GlueX-Doc-3189-v1(Draft)

Report on the
12 GeV Software and Computing Review

Jefferson Laboratory

10-11 November 2016

https://www.jlab.org/indico/event/183/overview

Committee members: David Brown (LBNL), Sergei Gerassimov (TU Muenchen/COMPASS), Emlyn Hughes (Columbia), Martin Purschke (BNL/PHENIX), Torre Wenaus (BNL/ATLAS) (chair)

December 31 2016

Table of Contents

Executive Summary

On November 10-11 2016 Jefferson Lab held its fourth Software and Computing Review for the 12 GeV program, following reviews in February 2015, November 2013 and June 2012. The review was convened by Bob McKeown with a review committee consisting of David Brown (LBNL), Sergei Gerassimov (TU Muenchen/COMPASS), Emlyn Hughes (Columbia), Martin Purschke (BNL/PHENIX), and Torre Wenaus (BNL/ATLAS) (chair). The committee was asked to review the state of software and computing developments for the 12 GeV program at Jefferson Lab, with particular emphasis upon the following (the last two are new to this review):

- Detector simulation, calibration, and event analysis
- Workflow tools for production analysis
- Computing plans, including projections for cores, disk, and tape for the next two years
- Software and computing management
- Performance of the Scientific Computing Systems
- Collaboration readiness to produce timely results for publication

The committee was asked to address the questions posed in the charge (included as an Appendix), keeping in mind the different timelines for the different halls, e.g. Hall D / GlueX has completed its engineering run phase and has been in production running since Fall 2016, while CLAS-12 is a few months out and should be ready for production running around Summer 2017.

As is invariably the case, the review was very enjoyable and well run; the panel thanks the organizers for another interesting and pleasant visit to JLab and the 12 GeV program.

The panel benefited from comprehensive and carefully prepared presentations and materials. And we came in right on time on the schedule!

Here follow the Committee's comments and recommendations across the areas covered by the review, organized as Observations, Findings and Recommendations. Observations are significant points of information highlighted by the panel as relevant to the Charge. Findings are inferences and conclusions drawn. Recommendations suggest a particular action or direction, often in light of a finding.

The first General section gives a summation of the panel's conclusions with respect to the charge points. The specifics behind the summation are found in the subsequent sections.

Comments and Recommendations

The following sections present the comments and recommendations of the Committee across the areas covered by the review: general issues, common Experimental Hall issues, Hall-specific issues, and management and computing.

General

Findings

- Offline software
    - Progress and readiness in offline software across the halls has been impressive. Runs in Halls A and D collected data in Spring 2016, and the preliminary results shown demonstrate their software and computing are functioning very well. CLAS12 has taken cosmic data and used that to validate their calibration. Software readiness was not an impediment. Hall C appears on track for readiness given their well understood and thoroughly validated software stack. Hall D's readiness for ramping up data taking and analysis is particularly impressive. Their use of large scale data challenges to exercise at scale and harden their software is commendable. They cite an expectation of submitting up to 7 papers for publication over the next year. Hall B's innovative offline software stack is impressive in its capabilities. The next 6-12 months will be key in ensuring the software successfully transitions to supporting physics running and analysis with broad participation from collaborators. Reconstruction and calibration challenges in December are important steps towards this but the schedule is tight; many items on the plan are due mid-November, only a few days after the review. An end to end data challenge-like test from event generation through analysis is planned for March, sufficiently in advance of running to allow addressing issues arising so long as the March date holds.
    - User support across the halls appears good, with regular meetings, documentation, and well-attended analysis workshops supporting the engagement of the collaborations in software and analysis. Hall D has seen analyzers emerging unbidden to pursue studies on their own initiative, exactly as it should be. Hall B showed a plan to create teams to manage the engagement of analysis with software.
    - The level of software commonality appears good. Halls and CC are sharing and co-developing software in many ways. We heard about important new or matured ones, including SWIF and RCDB. Community tools like ROOT are leveraged well. Shared, distributed computing is being used where appropriate, OSG for processing and Globus Online for data movement.
- Management & Computing
    - Addressing recommendations of last time:

        - "It would be beneficial if a professional code analysis tool such as Coverity would be licensed and made centrally available." Acquiring a Coverity license is driven by requests from user communities and no Coverity license request has been received by CC.
        - Everyone is using nightly builds and there has been commendably broad adoption of modern best practices software management tools.
        - Timestamps vs. run numbers has been addressed. Run number is preferred and primary, but there are complicating cases.
        - Exploring the use of analysis trains has been fully accepted. Hall D has implemented and is using them effectively, Hall B has plans to adopt them. The development has proceeded in close collaboration with CC, the SWIF tool being the basis.
        - Geant4 migration is well advanced.
        - A data management/cataloging tool is in place where most needed. In Hall D, the tool is not used yet because the scale thus far doesn't demand it.
        - On computing operations teams, we heard about teams being established in Halls B and D to manage the process from data taking to analysis.
    - Staffing levels seem appropriate. Optimization of how personnel are used in the CC led to the move of an FTE from development to operations support.
    - Management structures and processes continue to be working well. Engagement between CC, the collaborations and users appears good. Close consultation between CC and the Halls on requirements and optimizing the use of resources continues. The Halls are instituting/have instituted teams to oversee and execute analysis with the arrival of data taking.
    - Processes to manage risk, provide contingency, and manage the still evolving 12 GeV program schedule appear to be operating well. Careful tracking and provisioning of computing, storage and networking requirements continues. The risk represented by a single 10G external link has been ameliorated with a second 10G link. Cost management continues to be good, JLab computing appears lean, efficient and cost effective. Technology tracking and performance measurement continue to receive close attention, exposing for instance the relative inferiority of the KNL architecture for our workloads. Purchase planning proceeds in light of such findings.
    - Computing requirements are well stated and well justified. Areas potentially amenable to shrinking CC requirements are identified, such as Hall D simulation.
    - Transitioning to operations has made a smooth beginning in the now-running Halls and in CC.
- Readiness for producing publishable results
    - Hall A and Hall C present an expectation for submitting first publications 12-18 months after data taking. Software and computing is only a small contributor among the factors that set this timeline.

- Hall D presented a notable and commendable credo, that the “Collaboration has a vision that early physics results are important and benefit everyone.” Software and computing could have a deleterious effect on achieving this if preparations were not careful and comprehensive, given their requirements; in fact their careful preparations are consistent with the credo. They estimate as many as 7 papers to be ready for submission over the next year.
- Hall B presents their software to be in a good state of readiness; the challenges taking place between now and March will subject that expectation to an appropriate level of testing.
- The collaborations have begun organizing themselves for an active analysis program with working groups covering the processing chain and analysis activities.
- There are not visible gaps in computing resources that would impede the timely production of publishable results, beyond the known challenges of delivering all of GlueX’s projected simulation needs, for which ameliorating measures (in particular running part of it offsite) are available.

Experimental Halls - General

Recommendations

We heard about the concept of the Analysis Trains, which are used by various experiments. The trains facilitate an efficient access to the data, essentially pooling requests for the same set of files in a tape-centric way to minimize the number of file retrievals. We recommend to add the concept of a stable run list either to the Analysis Train (and so make them “named trains”), or as a separate tool. The advantages:

- This gets a list of good runs carefully vetted by many people, instead of by individuals, which helps eliminate runs/files with not readily recognizable problems from entering the analysis chain.
- The actually analyzed set of events remains stable across different train rides by the *same* analyzer, making results easier comparable. Differences in the results are then due to changed code, cuts, and so on, but not due to increased (or decreased) statistics, or a different set of selected runs/files.
- The fact that *different* analyzers always see the same set of events when using the same train makes cross-checks between different analyses projects easier, and makes results/quantities transferable to some extent (for example, one could get the mean of some distribution from another analysis if needed).
- Some quantities (e.g. multiplicities, events passing a particular set of standard cuts, etc) are then well established and known beforehand.

Once a list of runs/files has been established, it is kept stable, and new named train is made if the list needs to be changed. In this way one can again gauge the differences between the old and the new dataset.

Hall A and C

Observations

- SBS in Hall A just passed a successful DOE review this month.
- Hall A had its first physics run in 2016. A factor of 10 increase in rate is predicted for 2018.
- Computing needs for current and near-term future experiments in Halls A and C are modest and can be easily handed by the available resources at JLab.
- The PODD C++ framework is a one-pass tool which produces TTrees for analysis and calibration directly from either raw or simulated data.
- Hall A is using the common SWIF workflow management tool developed by CC based on requirements from Hall D.
- Github is used to distribute and organize simulation and reconstruction software.
- Redmine is being considered as a bug-tracking tool as it has more features than the github tool.
- Multi-arm configuration readout of Hall A is done with separate DAQ systems, which have different run numbers. This limits using the run number as a unique conditions index. Synchronization with timestamp is less ambiguous.
- Thorough comparisons between results of the Fortran-based ENGINE and the C++ based HCANA framework and modules using HMS data show excellent consistency.
- Geant4 is used for optical photon propagation in some SBS detectors. A parameterized response function can also be used.
- GEM response is simulated after Geant4 and includes the details of ion drift and related processes not described in Geant4.
- Several passes through the data are generally required to address calibration updates. These are included in the computing resource need estimates.
- Hall C HMS spectrometers are well understood from the 6 GeV program, and no major software or simulation (i.e. G4) upgrade is implemented, although beamline simulation and subsystem studies with G4 are being used.
- SHMS Q2 was recently installed and cool down performed. Q3 is still in Europe but ready for shipping and should arrive in December.
- Analysis macros written in Paw Kumac and Fortran for HMS are being converted to ROOT C++ by users for analysis of SHMS.
- SIMC has been validated against HMS results produced in 1999-2000.
- Historically, results from Hall C experiments have been published typically between 2 and 3 years after the experiments completed.
- Hall C begins 12 GeV operations in Spring of 2017.

Findings

- SBS aims to begin first data collection in Spring 2019. This represents a quite aggressive schedule.
- Full Geant4 MC for the SBS has been implemented and used for detector response studies. The detector will need to operate in a high background environment with no magnetic field using primarily ~100,000 channels of GEM detectors for tracking. Data rates are estimated to be 500 kHz/cm2, but the uncertainty in this rate is likely to be large. Approximately 3 PB of raw data will be collected for the already approved experiments.
- The same GEM detector package may get tested in the existing HRS (effectively a low background commissioning) during the A1n experiment, if this experiment runs in 2018. The uncertainty in the A1n schedule is driven primarily by the success of 3He target R&D program.
- The SBS group is only recently transitioning from a hardware focused effort to software and analysis. The group is planning to implement end to end analysis for the simulation.
- There appears to be no clearly defined data products for instance from GEM tracking that can be saved on disk. This could limit multi-pass processing options.
- The documentation and tests developed for the Hall C comparisons of Engine and HCANA are a valuable resource for future experimenters.
- SHMS appears to be on track for installation and commissioning for Spring 2017.
- Hall A and C software upgrades for 12 GeV running are complete and validated.
- The Hall A and C collaborators aim to release publishable results 12 to 18 months after an experiment has finished data collection. Historically, if an analysis develops difficulties, the timescale for release can be significantly longer.
- The recommendations from the previous review have been largely addressed. In particular nightly builds and automated monitoring are now standard. Using run number as a conditions key is not practical in dual-arm experiments due to legacy and DAQ issues.
- Given the operation of a new tritium target (2017), a challenging pair of parity violation experiments (2018), the possibility of running a polarized 3He target at high current with new GEM detectors (2018), the installation of the SBS (~2018), and participation in long term highly ranked programs such as MOLLER and SOLID in Hall A as well as the fact that the Hall A and C staffs work simultaneously in both halls, there is a real concern that the staff is oversubscribed.

Recommendations

- Continue to aggressively enlist the user community to commission the upcoming experimental program.
- Investigate whether substantially higher backgrounds in SBS than expected would alter projected computing needs or required software capabilities.

Hall B

Observations

- The CLAS12 experiment anticipates a beam delivery milestone in March, solenoid installation in April, and physics start in Fall 2017. Detectors are done, the torus magnet is done and tested at nominal current, while the solenoid is still to be completed.
- The magnetic field map has just been measured.
- ClaRA is the Java-based event reconstruction and analysis framework. Reconstruction packages are plugins to ClaRA (services), implemented as dynamically linked libraries.
- Track finding is based on the Kalman filter technique.
- GEMC is the C++ Geant4 based simulation package.
- Experiment setup information and detector geometry descriptions are stored in a DB (CCDB) and is common for reconstruction, simulation and analysis.
- Detector calibration and monitoring software is in an advanced stage. CCDB is also used as a storage and backup tool for calibration information (the same software as for Hall D).
- The Java based GUI and event display for every component of ClaRa software were developed in-house.
- The I/O package in Java is a common interface for decoding from DAQ, and for event I/O.
- Coverity is being considered for code testing and debugging. FindBugs is being used currently.
- The code versioning system is git.

Findings

- Background event overlay in CLARA or GEMC are foreseen but not yet implemented. This limits the options for producing realistic samples in the near-term simulation challenges.
- Detector calibrations using cosmic rays have been exercised, and improved performance of the SVT tracking residuals based on these have been demonstrated. This is an important proof of readiness.
- The planned calibration constant and reconstruction challenges to occur in December of this year, together with the full end to end exercise from simulated raw data to physics results planned for early Spring 2017 for three commissioning analyses, will be an important milestone for evaluating readiness for the Autumn 2017 run.
- It is essential that the schedule for these challenges hold, to allow time to react to what is learned before data taking begins.

Recommendations

- Continue to carefully track software milestones, and maintain an achievable plan, leading up to the commissioning/physics run in 2017.

Hall D

Observations

Hall D / GlueX detailed their plans and setup for data taking. From a high-level view, this came across as a very modern and mature setup that "does things right". We noted a comprehensive list of standard concepts and approaches were being applied.

- Successful engineering run last spring, physics running starts ~next week.
    - Spring: 800MB/s. 37TB data sample.
    - Fall: 2000MB/s. ~100MB data sample. (1PB raw).
- Nightly rebuilds of the entire codebase.
- Checks with standard tools such as valgrind.
- Good monitoring QA process
    - Disk resident monitoring samples of new data.
    - Automated monitor train run on first 5 files from each run, to validate promptly.
    - Periodic checks of the results with well-known input files.
    - Archiving of monitoring output for later "forensics" if needed.
- Automated calibrations.
- The goal is to be able to start production 2-3 weeks after the data have been taken. Expect to exercise this in Fall 2016.
- Partial Wave Analysis software (M-tools) exists and had been tested (without acceptance corrections yet).
- REST is a compact self-described flexible file format (XML-based) for reconstructed events (currently 37 TB has been produced). REST files have been distributed to university collaborator sites.
- Multithreaded Geant4 is complete and in beta. Found numerous G4 bugs and fed them back to developers, expectation is that they will be in G4 10.3. In the meantime, an in-house patched release is used.
- A lot of work has been done in tuning the simulation to match detector performance.
- The detector response simulation (MCSmear) has been implemented as independent module (between GEANT4/3 and reconstruction in the simulation chain).
- There was a recent analysis workshop with 60 people, some offsite. Recorded for sustained use. Intent is to hold them as needed as the software evolves. Went 3 years without one because the software was stable.
- Working groups established for physics, analysis, calibration, offline software to get the physics out quickly.
- The collaboration has worked closely with the computing center on developing CC's SWIF. SWIF's requirements came from Hall D.
    - Analysis train has ~15 users with 50 plugins so far, many channels. Good participation.

- Very close to first publications. Including the Fall 2016 run, expect 7-10 publications in the next year.
- A very impressive study was presented on high intensity running and L3 trigger, learning from the Spring 2016 run, discovering high rate issues and addressing them with existing hardware.
- L3 will be able to run in pass-through mode to measure its effects and validate that biases aren't being introduced.

Findings

- People are visibly diving into the data and coming out with results.
- The collaboration has a commendable credo: "Collaboration has a vision that early physics results are important and benefit everyone."
- The reconstruction rate was 22Hz/core previously, now is 9.8Hz/core. The slowdown is due to adding more hypotheses -- e+/e-, antiprotons, … -- and is understood. Physics is driving that more things need to be looked at. This does not blow out the expectations of getting adequate computing resources. But simulating 2x the data doesn't fit into the CC budget. The Geant3 simulation speed was 18 Hz, while the Geant4 based simulation is 9Hz, which is not an unusual differential in G3/G4 speed.
- Important lessons learned from experience to date were given, with which the panel concurs:
    - LARGE data challenges are important. 25 B events in the latest DC. Learned much about managing data sets.
    - Work closely with CC.
    - Cannot monitor enough.
    - Must have automatic online data monitoring, make it easy to spot problems.
    - Identify physics that can be analyzed quickly, even with minimal understanding of detectors, allowing to validate procedures as early as possible.
- The extrapolations from early engineering runs show that the overall and per-crate data rates are higher than the available bandwidths, which are typically 1Gb/s uplinks from the crates. A test is planned if an upgrade to 10Gb/s links would actually increase the throughput or merely expose the next-in-line bottleneck. To address the problem more generally, a Level-3 trigger will be run to trim the data rate down below the limits.

Recommendations

We recommend to explore the possibility of trading CPU power used for data *reduction* for data *compression*. Data files on disk were found to be gzip-compressible by about a factor of 2, so substantial gains could be realized by compressing the data stream at an early stage with an efficient algorithm. In this way, more data could get through the bottlenecks, which will increase the statistics for physics processes for which no fast trigger algorithm can be developed, and in general relax the demands for the L3 reduction factors.

While it was noted that this will not reduce the tape usage because the tape drives already perform what amounts to a late-stage compression, the compression at an early stage will reduce the overall throughput upstream. Other experiments have found that the reduced data sizes per event speed up the later analysis, because I/O is usually a major component of the overall analysis time. Savings here usually overcompensate the penalty for having to decompress the data.

Management and Computing

Observations

- Equipment replacement cycle is not fixed at 4 or 5 years, equipment is kept on the floor while it is cost effective to do so, so long as space/power/cooling is available.
- The farm delivered 38M core hours in FY16.
- Moving resources between the farm and LQCD on an accounted loan basis continues to be very effective for optimizing and maximizing the use of facilities. Resources are regularly shared at the 1000 core level.
- Lustre disk space is now managed automatically to keep utilization high while avoiding filling the disks (auto deletion of the least recently used data).
- The Data Center move in 2017 will add generator backup to the existing UPS backup. Scheduling unavoidably places the tape silo move during the Spring 2017 run; disk buffers will be used to avoid impact on running.
- The addition of 200 TB/year of disk appears adequate to meet current and future needs.
- CC and Hall D have worked in close concert on developing and commissioning CC's SWIF workflow manager. Now in a requirements gathering phase for SWIF version 2.
- A planned reprioritization of CC manpower replaced a developer position with higher priority operations support.
- The processing needs of the different halls vary dramatically. Hall A and Hall C processing needs are of the order of the laptops in the review room. Hall D followed by Hall B dominate the Data Center processing requirements.
- Hall D simulation is both the largest processing component and the component that is most amenable to compressing to accommodate capacity constraints. Simulation can be run effectively off site, and the simulation to real data ratio may be amenable to tuning. GlueX uses a ratio of 2 at present; we heard that a ratio of 1.5 at least for some studies is under consideration.
- GlueX's requirements have evolved due to larger event sizes and processing times.

Findings

- Experiments now present their CPU requirements in terms of core-week needs, which seems a clear and transparent means of communicating needs.
- The addition of a second 10G link of off site connectivity, with SciComp getting a 50% share of the new link, is a very good development.
- GlueX's use of Geant4 MT makes their simulation suited to some HPC resources and particularly Cori at NERSC (modulo architectural inefficiencies in its KNL based Phase

2). ATLAS in particular has worked closely with Geant4 to commission Geant4 V10 on Cori and improve its performance. GlueX might consider putting in a proposal for Cori time at some point.

- Projected core count is below expected needs in 2018. Some purchasing has been delayed due to budget uncertainties.

Recommendations

- Consider formally including off site computing resources in the computing planning, particularly for Hall D. Predictable availability of offsite resources (OSG, HPCs) would contribute directly to controlling Data Center capacity requirements.

Appendix: Agenda

Thu Nov 10

08:00 - 08:40 Executive Session

08:40 - 09:00 Introduction and 12GeV Overview

Speaker: Allison Lung (JLab)

09:00 - 09:30 Halls A and C Overview and Progress

Speakers: Ole Hansen (Jefferson Lab), Mark Jones (Jefferson Lab)

09:30 - 10:00 Hall B Overview and Progress: CLAS12 Software Overview and Progress

Speaker: Veronique Ziegler (Jefferson Lab)

10:15 - 10:45 Hall D Overview and Progress: GlueX/Hall D: A Blueprint for Physics

Speaker: Prof. Curtis Meyer (Carnegie Mellon University)

10:45 - 11:15 Computing Overview

Speakers: GRAHAM HEYES (Jefferson Lab), Sandy Philpott (Jefferson Lab)

11:15 - 11:45 Halls A and C Deep Dive: Readiness for first Hall C data analysis and publication

Speaker: Dipangkar Dutta (Mississippi State University)

11:45 - 12:15 Halls A and C Deep Dive: Super Bigbite Spectrometer Software Overview

Speaker: Seamus Riordan (Stony Brook University)

12:15 - 13:30 Working Lunch (Executive Session)

13:30 - 14:15 Hall D Deep Dive: GlueX Data Analysis

Speaker: Dr. Paul Mattione (JSA)

14:15 - 15:00 Hall D Deep Dive: High Intensity Running in Hall D

Speaker: David Lawrence (Jefferson Lab)

15:30 - 16:15 Hall B Deep Dive: CLAS12 Online/Offline Software Tools 45' ( )

Speaker: Gagik Gavalian (Old Dominion University)

16:15 - 17:00 Hall B Deep Dive: CLAS12 Analysis Readiness 45' ( )

Speaker: Raffaella De Vita (INFN - Genova)

17:00 - 18:00 Executive Session

18:00 - 18:45 Questions

Fri Nov 11

08:00 - 10:00 Q&A

10:00 - 13:00 Executive Session - Prepare draft (Working Lunch)

13:00 - 13:30 Close Out

Appendix: Charge

The committee is asked to review the state of software and computing developments for the 12 GeV program at Jefferson Lab, with particular emphasis upon

- Detector simulation, calibration, and event analysis
- Workflow tools for production analysis
- Computing plans, including projections for cores, disk, and tape for the next two years
- Software and computing management
- Performance of the Scientific Computing Systems
- Collaboration readiness to produce timely results for publication

The committee is asked to address the following questions, keeping in mind the different timelines for the different halls (e.g. Hall D / GlueX has completed its engineering run phase and is since Fall 2016 in production running; CLAS-12 is a few months out, and should be ready for production running around Summer 2017):

1. Offline Software: Detector Simulation and Analysis

a. Are the halls making appropriate progress towards having their simulation, calibration and analysis software ready? Are they meeting their previously set milestones?

b. Have an adequate set of milestones been identified, and an appropriate set of tests been incorporated into the milestones, to measure progress towards final production running?

c. Are the halls doing the right level of at-scale testing of each of simulation, event reconstructions, and physics analysis appropriate to the time before engineering and physics running?

d. Are the halls getting users engaged at an appropriate level to demonstrate usability and readiness from a user's perspective? Have the collaborations identified effective and appropriate mechanisms to support utilization of the software by the entire collaboration? Is the level of user documentation appropriate for this point in time?

e. Are appropriate efforts towards software commonality being made across the halls and/or with the wider HE/NP communities?

2. Management

a. Did the halls respond appropriately to the recommendations of the last review?

b. Are staffing levels for software development and documentation appropriate?

c. Are the current management structures and processes well-matched to the needs of the collaborations (including users)?

d. Are there appropriate contingency and risk-management processes in place? Have risks been appropriately identified?

e. Are reasonable change control processes being used to address scope and milestone changes?

f. Are there adequate plans for transitioning from a development phase into a deployment and operations phase? Are the timelines appropriate?

3. Computing and Networking

a. Are the requirements for computing, storage and networking well stated and well justified? Are all of the assumptions clearly stated, and are all of the units clearly defined (e.g. "E2670 v4 core" vs "core")?

b. Are the computing and networking plans of the laboratory well matched to the requirements? Are they cost effective, and are budgets appropriate for these plans?

4. Readiness for producing publishable results

Given the anticipated dataset for the first year of production running:

a. Do the collaborations have all software elements in place to produce timely results on this dataset?

b. Are the collaborations properly organized, with working groups engaged in relevant simulations, calibrations, and higher level physics analysis to effectively and efficiently analyze the expected dataset?

c. Are there any identifiable gaps in computing resources that would impede the timely production of publishable results?