - Sam dataset definitions created.
    - File transfer service configured, and mc files uploaded to enstore.
  - Able to read data files from sam datasets in art programs (ifdh_art product).

MC Production

- So far microboone has done two major MC productions.
- Following table is based on information in sam database.

Production	Files	Events	Size (Tb)	Event Size (Mb/Event)
Fall 2012	290	290,000	0.71	2.4
Summer 2013	270	340,000	1.82	5.4

- Event sizes for complicated events (e.g. including cosmics) can be much larger (up to ~40 Mb/ev). Event size may expand further due to more realistic noise simulation.
- Typical cpu time is 5-10 min/ev (may expand).

Estimated MC Needs for 2014

- Enstore.
  - (1,000,000 events) * (20 Mb/ev) = 20 Tb.
- Production cpu.
  - (1,000,000 events) * (600 sec/ev) = 7000 cpu-days.
- Above estimates are highly uncertain.

Grid and Batch Computing

- Microboone currently runs batch jobs using microboone and opportunistic resources in FermiGrid.
  - 500 microboone batch slots in GP farm.
  - Job submission via jobsub.
  - We currently only submit batch jobs to nodes that have microboone bluearc disks mounted.
    - This needs to change.
- In September, microboone was approached by SCD grid department, and offered help in getting our batch jobs running on OSG.
  - The success of this operation requires us to make our offline environment more portable.
  - Aiming for mid-November.

Portable Environment

- We want to develop a portable offline environment.
  - Distribute larsoft and dependent products (art, geant4, genie, root, etc.) as compiled binaries.
    - Maybe in multiple flavors.
    - Distribute via cvmfs (upload to oasis cvmfs server).
  - Portable ups initialization.
    - Reconfigured larsoft should be packaged as “relocatable ups” products.
- Use cases for a portable environment.
  - OSG.
  - Remote institutions.
  - Personal computers and laptops.
    - Ability to do development on a computer with a local disk is a plus.