CLAS12 Software Readiness Review - Replay

Jerry Gilfoyle (for the CLAS12 Software Group)

Physics Department, University of Richmond, Virginia

Outline:

0. History
1. Introduction
2. Software Framework
3. Management
4. Requirements $\leftarrow$ Main focus
5. Manpower
6. Conclusions
7. Follow-up

History - 1

Information Technology for the 12 GeV Era - Internal Review

May 20, 2011

Thomas Jefferson National Accelerator Facility, Newport News, VA

CIO Message

Colleagues,

The one day internal CIO review of most aspects of Information Technology that impact preparations for and initial running of the 12 GeV science program will take place Friday, May 20. This review is intended to get a good understanding of progress towards IT in the12 GeV era, and discover if there are areas that might need increased effort in the coming year. Thank you in advance for working this important area of the Lab. Please let me know if you have any questions or comments.

Thanks,

Roy

Review Panel:

Chip Watson, Chair (JLab, Deputy CIO)	Cortney Carpenter (W&M, CIO)
Graham Drinkwater (ATG)	Brad Sawatzky (Hall C IT)
Richard Jones (UConn, GlueX)	Karl Slifer (UGBOD IT rep)

History - 2

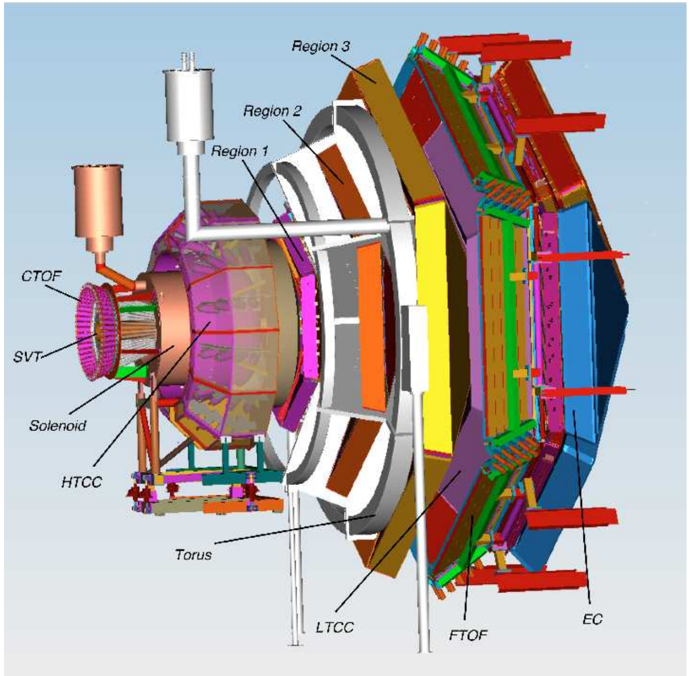
8:30	Welcome and Charge – Roy Whitney
8:40	Data Acquisition – Graham Heyes
9:00	Hall A Simulation & Analysis – Ole Hansen
9:10	Hall B Simulation & Analysis – Jerry Gilfoyle
9:35	Hall C Simulation & Analysis – Steve Wood
9:45	Hall D Simulation & Analysis – Mark Ito
10:10	Physics Summary – Graham Heyes
10:20	Break
10:35	Accelerator Controls – Matt Bickley
10:55	Accelerator Physics, FEL, future Light Source needs – Wes Moore
11:15	Scientific Computing and Lattice QCD – Sandy Philpott
11:35	Management Information Systems – Kari Heffner
11:45	Computer Networking & Infrastructure – Andy Kowalski
12:15	-- lunch --
4:00	Closeout

Introduction

- CLAS12 - Large acceptance spectrometer based on CLAS6.
- Luminosity increases by a factor of ten over CLAS6 ($L = 10^{35}\ cm^{-2}s^{-1}$).
- Software Goal:

  Ready to analyze data at turn on (October, 2014).

- Software development is far along.
- Planning has been ongoing.



	Forward	Central
θ_{track}	$5° - 40°$	$35° - 135°$
θ_{photon}	$2.5° - 40°$	
$\Delta p/p$	< 0.01	< 0.05
$\Delta\theta$	$< 1\ mr$	$< 10 - 20\ mr$
$\Delta\phi$	$< 3\ mr$	$< 3 - 5\ mr$

Software Framework

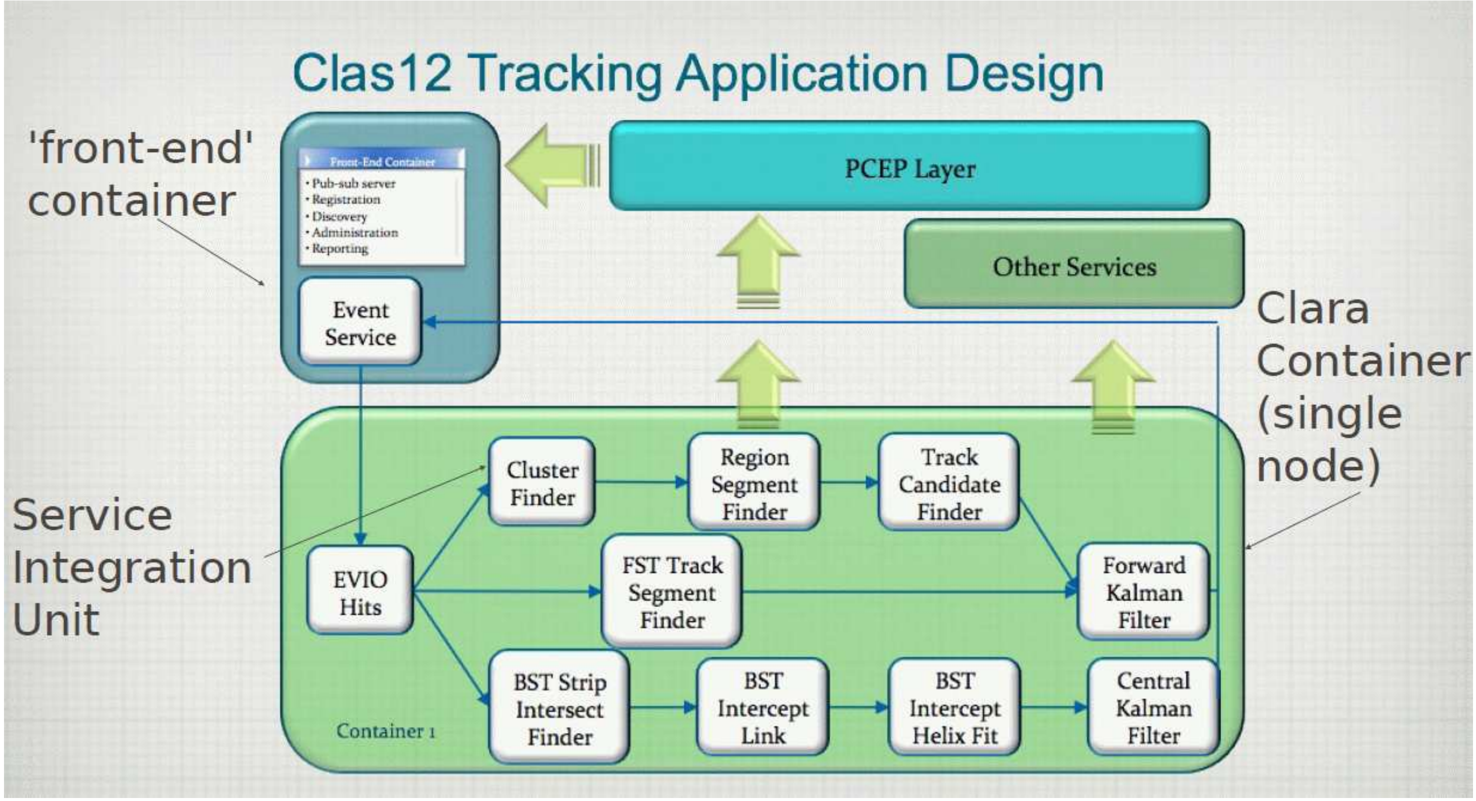
- CLAS12 Reconstruction and Analysis Framework (CLARA)
  - Service Oriented Architecture - build/maintain complex, distributed software system.
  - Example: CERN Technical Infrastructure Monitoring.
- SCons
  - Open Source software construction tool.
  - Improved, cross-platform substitute for Make.
- SVN
  - Open source software versioning and revision control.
  - Successor to CVS.
- Already adopted SCons and SVN for CLAS6.

CLARA

- Service Oriented Architecture (SOA) for physics data processing - multi-threaded, distributed environment.
- The fundamental unit is the ‘Service’ - physically independent software programs with a common interface.
- Services are loosely coupled, and may participate in a variety of algorithms.
- Interface.
  - Specifies a set of methods an object can perform but not the implementation of those methods.
  - Promote flexibility and reusability in code by connecting objects in terms of what they can do rather than how they do it.

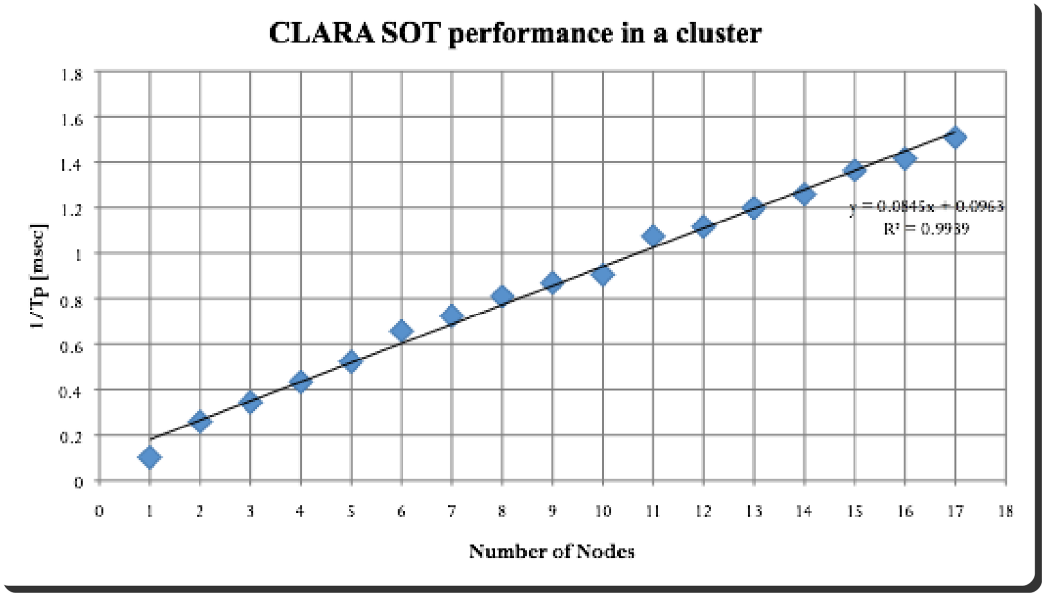
CLARA/SOA Example - 1

- Service Integration Unit - allows user applications to be presented as CLARA services.
- PCEP layer - Physics Complex Event Processing.
- Services can originate on different nodes.



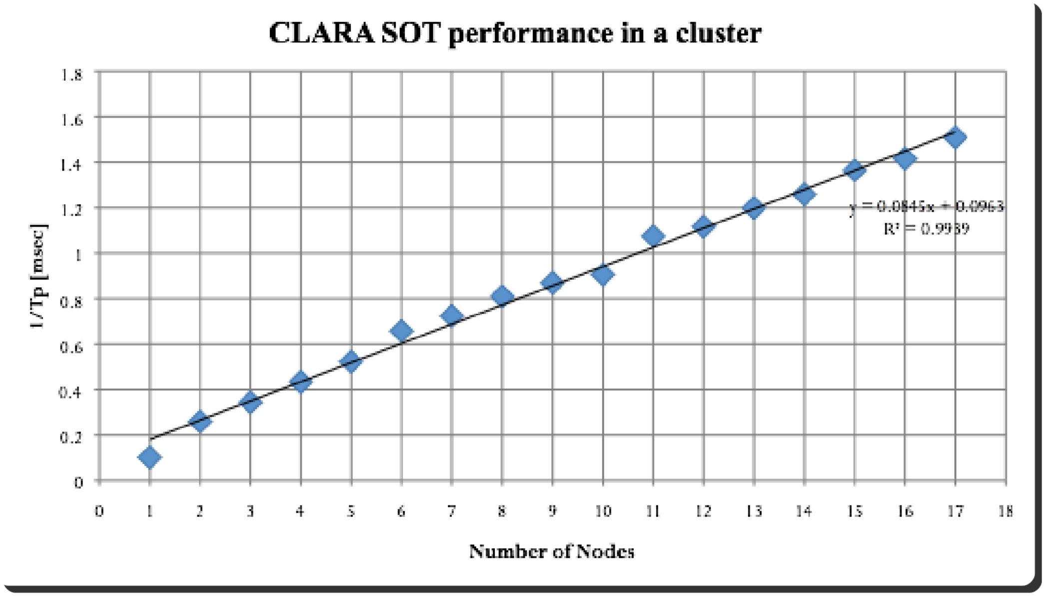
CLARA/SOA Example - 2

- Testing the CLAS12 tracking service.
- Tested on Spiderwulf - University of Richmond Nuclear and Astro-Physics Cluster: 17 nodes, Xeon, 2×6 Westmere nodes.
- Electron events generated from CLAS12 simulation *gemc*.



CLARA/SOA Example - 2

- Testing the CLAS12 tracking service.
- Tested on Spiderwulf - University of Richmond Nuclear and Astro-Physics Cluster: 17 nodes, Xeon, 2×6 Westmere nodes.
- Electron events generated from CLAS12 simulation *gemc*.



ONGOING ISSUE: CLARA access at JLab blocked by security barriers.

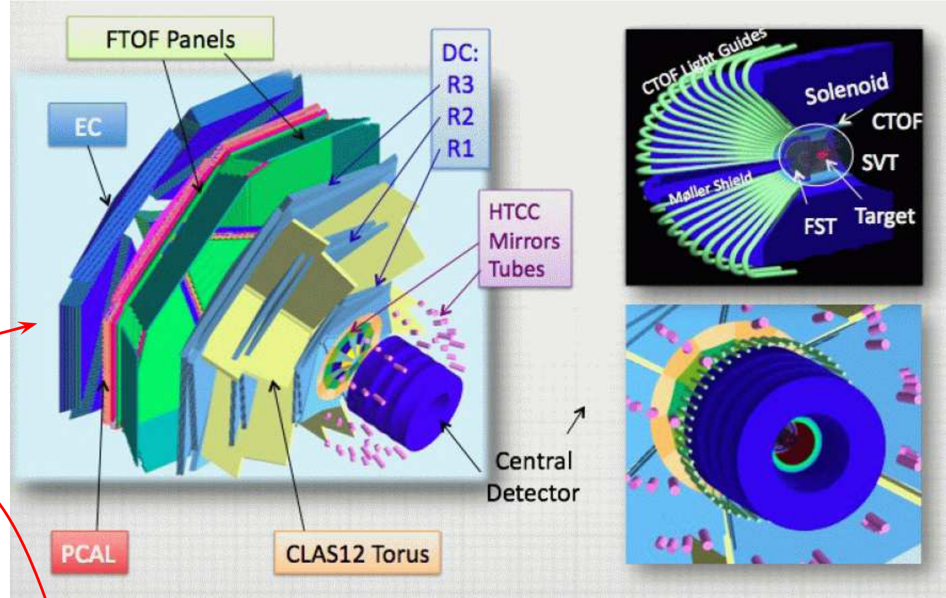
Management

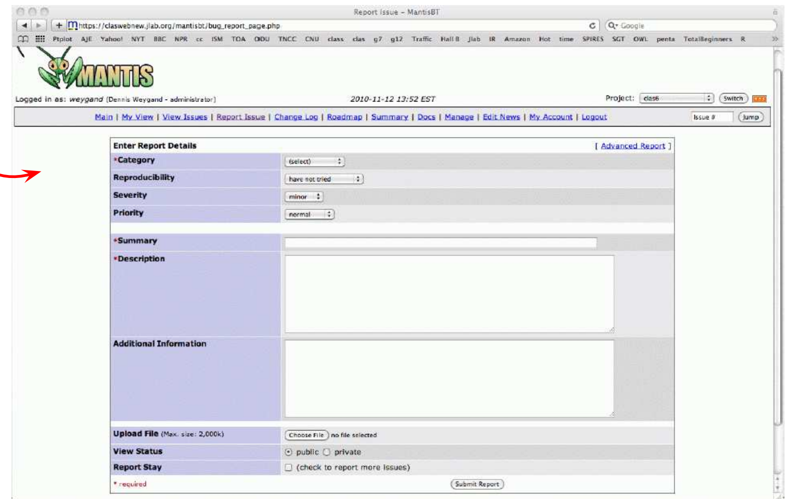
- CLAS12 Software Group (leader: Dennis Weygand).
- Wiki

- Tutorials to set up services in C++ and Java.
- Collaborations with Hall D and DAQ group.
  - JANA
  - Database
  - Event display
  - EVIO
  - cMsg

Management

- Tools:
  - Interfaces of calibration database to JANA, C++, MySQL, ...
  - Simulation: *gemc*
  - Bug reporting - Mantis
- Policies in place:
  - Regular builds of CLAS6 and CLAS12.
  - test histograms.
  - data challenges.





Requirements

- Calculations follow format from Graham Heyes.
- Assume an October, 2014 start date.
- Will present the major assumptions and results for:
  - data acquisition
  - calibration
  - simulation
  - reconstruction (formerly analysis in spreadsheet)
  - analysis

Ratio of Simulation:Reconstruction:Analysis

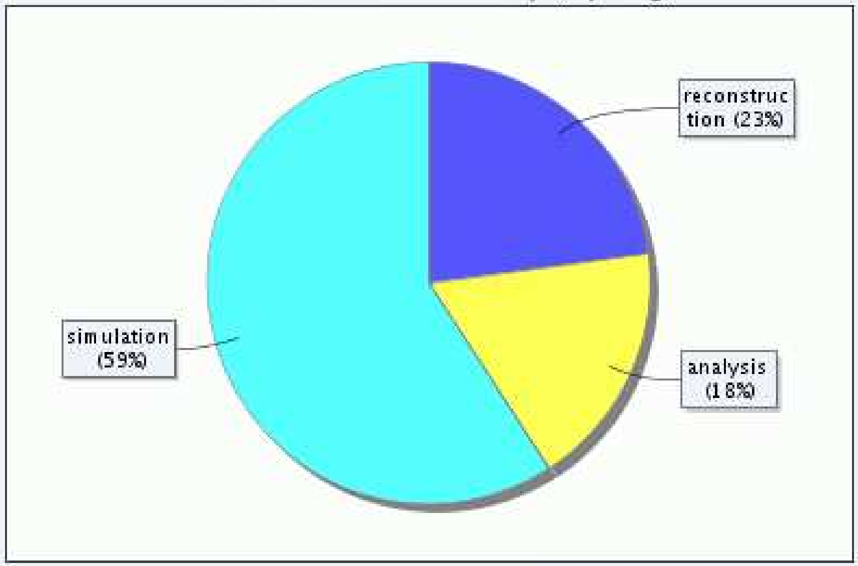
Process hours from Computer Center for the last year.

Project View | Job Type View | Job Status View | Job Walltime View

Farm Job Stat (Job Type View)

From: 2010-04-19 (yyyy-mm-dd) **To :** 2011-05-19

Process Hour Summary (by org)



Job Type	Total Job Count	Used Process Hours
debug	36,432	5,511.1
reconstruction	1,182,092	1,557,314
one_pass	9,441	1,647.5
analysis	2,002,900	1,221,338.2
test	36,432	5,511.1
simulation	1,046,276	4,033,710.9
Total	4,313,573	6,825,032.7

Last updated: Thursday May 19, 11:14:03 EDT 2011

Data Acquisition

Event rate	10 kHz	Weeks running	30
Event size	10 kBytes	24 hour duty factor	60%

$$\text{Data Rate} = \text{Event Rate} \times \text{Event Size} \tag{1}$$
$$= 100 \text{ MByte/s}$$

$$\text{Average 24-hour rate} = \text{Data Rate} \times \text{24-hour duty factor}$$
$$= 60 \text{ MByte/s}$$

$$\text{Events/yr} = \text{Event Rate} \times \text{Weeks Running} \times \text{24-hour duty factor} \tag{2}$$
$$= 1.1 \times 10^{11} \text{ Events/yr}$$

$$\text{Data Volume/yr} = \text{Events/yr} \times \text{Event size} \tag{3}$$
$$= 1100 \text{ TByte/yr}$$

Calibration - 1

CPU-time/event-core	67 ms	Desired processing time	20 min
Data set size	2 GB	Event size	10 kBytes
High-priority fraction	1%		

$$\text{Events/priority data set} = \frac{\text{Data set size}}{\text{Event Size}} = 2 \times 10^5 \text{ events} \tag{4}$$

$$\text{CPU time/priority data set} = \text{Events/priority data set} \times \text{CPU-time/event-core} = 1.3 \times 10^4 \text{ s} \tag{5}$$

$$\text{Cores/data set for priority} = \frac{\text{CPU time/priority data set}}{\text{Desired processing time}} = 11 \text{ cores} \tag{6}$$

Calibration - 2

CPU-time/event-core	67 ms	Events/yr	1.1×10^{11}
Data passes	5	Desired processing time	20 min
Data fraction used	10%	Output size/input set size	5
Data set size	2 GB	Event size	10 kBytes

$$\begin{aligned} \text{Output data set} &= \text{Data set size} \times \text{Output size/input set size} \\ &= 10 \text{ GByte} \end{aligned} \tag{7}$$

$$\begin{aligned} \text{Medium-priority CPU time/yr} &= \text{Events/yr} \times \text{CPU-time/event-core} \times \\ &\quad \text{Data fraction used} \times \text{Data passes} \\ &= 3.6 \times 10^9 \text{s} \end{aligned} \tag{8}$$

$$\begin{aligned} \text{Cores for non priority} &= \frac{\text{Medium-priority CPU time/yr}}{\text{one yr in seconds}} \\ &= 116 \text{ cores} \end{aligned} \tag{9}$$

Simulation - 1

CPU-sim-time/event-core	400 ms	Fraction to disk	2%
Sim-events/yr	7×10^{10}	Fraction to tape	10%
Output event size	50 kBytes		

$$\begin{aligned}\text{CPU-time/yr} &= \text{CPU-time/event-core} \times \text{Sim-events/yr} \\ &= 3 \times 10^{10}\ \text{s}\end{aligned} \tag{10}$$

$$\begin{aligned}\text{Dedicated farm cores} &= \frac{\text{CPU-time/yr}}{\text{one yr in seconds}} \\ &= 828\ \text{cores}\end{aligned} \tag{11}$$

$$\begin{aligned}\text{Work disk} &= \text{Sim-events/yr} \times \text{Output event size} \times \\ &\quad \text{Fraction to disk} \\ &= 65\ \text{TBytes}\end{aligned} \tag{12}$$

Simulation - 2

CPU-sim-time/event-core	400 ms	Fraction to disk	2%
Sim-events/yr	7×10^{10}	Fraction to tape	10%
Output event size	50 kBytes		

$$\begin{aligned}\text{Tape storage} &= \text{Sim-events/yr} \times \text{Output event size} \times \\ &\quad \text{Fraction to tape} \\ &= 326 \text{ TBytes/yr}\end{aligned} \tag{13}$$

$$\begin{aligned}\text{Average bandwidth} &= \frac{\text{Output event size} \times \text{Dedicated farm cores}}{\text{CPU-sim-time/event-core}} \\ &= 104 \text{ MByte/s}\end{aligned} \tag{14}$$

Reconstruction - 1

CPU-data-time/event-core	67 ms	Output size/input size	2
Data passes	1.7	Output fraction on work disk	10%
Event Size	10 kBytes	Events/yr	1.1×10^{11}
Data volume/yr	1.1 PBytes/yr		

$$\begin{aligned}\text{CPU time per yr} &= \text{Events/yr} \times \text{CPU-data-time/event-core} \\ &\quad \times \text{Data passes} \\ &= 1.2 \times 10^{10}\ \text{s}\end{aligned} \tag{15}$$

$$\begin{aligned}\text{Dedicated farm cores} &= \frac{\text{CPU time per yr}}{\text{one yr in seconds}} \\ &= 393\ \text{cores}\end{aligned} \tag{16}$$

$$\begin{aligned}\text{Cooked data to tape} &= \text{Data Volume/yr} \times \text{Data passes} \\ &\quad \times \text{Output size/input size} \\ &= 3700\ \text{TByte/yr}\end{aligned} \tag{17}$$

Reconstruction - 2

CPU-data-time/event-core	67 ms	Output size/input size	2
Data passes	1.7	Output fraction on work disk	10%
Event Size	10 kBytes	Events/yr	1.1×10^{11}
Data volume/yr	1.1 PBytes/yr	Dedicated farm cores	393

$$\begin{aligned} \text{Disk storage} &= \text{Cooked data to tape} \times \text{Output fraction on work disk} \\ &= 370 \text{ TByte} \end{aligned} \tag{18}$$

$$\begin{aligned} \text{Average bandwidth} &= \text{Event size} \times (1 + \text{Output size/input size}) \times \\ &\quad \frac{\text{Dedicated farm cores}}{\text{CPU-data-time/event-core}} \\ &= 176 \text{ MBytes/s} \end{aligned} \tag{19}$$

Post-Reconstruction Analysis

CPU-data-time/event-core	67 ms	Fraction of desired events	20%
Data passes	10	Work disk space	370 TBytes
Events/yr	1.1×10^{11}		

$$\begin{aligned} \text{CPU time per yr} &= \text{Fraction desired} \times \text{Events/yr} \times \\ &\quad \text{CPU-data-time/event-core} \times \\ &\quad \text{Data passes} \\ &= 1.5 \times 10^{10}\ \text{s} \end{aligned} \tag{20}$$

$$\begin{aligned} \text{Dedicated farm cores} &= \frac{\text{CPU time per yr}}{\text{one yr in seconds}} \\ &= 463\ \text{cores} \end{aligned} \tag{21}$$

Requirements Summary

	Cores	Disk (TByte)	Tape (TByte/yr)
DAQ	-	-	1100
Calibration	127	-	-
Simulation	828	65	327
Reconstruction	393	370	3700
Analysis	463	370	370
Sum	1811	805	5497

CLAS/CLAS12 Software Manpower (Preliminary)

	Function	Name
1	Management and Framework (CLAS)	Weygand, Gyurjyan, Heddle
2	Management and Framework (others)	Wolin, Lawrence, Abbott, Timmer, Lee
3	Core Developers (CLAS)	Ungaro, Gilfoyle, Wood, Procureur, Goetz
4	Developers (undergraduates)	Paul, Carbonneau, Frasier, Moog, Musalo, ...
5	Users	$\approx$ 10 FTEs listed in SoS statements

Names listed in rows 1-3 provide $\approx$ 5 FTEs focused on CLAS12 software.

Conclusions

- Software framework is being developed; considerable progress in last year - CLARA, svn, SCons.
- Management tools are in place and a core group exists - exploiting overlaps with DAQ and Hall D.
- DAQ - $\approx 10^{11}$ events/yr $\rightarrow 1$ petabyte/yr.
- Calibration - about 130 cores required.
- Simulation - 276 cores required with 109 TBytes/yr of tape storage.
- Reconstruction - about 400 cores required with 3.7 PByte/yr of cooked data to tape.
- Post-reconstruction analysis - about 460 cores required.
- Manpower is still limited to a small group of core developers.

Conclusions and Questions

- Software framework is being developed; considerable progress in last year - CLARA, svn, SCons.
- Management tools are in place and a core group exists - exploiting overlaps with DAQ and Hall D.
- DAQ - $\approx 10^{11}$ events/yr $\rightarrow 1$ petabyte/yr.
- Calibration - about 130 cores required.
- Simulation - 828 cores required with 91 TBytes/yr of tape storage.
- Reconstruction - about 400 cores required with 3.7 PByte/yr of cooked data to tape.
- Post-reconstruction analysis - about 460 cores required.
- Manpower is still limited to a small group of core developers.

- Ratio of simulated events to data collected?
- Speed of simulation?
- Effect of user computing resources?

Pre-Closeout Questions - 1

On Fri, May 20, 2011 at 2:13 PM, Chip Watson <watson@jlab.org> wrote:
For each hall:
1. Are there any online and/or offline readiness reviews planned?

Not at this time. We (the CLAS12 Software Group) have started to discuss holding another workshop in a year. We held a workshop last year where we received feedback from experts from BNL and FNAL, but there was no written report.

2. Justify the ratio of required tape storage to disk storage.

Experience with the existing farm.

Pre-Closeout Questions - 2

3. Explain the ratio of reconstructed (cooked) to raw event size.
What is added / dropped?

The factor of two we quote is a compromise between electron and photon running. For electron running in CLAS6 we typically have good electrons in about one-third of the triggers. We keep the raw data for these analyzed events and also derived quantities like drift times, deposited energies, uncertainties, etc which can be 1-2 times the raw event size. For example, each drift chamber wire will have a drift time derived from tracking which adds a large amount of data to each event since much of the event data comes from the drift chambers. Thus, for electron running the output events are much bigger than the input event size, but we only keep about one-third of the raw triggers. We end up keeping about the same volume of cooked data as the raw data. For photon running we keep a much higher fraction of the triggers and, as with electron running, write out the raw data plus derived quantities. This step again makes the final event size several times bigger than the raw event size.

Pre-Closeout Questions - 3

4. Are their budgets in place for online (counting house) computing infrastructure?

There is a budget of $100k allocated for equipping the counting house, but a detailed plan for the hardware has not been developed yet. It will done later when there is a clearer view of all the technologies (electronics, software, etc.) that will be needed in the counting house.

5. What capabilities will exist or are planned for remote researchers to participate in commissioning and running?

We will have monitoring capabilities similar to what is in place now with CLAS6. See, for example the page below which is one of the main monitoring tools and is accessible remotely. It monitors simple quantities like the size of each event and also does reconstruction on a subset of the data to obtain higher level results like the number tracks per event, hits per track, drift chamber residuals, TOF rates, and so on.

http://clasweb.jlab.org/cgi-bin/ONLINE_TIMELINE/timeline_frames.pl

Pre-Closeout Questions - 4

Vardan Gyurjyan developed a system for a more sophisticated monitoring (the E-counting house) for CLAS6, but it was not used much. The link below is to the site for this project.

http://clasweb.jlab.org/clasonline/rc/hallB/e-cr.htm

Report - 1

General Findings and Recommendations

- ‘no critical show-stoppers at this stage’.
- ‘Speaking generally, the software development groups for all halls appeared somewhat understaffed.’
- ‘None of the halls have online or offline readiness reviews planned at this time.’ Data challenge of entire online$\rightarrow$offline chain recommended.
- ‘Establish a more formal joint effort to ensure that analysis software is ready to meet the experimental requirements.’

Report - 2

Hall B Findings and Recommendations

- ‘Their plans and progress generally appear to be in good shape, however manpower was noted as a potential concern (in particular, there is a fairly small group of core software developers).’
- With regard to CLARA the committee had questions about reliability. ‘It was not clear if the security, bandwidth, latency, uptime, and other associated issues have been fully considered.’.
- ‘running CLARA with components distributed across a wide area network is not likely to be a particularly high performance choice’.
- ‘the cost of disk and tape will be non-negligible, and the presented plan does not yet include keeping a duplicate of raw data.’

Closeout - 1

Summary of Observations

- Effort in preparing talks is appreciated!
  - All presenters appeared open in their presentations and self evaluations
  - Presentations contain lots of useful details
- No show stoppers… everything is either in good shape, or could be put into good shape with appropriate actions
- Some areas need improvement in staffing and/or management

Closeout - 2

Concerns (1)

Halls High Level:

- 12 GeV computing requirements don't fit into constant effort of base program (not funded by 12 GeV project), and are assumed to be covered by the base program
  - don't have numbers to say in what year the 50-50-50 fails to meet requirements (but it will)
- Software: No common process for defining requirements, no common management structure
- 4 halls not sharing much software

Closeout - 3

Concerns (2)

- Hall B:
  - Distributed processing / analysis plans: creates dependencies on remote services (24/7 operations requirement to keep it all up?)

Halls do not yet have robust plans for testing and reviewing readiness to operate.

Identification of risks, and addressing risks, still needs to be done.

WAN

Plans for WAN upgrades / bandwidth availability above a second 10g link are fuzzy (budget if students watch lots of movies?); also plans for 2nd independent path

Videoconferencing is now a critical service: seeing who is speaking matters, following cursor is also useful (desktop sharing); self-support won't be adequate

Additional Slides

Simulation - 1b: Sim events = data events

CPU-sim-time/event-core	400 ms	Fraction to disk	2%
Sim-events/year	10^{11}	Fraction to tape	10%
Output event size	50 kBytes		

$$\begin{aligned} \text{CPU-time/year} &= \text{CPU-time/event-core} \times \text{Sim-events/year} \\ &= 4.4 \times 10^{10}\ \text{s} \end{aligned} \tag{22}$$

$$\begin{aligned} \text{Dedicated farm cores} &= \frac{\text{CPU-time/year}}{\text{one year in seconds}} \\ &= 1,381\ \text{cores} \end{aligned} \tag{23}$$

$$\begin{aligned} \text{Work disk} &= \text{Sim-events/year} \times \text{Output event size} \times \\ &\quad \text{Fraction to disk} \\ &= 109\ \text{TBytes} \end{aligned} \tag{24}$$

Simulation - 2b: Sim events = data events

CPU-sim-time/event-core	400 ms	Fraction to disk	2%
Sim-events/year	2×10^{10}	Fraction to tape	10%
Output event size	50 kBytes		

$$\text{Tape storage} = \text{Events/year} \times \text{Output event size} \times \text{Fraction to tape} = 544 \text{ TBytes/year} \tag{25}$$

$$\text{Average bandwidth} = \frac{\text{Output event size} \times \text{Dedicated farm cores}}{\text{CPU-sim-time/event-core}} = 173 \text{ MByte/s} \tag{26}$$