DAQ Group report

June 13th 2017

- Group highlights
- CLARA - NASA collaboration.

Ed Jastrzembski

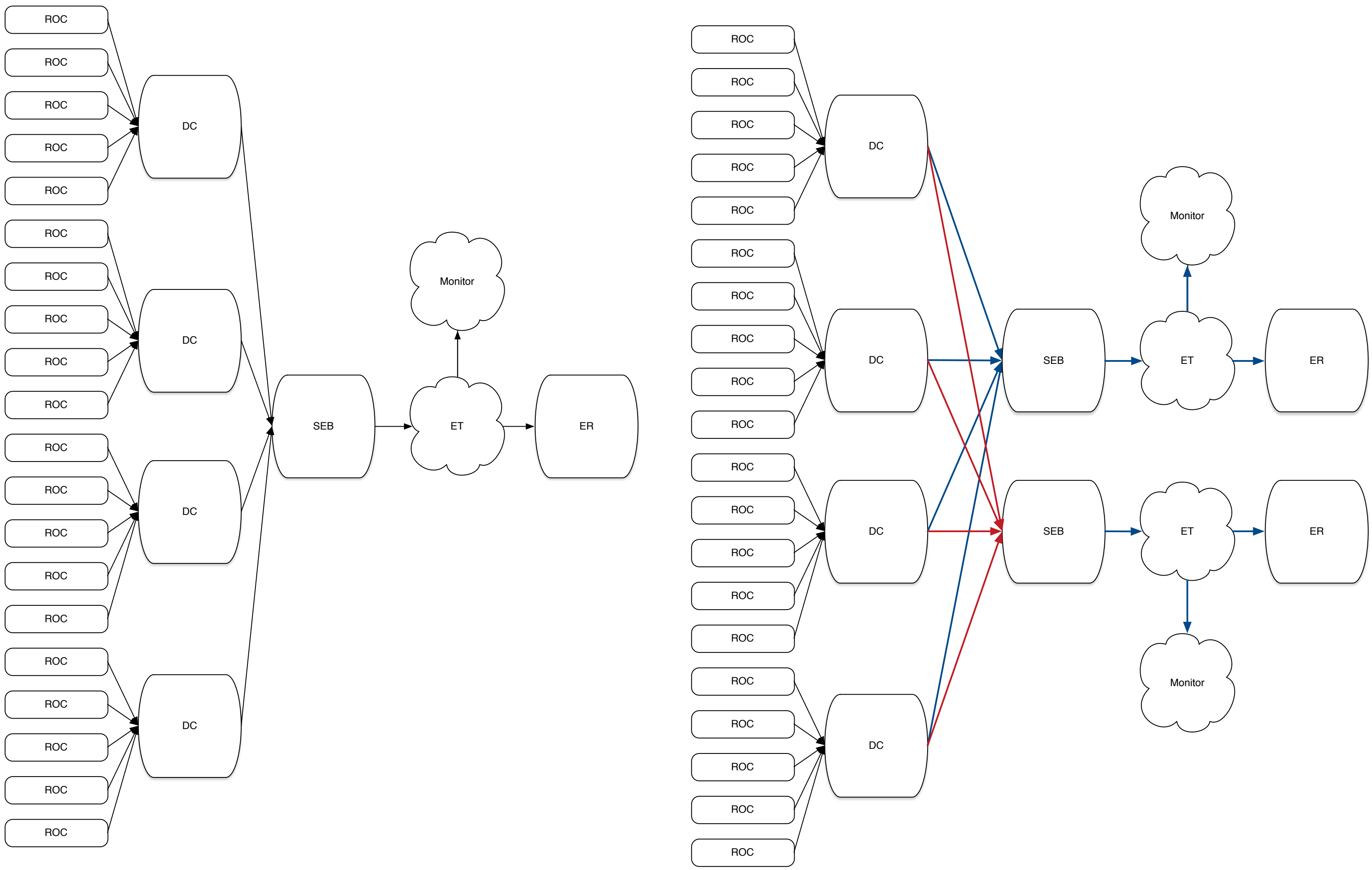
- Approached Thia and discussed the idea of assigning Ed to work closely with halls A and C on future experiments.
  - Coordinate electronics design, development, procurement, deployment.
  - Liaison between halls A & C, DAQ and Fast Electronics.
  - Aim towards consistency between experiments.
- Met to discuss TDIS, aim to use TDIS to explore ideas for later detectors such as SoLID.
  - Thia has started monthly TDIS meetings to get the ball rolling.
  - Ed is doing a lot of research into TDIS, TPC readout schemes etc.

Carl Timmer

- GLUEX With intensity running review:
  - July 2016 review highlighted issues with running at high intensity.
  - Event size dependency on intensity threatened to produce unacceptable data rates.
  - GLUEX put in a lot of work to understand triggers and thresholds and bring data rates down to a manageable size ~1.5 Gbyte/s at $5x10^7$ γ/s.
- Carl has been working on improving stability and throughput at high rates.
  - Spring 2016 the maximum stable throughput was ~300 Mbyte/s
  - Fall 2016 this was increased to ~1 Gbyte/s to disk (1.2 without disk, build rate).
  - Spring 2017 - dual parallel data streams, build rate ~1.4 Gbyte/s per EB 2.7 total.
    - Can't write to disk at this rate since GLUEX would need a second RAID.
- Data compression - savings from compressing data, which in GLUEX case can be 50% of the original size, offset by the effort (cpu power) needed to compress. Using current technology, hardware and algorithms, a single CPU core can compress a data stream at about ~300 Mbyte/s, which is a useful rate. Carl is looking at how to add to CODA.
  - Gagik added data compression to CLAS12 by writing EVIO data to a compressed format.
  - Carl and Gagik will look at incorporating this into EVIO so that all halls can make use.

Parallel Event Building

- On the left is the scheme used in GLUEX so far.
- For high intensity running the SEB and disk are a bottleneck, this was anticipated.
- Reconfigure to use two SEB and two disks, throughput doubles as expected.



Carl and David A.

- David Abbott has been busy testing Carl's code updates on DAQ group machines then in hall-D.
- We still have tricks up our sleeves:
  - Direct access to Infiniband.
  - Improve our code that sends from the Data Concentrator to Secondary EB.
- Why would we do this?
  - We have demonstrated that we can take data at 2.7 Gbyte/s with two parallel SEBs and 1.2 Gbyte/s with one SEB.
  - GLUEX high intensity needs ~ 1.5 Gbyte/s.
    - Needs two SEB nodes and two RAID disk arrays.
    - Complex system and complex things tend to break more easily.
  - It would be nice to get a little boost so that we can do the 1.5 Gbyte/s with only one SEB and RAID if needed.

Direct access to Infiniband (Abbott)

- CODA was written to use TCP network protocol as the underpinning of data transport.
  - Advantage of being a reliable protocol with “back pressure”
  - Keep sending on a socket, if the other end can’t cope the sender blocks until the data finally moves.
  - Unless the connection breaks no data is lost.
- For high bandwidth applications we use InfiniBand hardware. TCP is implemented as a layer on top of the InfiniBand core.
- However, InfiniBand is reliable at the hardware level.
- The TCP software protocol provides communications reliability that is redundant. We could add code to access the SDP Kernel driver and cut out the TPC overhead.
  - Major code change that is InfiniBand specific.
  - Try to avoid but it is possible if needed.

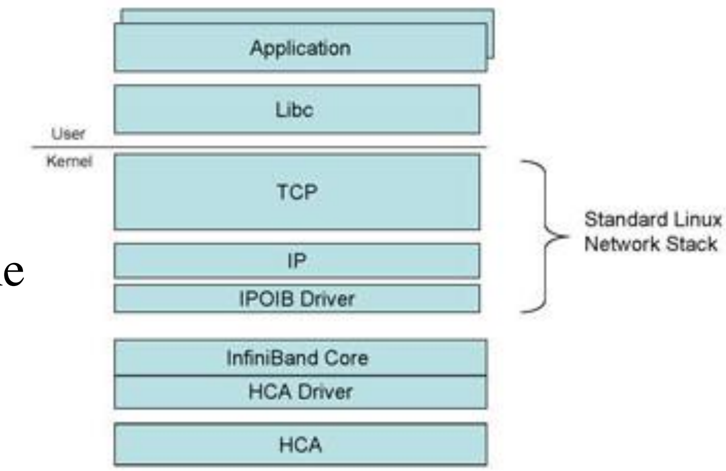


Figure - 2

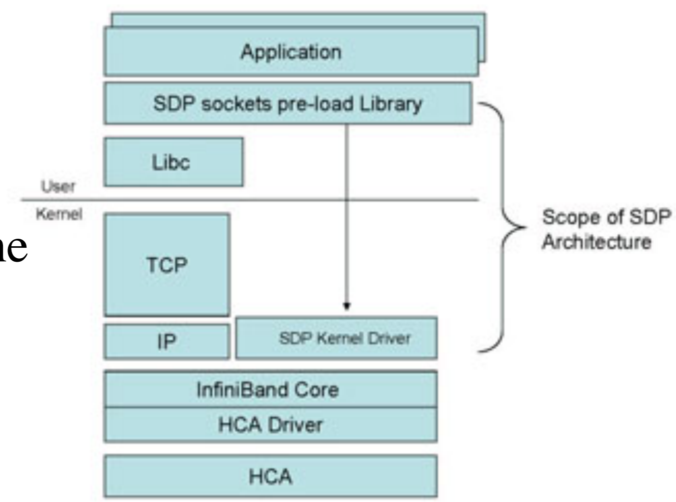
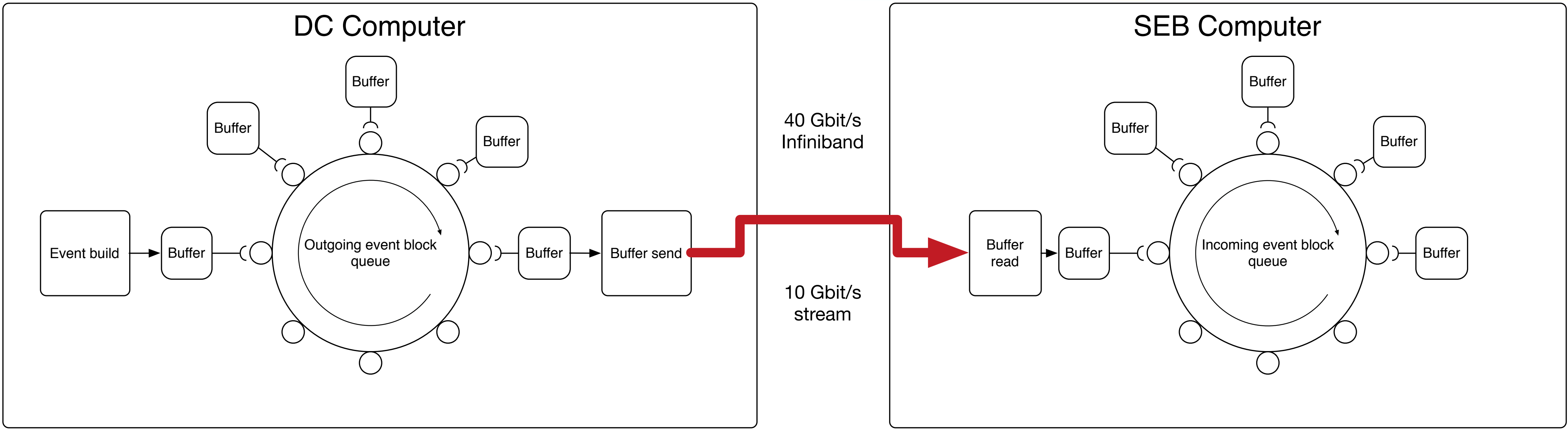


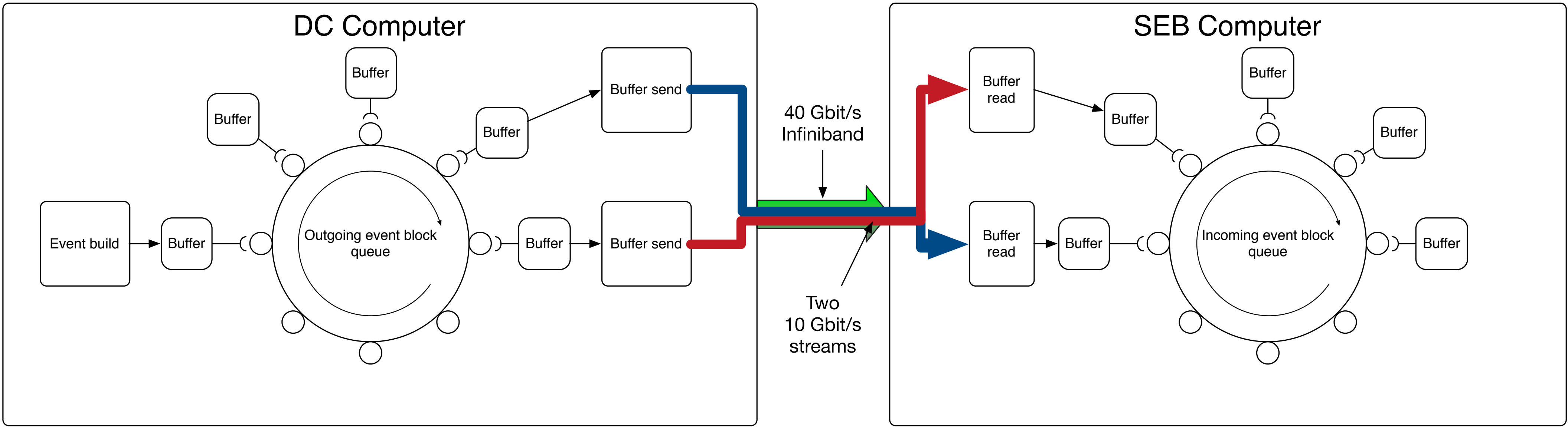
Figure - 3

Improve DC to SEB flow



- We have a nice fat pipe, 40 Gbit/s Infiniband but struggling to get above 10 Gbit/s Why?
  - Computers on either end are not overloaded.
- Answer:
  - Send thread pulls buffers of data off a ring queue and sends on the network.
  - Read thread pulls buffers off the network and puts them on a ring queue.
  - Neither of these threads is overworked.
  - “Under the hood” the OS assigns a single compute core to do the sending, this core is overworked!

Solution



- Assign two threads to sending and two to receiving and send over two TCP/IP sockets.
  - Forces the OS to assign a compute core to each socket doubling the compute power available for sending and receiving.
  - This is a simplification, care needs to be taken that the buffer ordering is preserved etc.
- Carl has coded this and the throughput of the links between the parts of the distributer parallel event builder are boosted in performance.
- Carl and David are now working on getting all of this to function stably at high rate.

Computing

- In the last few computing reviews we have shown the lab building up local computing resources based on 12 GeV requirements.
  - GLUEX have a better handle on the high luminosity requirements.
  - CLAS12 making very good progress.
  - Compression will lower network bandwidth and disk requirements.
  - Everything scales with weeks of running which have been and will be lower.
  - No funds to add cores in FY17.
  - Almost certainly nothing in FY18.
- Chip and I are working with the halls with various scenarios for FY18.
  - What must be done at JLab? Calibration, reconstruction…
  - How far can we expand use of OSG?
  - Can we push work to a DOE center, NERSC, ORNL etc?
  - Is it cheaper to pay for commercial cloud, Amazon, Google etc?
- Starting technology pilots.

Workshop

- Second "Future Trends in Nuclear Physics Computing" workshop held in May.
  - Four talk symposium to open the event.
  - Fifteen talks over the next 2.5 days.
  - Fifty six registered participants.
  - We are working on a workshop report that should be complete in the next month.
  - General themes:
    - Many groups using cloud, grid or large data centers.
    - Distributed data management to get the data to the compute resource.
    - In all cases people want to hide this infrastructure so access is transparent.
    - Big monolithic software frameworks are being replaced by more modular frameworks, Lego™ was mentioned a lot.
      - Example, one algorithm may work best on a laptop or local resource but a different one be better for a NERSC supercomputer.
    - AI and machine learning are advancing to the point where there is a lot of interest in making use.

NAIADS

NASA INFORMATION AND DATA SYSTEM FOR EARTH SCIENCE DATA FUSION AND ANALYTICS

NASA challenge

Next 10 years: JPSS-1/2/3, GOES-R, PACE, TEMPO, CLARREO;

New Decadal Survey and Venture Class missions;

Missions last longer, sensors become more complex;

The Climate Model outputs oversize the observations (currently 10s of PB);

The OSSEs are becoming high priority (large data volume).

The next DS push for the Integrated Earth Observing System

Observational data from multiple geo-satellites must be merged together to build an event of interest

Existing data fusion and data processing solutions are IO limited

Looking for new approaches

Proposal

NASA Langley Interest in JLAB CLARA framework.
V. Gyurjyan, et al. "CLARA: A Contemporary Approach to Physics Data Processing" Journal of Physics: Conference Series **331** (2011) 03201

Meetings at the administrative and technical levels

Created inter-agency collaboration

Submitted the proposal in response to NASA Research Announcement NNH14ZDA001N – AIST. Research Opportunities in Space and Earth Sciences (ROSES) 2014, A41: Advanced Information Systems Technology

NASA Information And Data System (NAIADS) for Earth Science Data Fusion and Analytics

Dr. Constantine Lukashin, Principal Investigator
NASA Langley Research Center, Hampton, VA

Aron Bartle, Co-Investigator
Mechdyne, Virginia Beach, VA

Dr. Vardan Gyurjyan, Co-Investigator
Thomas Jefferson National Accelerator Facility, Newport News, VA

Dr. Carlos Roithmayr, Co-Investigator
NASA Langley Research Center, Hampton, VA

Dr. Jun Wang, Collaborator
University of Nebraska, Lincoln, NE

Chris Currey, Collaborator
NASA Langley Research Center, Hampton, VA

Dr. Gagik Gavalian, Collaborator
Thomas Jefferson National Accelerator Facility, Newport News, VA

John Kusterer, Collaborator
NASA Langley Research Center, Hampton, VA

Proposal submitted in response to NASA Research Announcement NNH14ZDA001N – AIST Research Opportunities in Space and Earth Sciences (ROSES) 2014, A41: Advanced Information Systems Technology

July 9, 2014

Award

NAIADS proposal was one of 24 projects selected for federal funding, with $1M **budget for 2-year period.**

Started 2015 and ended January 2017.

NASA Science Mission Directorate
Research Opportunities in Space and Earth Sciences – 2014
NNH14ZDA001N-AIST
A.41 Advanced Information Systems Technology (AIST)

NASA's Science Mission Directorate, NASA Headquarters, Washington, DC, has selected proposals for the Advanced Information Systems Technology Program (AIST- 14) in support of the Earth Science Division (ESD). The AIST-14 will provide technologies to reduce the risk and cost of evolving NASA information systems to support future Earth observation and to transform those observations into Earth information.

Through ESD's Earth Science Technology Office a total of 24 proposals will be awarded over a 2-year period. The total amount of all the awards is roughly $25M.

The Advanced Information Systems Technology (AIST) program sought proposals for technology development activities to enable science measurements, make use of data for research, and facilitate practical applications for societal benefit by directly supporting each of the core functions within ESD: research and analysis, flight, and applied sciences. The objectives of the AIST Program are to identify, develop and demonstrate advanced information system technologies that:

Reduce the risk, cost, size, and development time of ESD space-based and ground-based information systems,

Increase the accessibility and utility of Earth science data and models, and

Enable new Earth observation measurements and information products.

A total of 124 proposals were evaluated of which 24 have been selected for award.

EOS Data Profile

Earth Observation Data

- Expected volume of used data products: exceed 100 Petabytes in 10-15 years.
- Data is distributed nationally (NASA, NOAA), and internationally (ESA, etc.).
- NASA observational data format is standard HDF (Hierarchical Data Format)
- NOAA observational data format is standard NetCDF (Network Common Data Form)
- Observations from different sensors are NOT synched / merged !
- Each sensor/mission has a separate data product line !

Climate Model Data

- Expected volume of CM outputs: may exceed Exabyte in 10-15 years.
- Data is distributed nationally (NASA, NOAA, DoE), and internationally (UK).
- Multiple Climate & Earth System models (30 – 50).
- Perturbed Physics Ensemble (PPE) for a single model: about 5 Petabytes.
- Climate Model output format is standard NetCDF

EOS Requirements

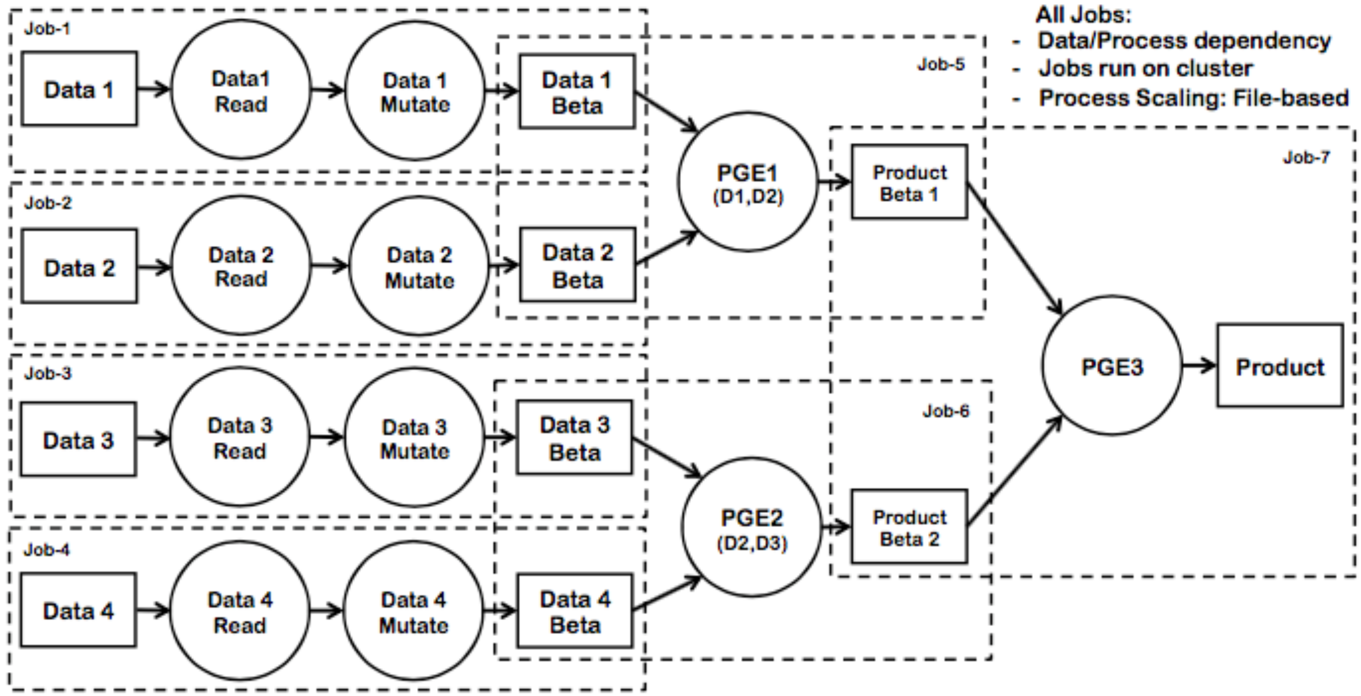
Maximize information content and science output

Major Challenges

- IT infrastructure is not optimized for the task: data is distributed, slow connections, security, etc.
- Traditional PGE-based workflow is I/O bound (job-per-file-per-CPU)
- Traditional configurations are network bound for heavy data handling
- Time and resource intensive, the costs are very high.
- Opportunities are not realized: e.g. MISR/MODIS, A-train Level-1, etc.

Traditional data processing diagram

Traditional data processing approach is slow and inefficient



- Product Generating Executable (PGE) based
- File-by-file multiple batch jobs
- A lot of intermediate data (beta);
- Process is I/O limited, scaling is problematic.

NAIADS Objectives

IO Optimizer: minimize data persistency

Networking: local and distributed data-Event building.

Distributed software: move code to data, not otherwise.

New workflow and scaling: Complete in-memory data-Event streaming, massive scaling;

Multi-lingual: Support new and heritage codes, optimal language for given service.

Flexibility: Re-configurable to multiple applications (fusion, processing, data mining, etc.). Micro-services approach

Adaptability: supporting various hardware configurations (cluster, cloud, servers, etc.), and various file systems: NFS, GPFS, Apache Hadoop / Spark, etc.

Portability: Support various OS platforms (Unix, Linux , MacOS, ...)

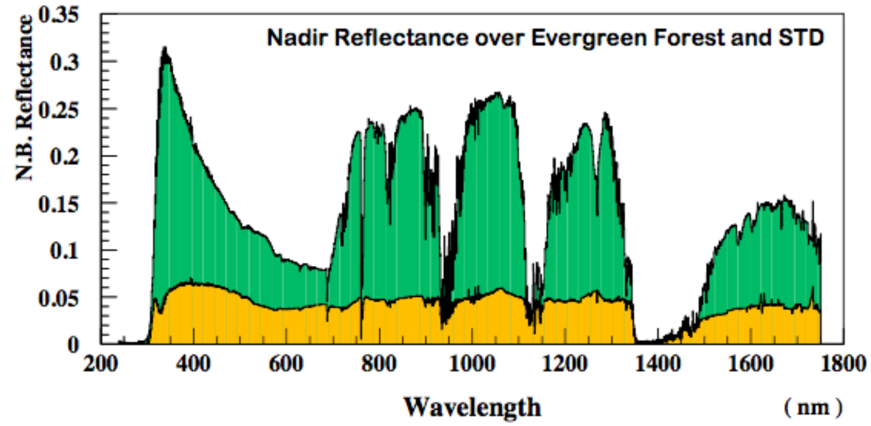
Data format agnostic: I/O, transient data, and metadata (NetCDF, HDF)

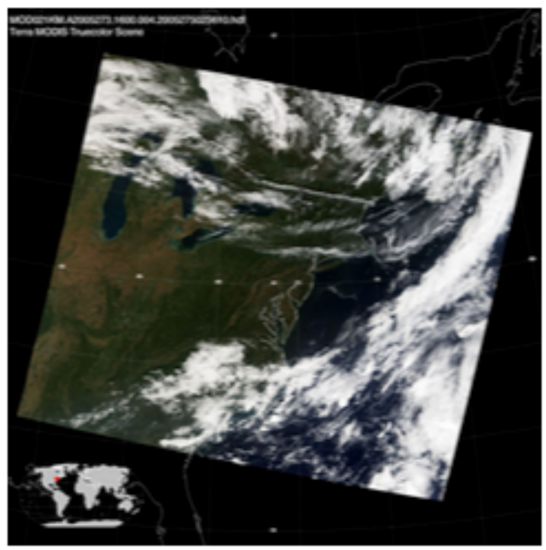
Provenance: track data at any stage of processing.

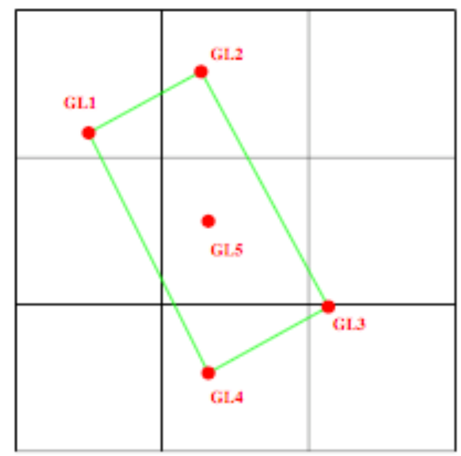
CLARA framework was chosen to meet NAIADS design objectives

Application: SCIAMACHY/MODIS/ECMWF Data fusion and processing

- OBJECTIVES:
  - To demonstrate NAIADS approach and full functionality using existing data;
  - To benchmark NAIADS performance; Available data: **9 years** of near-coincident measurements of from SCIAMACHY and MODIS; **Create new fused SCIAMACHY/ MODIS/ECMWF data product (requested by a number of projects).**
- SCIAMACHY Level-1 Data:
  - Spectral measurement for every footprint: 30 km x 230 km; Swath 950 km (4 footprints) from 10 AM Sun-synch orbit.
- ECMWF Data (re-analysis):
  - Gridded (0.125o);6 weather parameters; Map every 6 hours;
- MODIS/Terra Level-2 Data:
  - Level-2 Cloud and Aerosol Data Spatial scale: 1 / 5 km and 10 km spatial; Swath 2300 km (global coverage daily); 10:30 AM Sun-synch orbit.



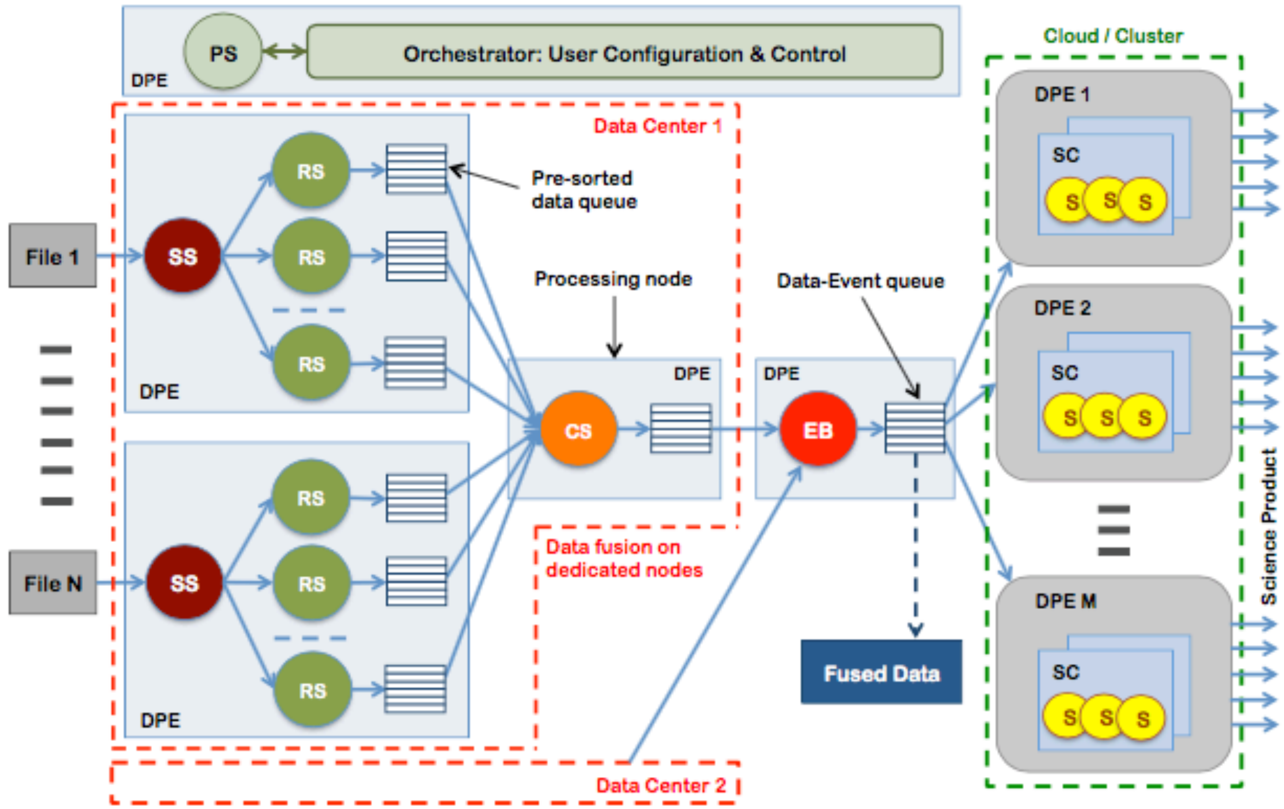




SCIAMACHY FOOTPRINT 30 x 230 km

NAIADS Architecture

- PS: multi-sensor coincident data predictor service.
- SS: into-memory fast data staging service (multi-file).
- RS: parallel from-memory data reader service (pre-sorting).
- CS: data concentrator service in a data center (IO/network optimization).
- EB: complete data-Event Builder (adaptation to algorithm).
- Scaling: data-Event streaming to Cloud with minimized IO.



NAIADS deployment on AWS

AWS c4.8xlarge instances, 36 vCPUs ~= 18 physical cores

Staging data from AWS S3

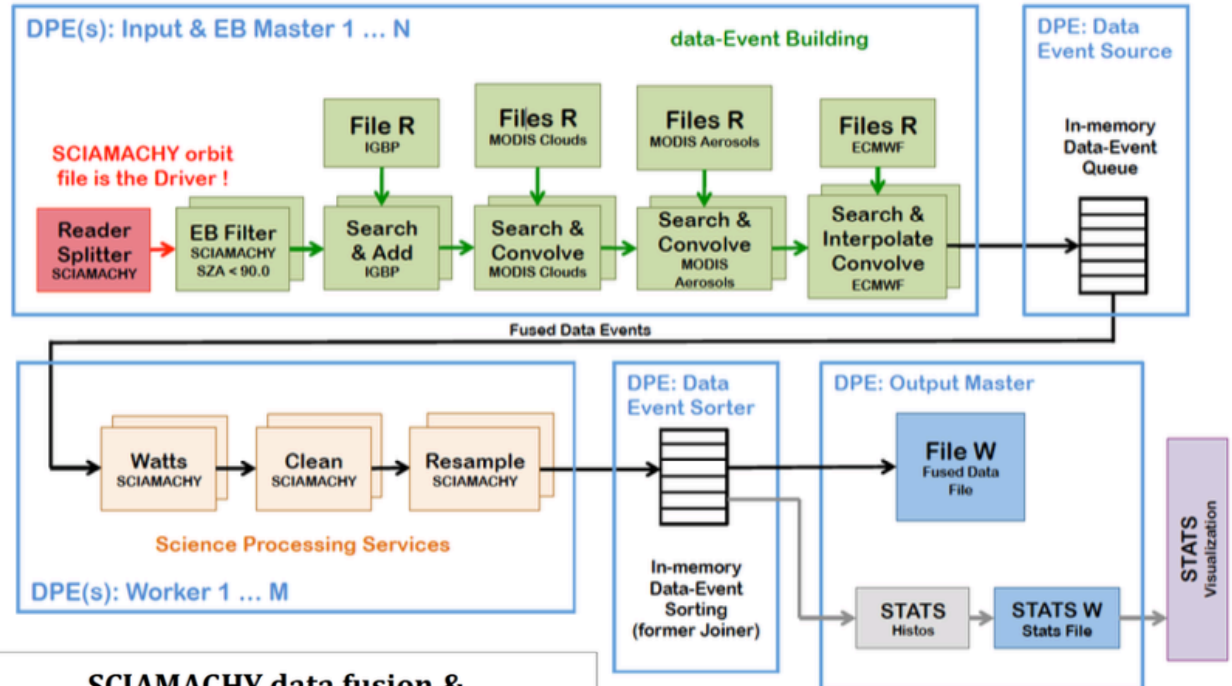
Data processing rate based on average workflow execution over 10 SCIAMACHY files.

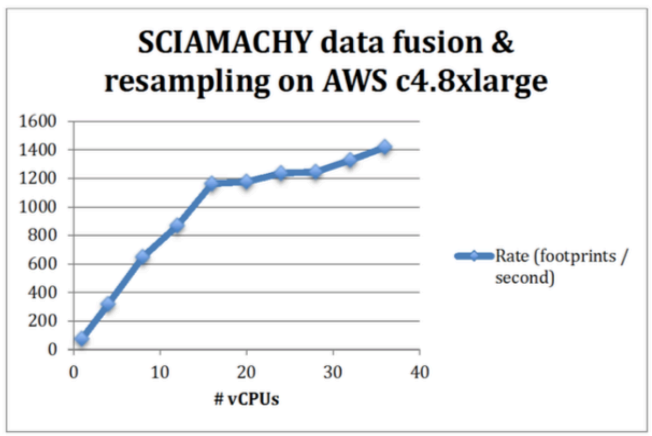
Vertical scaling up to 1.4KHz on single AWS node

Data processing continuous web monitoring

9 years of data has been processed. More data is currently being processed.

Successfully met project goals. Demonstrated what is possible.





Technology Transfer

New interagency agreement between NASA LaRC and JLAB DOE to use CLARA for multi-satellite data processing.

Total budget $250K for the time period from July 1st 2017 to January 1st 2018. JLAB will get ~25% of the total budget.

Project PI: Paul W. Stackhouse
paul.w.stackhouse@nasa.gov

NASA LARC D&F TEMPLATE

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA)
LANGLEY RESEARCH CENTER
DETERMINATION AND FINDINGS (D&F)

Authority to use an Interagency Acquisition (IA) pursuant to Federal Acquisition Regulation (FAR) 17.5 and The National Aeronautics and Space Act (51 USC 20113(e)) as implemented by NASA FAR Supplement (NFS) 1817.5 and The Economy in Government Act (31 USC 1535)

FINDINGS

1. NASA Langley Research Center's (LaRC) "Requesting Agency" Science Directorate has a need to enter into an Interagency Acquisition (IA) with the Department of Energy (DOE), Thomas Jefferson National Accelerator Facility (Jefferson Lab) "Servicing Agency" for the demonstration and extension of recently developed Earth Science data processing middleware to a new application.

The proposal – "NASA Information And Data System (NAIADS) for Earth Science Data Fusion and Analytics" – was selected for funding by the NASA Earth Science Technology Office's (ESTO) Advanced Information Systems Technology (AIST) program. The NAIADS project was a 2-year collaborative effort between the NASA LaRC and Jefferson Lab teams focusing on development of a new and advanced framework for Earth Science data fusion. The original interagency agreement between NASA and DOE under which NAIADS was developed and tested was established in 2015 and extended through March 15, 2017. The AIST program approved the application of the new advanced framework for an additional Earth Science data fusion case study. This new interagency agreement will be established to support this statement of work, which provides for an additional case study of the NAIADS framework requiring work through January 2018.

Description of Work: The scope and objective of this effort is to adapt and test the NAIADS software framework developed under the initial interagency agreement with a more complicated Earth Science data fusion data production project at NASA LaRC. The CLARA (CLAS12 Reconstruction and Analysis) is a Physics Data Processing (PDP) application framework developed for CLAS12 experiment at Jefferson Lab[1]. This CLARA framework, developed by the Jefferson Lab team serves as the data stream processing framework for the NAIADS implementation. Both institutions will benefit from implementation of such services as into-memory data mapping, in-memory read parallelization, on-line web services for data processing control and monitoring, and data analytics (filters, histogram and N-Tuple building, fitting). The NAIADS software includes the following features a) multi-sensor data-Events building, required for advanced science algorithms; b) data fusion with data-Event granularity for large and distributed data sets; c) massive process scaling on multi-CPU computing systems by efficiently utilizing multicore architecture systems, by data streaming, and minimizing the IO;