CLAS12 software session

March 6, 2018

Outline

- Software group organization
- 12-GeV experimental computing review
- One year of software developments
- Data processing workflow
- Computing resources and infrastructure
- Documentation
- Work plan for upcoming months

Software in the CLAS Collaboration

At the November meeting:

- Ad-Hoc Committee composed by D. Ireland (Chair), D. Heddle and S. Kuhn charged to discuss the role of the Software Group in the Collaboration and the possibility of electing a Software Coordinator:
  - Software Coordinator should be created as a role within the CLAS Coordinating Committee (CCC)
  - A Software Working Group (SWG) should be constituted on the same level as the current three physics working groups
  - The Software Coordinator will be the chair of the SWG
  - All CLAS collaborators may be associated with this new group by opting in
- *Ad-Hoc Committee charged to elaborate a proposal on how to integrate these changes in the charter*
- *Text available at: https://www.jlab.org/Hall-B/secure/claschair/software/swg_change/*

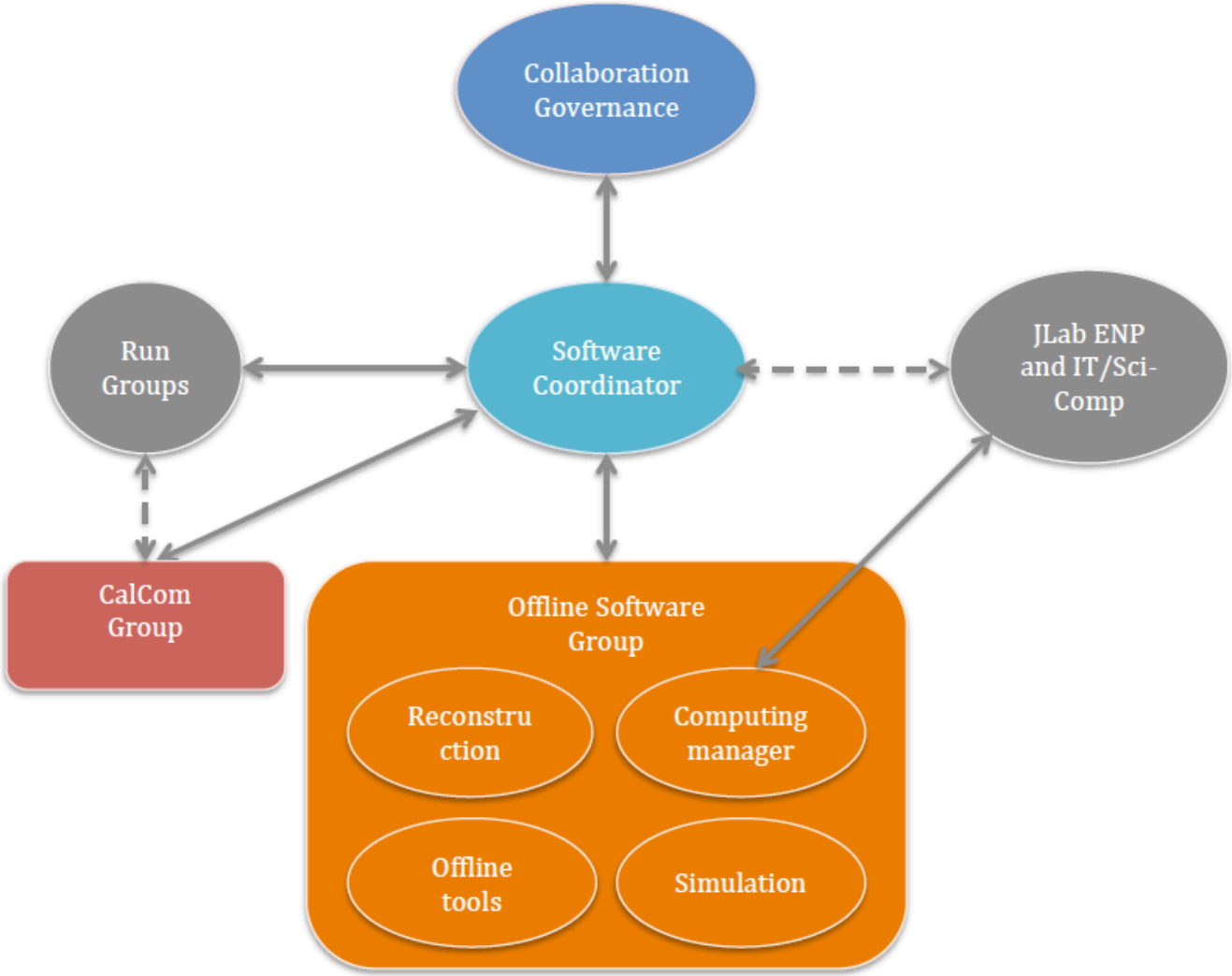
At this meeting:

- Text of charter change made available
- Discussion on proposed text and software coordinator election process on Friday:
  - Should the Software Coordinator be elected by the whole Collaboration or by the Software Group members?
- If endorsed, vote at the following meeting
- If change approved, proceed with the election of software coordinator in the summer to make the Software Coordinator take the position on September 1st

Software organization

Software Coordinator:

- liaison between the software group and the CLAS12 Run Groups
- liaison between the software group and the CLAS Collaboration Governance
- define priorities for the support of data taking, calibration, processing and analysis based on the input from the Collaboration Governance and the Run Groups
- define in collaboration with the CLAS12 Software Leaders the software development strategy and plan, addressing the identified priorities.
- liaison with the CalCom group for software and computing aspects of the calibration and commissioning tasks
- monitor the offline software development progress and propose corrective actions when needed



Software organization

- **Reconstruction (V. Ziegler)**: lead and supervise the development, implementation and validation of algorithms for CLAS12 reconstruction
- **Simulation (M. Ungaro)**: lead and supervise the development and maintenance of the CLAS12 simulation framework to support running onsite and offsite
- **Offline Software and Analysis Tools (G. Gavalian)**: lead and supervise the development and maintenance of common software (data formats, I/O, geometry, database, monitoring/calibration tools,) and analysis tools
- **Offline Computing Manager (N. Baltzell)**: lead and supervise the development of workflow and tools for data processing, define the strategy and plan for resource allocation providing guidance to Run Group chefs and analysis coordinators; be the liaison between Hall B/CLAS12 and JLab ENP and IT division, in particular Scientific Computing, for day-to-day operation

For all:

- Define in collaboration with the software coordinator and leads strategies and plans
- Provide guidance to users contributing to the development
- Maintain documentation

Software review

- JLAB 12 GeV software and computing review scheduled for November 27-28 2018
- Hall B/CLAS12 scheduled to give two talks:
  - Overview and Progress (Stepan): schedule for next 5 years, path to publication, computing requirements, etc...
  - Deep Dive (Raffaella, Graham): software status, software organization, recent successes, etc...

Recommendations (for the lab)

- Prepare to support increasing interest in machine learning and modern data science tools, possibly in collaboration with other labs to leverage existing solutions
- Consider increasing the central support for offsite resource access, especially for OSG and data transfers, leveraging work already done by GlueX and CLAS12 and at other laboratories

Observations (on the Hall B/CLAS12)

- About 20 bullets points
- Two in bold:
  - Run Group A was able to process (calibrate, reconstruct, produce DSTs) within three months of collecting data. Significant physics results are planned for fall 2019. First results on 10% of the data were shown at DNP2018 (Congratulations!)
  - Overall, CLAS12 is on track to produce timely and important science and is leveraging expertise and processes from other experiments and the lab itself

One year of software development

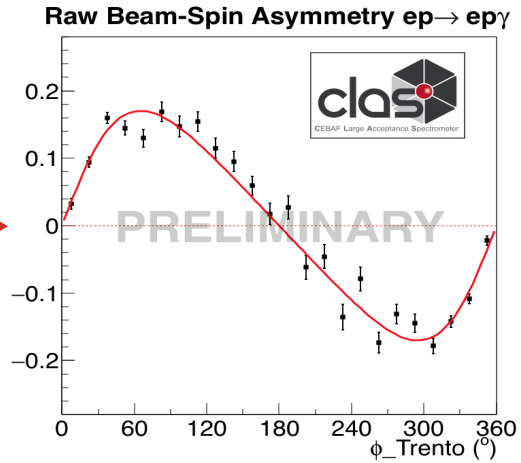
A very incomplete list…

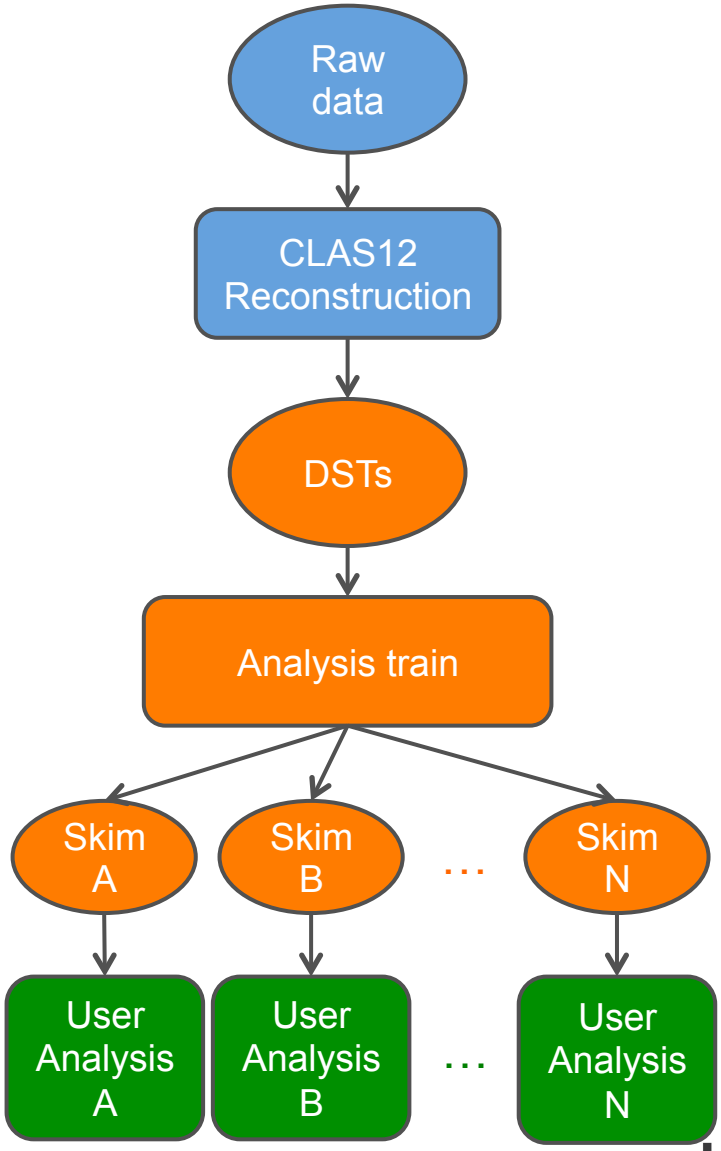
- Reconstruction:
  - Updates/improvements to all detector reconstruction packages
  - Major improvement of reconstruction speed (x3)
  - Completion of event building scheme (forward/central, charged/neutral particles, multiple software triggers)
  - Improved reconstruction performances (efficiency, resolution, …)
  - Roads development for tracking trigger
- Simulations:
  - Continuous updates/improvements to detector descriptions (geometry and response)
  - Docker-based distribution for easy deployment and running
  - Support to use offsite resources
- Offline tools:
  - Continuous support to calibration, reconstruction, analysis needs
  - Important upgrades to handle new bit-packed fADC and MM raw data
  - Implementation of analysis trains for skimming
  - Update of swimming package to support non-symmetric maps and shifts
  - New HIPO4 data format for increased performances and functionalities
- Software management and processing tools:
  - (More) robust software management policies and release validation
  - SWIF based workflow for massive decoding
- Framework (CLARA):
  - Implementation of thread affinity for improved performances and optimal utilization of farm nodes
  - (Countless) updates to provide new functionalities
  - Support to Hipo3 and Hipo4
  - Shared installation

Data processing workflow

- Input: raw EVIO files from DAQ
- Decoding to HIPO files, implementing translation tables and fADC pulse analysis
- Event reconstruction using COATJAVA packages running in CLARA and producing HIPO DSTs
- Use analysis trains to skim different event topologies and produce separate HIPO files
- Skimmed files distributed to users for physics analysis

Exercised for DNP





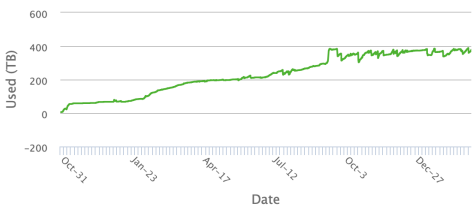
Computing resources and infrastructure

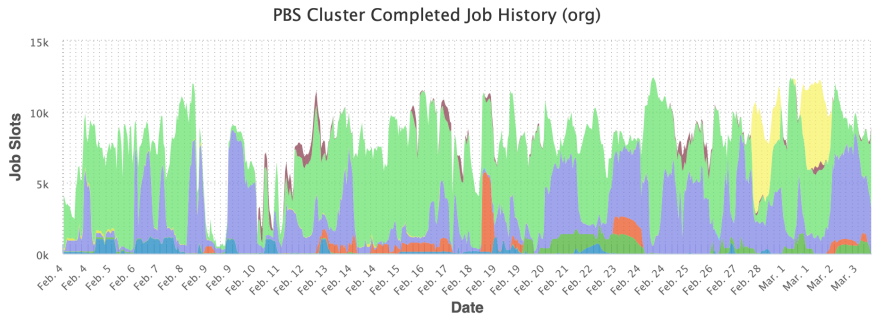
- Work in progress to improve the utilization of the allocated computing resources:
  - Disk space usage by Run Group and Users (cleanup, “virtual” quotas)
  - Use of SWIF workflows for smart use of farm and disk
  - Optimization of CLARA configuration for running on the farm
- Increase of computing resources in the near future:
  - 2X increase of disk space
  - Batch farm expansion
- Upgrade of batch farm software from PBS to SLURM

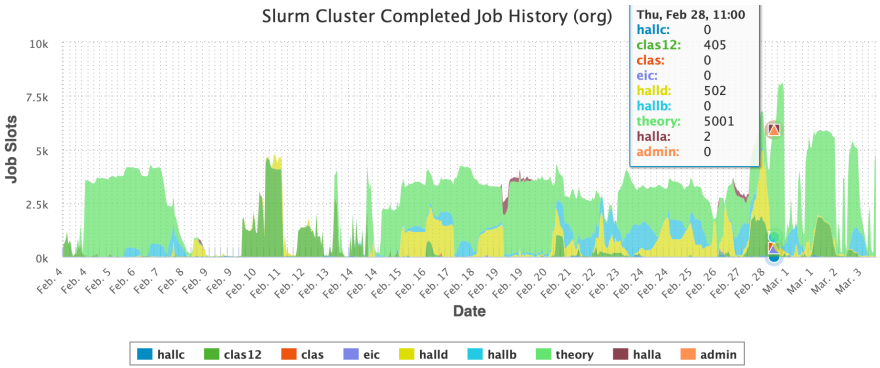
Scicomp New /work Disk Pool Usage

Project Name	Quota (GB)	Used (GB)
general	8,191	5,862
halla	51,199	29,319
hallb	230,399	205,353
hallb/clas	35,839	30,654
hallb/clas12	174,079	152,794
hallb/general	25,599	21,905
hallc	40,959	13,123
halld	165,868	133,593
	496,616	387,250

Last updated: 2019-03-04 17:48:37







Use of offsite resources

Additional computing power to support CLAS12 data analysis by exploiting offsite resources:

- Simulations:
  - Open Science Grid (OSG)
  - MIT computing farm
  - Support to GW users in utilizing local computing cluster
  - …
- Reconstruction:
  - Significant allocation at National Energy Research Scientific Computing Center (NERSC) for GlueX and CLAS12
- Software distribution:
  - Docker container transformed into singularity image
  - Share via CernVM File System (CVMFS)

Documentation

- New centralized software wiki:
  - Single “portal” for all relevant information and specific web or wiki pages
  - Intended for “official” CLAS12 software
  - Replace previous individual pages and links including obsolete CLAS12 wiki*
  - Work on updating and completing underlying documentation will continue

https://clasweb.jlab.org/wiki/index.php/CLAS12_Software_Center

*all relevant links moved elsewhere and page to be dismissed (detailed information will be circulated soon)

Work plan for next month

Focused on preparation for pass1 cooking

Data format:
- complete transition to Hipo4 (update of trains)
- first production release based on Hipo4

Geometry:
- Use target offset to shift CD detectors
- Implement alignment tables for FC detectors

Simulations:
- Infrastructure to submit simulations to MIT farm
- RG dependent geometries
- GEMC configuration in Evio file with Json format

Data processing tools:
- Extend SWIF workflow functionalities

Reconstruction updates:
- DC Tracking:
  - Beam x/y offsets validation
  - Update of trajectory bank
  - Validation of dc wire distortion
  - Tracking efficiency improvements (ongoing, will extend beyond march)
- CVT:
  - Efficiency studies and improvements
  - Use of alignment in tracking
- ECAL:
  - Logarithmic weighting in cluster position
  - Validation of moments calculation
  - Edge distance calculation for fiducial cuts
- TOF:
  - Position dependent TW for FTOF
  - Position dependent time correction for CTOF
- FT:
  - FTT reconstruction
  - FTC TW correction
- HTCC:
  - Hit bank with pointers between clusters and ADCs
- EB:
  - trajectory bank information used for hit-matching
  - track hit matching to allow many-to-one relations
  - save FT detector infos in REC:Calorimeter, REC::Scintillator banks
  - FT-based start time
  - Vertex correction to start time

Session agenda

CLAS Collaboration Meeting - CLAS12 Software

Convener: Raffaella De Vita (INFN - Genova)

Location: F113 - Remote connection via https://bluejeans.com/766870016

08:30 **Introduction and news** *25'*
Speaker: Raffaella De Vita (INFN - Genova)

09:00 **Reconstruction status, recent upgrades and development plans** *30'*
Speaker: Veronique Ziegler (Jefferson Lab)

09:35 **EB updates and development plans** *20'*
Speaker: Nathan Baltzell (Jefferson Lab)

10:00 **Coffee break** *30'*

10:30 **Clara: new features and shared installation** *20'*
Speaker: Dr. Vardan Gyurjyan (Jefferson Lab)

10:55 **Data processing tools** *20'*
Speaker: Nathan Baltzell (Jefferson Lab)

11:15 **CLAS12 simulations: use of off site resources** *20'*
Speakers: Dr. Maurizio Ungaro (Jefferson Lab), Sangbaek Lee (MIT), Robert Johnston (MIT), Patrick Moran (MIT)

11:40 **The new HIPO4 format** *20'*
Speaker: Gagik Gavalian (Jefferson Lab)

12:05 **The CLAS12TOOL analysis package** *20'*
Speaker: Dr. Derek Glazier (University of Glasgow)

+Workshop on Friday!