Hall-D/GlueX Software Status

12 GeV Software Review III
February 10, 2015
Mark Ito

Data Challenges (DCs)

- DC2 - March/April 2014
  - 10 billion events with EM backgrounds included - OSG, JLab, MIT, CMU, FSU
  - 4500 Concurrent Jobs at JLab
  - 11,000 Concurrent Jobs on the OSG
  - Well under 0.1% failure rate
- DC3 - January/February 2015
  - Read data in raw-event format (EVIO) from tape and produce DST format (REST) files.
  - JLab only
  - Test throughput from Tape Library
  - Run Multi-threaded jobs

Computing Model and Requirements

- Estimate of cores needed reduced by a factor of 3 with new Haswell chip
- Real data tests show single-core event reconstruction rate of 30 Hz, a factor of 20 compared to basis of original estimate, but…
  - "junk" events not accounted for (event fraction, reconstruction rate)
  - Photon beam spectrum not realistic (bremsstrahlung vs. coherent bremsstrahlung)
- Realistic event size? (waveform data, header words)

Offline monitoring

- Weekly reconstruction pass through data
  - All data as of Friday afternoon
  - Online monitoring plots reproduced
  - Full reconstruction with updated reconstruction code, constants
  - Skims of raw events done for calibration
  - DST data produced (REST format)
  - Web site for browsing results
  - REST data good enough to observe multi-particle final states (omega to pi+pi-pi0 with kinematic fits)
  - Continuing on a bi-weekly schedule

Calibration Committee

- Bi-weekly meetings, chaired by Sean Dobbs of Northwestern
- Preliminary list of constants compiled in advance of run, used to guide activity
- Calibration procedures still being developed
- Some progress:
  - Basic timing offsets
  - Global energy scale for calorimeters determined

Calibration Database Experience

- Fully integrated into reconstruction
- Near-complete migration of constants into database
- All detector groups making contributions (no known rogue systems)
- SQLite form of database as alternate to MySQL/MariaDB
  - Complete history and versioning support
  - Solves:
    - Distribution (remote sites, network-challenged computing)
    - Server contention from farm usage
  - Drawback: no automated back-annotation

Data Distribution

- Used OSG SRM to archive DC2 OSG results to JLab Tape Library
- Raw data shipped to CMU using Globus Online
- Fall 2014 REST data only [?] MB total

Data Management

- Raw data stays on JLab Tape Library
- Reconstructed REST-formatted data compact
  - Keep as much as possible disk-resident at Jlab
  - Distribute most (all?) to collaborating institutions
- Data Catalog/Tracker needs development
  - Existing package? Develop one?
  - Crucial dependence on remote data deployment goals

Online Conditions Database

- Database to store online run conditions, e. g., magnet current settings, configuration files, etc.
- Two were deployed
  - one looks great (API, jquery-based web interface)
  - one was useful (hand work required, HTML-base web interface, CSS-free, Javascript-free)
- Effort underway to consolidate and expand.

Summary

- The collaboration feels that we reached a number of milestones that we did not expect to see until we were well into the April 2015 run.
- Software performed all critical tasks needed to support detector commissioning and to see physics signals.
- Real data exposed areas requiring further work

To be added to the talk!?

- break-out requirements from Graham's talk
- code profiling not done
- never ending data challenge never started
- Geant4 not there yet
- real data: number of events, disk footprint, processing time
- Source code management/review needs further development