

ClaRA & JANA

Vardan Gyurjyan
DAQ group

Software Framework



- Out of sight
 - The core of a framework, architecture of a software system
 - component of the framework that manages developer provided code
- For developers
 - Represent those parts where the programmers using the framework add their own code to add the functionality specific to their own project.



Software application

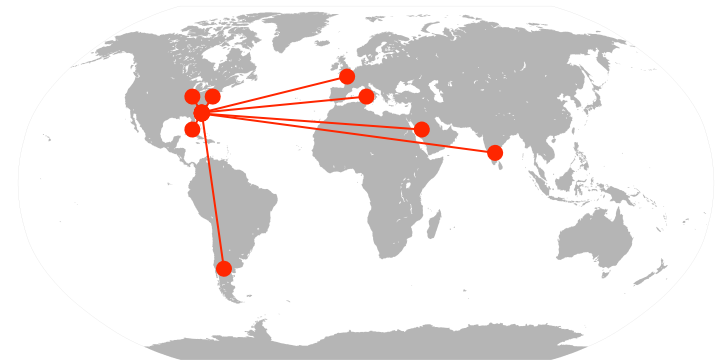
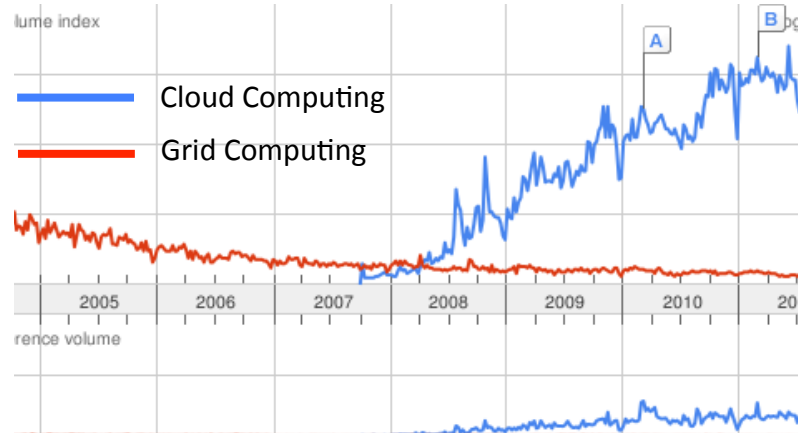
Similarities, yet...

- Both are **multi-threaded** physics data processing application development frameworks
- For developers
 - JANA: **Data** and **Algorithm** classes (C++) to create data and factory objects.
 - The factory needing a type of data object as input doesn't need direct knowledge of the factory that actually generates those objects.
 - ClaRA: **Jservice** abstract classes (Java, C++, Python) to create application modules (**services**) (for e.g. TOF, EC, CPT, etc...)
 - Services are autonomous (loosely coupled: can be run as a stand alone applications or be assembled in a larger application)
- For users
 - JANA: **JEventProcessor** Application is a single process that must be compiled.
 - ClaRA: **JOrchestrator** Application is designed by linking modules together (no coding and/or compilation).
 - > *clara r-tof-ec-ct-ft-w -l flist.xml*
 - ClaRA user:
 - Use of services developed in different languages in a single application (e.g. HPS reconstruction in Java)
 - Ease of optimization, maintenance and contribution.



The Difference

- Conceptual difference
 - ClaRA is based on SOA (Service Oriented Architecture)
 - JANA is based on OOA (Object Oriented Architecture)
- ClaRA vs. JANA
 - **ClaRA application: multi-process & multi-threaded.** The ClaRA service can be a process or a thread within a single process that communicates with other services through message passing. The message initiates the service execution.
 - **JANA application: single-process & multi-threaded.** In JANA all the data and factory objects of an application are running in a same thread, sharing thread memory space (thread stack and thread controlled heap), as well as thread's set of machine registers. Pointers are used to pass data between data and factory object.
- As a result ClaRA presents
 - Environment for cloud computing and distributed data handling.
 - Location independent resource pooling.



ClaRA cloud processing.

CLAS data from the ODU data-center is being actively analyzed by users from Scotland, Germany, Chile, India, Israel, MIT, FSU, ODU and others.

Side by Side

Functionality	ClARA	JANA	Comments
Multi-threaded	✓	✓	Equal amount of effort is required to be implemented by a developer to write multithreaded code in both frameworks.
Modularity	✓	✓	Both OOA and SOA present modularity
Loose coupling of modules	✓		ClARA modules (services) are less coupled than JANA modules (factories). As a result ClARA modules are more easy to debug, test and maintain.
Multi-lingual	✓		Modules can written in Java, C++ and Python
Traditional batch processing	✓	✓	
IO/application separation, and application persistence	✓		JANA has a File-Job correspondence. JANA job is an application. File ends – job ends. ClARA job is a Node. ClARA application is persistent.
Cloud (distributed) processing	✓		
Facts			<p>ClARA based distributed data processing environment (Data Mining Project) Manpower: Single physicist <1 year . Status: Actively in use. User comments: “very easy to use”. Review comment: “forward thinking” , “interesting solution for the large data preservation, and distributed processing.”</p> <p>Clas12 event reconstruction system (SOT project) Manpower: Single students <1.5 year. Status: Retired. Gradual replacement with new services Review comment: “Innovative framework”.</p>

Collaboration

- ClARA service engines can be developed using Jana tools and libraries.
 - ClARA C++ services presented as JANA factories can run within the JANA framework.
 - only ClARA C++ service engines can be presented as JANA factories (no support for other languages in JANA)
- JANA factories can extend ClARA interface, thus
 - being able to be used both within the JANA framework and be deployed within the ClARA cloud computing environment for distributed data processing.
 - Within ClARA environment JANA factories will collaborate with available third party services (possibly written other than C++ based technologies)

Note. In order JANA to be distributed it must use external middleware (e.g. DAQ distributed middleware for L3 or data quality monitor, ClARA or any other cloud middleware.)



Hall B and D frameworks

ClaRA

CLARA
CLAS RECONSTRUCTION AND ANALYSIS
FRAMEWORK

Users Manual



Thomas Jefferson National Accelerator Facility
200 Jefferson Avenue, Newport News, Virginia, 23606

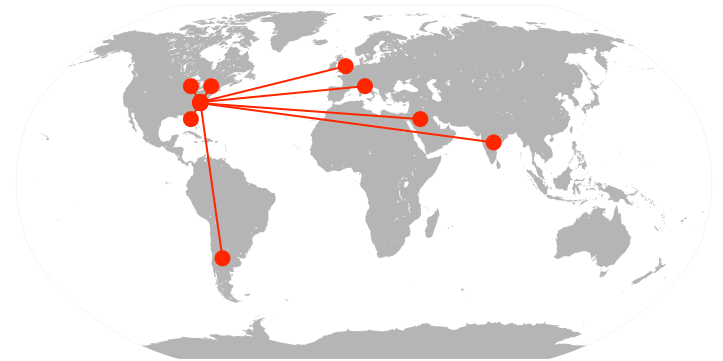
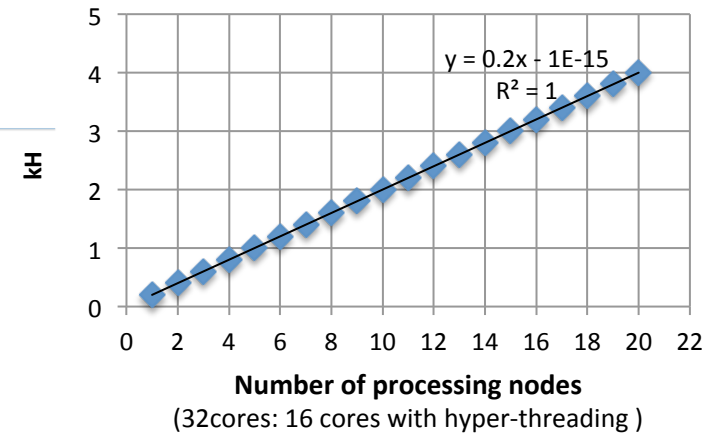
- Cloud computing framework.
 - Implements SOA architecture
- Data processing major software components as services (SaaS)
 - Multilingual support
 - Services can be written in C++, Java and Python
- Data (storage and persistency) as a services (IaaS)
- Supports both traditional and cloud computing models
 - Single process as well as distributed application design modes
 - Centralized batch processing
 - Distributed cloud processing
- Utilization of multicore processor systems
 - Built in Multi-threading of a user service
 - Requires thread safety of a service code
- Ability to expand computing power with minimal capital expenditure
 - Dynamic elasticity.
 - Utilization of IT resources of collaborating Universities.
 - Take advantage of available commercial computing resources.
 - On-demand data processing.
 - Location independent resource pooling.
- In short
 - simplifies user contribution in developing physics data processing applications
 - prevents future chaotic software fragmentations
 - provides an environment for cloud computing and distributed big data handling

Publication:

V. Gyurjyan, et al. "CLARA: A Contemporary Approach to Physics Data Processing
Journal of Physics: Conference Series **331** (2011) 032013

CLAS12 Reconstruction Rate

JLAB batch farm processing

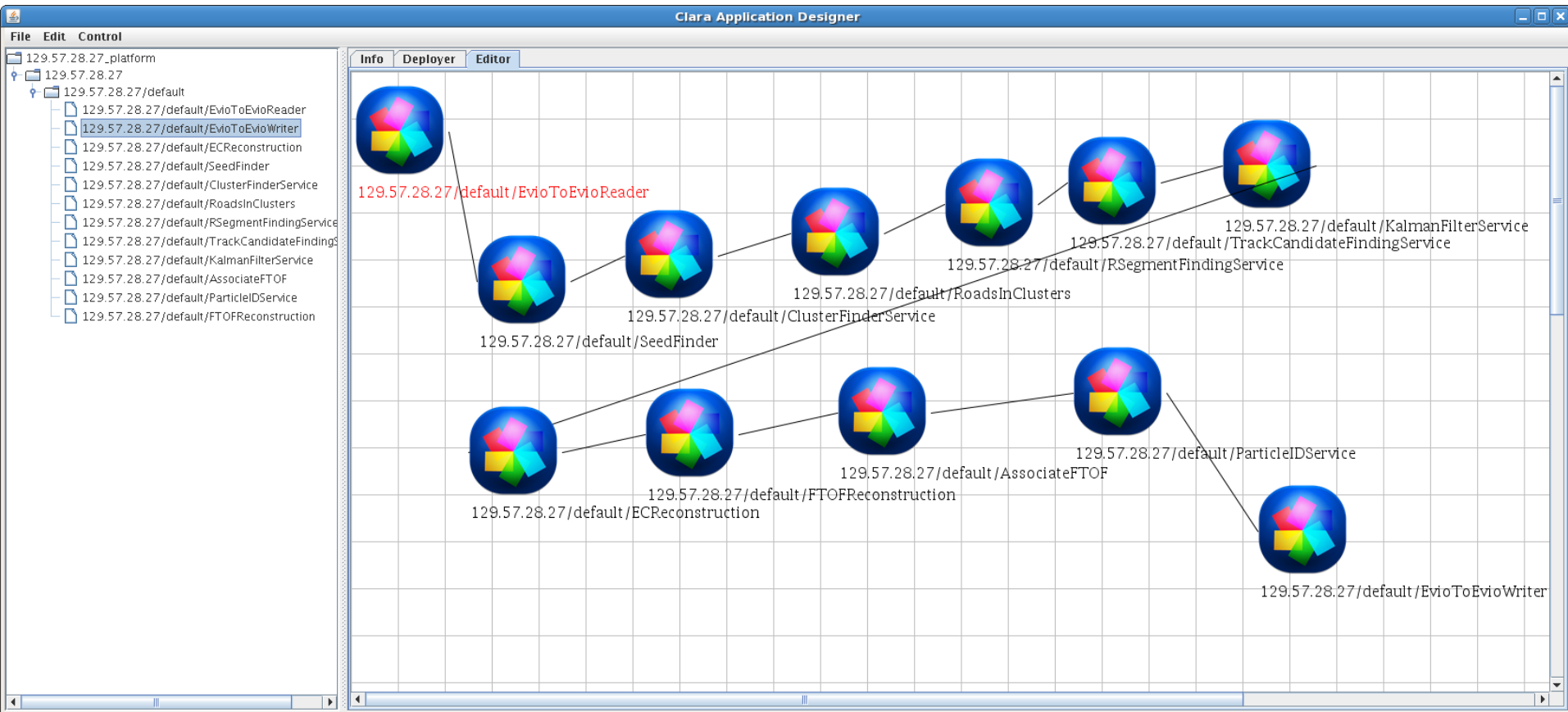


Cloud processing.

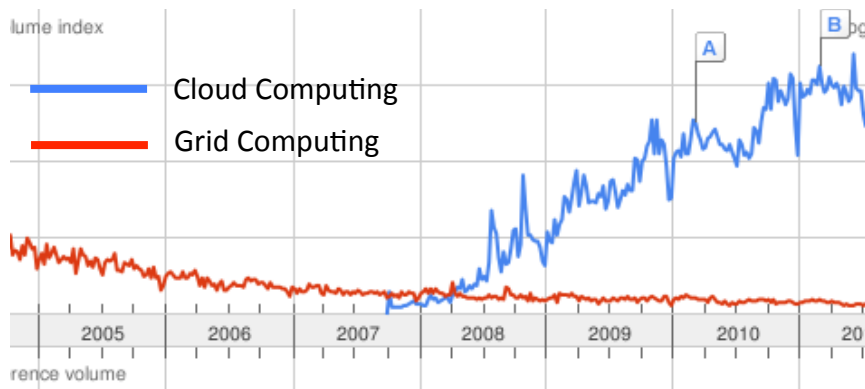
CLAS data on the ODU data-center is being actively analyzed by users from Scotland, Germany, Chile, India, Israel, MIT, FSU, ODU and others.

Thank You.

Application Graphical Designer



Advanced Computing Emerging Trends



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

Helix Nebula project at CERN (02 Mar 2012 V3.co.uk digital publishing company) .

CERN has announced it is developing a pan-European cloud computing platform, with two other leading scientific organizations, which will take advantage of the increased computing power cloud systems offer to help scientists better analyze data. Interview with the Frédéric Hemmer, head of CERN's IT department. "CERN's computing capacity needs to keep up with the enormous amount of data coming from the Large Hadron Collider (LHC) and we see Helix Nebula, the Science Cloud, as a great way of working with industry to meet this challenge."

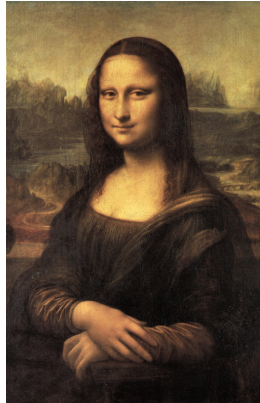
Magellan project at US Department of Energy.

USDE will spend \$32 million on a project that will deploy a large cloud computing test bed with thousands of Intel Nehalem CPU cores and explore the work of commercial cloud offerings from Amazon, Microsoft and Google for scientific purposes.

Clara and Clara-based effort

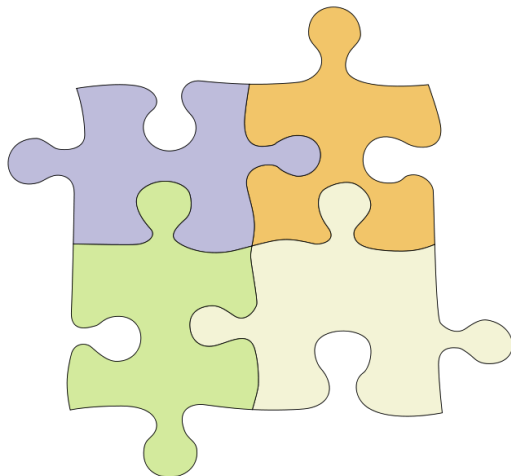
Project	Start Date	Status	Number of Lines	Comments
ClaRA	2009	Maintenance	23611	Cloud computing framework
EC reconstruction	2012	Maintenance	8734	ClaRA services that implements the EC reconstruction algorithms and map utilities interfaces to databases
PCAL reconstruction	2013	Active	2864	ClaRA services that implements the PCAL reconstruction algorithms
TOF reconstruction	2012	Commissioning	17084	ClaRA services that implements the TOF reconstruction algorithms for the FTOF and the CTOF and interfaces to the CLAS6 configuration database for the FTOF
FTOF calibration	2013	Active		FTOF calibration services
TRAC (Track Reconstruction Application for CLAS12)	2012	Active	16401	Contains packages for track reconstruction algorithms and services for tracking in the central and forward detector systems, respectively
Forward Tagger Reconstruction	2013	Active	1305	ClaRA services that implements the FT Calorimeter reconstruction algorithms
Geometry Service	2012	Active		Service to provide simulation and reconstruction clients with detector geometry
Histogram Services	2013	Active	7269	Statistical data histogramming for reconstruction output validation
Event Builder	2013	Active	1197	Event Builder Service
Persistency Services	2013	Commissioning	6048	Contains various utility services such as magnetic field, banks and event reader/writer services, etc.
Data Mining	2012	Maintenance		
SOT	2010	Retired	177663	Charge particle tracking, TOF, EC, PID, Histogram, etc.

ClaRA application design



Programmer

Designer
without programming skills

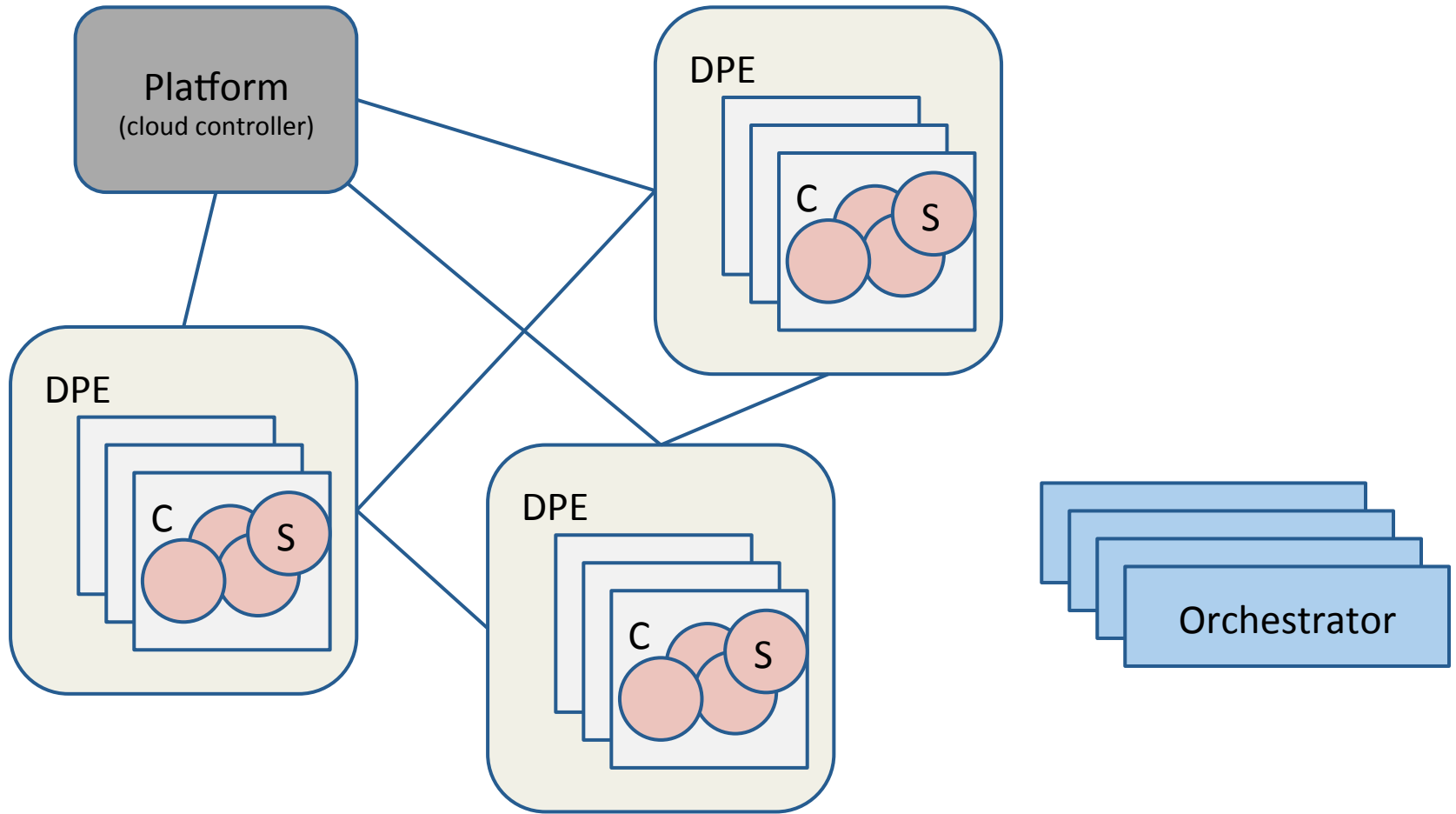


Services

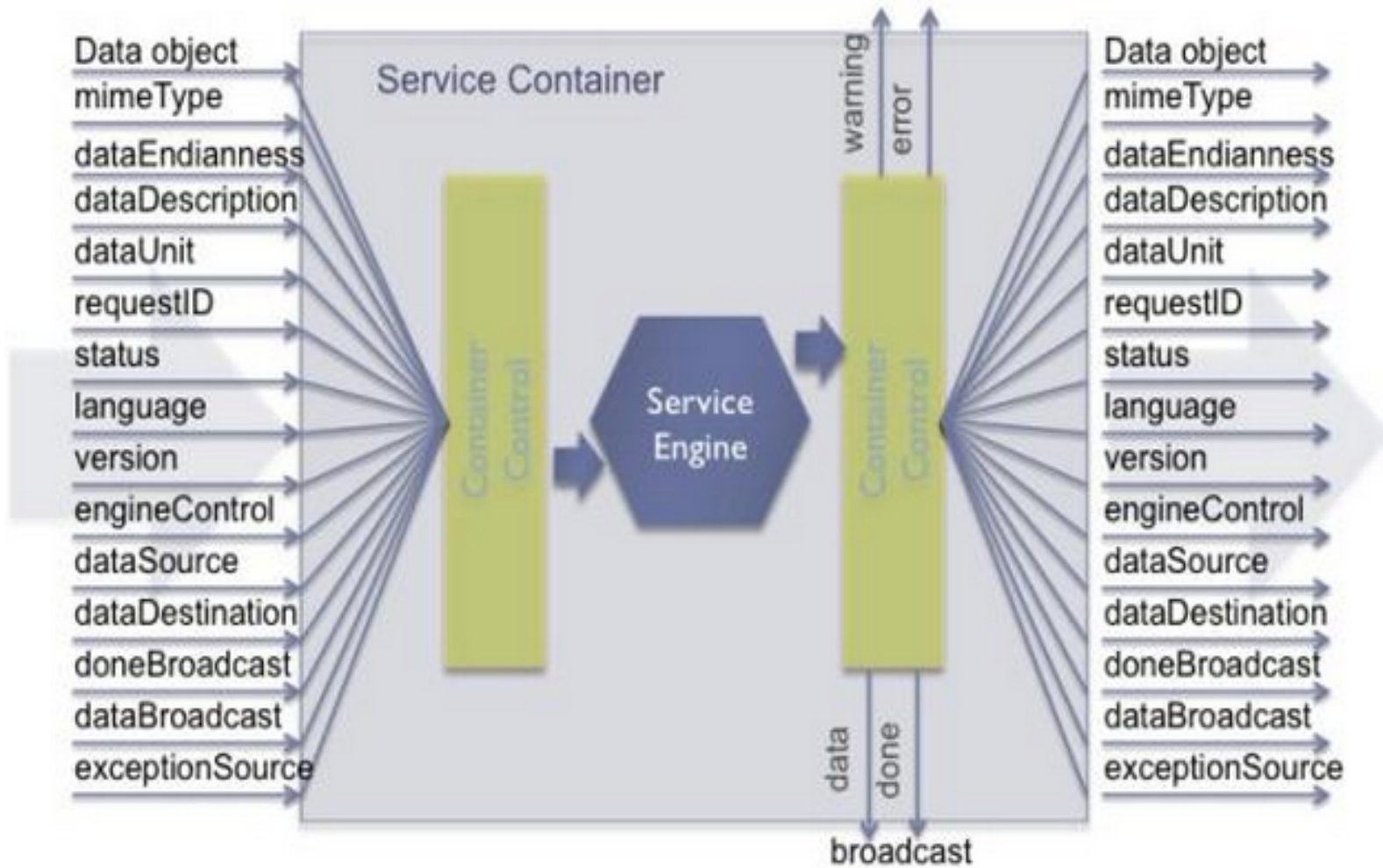


PDP Applications

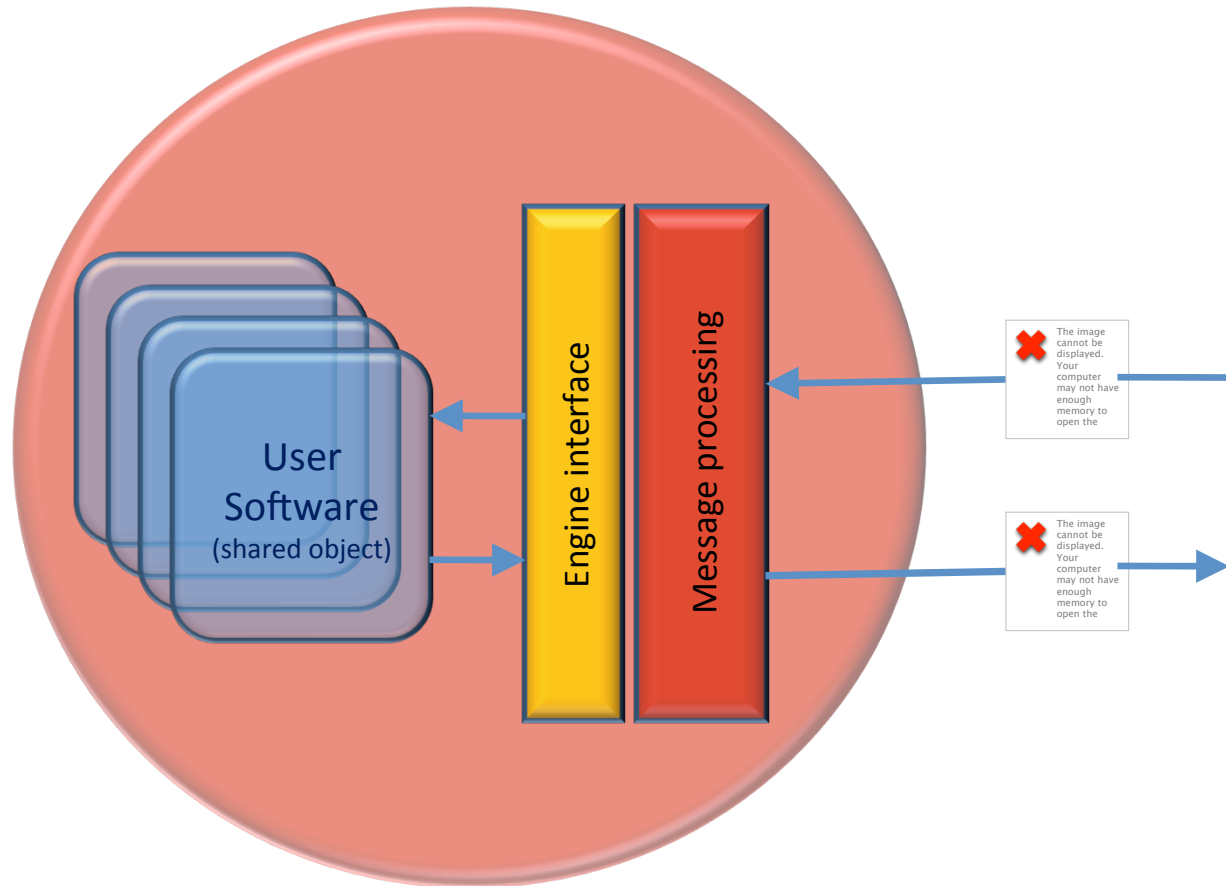
ClaRA Components



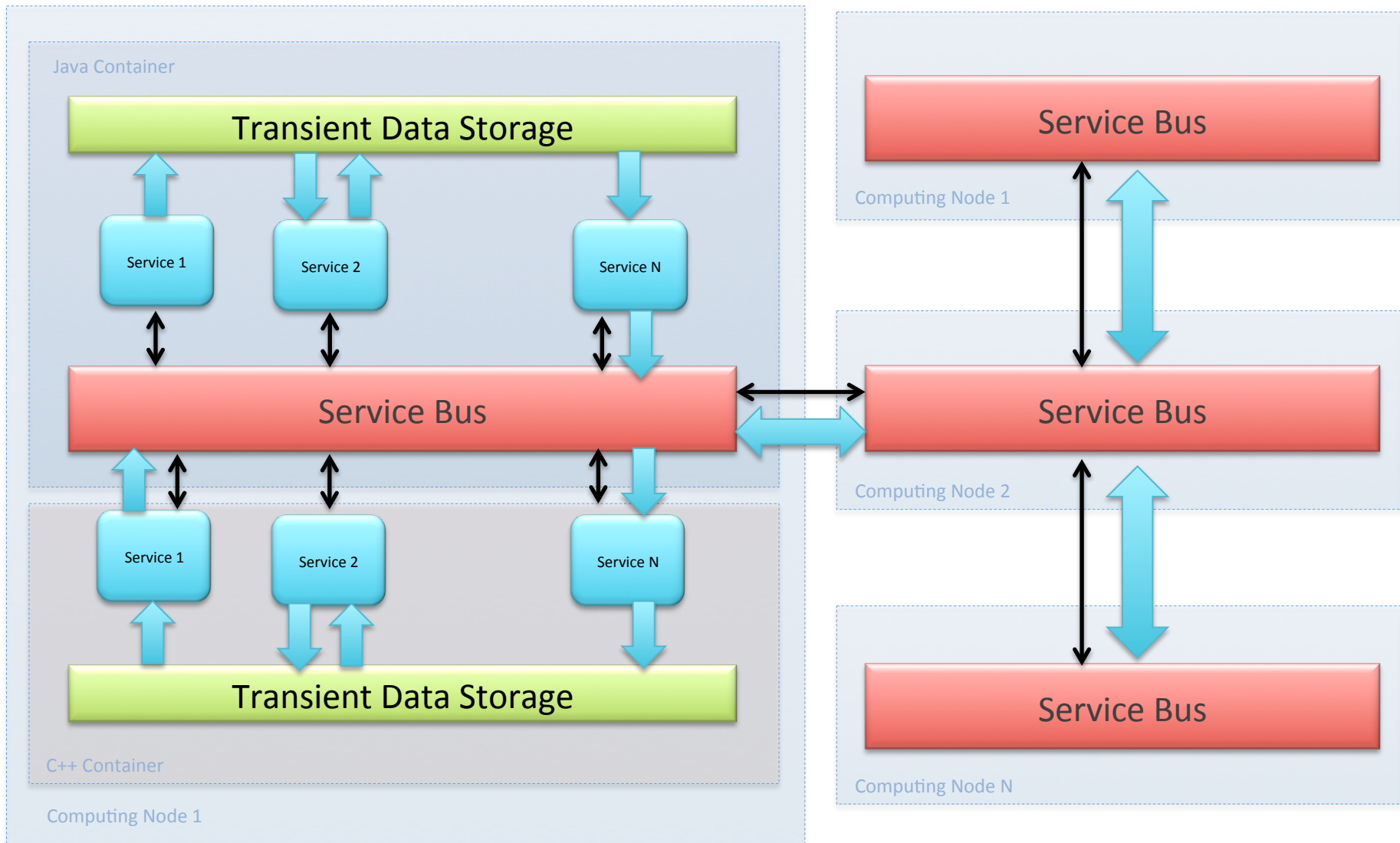
Transient Data Envelope



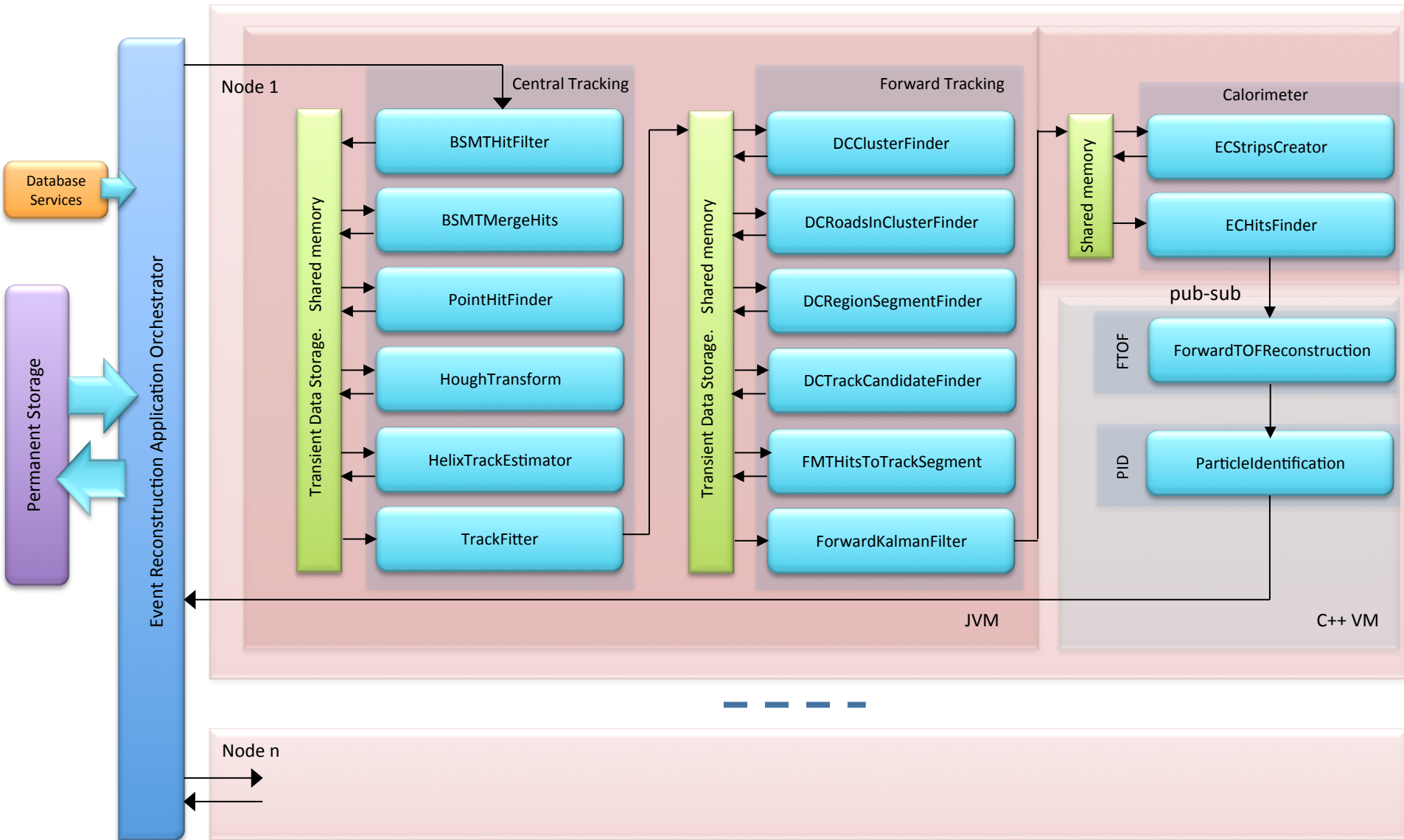
ClaRA SaaS Implementation



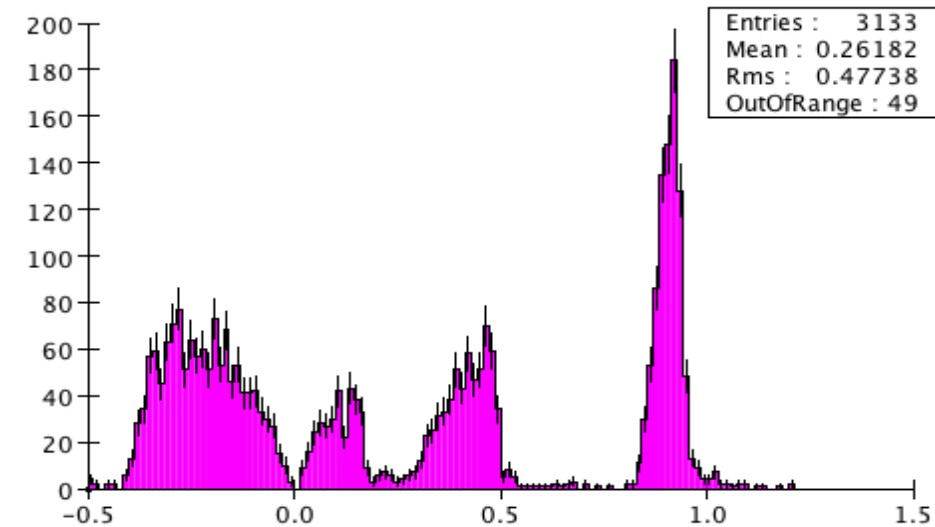
ClaRA Cloud Formation



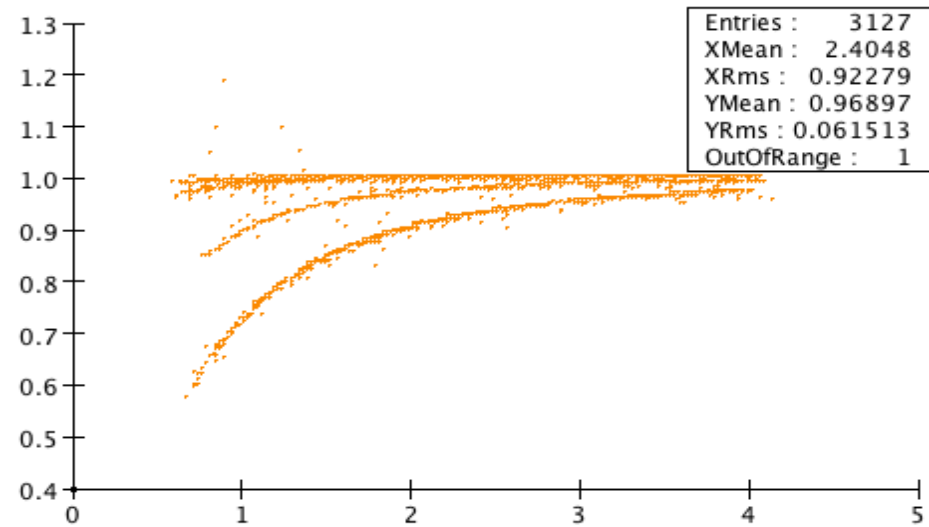
Event Reconstruction



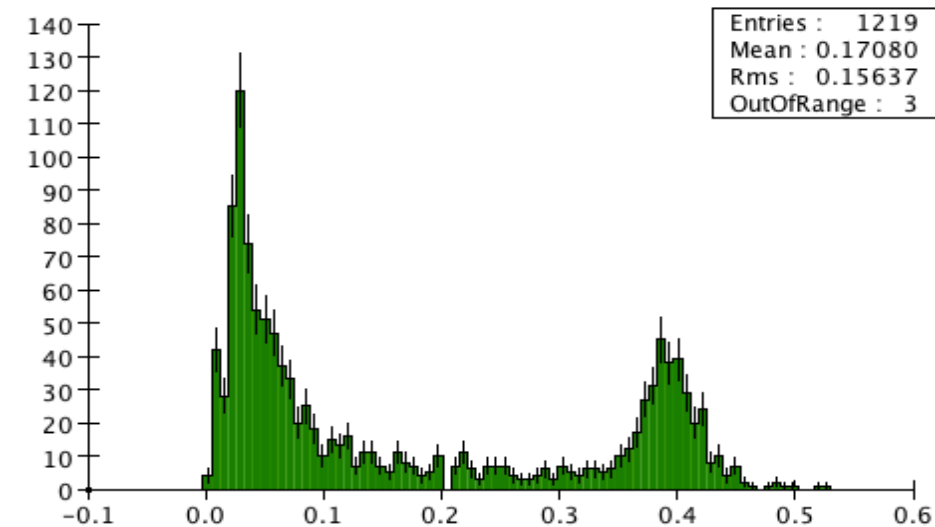
Mass Spectra



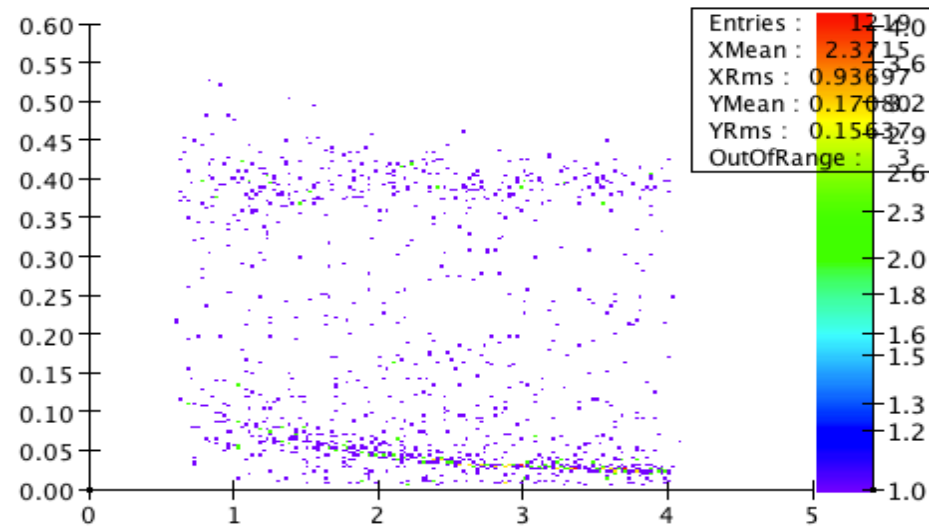
Beta vs. Mometa



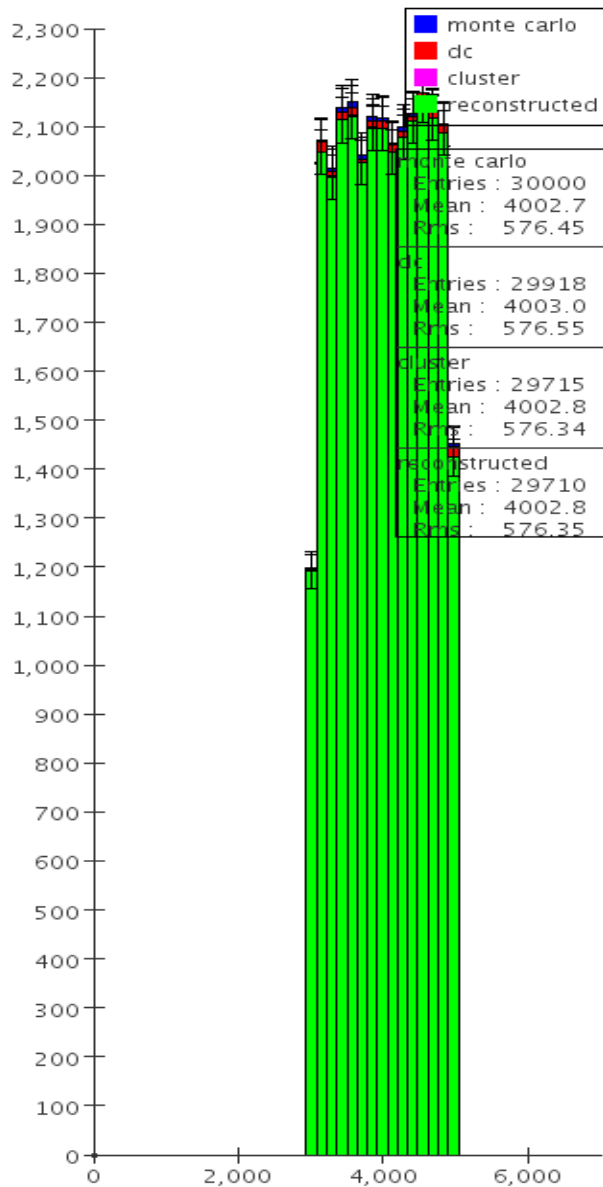
EC Sampling Fraction



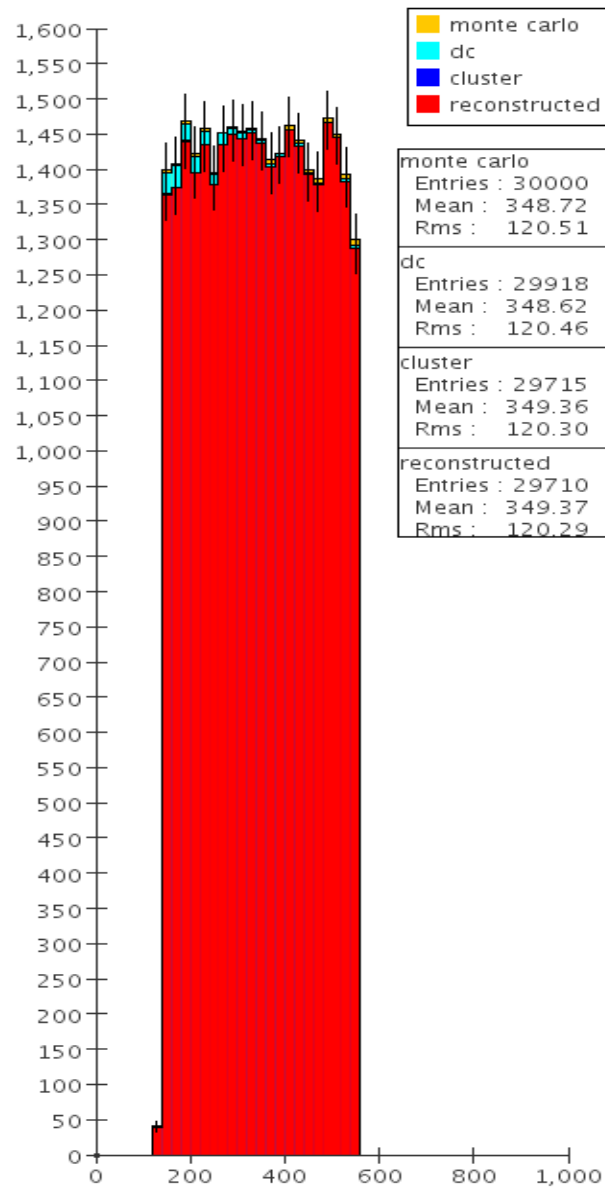
EC Edep vs. Mometa



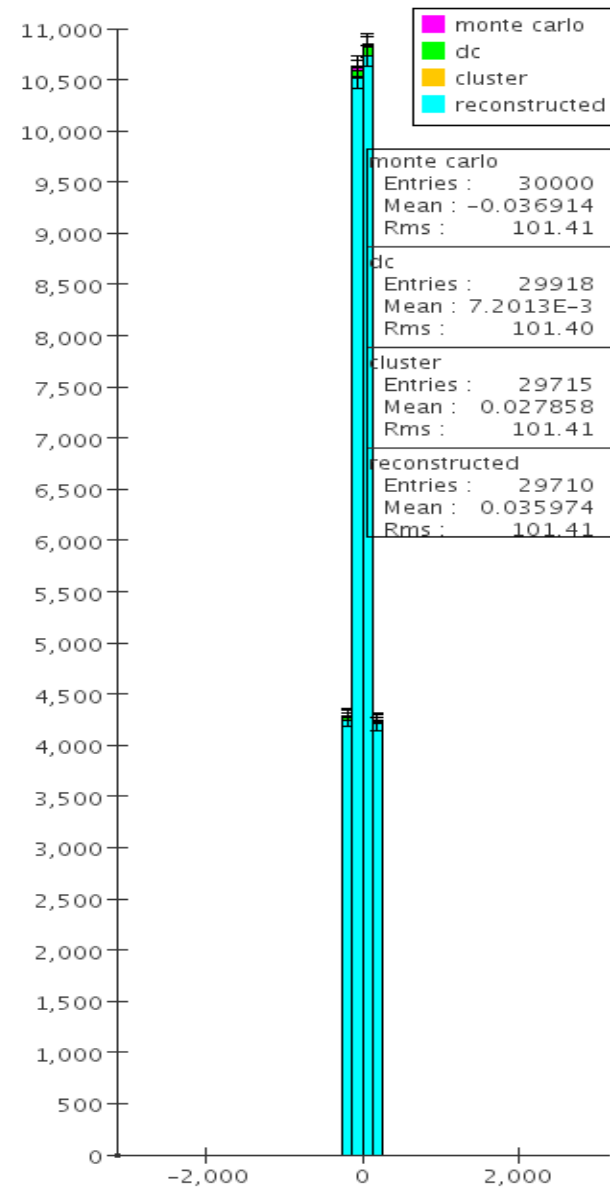
momentum



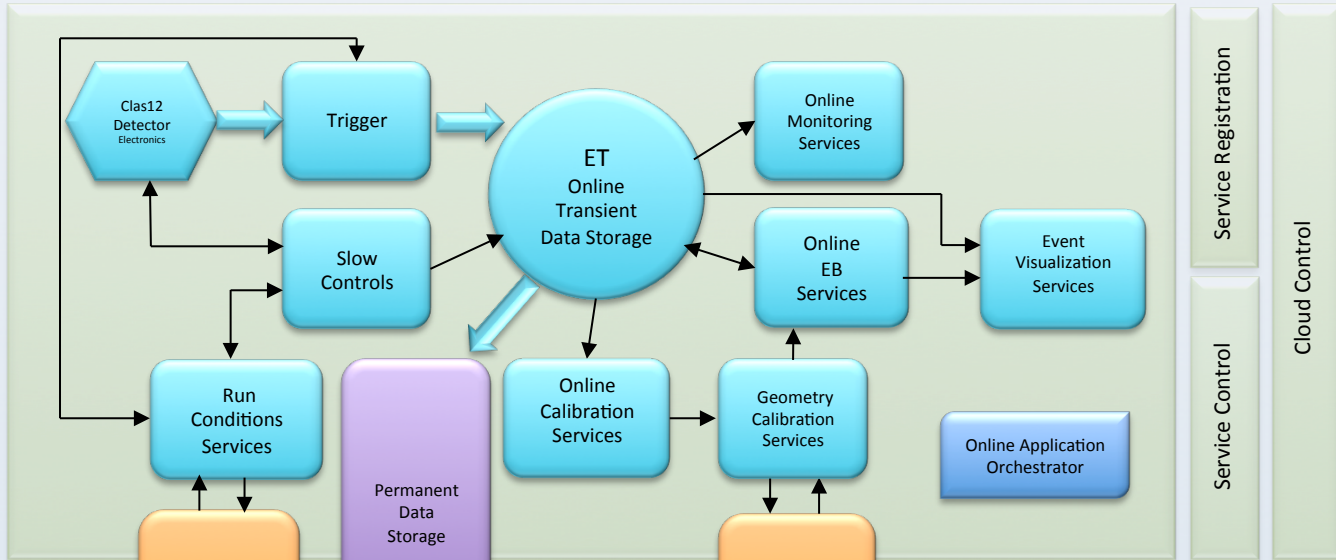
theta



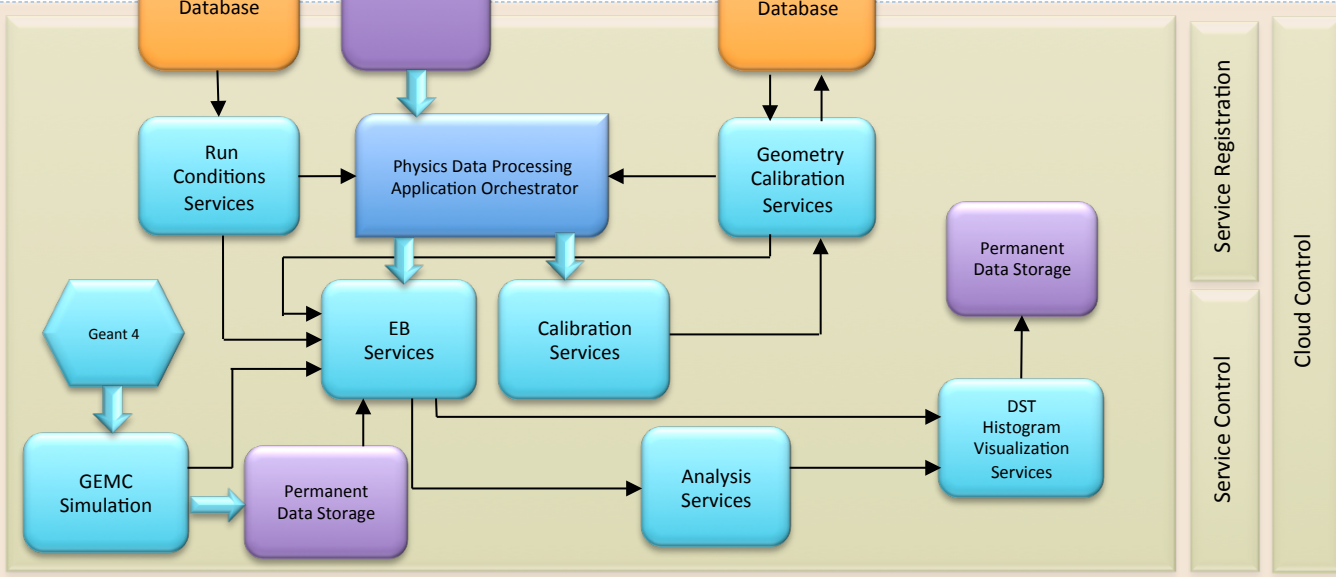
phi



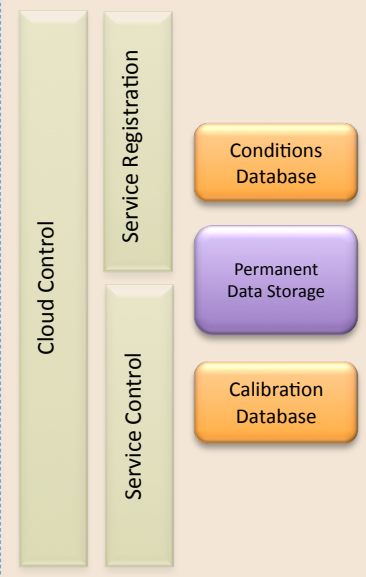
Online Cloud



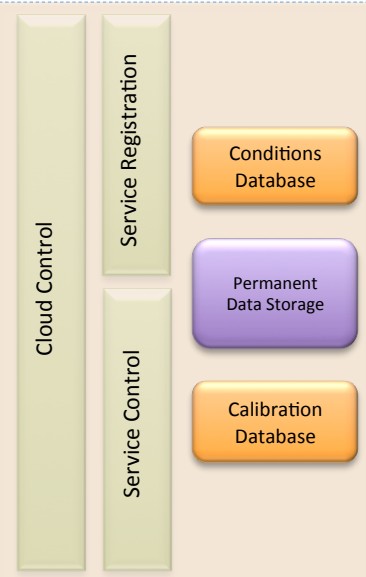
Offline JLAB Cloud



Offline University Cloud 1

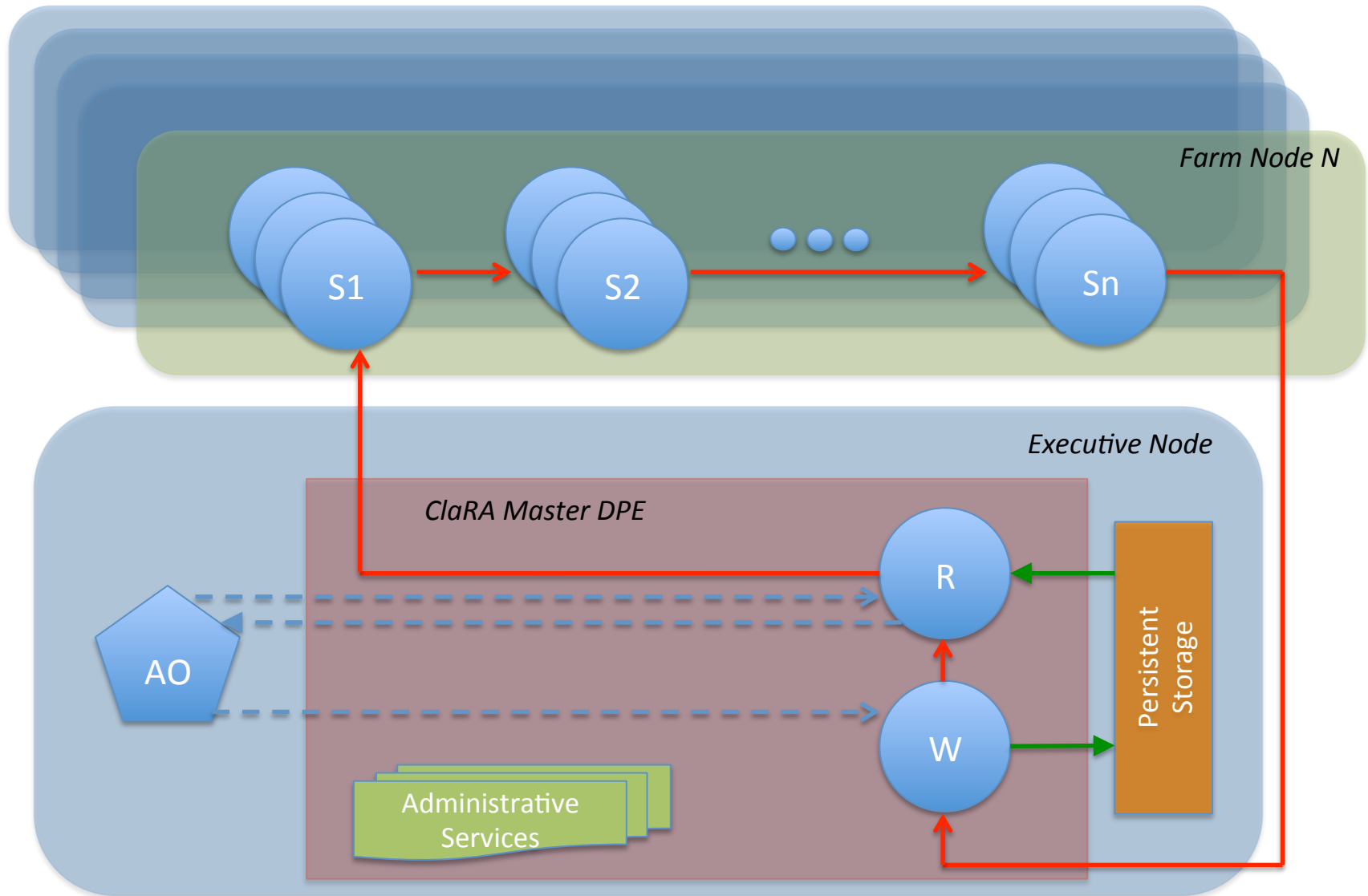


Offline University Cloud n



Cloud Scheduler

Single Data-stream Application

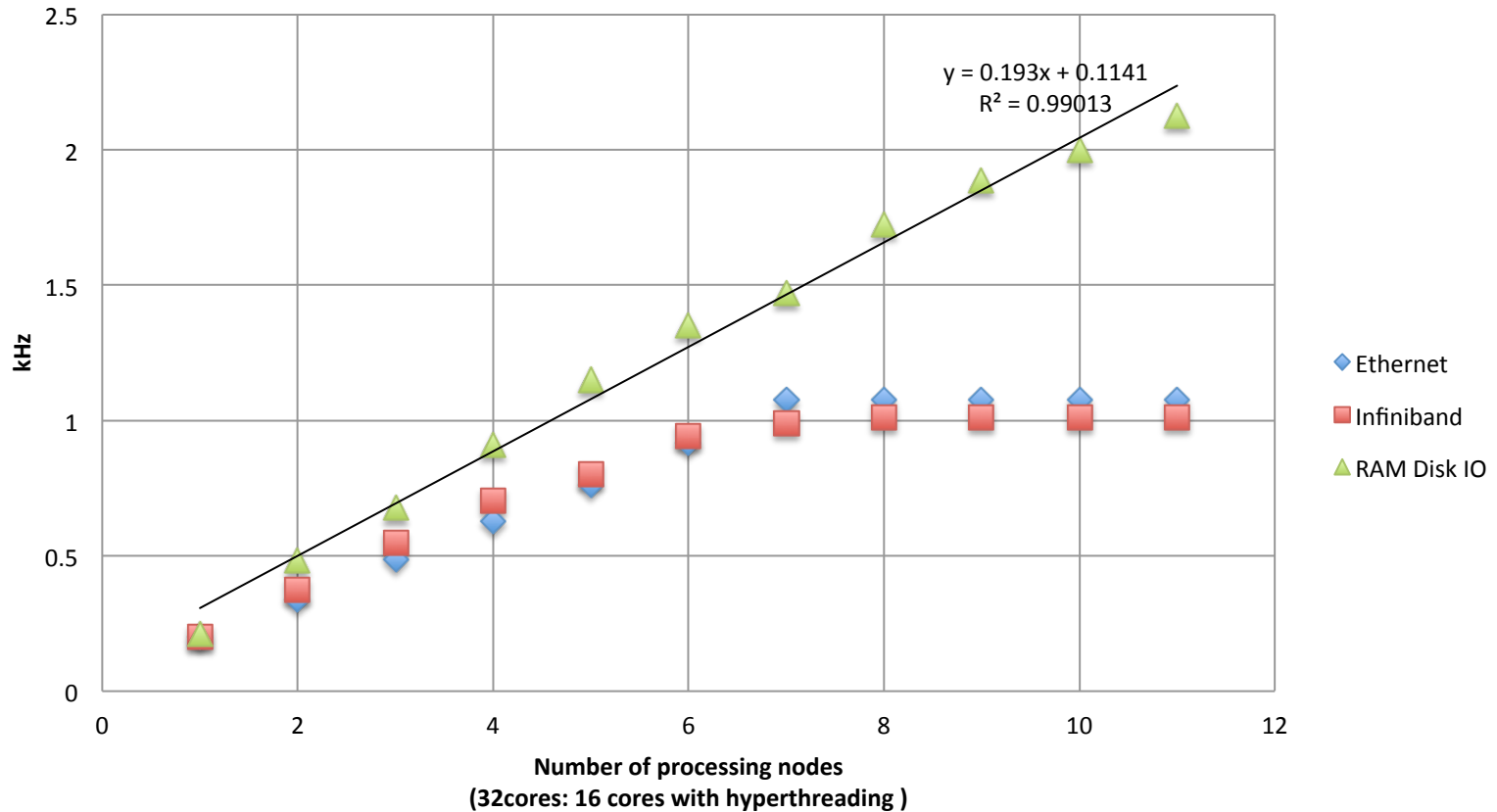


Single Data-stream Application

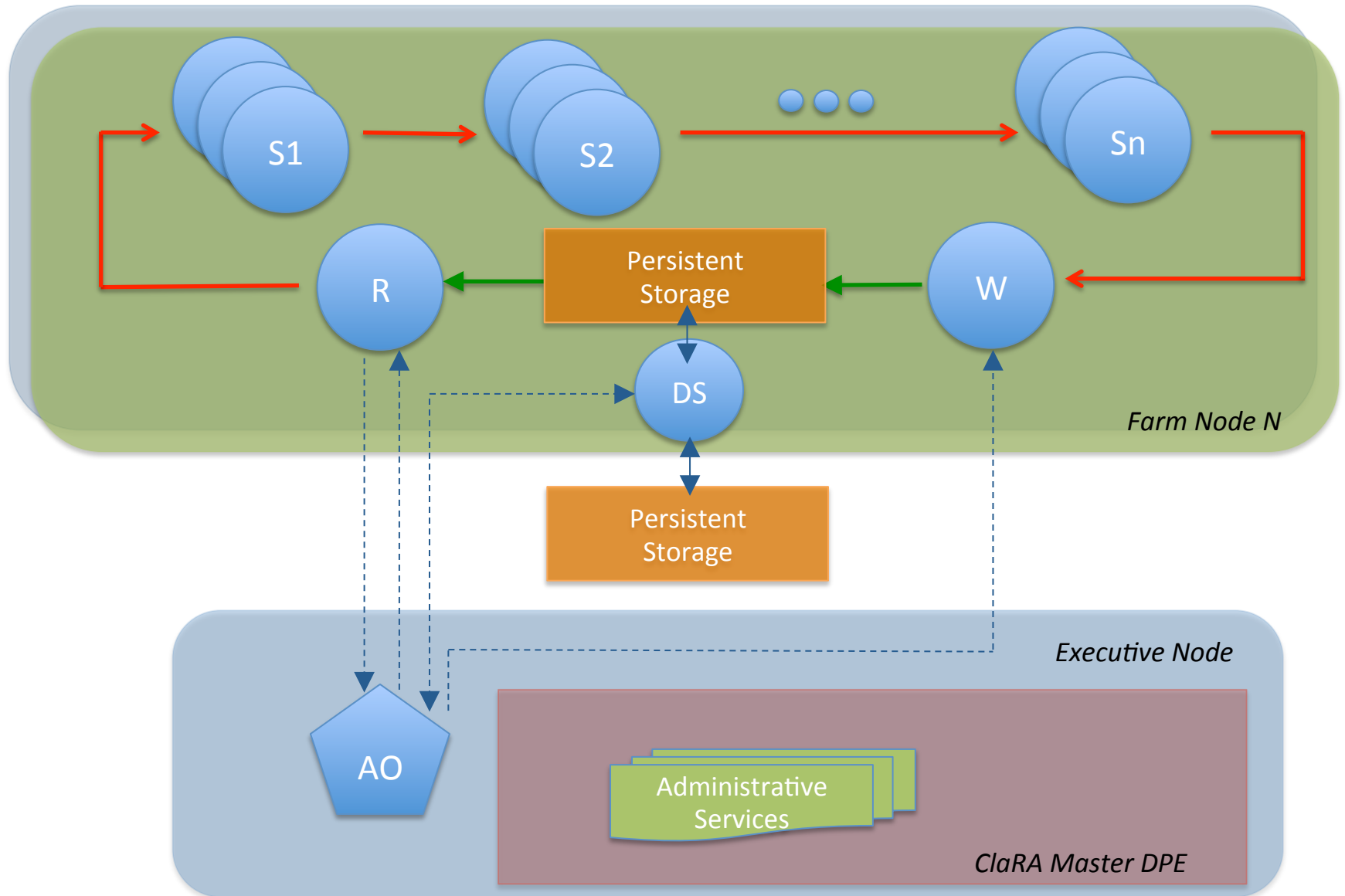
Clas12 Reconstruction: JLAB batch farm

Data Processing Rate

Single datastream



Multiple Data-stream Application



Multiple Data-stream Application

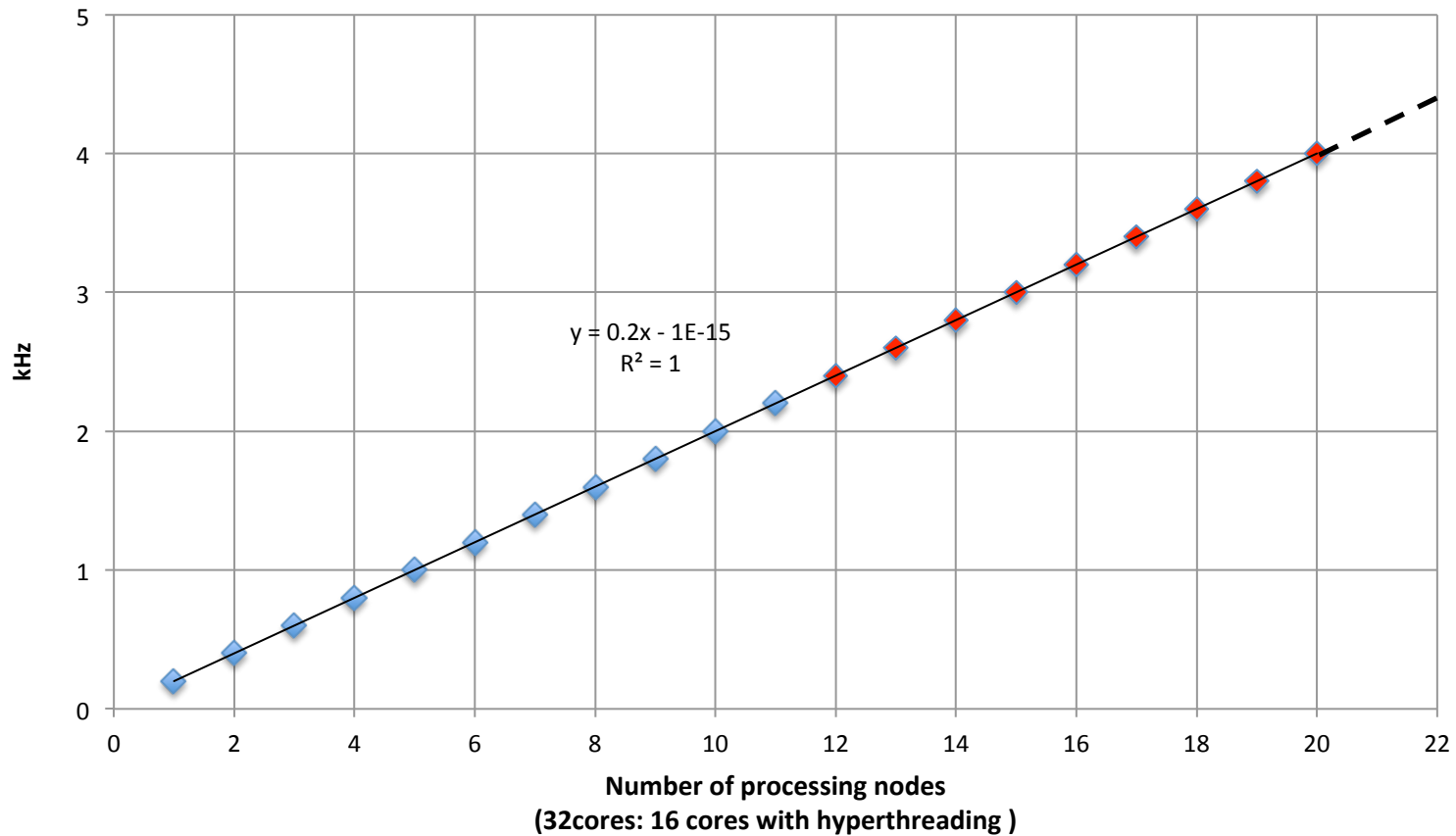
Clas12 Reconstruction: JLAB batch farm

Data Processing Rate

Multiple datastreams

One data-file per processing node

Data-file contains 10K events



Service Bus Performance measurements

Producer -> Java Server -> Consumer, 1Gigabit Ethernet

