

Report from EICUG Computing Coordination Group

Graham Heyes

EIC Computing Coordination Committee - introduction

- Liaisons for EIC computing resources from labs have been meeting for several years
 - Jerome Lauret (BNL) and Graham Heyes (JLab)
- Added Andrea Bressan – represent INFN, EICUG, and EIC Software WG
- During the detector proposal development process added reps:
 - ECCE – Cameron Dean
 - ATHENA – Wouter Deconinck
- Purpose of the committee was to:
 - Coordinate development and deployment of computing resources at BNL and JLAB
 - Provide a mechanism to include resources from other labs etc
 - Gather requirements
 - Monitor usage
 - Provide help when help is needed

EIC Computing Coordination Committee – charge

In 2022 the charge was reconfirmed with some refinement:

- Coordination of resources among EIC computing efforts
 - Assure computing infrastructure follows DOE and host lab mandates and guidelines
- Keep record of required and available resources as well as their usage
- Access point for institutions that intend to contribute computing resources
 - Propose level to include representatives of major resource provider contributors
- Start infrastructure assessment to fold in international computing resources
- Pursue and develop in-common data flow and management (for detector-1 and potential detector-2).

EIC Computing Coordination Committee - membership

- Contact persons/liaisons for EIC computing resources from labs (2)
 - Jerome Lauret (BNL) and Graham Heyes (JLab) – co-chairs
 - Contacts for EIC Project Detector, also listed as contacts at EICUG web pages:
<http://eicug.org/web/content/electron-ion-collider-users-group-eicug>
- Links to EICUG Software/AI Working Group (2)
 - Andrea Bressan, Cristiano Fanelli
- Computing/Software Detector-1 Working Group (+1)
 - (Andrea Bressan, Cristiano Fanelli,) David Lawrence
- Simulation, Production and QA Working Group (1)
 - Wouter Deconinck

EIC Computing Coordination Committee - timeline

- Regular meetings of CCG this year to accompany detector-1 consolidation, the path to detector-2 process, and preparations towards pre-Technical Design Report
 - Meet nominally every four weeks – ad hoc as required
- Structure and responsibilities to be revisited early 2023 after formal Detector-1 Collaboration is formed and governance structure is defined
 - Computing is not a project responsibility - labs would provide support
 - Respective lab managements have discussed and agreed with the above timeline

JLab improvements since last report

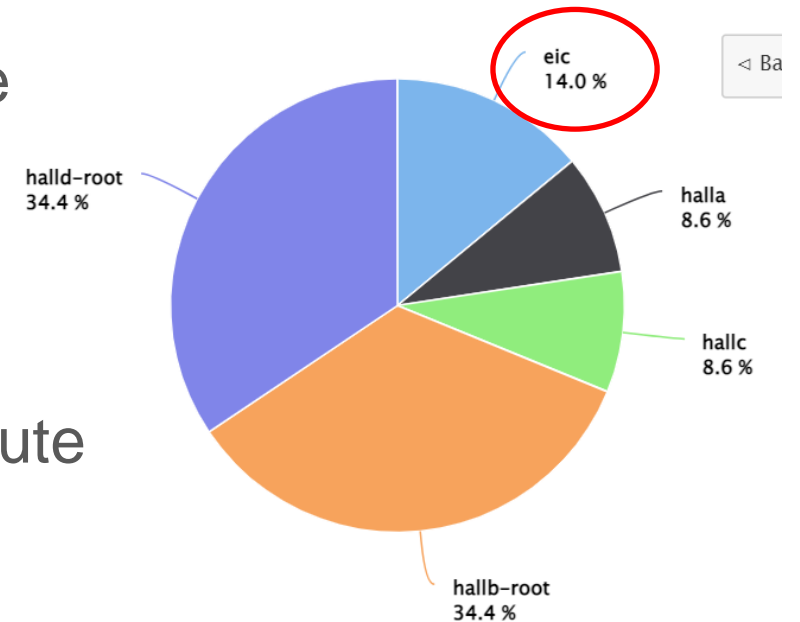
- JLab has upgraded to OSG 3.6
 - Token-based authentication, dropping GSI
- Finalizing additional OSG submitter nodes (scosg2201 gluex , scosg2202 clas12)
 - This leaves scosg20 as dedicated target for EIC jobs.
- Currently still have XRootD read-only access to EIC disk storage
 - Close to having SciToken based mechanism for read/write access
 - Using CoManage to control approval of access
 - Will allow access control for both BNL and JLab identities using CILOGON and the JLab Shibboleth/Single Sign on providers
 - This is running now in our testbed,
 - Beta testing in August
 - Wider testing after that.

JLab improvements since last report

- Hired Brad Sawatzky as “Nuclear Physics Computing Lead”
 - Coordinate computing resource use and other access of NP computing at JLab.
- Brad is looking at RUCIO and has been talking with Markus and others
- Tentative plan is a pilot RUCIO install using repurposed hardware later this summer
 - Use to inform budget decision for production system in FY23

JLab improvements since last report + upcoming

- Started using ServiceNow to add Knowledgebase articles to replace outdated documentation
- Improvements to reporting WebApps at scicomp.jlab.org
- \$300k procurement to add dual AMD Milan compute nodes to cluster
 - EIC has a 14% fair share allocation, adding nodes boosts resources available to EIC.
- Four additional tape drives have been installed total of 24 LTO8s.
- ESNNet 6 hardware is installed at JLab currently dual 20 Gb/s soon upgrade to 100 Gb/s



BNL activities

- Computing efforts supported by a Program Development activity
"Data store and access infrastructure in support of EIC's science"
- Project goal: Deploy a reliable, easy access and scalable storage services in support of Electron Ion Collider (EIC) science and its future detector simulation needs.
 - Support for the simulation needs for the EIC by covering for 60% of the EIC science requirements in storage sustaining 50% of its simulation needs and +10% of user analysis
 - Build a solid foundation to an EIC computing infrastructure at BNL in support of data science
- Modest funding but allowing for dedicated hardware for the EIC over a 3 years plan.
 - Year #1: Dedicated HPSS infrastructure for the EIC + consolidation of the S3 storage (dedicated hardware)
 - Year #2: (a) Grow the storage by +0.5 PBytes (b) Enable Federated ID access to the storage © work with JLab's partners toward a truly global file system where storage could be added on either end
 - Year #3: (a) Grow the storage by +0.5 PBytes (b) Enable global distributed caching for our EIC storage facility i.e. automated data migration to remote cache would allow a fast, low latency access of cached dataset throughout the institutions participating in the EIC.

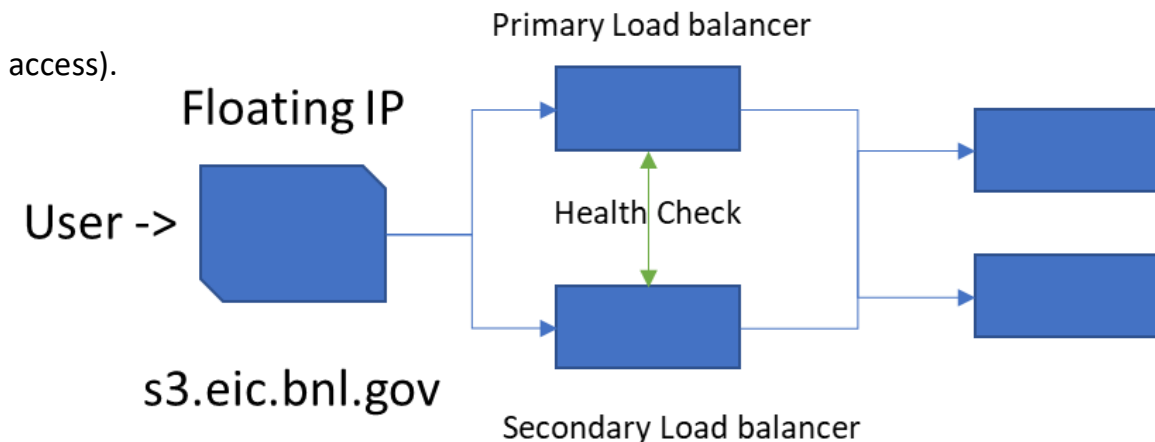
Where are we? What did we do?

- The planned mass archival system is fully operational - BNL now has a kernel for a production quality HPSS setup for the EIC
 - 3 dedicated LTO8 drives - Each drive has maximum throughput of 360 MB/sec
 - 300 TB HPSS disk cache
 - 180 LTO8 tapes. Each tape provides 12 TB of space for a total of ~2 PB
 - Perspective: The EIC has $< 1/2$ Pbytes of data at BNL (x2 with JLab) to save in HPSS. The HPSS dedicated archival storage service covers for today's needs.
- DataCaorusel service will be used for an efficient data retrieval of files from tape
 - Used at RHIC and in ATLAS
 - The first components are deployed and ready (the user based command line will be provided in 1-2 weeks from now)



S3 storage consolidation, to be completed in FY22

- We faced delivery delays in hardware so an early deployment was not possible.
 - However, all hardware was shipped a few days ago
 - We are now looking for a provisioning in early to mid August
 - During this work, the data in BNL/S3 will be moved to the new storage (downtime will be announced)
 - Users keys will be replaced, refreshed
- Our current setup:
 - Single MinIO NAS gateway being used on top of a Lustre filesystem to export the POSIX filesystem as S3 storage. Performance was limited to the single NAS gateway and the setup does not have server redundancy.
 - Extracting audit logs is not trivial. Performance – 10Gbps. Current Usage – 337TB and 23 million files.
- Planned with new hardware
 - HIGH Availability (HA) Proxy Load Balancing (no single point of failure or access). Expected performance would be quadrupled i.e. 40 Gbps.
 - Comprehensive logging would be added to get detailed per user/client statistics.



Conclusions

- The committee composition has evolved to reflect new detector1 groups
- The ECCG is the “one place to go” for computing requirement requests
 - Gives you and us the opportunity to stay aligned
 - Greatly helps the two labs to plan
- Remember that, more so than usual, providing computing resources takes time
 - Plan for and make requests as early as possible - little room for last moment provisioning of hardware
 - We plan to proceed with a new wave of requirement gathering once the detector #1 re-design is more mature
 - If there will be major course changes, please let us know
- Most of all, we are here to help you succeed with computing tasks