CLAS12 Software

D.P. Weygand

Thomas Jefferson National Accelerator Facility

Clas12

ClaRA
Simulation
- GEMC

CCDB
Geometry Service
Event Display
Tracking
- Socrat
- SOT
- Gen III

Event Reconstruction
Post-Reconstruction Data Access
- Data-Mining

Slow Controls
Documentation
- Doxygen
- Javadoc

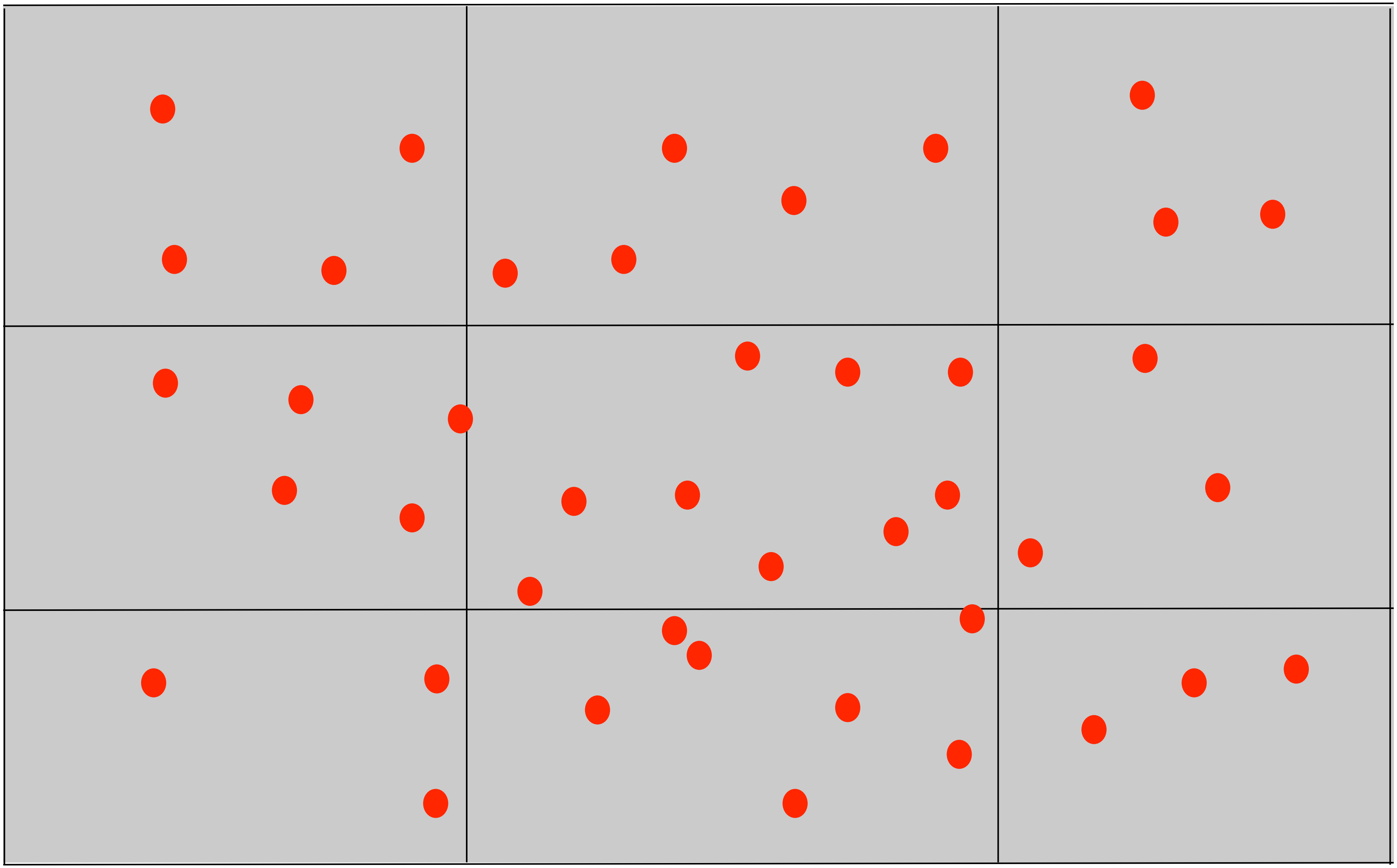
Testing/Authentication

Detector Subsystems
(Reconstruction and Calibration)
- EC
- PCAL
- FTOF
- CTOF
- LTCC
- HTCC

OnLine
Code Management
- SVN
- Bug Reporting
- Support

Visualization Services
Support Packages
Eg.
- CLHEP
- Root
- jHepWork

Why SOA? Software Entropy



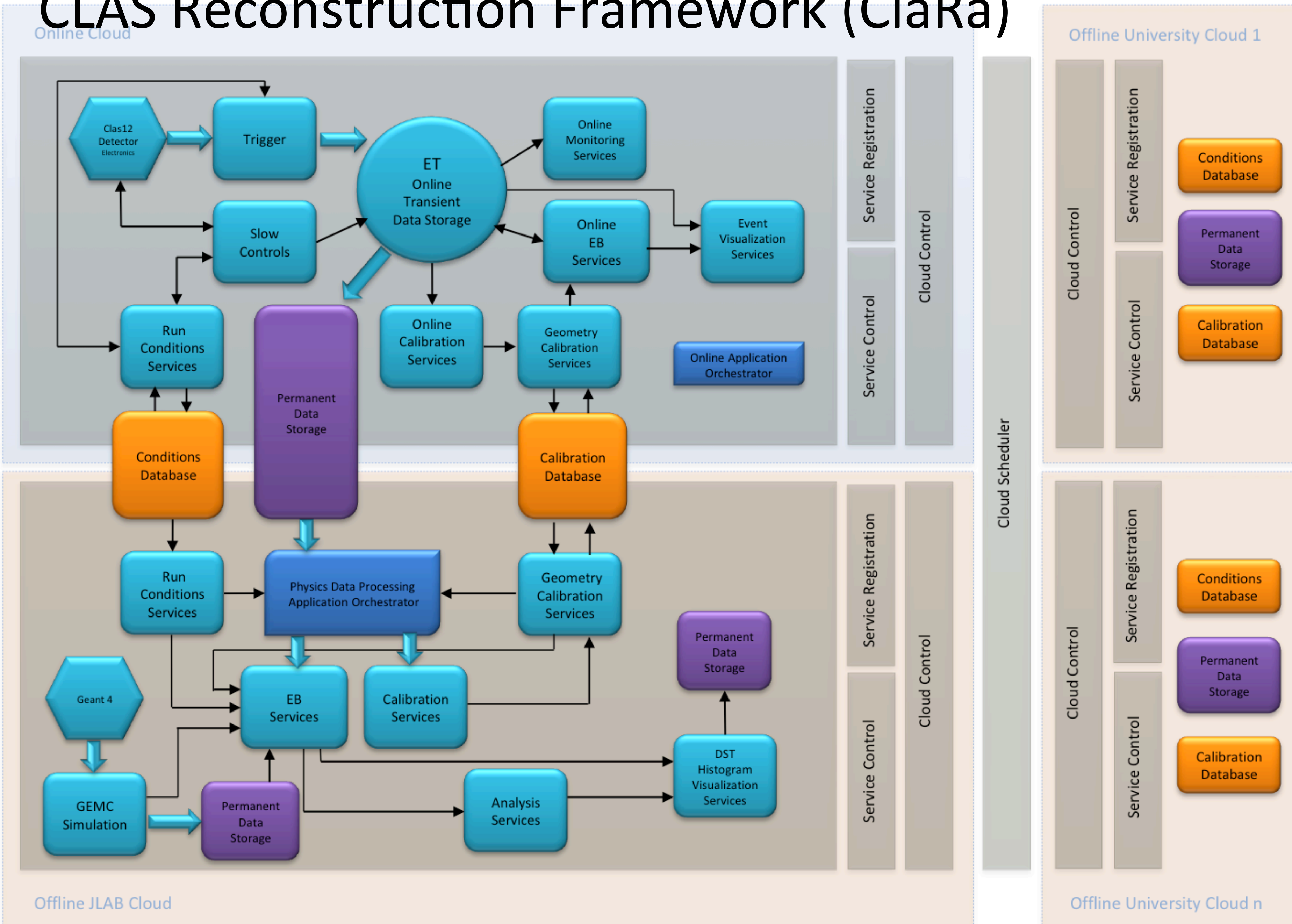
Computing Challenges and Complications

- Growing software complexity
  - Compilation and portability problems
  - Maintenance and debugging difficulties
  - Scaling issues
- Author volatility
  - Difficulty enforcing standards
  - Drop-out authors
- OS uniformity problems
  - Difficulty meeting the requirements by collaborating Universities
- Language uniformity problems
  - Contributions in various High Level Languages
- Steep learning curve to operate the software
  - Technology uniformity requirements
  - Users/contributors qualification
  - Coherent documentation software applications

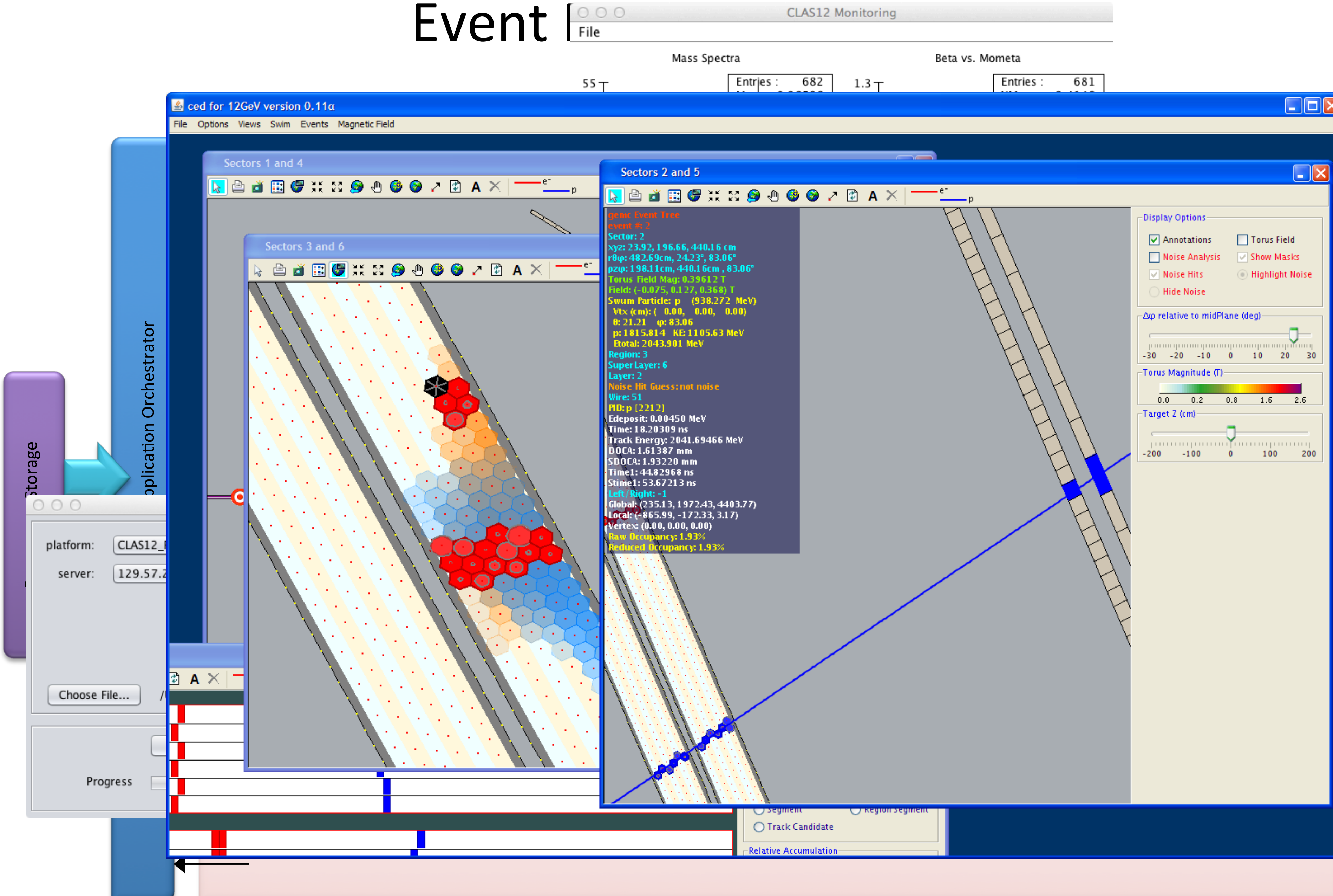
Long lifetime, evolving technologies

- Complexity through simplicity
  - Build complex applications using small and simple components.
- Enhance utilization, accessibility, contribution and collaboration
  - Reusability of components
  - Integration of legacy and /or foreign components
  - On-demand data processing.
  - Location independent resource pooling.
- Multi-Threading
  - Effective utilization of multicore processor systems.

CLAS Reconstruction Framework (ClaRa)



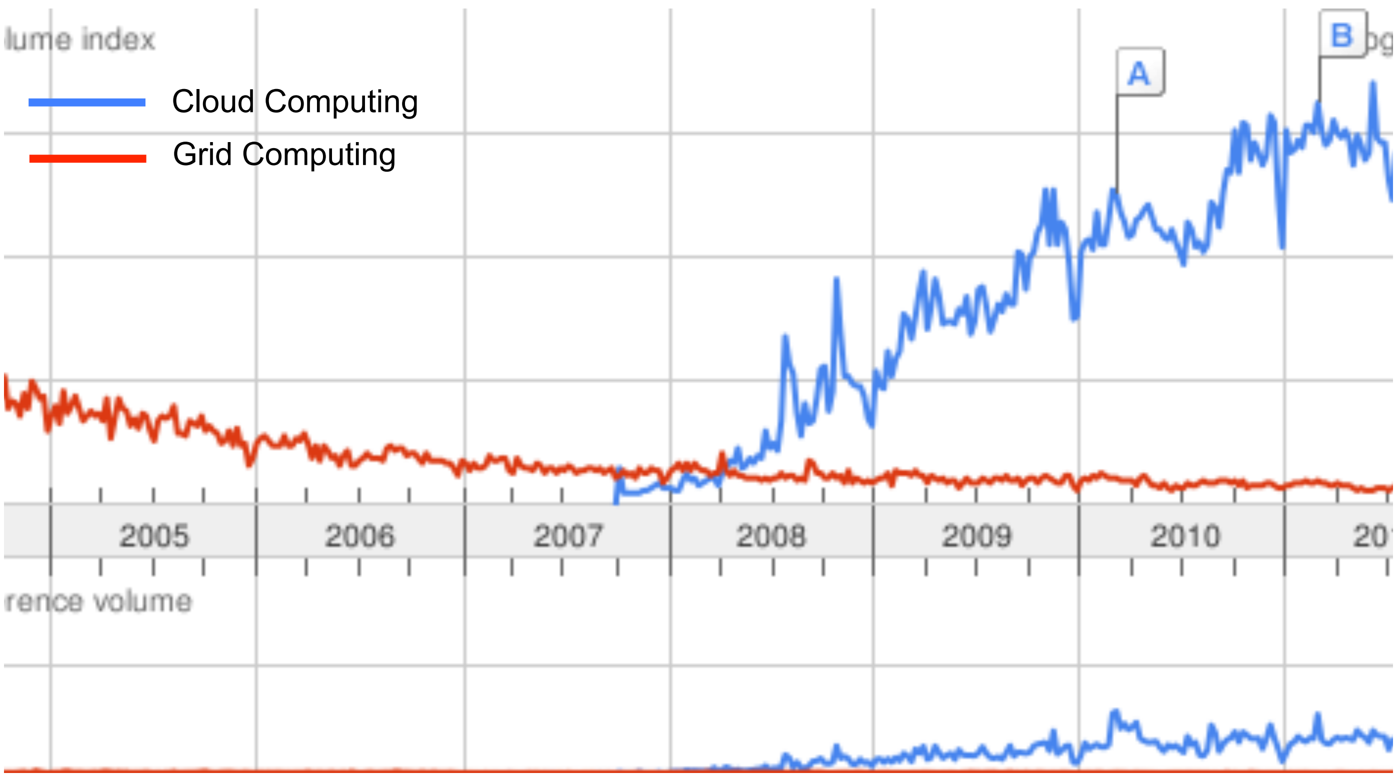
Event



CLAS12 Monitoring
File
Mass Spectra
Beta vs. Mometa
Entries : 682
Entries : 681
ced for 12GeV version 0.11α
File Options Views Swim Events Magnetic Field
Sectors 1 and 4
Sectors 3 and 6
Sectors 2 and 5
gemc Event Tree
event #: 2
Sector: 2
Torus Field Mag: 0.39612 T
Swum Particle: p (938.272 MeV)
Region: 3
SuperLayer: 6
Layer: 2
Noise Hit Guess: not noise
Wire: 51
PID: p [2212]
Edeposit: 0.00450 MeV
Time: 18.20309 ns
Track Energy: 2041.69466 MeV
DOCA: 1.61387 mm
SDOCA: 1.93220 mm
Time1: 44.82968 ns
Stime1: 53.67213 ns
Left/Right: -1
Vertex: (0.00, 0.00, 0.00)
Raw Occupancy: 1.93%
Reduced Occupancy: 1.93%
Display Options
Annotations
Torus Field
Noise Analysis
Show Masks
Noise Hits
Highlight Noise
Hide Noise
Torus Magnitude (T)
Target Z (cm)
Track Candidate
Relative Accumulation
Storage
Application Orchestrator
platform: CLAS12_
server: 129.57.2
Choose File...
Progress

Grid vs. Cloud Computing

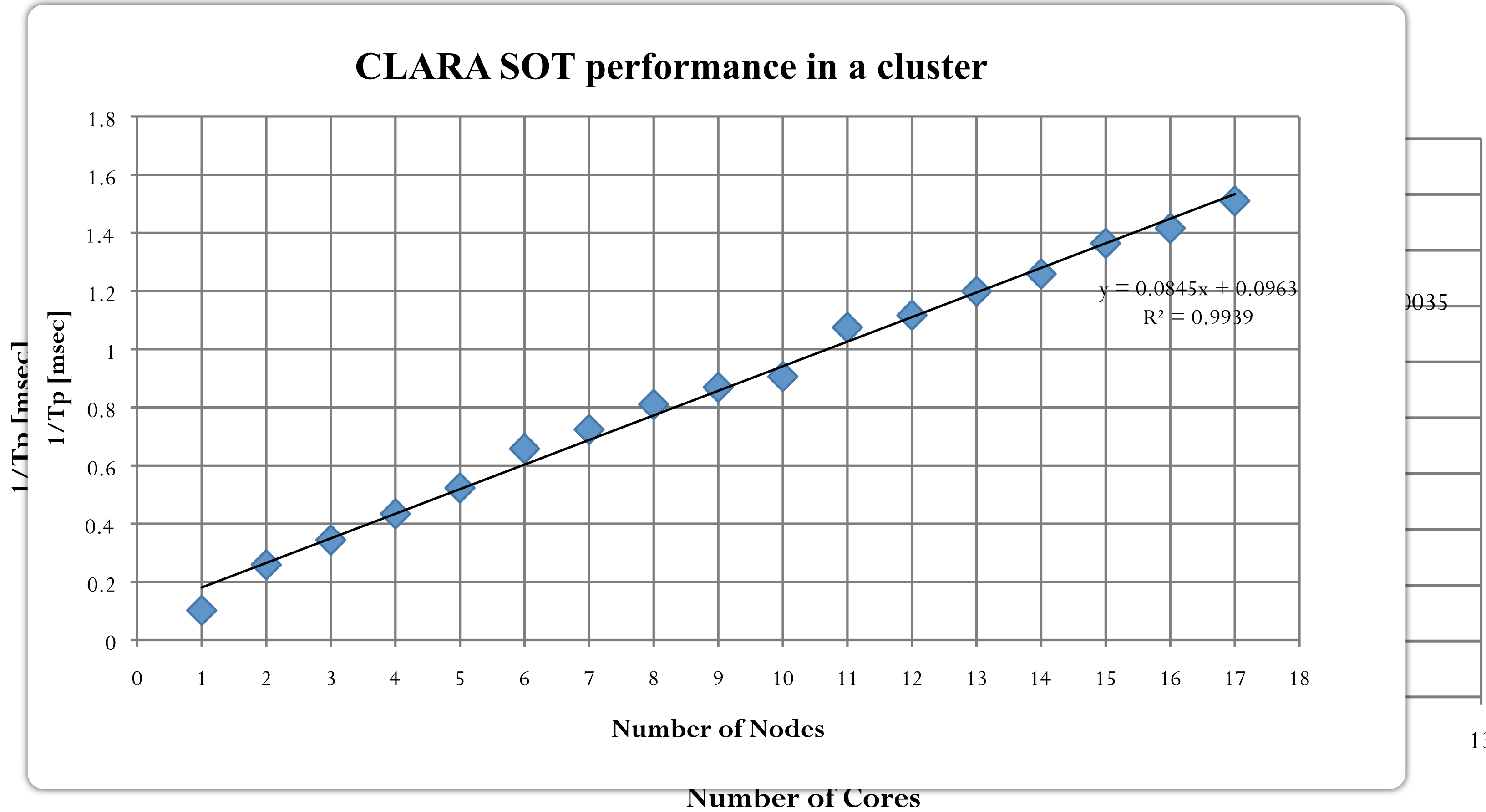
Google Trends



Google Trends.
Scale is based on the average worldwide traffic of **cloud computing** in all years.

MultiThreading with one 1 Task Service (SOT) running in a single node

Parallel distributed processing 17 node Xeon 2x6 Westmere CPU



9 8

Risks: SOA

- **Service communication latency**
  - **careful analysis of the criticality of a service, introduce built-in tolerance for variations in network service response times**
- **Evolutionary development: Building and updating continuously.**
  - **Management and administration. Strict service canonization rules**
- **The workloads of different clients may reside concurrently on the same service that can potentially introduce “pileups”.**
  - **Solve by introducing service access policies**
- **Network security**
  - **Client authentication and message encryption**

Event Reconstruction

Reconstrucion

Calibration

Tracking	Central part of the event reconstruction
EC/Pcal	e γ
TOF	K/π/p discrimination
HTCC	π/e discrimination
LTCC	K/π discrimination

Tracking

Reconstruction

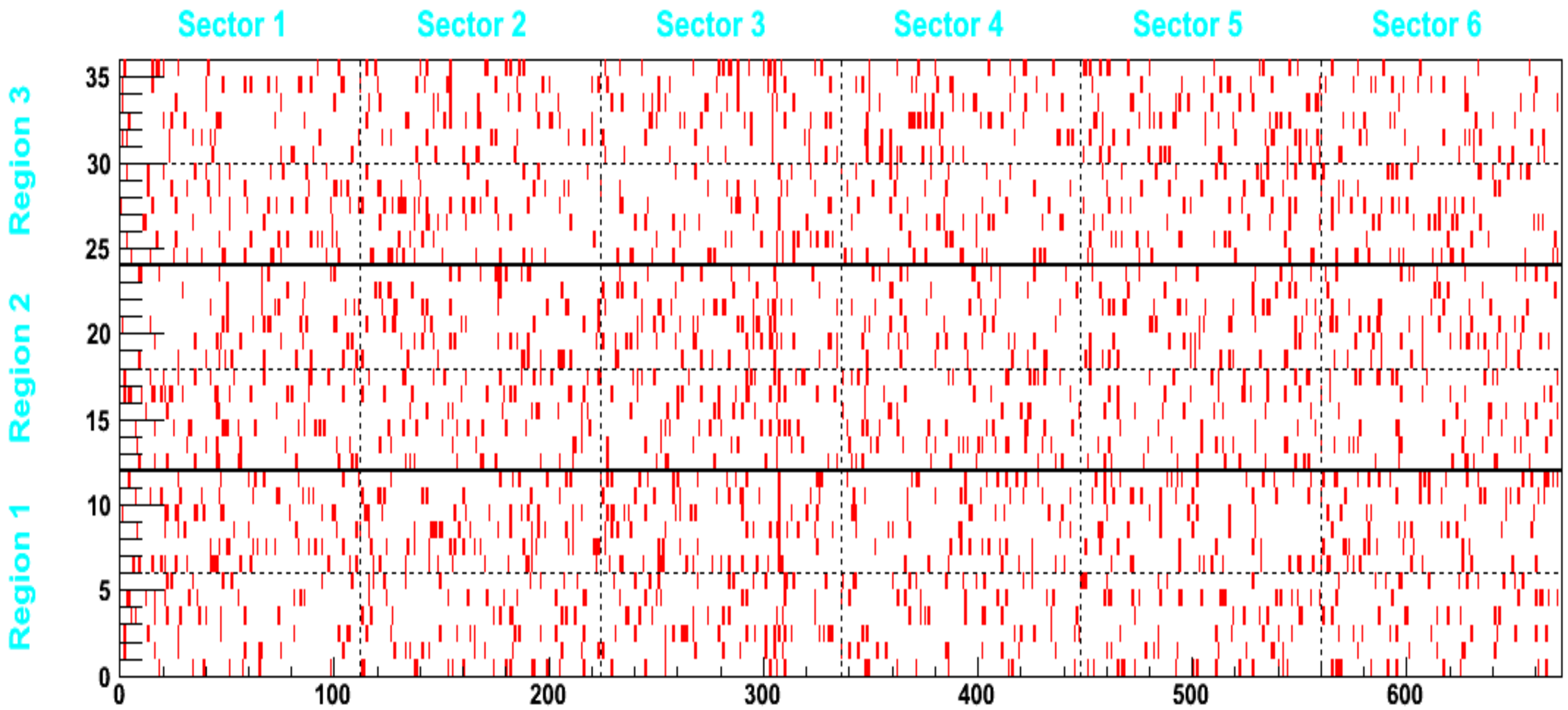
- No Legacy Tracking code from CLAS6
- First version was ROOT-based code: Socrat
- Socrat was rewritten (OO & SOA) → SOT
- Third generation tracking code currently under development

Calibration

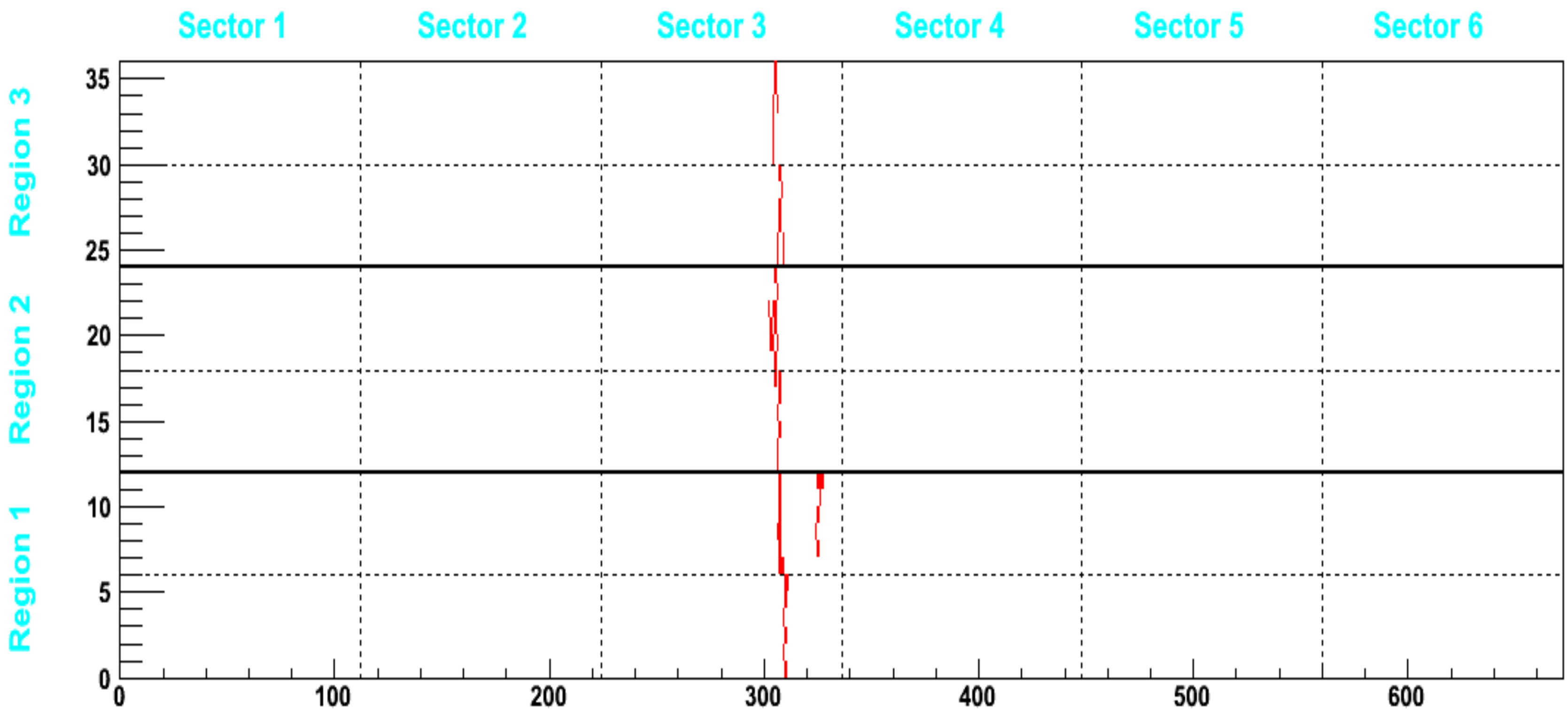
- Algorithms will be based on CLAS6 legacy
- Will be converted to Object-Oriented language and SOA compliant

Reconstruction in DC

Starting point (uncorrelated background, just for illustration):



Track finding

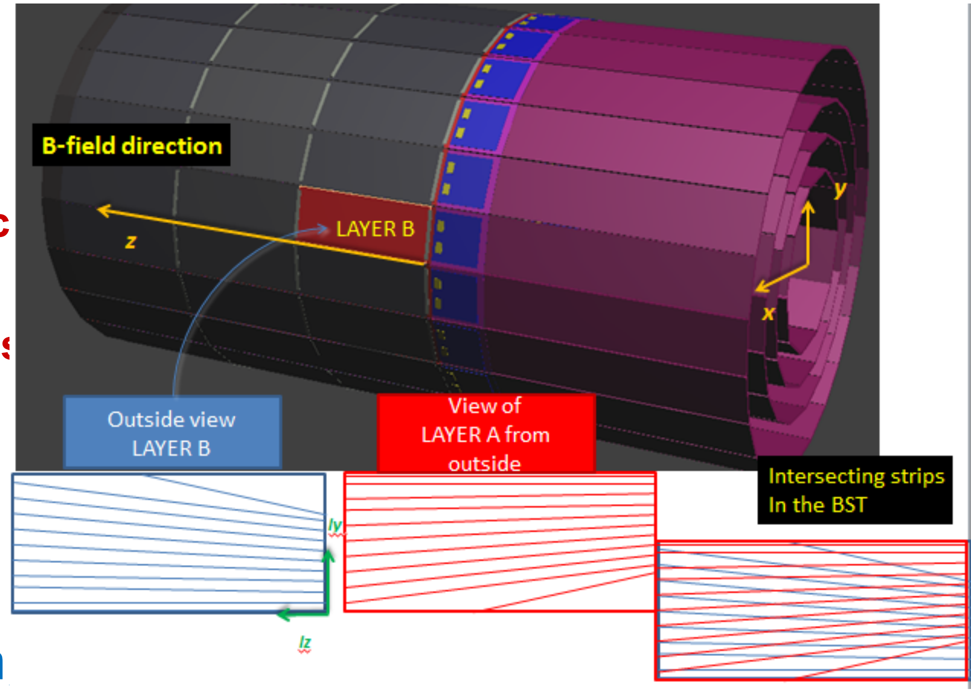


Corresponding structures in Socrat:
DChit, DCcluster, DCTrackSegment, DCTrackCandidate

3rd Tracking Strategies

➢ Central Tracking

- **Hit Recognition**
  - Central tracker strip ``intersection” determination
    - not on the same plane
    - i.e. trajectory-dependent
    - hence use iterative algorithm to improve hit position accurac based on track’s angle of intersection with BST plane



- **Pattern Recognition**
  - Hough Transform (✓)
  - Geometrical linking algorithm

    Look-up tables, … (to be implemented and tested)
- **Track Fitting**
  - Kalman Filter (✓)
  - Global Fitting Methods (to be implemented and tested)

EC/PCal

Reconstruction
- Legacy code from CLAS6 rewritten → SOA/Java
- Currently part of the reconstruction suite

Calibration
- Algorithms based on CLAS6 legacy
- Converted to Object-Oriented language
- SOA compliant

FTOF/CTOF

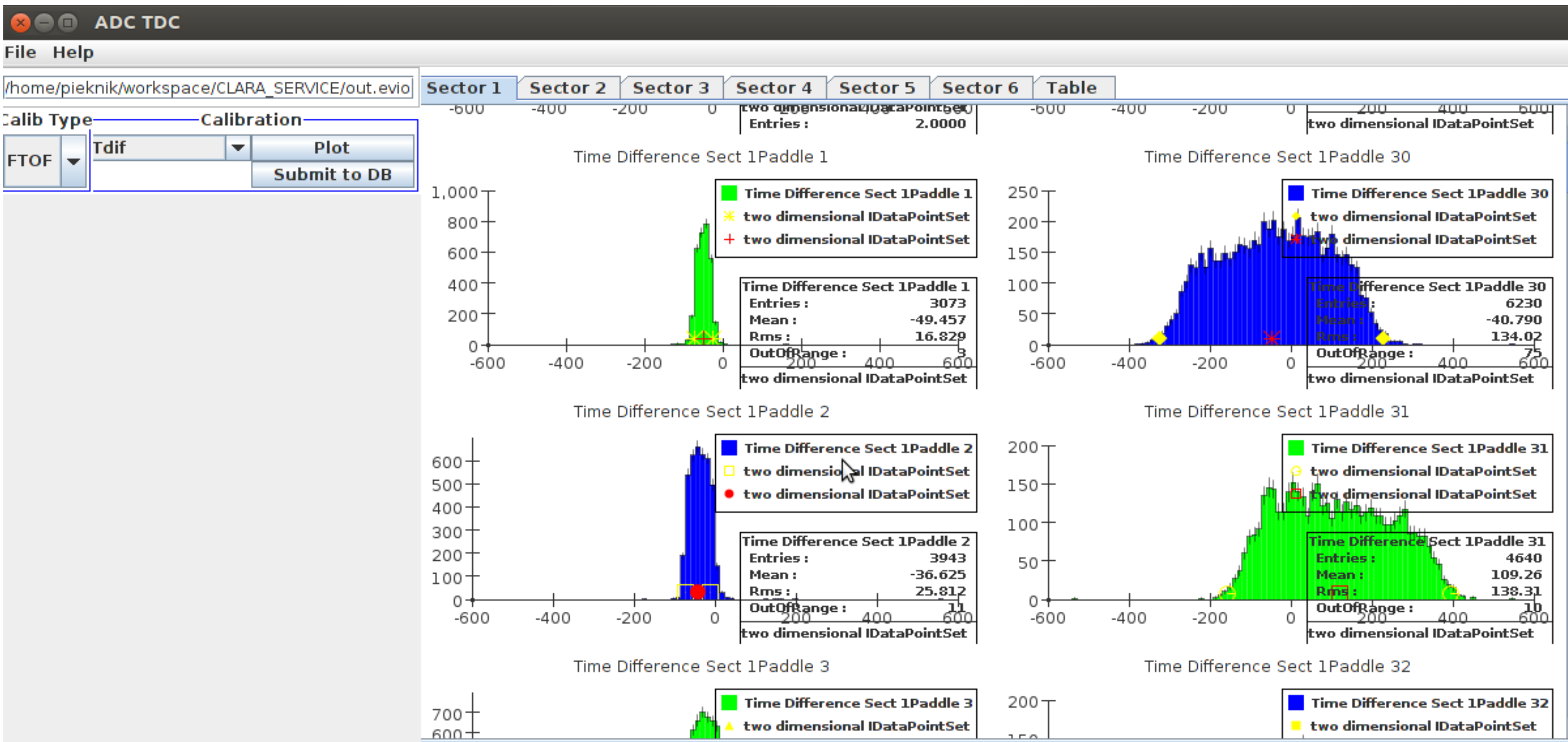
Reconstruction

- Legacy Algorithms from CLAS6
- Currently part of the reconstruction suite

Calibration

- Algorithms based on CLAS6 legacy
- Converted to Object-Oriented language
- SOA compliant

Sample from FTOF/CTOF Calibration Suite



HTCC/LTCC Software Status

No CLAS6 Legacy Reconstruction Code

- **HTCC simulation in GEMC is completed**
- **LTCC in GEMC is close to ready (M. Ungaro)**
  - **Software will be similar to what already exists for HTCC**
- **HTCC reconstruction code (clustering algorithm) is fully tested and working as a "plugin" to the "COAT" libraries**
  - **Fully tested with simulated data (gemc)**
    - **"Intermediate digitization": sector, ring, half, #ph.e., time**
      - **Need "Full digitization": crate, slot, channel, ADC, TDC to develop calibration procedure**
    - **Behavior as expected**
  - **PID performance evaluated (consistent with previous estimates from GEANT3)**

Place Holder for GEMC 3-4 slides and movie theme music

Simulation/GEMC

One of our most mature software projects

Code Management

Subversion:

Open source software community's replacement for cvs.

Has many of the same features and employs the same no-lockout paradigm.

Plug-ins are available for the popular integrated development environments, such as the widely used eclipse.

This allows one to check in, check out, track changes, and merge differences with mouse-clicks in a development environment rather than through a command line

Used by other Halls, DAQ

Code Release

GEMC: Healthy release cycle

Clara platform: Trivial to deploy

Reconstruction suite currently being patched

Universal Software Tools

Root
CLHEP: C++ tools for HEP

jHepWork/freeHep
COLT: Java implementation of CLHEP
JAIDA
jMath: CLAS java mathematical/physics tools

NumPy: Python Mathematical tools
SciPy: Python Numerical tools

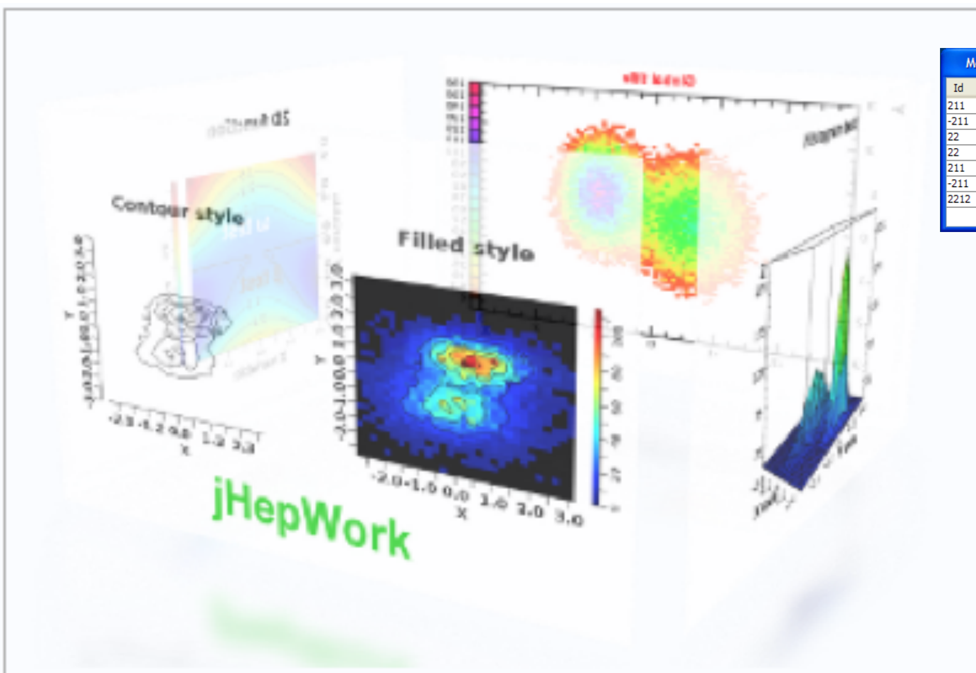
Geant4: Particle tracking through material

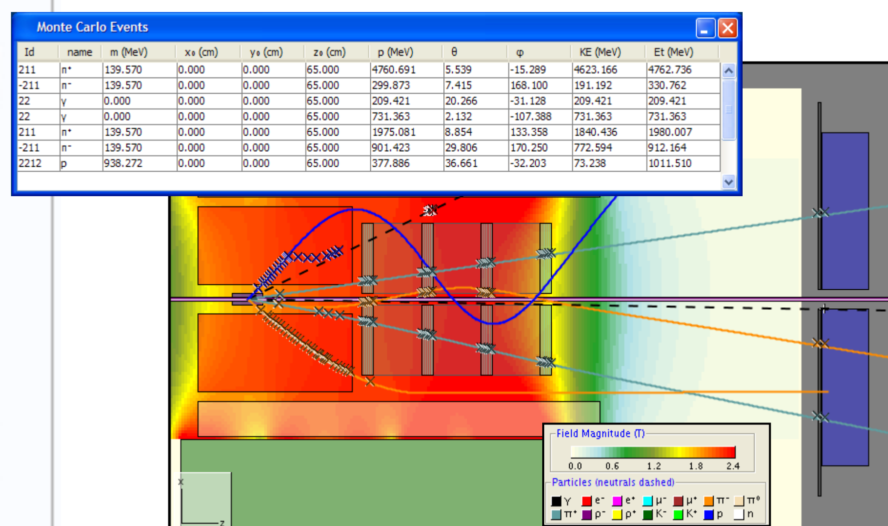
CCDB: mysql based calibration and conditions (and geometry) database (Shared with GlueX)

HDF5: NCSA developed data format

Standard IDE's: Eclipse and NetBeans

bCNU/jevio Event Display
(Shared with GlueX)





Quality Assurance and Authentication

As in the current CLAS software system, the standard reconstruction suite of services will be built daily, and the reconstruction package tested against a set of standard datasets. The output will be reviewed by a software package and checked against a standard, and crucial indicators of the reconstruction can be stored in a database and tracked over time. Dramatic changes in the program performance can then be easily identified, and with the software tracking provided by the Subversion code repository unanticipated code changes can be identified. In addition, at any level the individual code developer will be able to check any version against the standard suite.

At least three types of data: the first is pure simulation, that is monte carlo generated data through the CLAS12 detector without any detector resolution included. Reconstruction of this data set should return exactly what was input; any deviation is suspect and cause for special consideration. The second set of standard data will be a persistent monte carlo data set with full simulation, whose results should remain consistent with input parameters. Finally, varied sets of actual data, fully testing as completely as possible all aspects of the reconstruction software, will be utilized to track the code development.

Access to the Collaboration

Developers:

Java and C++ Abstract Classes allow simple implementation of services

"COAT" plugins allow service implementation into the reconstruction chain

Template examples in the repository

Data Analysis

Deployment of a Clara Reconstruction Platform

Reconstruction analysis will imply interaction with a cloud/clouds

Access to data: Building on ODU data mining application

Documentation

Primary Documentation is via CLAS wiki

Class docu
doxyge
Javadc

Bug Tracking v

Logged in as: *weygand* (Dennis Weygand - administrator) 2012-05-14

Main | My View | View Issues | Report Issue | Change Log | Roadmap | Summ

Unassigned [^] (1 - 1 / 1)

0000054	Lund files and luminosity [clas12] General - 2012-03-19 11:38

Resolved [^] (1 - 10 / 14)

0000051	make sure numbers are written on the mysql DB [clas12] gemc - 2012-03-05 08:13
0000006	Build system is out of date with documentation, and I am clueless [clas12] gemc - 2012-02-24 11:03
0000043	Need Paraboloids in GEMC [clas12] gemc - 2012-02-24 11:00
0000050	Reached limits of 40 max hierarchy for <>. Exiting. [clas12] gemc - 2012-02-21 10:47
0000048	name ordering in composition [clas12] gemc - 2012-02-03 10:32
0000033	length of $detector{"type"} is not large enough [clas12] gemc - 2011-03-30 10:44
0000029	problem with gemc_evio2root [clas12] gemc - 2011-03-04 13:07
0000027	non integer pid [clas12] gemc - 2011-02-22 19:41
0000020	harmless duplicate material definition in clas12/geo/sector/sector.pl [clas12] gemc - 2011-02-16 17:55
0000019	option on gemc website [clas12] gemc - 2011-02-14 13:47

CLAS12 CPU/Storage Requirements

	Cores	Disk (TByte)	Tape (TByte/yr)
DAQ	-	-	1100
Calibration	127	-	-
Simulation	888	65	327
Reconstruction	393	370	3700
Analysis	463	370	370
Sum	1871	805	5497

CLAS12 CPU/Storage Requirements

		Units	2015 CLAS12
1		Units	2015 CLAS12
2	**Acquisition**		
3	Event rate	kHz	10
4	Event size	k bytes	10
5			
6	Weeks running	weeks	30
7	24 hour duty factor	%	60.00%
8			
9	Data Rate MB/S	MB/s	100
10	Average rate over 24 hours MB/S	MB/s	60
11	Raw events per year	Evts/yr	1.1E+11
12	Raw data volume per year	TB/yr	1088.64

Reconstruction Analysis		
Time to process one event/core	ms	67
Number of passes through data		1.70
Output event size/input size	ratio	2.00
Percentage of output on work disk		10%
CPU time per year	s	1.2E+10
Dedicated farm cores		393
Cooked data to tape	TB/yr	3701
Disk storage, calculated	TB	370
Average bandwidth	MB/s	176

	Cores	Disk (TByte)	Tape (TByte/yr)
DAQ	-	-	1100
Calibration	127	-	-
Simulation	888	65	327
Reconstruction	393	370	3700
Analysis	463	370	370
Sum	1871	805	5497

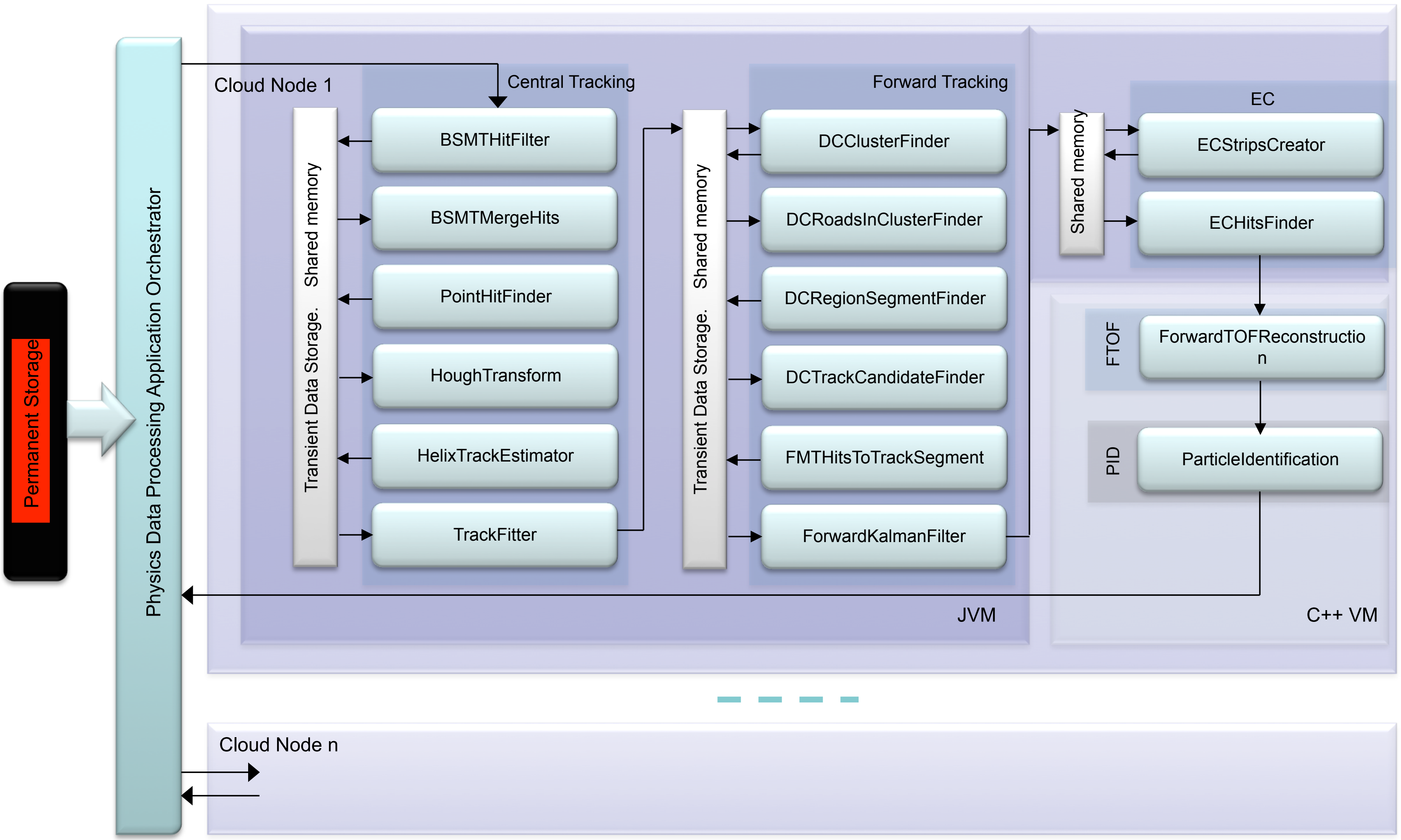
36	Output event size	k bytes	50
37	Fraction stored to disk	%	2%
38	Fraction stored to tape	%	10%
39	Multiplicity		1
40	CPU time per year	s	2.8E+10
41	Dedicated farm cores		888
42	Work disk calculated	TB	70.
43	Tape storage calculated	TB/yr	350.
44	Average bandwidth	MB/s	111

Core Summary		
DAQ	Cores	
Calibration	Cores	127
Simulation	Cores	888
Reconstruction Analysis	Cores	393
Post-Reconstruction Analysis	Cores	463
Physics Analysis	Cores	380
Sum	Cores	2,251

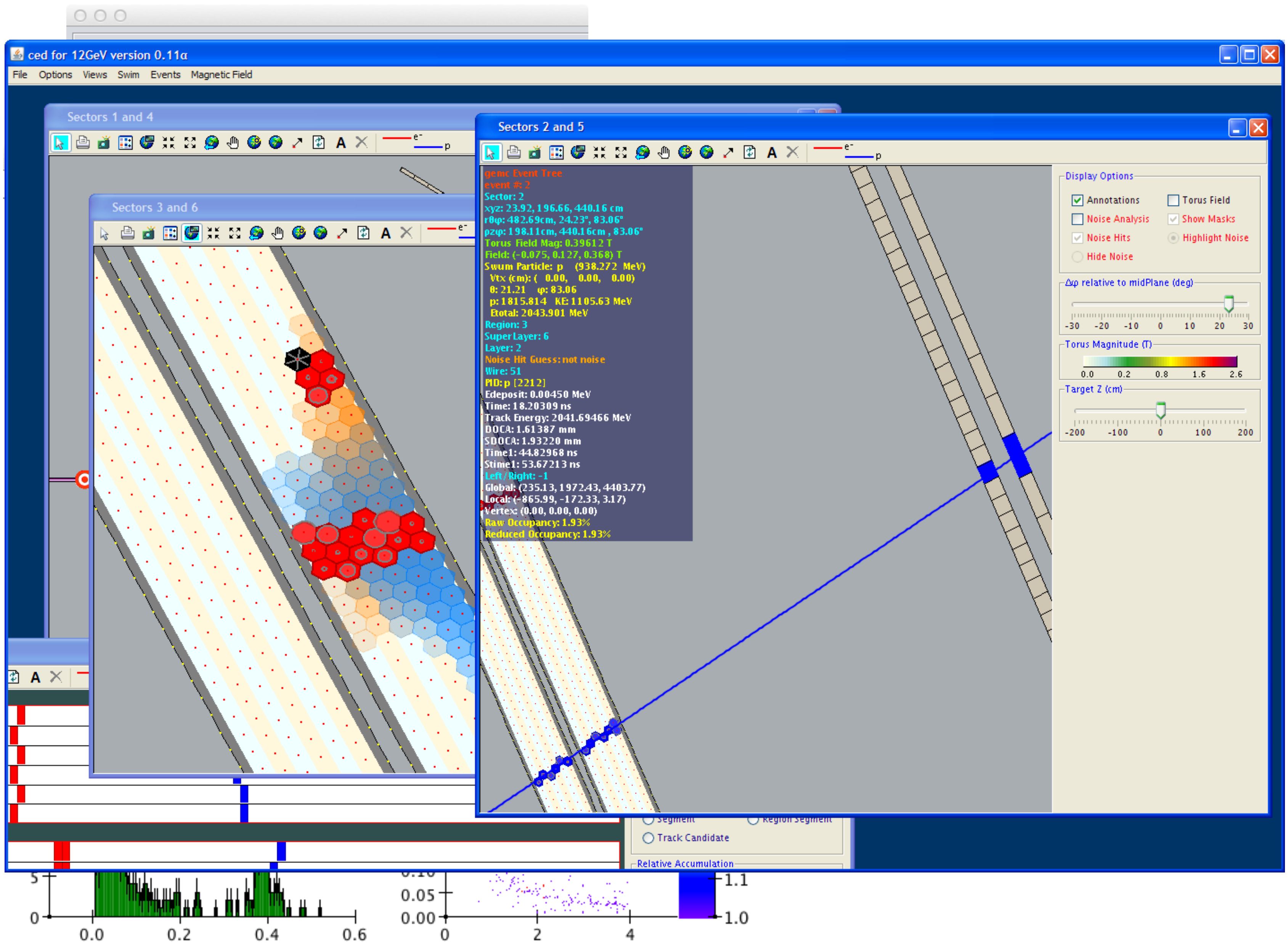
What is the state of simulation, data acquisition, calibration and analysis software, including usability and readiness from a user's perspective? Are the software plans complete, and is the scope appropriate?

Is there adequate progress in software maturity, and is there a defined set of goals leading towards full readiness ahead of production running?

Event Reconstruction



Event Reconstruction



Place Holder for Documentation

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

Place Holder for Project timeline

To what extent will software tools and components common across the halls and/or with the wider HE/NP communities be utilized? Are efforts towards commonality being made?