Jefferson Lab Scientific Computing Infrastructure Update

CLAS Collaboration Meeting

July 2025

Brad Sawatzky

Tuesday, July 8, 2025

Jefferson Lab

U.S. DEPARTMENT OF ENERGY | Office of Science

JSA

Jefferson Lab's High Throughput Computing – The Farm

- Farm has many components
  - ~30000 compute threads
  - ~11 PB Lustre
  - ~5 PB NFS/XRootD (ZFS)
  - ~100+ PB of (online) Tape
  - Consumes >400kW
  - GPU nodes available as well
  - Interactive nodes (ifarm240x)
- Growth is $$$ and based on projections from Halls
  - Expenditures generally switch between storage + CPU every other year

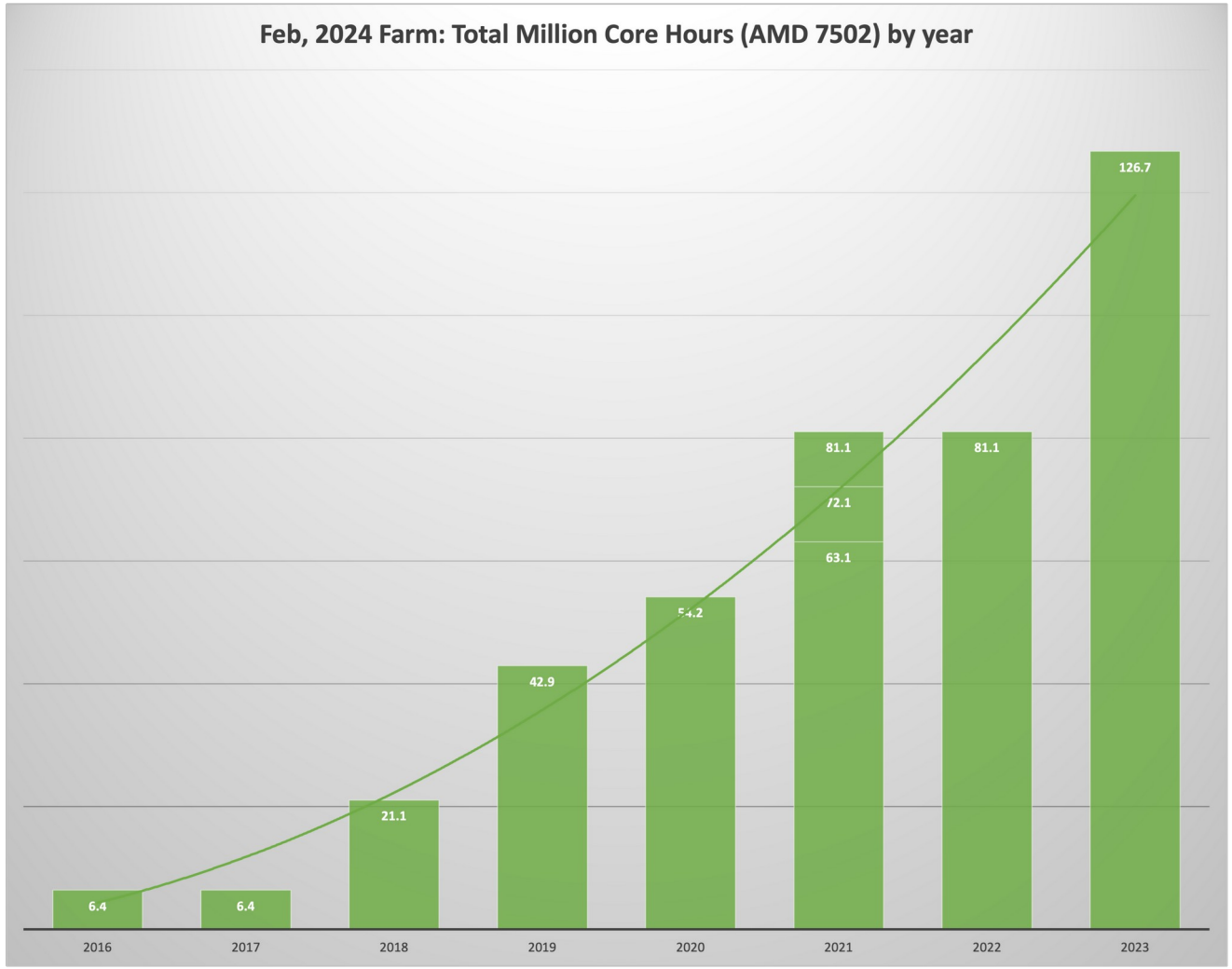
    (FY25 is 'Compute' year; FY26 is 'Disk')
  - Computing projections assessed annually; next one Aug 2025

Jefferson Lab's High Throughput Computing – The Farm

- FY25 +1600 AMD EPYC 9354 "Genoa" cores
  - 4 GB/thread
  - Installation in Fall 2025
- FY23 3072 EPYC 7763 AMD "Milan" cores
- FY19 – FY21 Were EPYC 7502 "Rome"
- GPU Node available but underutilized
  - 20 A800 (40 GB) cards
  - 18 A100 (80 GB) cards

- Planning for Disk purchase in FY26 (tentative)

  → +10PB of /cache, /volatile (Lustre)



Jefferson Lab

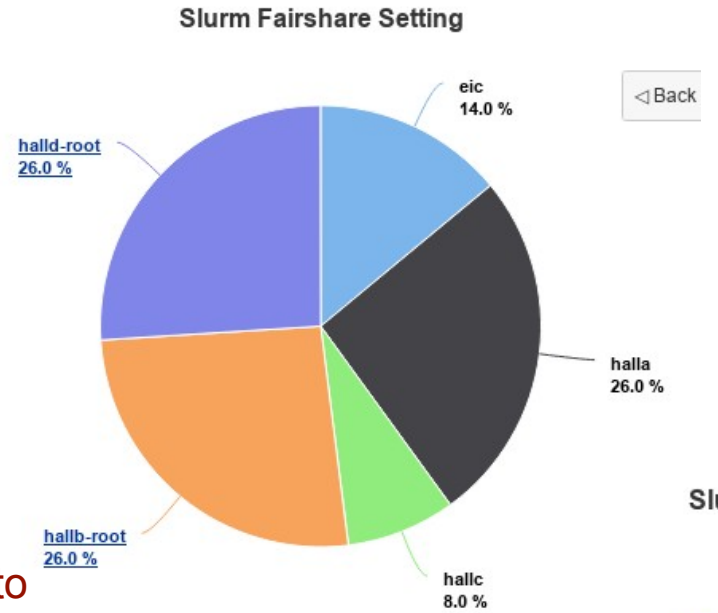
Disk Storage Areas and Their Uses

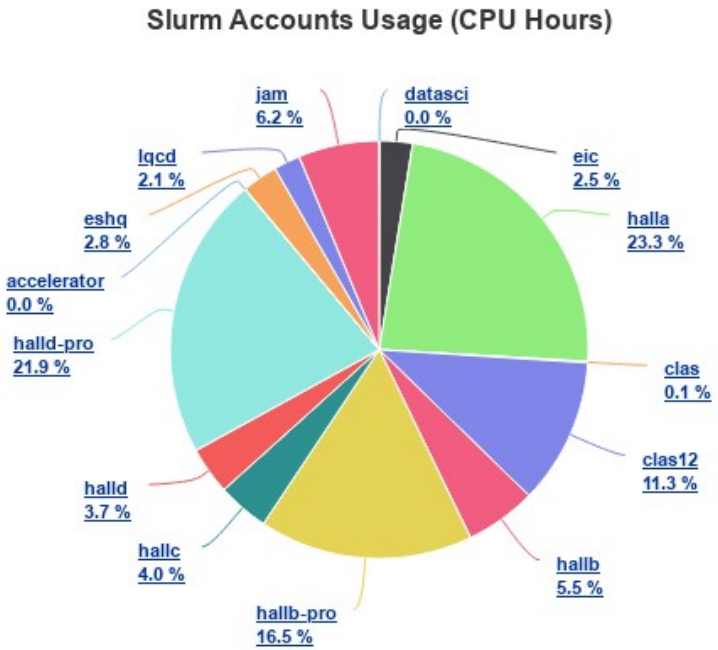
- *There is strong demand for more disk. This is an explicit priority for for us.*
  - High performance, reliable disk and associated infrastructure is still expensive and has been long lead…
  - We attempted to get ahead of the curve on disk in FY24 but JLab cancelled a planned Lustre disk purchase (We will try again in FY26)
  - NEW CephFS disk storage (standard POSIX) will add to /work in FY25 (~1 PB)
- Lustre Storage is good for large files, streaming, large block I/O, production farm runs.
  - Lustre is *not* good for small files, high IOPS, and frequent metadata operations (worst case: open, write 1kB, close, repeat)
- /work will not scale for large farm campaigns.
- Node-local /scratch is good for jobs with high IOPS to working files.
  - *Note: SWIF-declared MSS files are automatically copied to node-local working directory*
  - *Old GlueX wrappers still doing this manually (watch for old code and update it)*

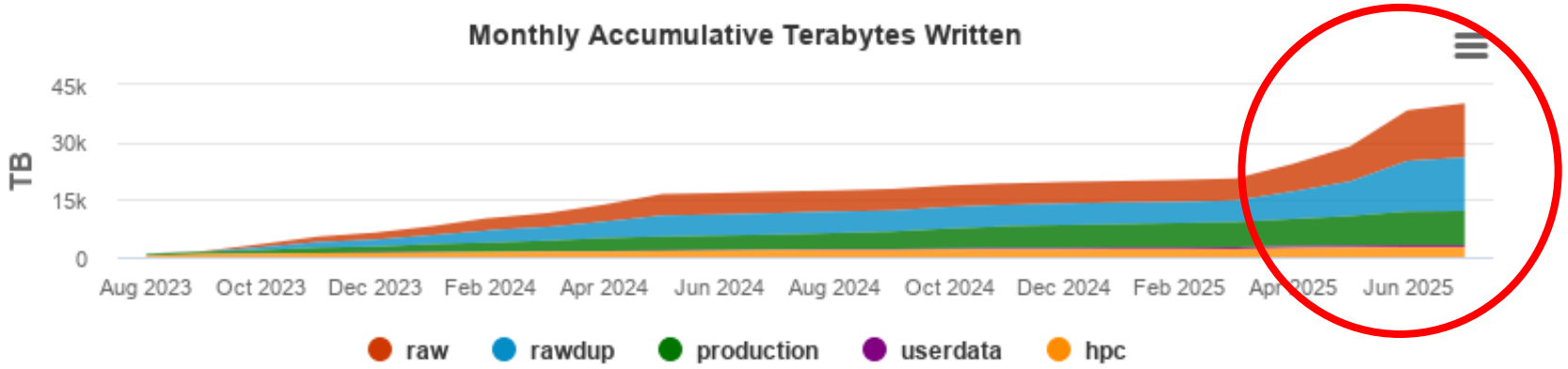
Path	Best Use	FS Type	Deletion	backup
/cache	Bulk I/O, Migration to tape	Lustre	Once on tape	/mss
/volatile	Bulk I/O Temporary storage	Lustre	auto	NO
/work	Source code, DB files, exe's, etc. User Managed	NFS+ ZFS	manual	NO
/home	Dot files, personal documents, etc	NFS ssd	manual	YES
/farm_out	Farm job stdout/stderr	NFS ssd	auto	NO
/group or /scigroup	Source code Papers, thesis, analysis scripts	NFS ssd	Manual	YES
/scratch	Farm job I/O to node local disk	ssd	auto	NO
/u/scratch	*CUE scratch. Deprecated (Unavailable on el9)*			
/cvmfs	Software stack. Configuration.			

Jefferson Lab's High Throughput Computing – The Farm

- The farm is routinely busy
  - Utilization generally over 80%
- FairShare allocations are used to balance consumption between the halls
  - A,B,D ~26% each
  - C ~ 8%
  - EIC ~14%
  - Bursts beyond share when cycles are free
  - CLAS routinely claims free cycles.
  - NB: Hall A data volumes and analysis requirements rose to Hall B/D levels with the SBS program in 2025
- Data R+W to tape exceeded 13 PB (!) in June 2025



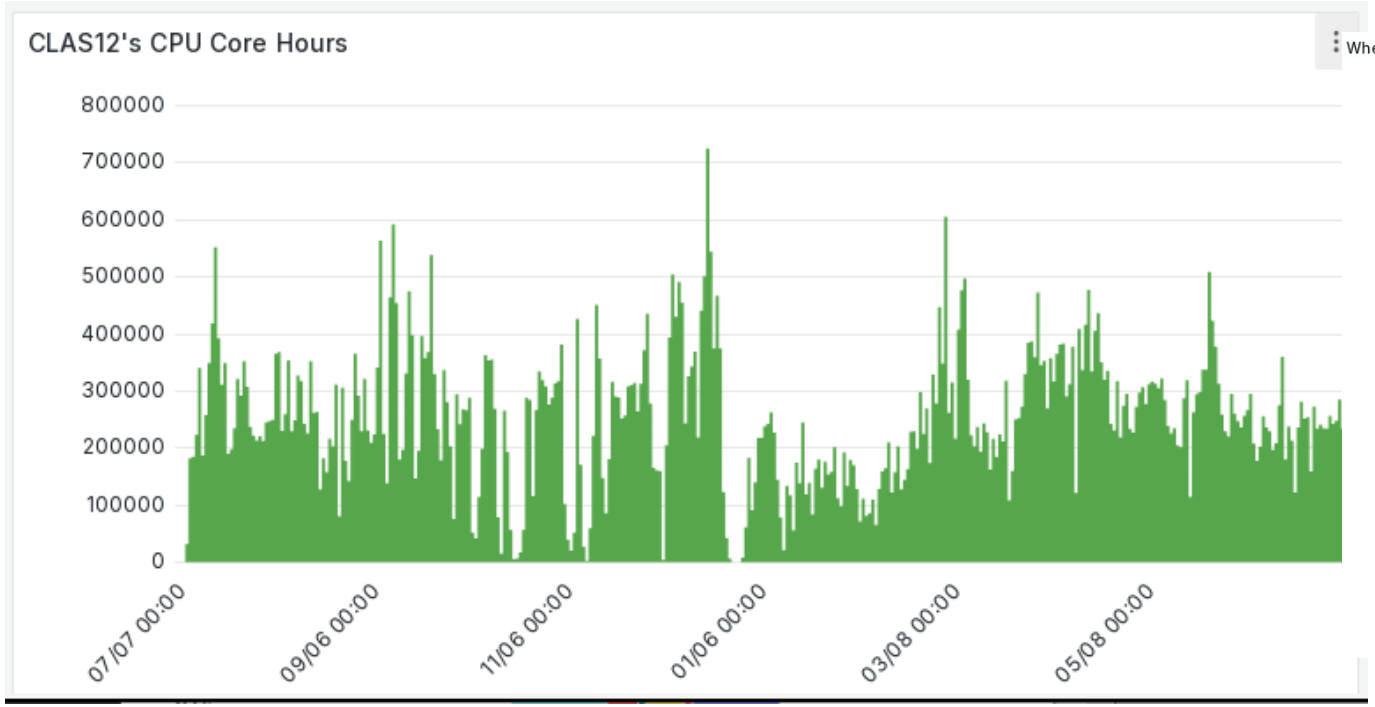


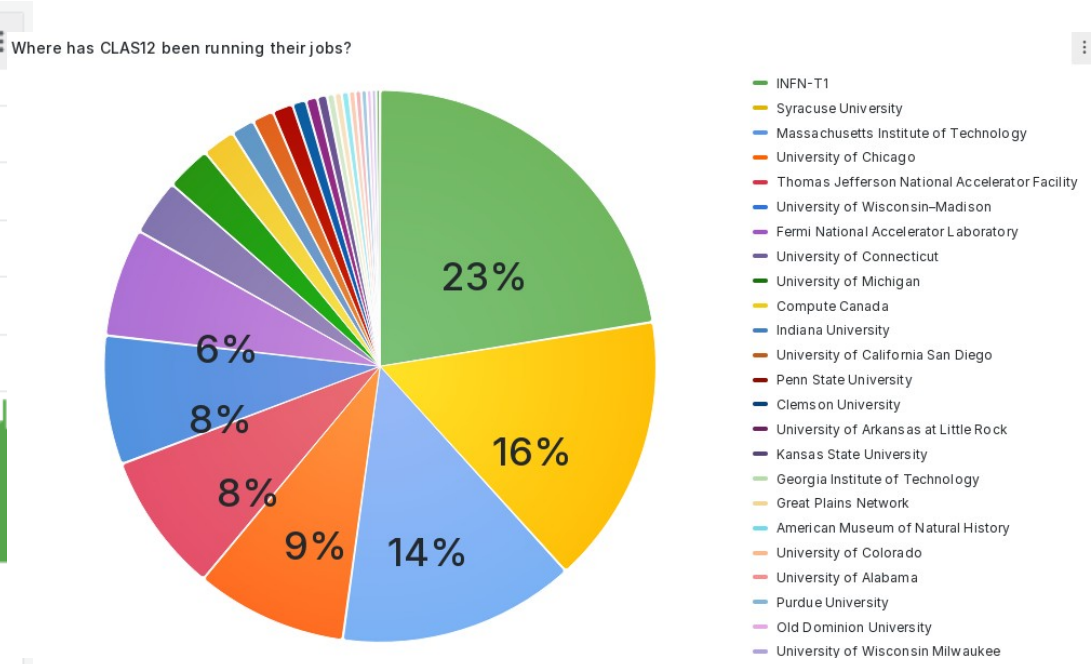


Jefferson Lab

Open Science Grid Processing

- The Open Science Grid continues to be a significant resource for Monte Carlo Simulation Compute Cycles. GlueX and CLAS12 are significant consumers of CPU cycles.
- EIC runs the bulk of their simulation on the OSG
- Encouraging Hall A (SBS, MOLLER) to follow your lead!

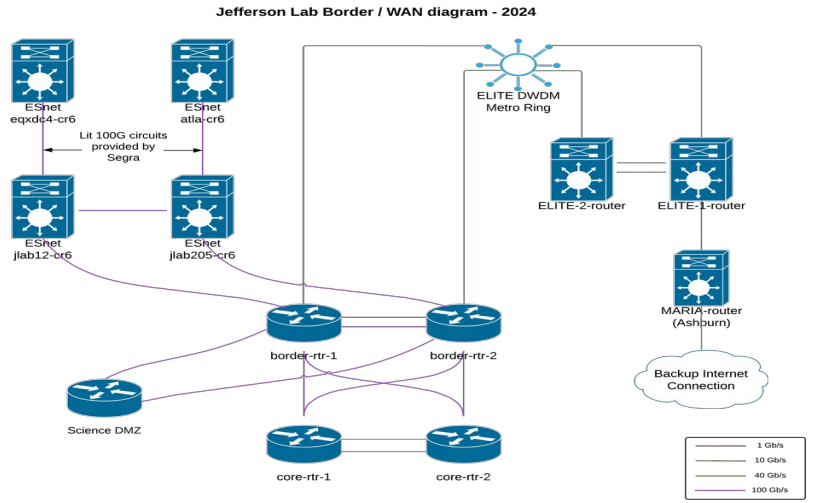


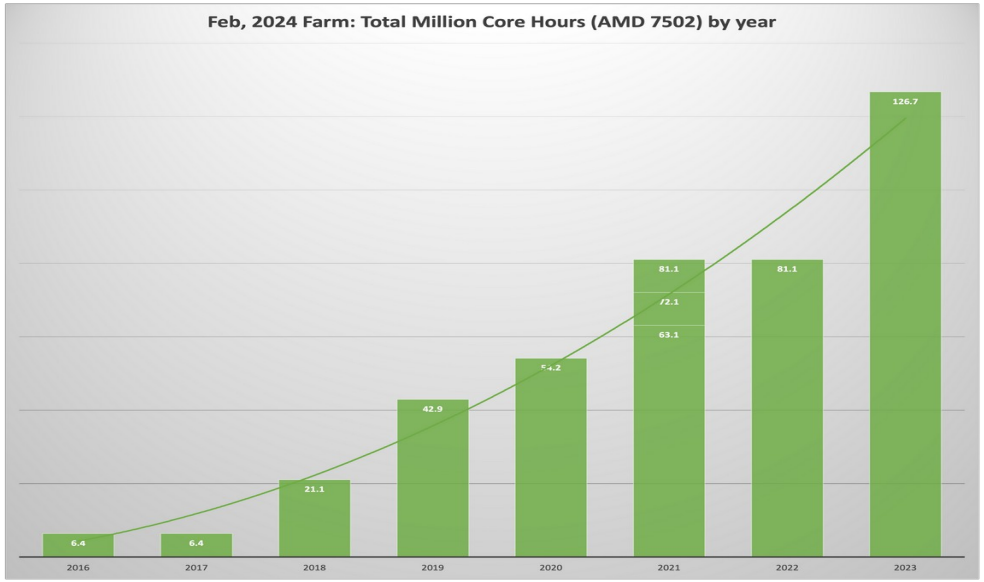


Jefferson Lab

Recent and Near-term Infrastructure Updates (HW) : 2025

- "Farm25" CPU node installation (late-fall)
  - This will require ~1day downtime for Farm in Oct/Nov. We will schedule/plan with your Hall Compute Coordinator!
- 20 new A800 GPUs installed (done)
  - Currently underutilized!
- "/work" disk increase (early-fall)
  - All NVMe disk; CephFS backed
  - Onsite S3 storage if there is interest(?)
- +4 tape drives brought online (done)
  - Tape system throughput is ~12.5 GB/sec
  - Significant optimizations to SW and HW backing Jasmine in May/July
- Tentative plan is to buy Lustre disk in FY26 (/cache, /volatile)
  - Guided by Hall needs/requests passed on by Hall Compute Coordinators





Jefferson Lab

Hall ESX Virtual Machine Cluster

- VMs within the Hall experimental enclave are available
  - High-uptime infrastructure explicitly targeted to support Hall operations, etc.
  - Update policy, downtime scheduling for VMs defined by Hall Compute Coords as usual.
  - Ex: Slow control systems
    - EPICS softIOCs
    - Windows/Rockwell control systems
      - ie. Hall C: cmagnets, skylla10
    - PXE boot hosts/services
  - Ex: Data-base hosts
    - RCDB, CCDB hosts
  - Ex: “Remote CH” support hosts
- VM hosts functionally operate within the Hall subnets (no firewall issues)
  - Direct access within respective subnets
  - 2-factor hop (via. hallgw, etc) as with any existing Hall hosts
- Take advantage of VM flexibility
  - snapshotting / backups
  - auto-failover on HW issues
  - advantages wrt “cloning”, load balancing, etc
- Keep this in mind for future deployments, HW upgrades, etc
  - Cheaper / better than HW for many applications

Infrastructure Updates (SW): 2025

- code.jlab.org (GitLab)
  - CI/CD
  - Container registry
  - JLab GitHub Org will remain while cost-effective
- CVMFS 'Stratum 0' for JLab
  - /cvmfs/jlab.opensciencegrid.org/ ← *NEW*
  - will replace /cvmfs/oasis.opensciencegrid.org/jlab/ at some point
- Kubernetes for workflows that don't fit Batch model
  - OpenShift 'enterprise' K8 platform rolled out in 2024; it was aggravating...
    - required significantly more 'backend' work than advertised
    - CI/CD now stable; other (internal) projects being deployed; adding some hardware to the cluster
  - more conventional K8 deployment being rolled out in TestBed
    - This is probably where Users should start

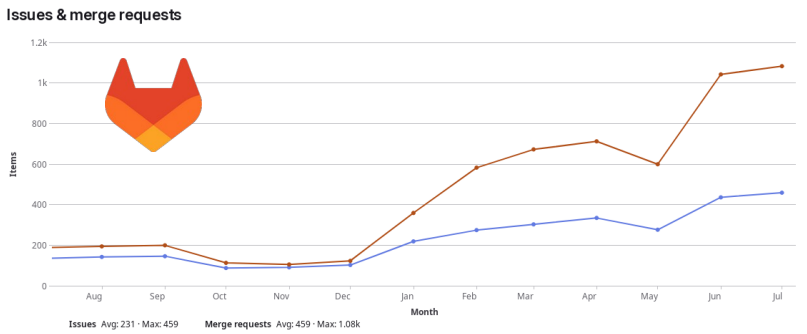
- Rucio
  - Distributed (large-file) data management framework
  - EIC simulation campaigns in full production
  - MOLLER, GlueX planned next
  - Swif+Rucio file URI integration in progress
    - rucio://…..

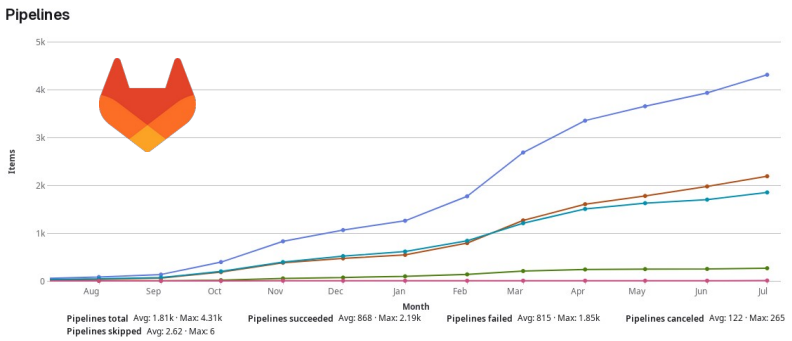
- JLab Research DB
  - "One stop shop" to locate data, publications, workflow information, logbook references, etc…
  - *Ties into DOE and JLab Data Management policy changes that will be announced Fall 2025*

code.jlab.org (GitLab Service)

- GitHub is getting $$$
  - CI/CD, storage, etc are all metered costs
  - JLab is on a ‘legacy’ license model for now but limitations are frustrating
- JeffersonLab GitHub Org will be maintained as-is
  - BUT goal is for code.jlab.org to be a “value-added” proposition

- code.jlab.org (GitLab instance)
  - JLab run/managed
  - Open / Offsite access
    - Federated logins avail.
  - CI/CD and Storage can leverage our Farm
  - Built-in Container Registry
  - Supports several Data Management requirements important to JLab / PhysDiv





Jefferson Lab

GSPDA Mini-Software Workshops

- Thanks to Grad Students PostDoc Assoc. for organizing!
  - Kate Evans (W&M)
  - Cameron Clarke (JLab)
- Targeting new JLab grad students and Farm users
  - ifarm/Farm basics
  - worked examples using git, slurm, swif, containers, etc
- JSA support could be better
  - Please advocate for funding in FY25!
  - Contributed talks would be great!

Thursday, May 22, 2025

9:00 A	**Welcome**
9:10 A	**Lecture: Computing Farm Basics and Best Practices** - Brad Sawatzky (Jefferson Lab)
10:30 A	**Break**
10:45 A	**Lecture: Accessing the Fa...**
11:15 A	**Hands-On: Accessing the Farm from Your Computer**
12:00 P	**Networking Luncheon**
1:30 P	**Hands-On: Navigating the Farm and Moving Files**
2:30 P	**Break**
2:45 P	**Hands-On: Using Git and Executing a Basic Script** - Cameron Clarke (Jefferson Lab) Katherine Evans (College of William & Mary)

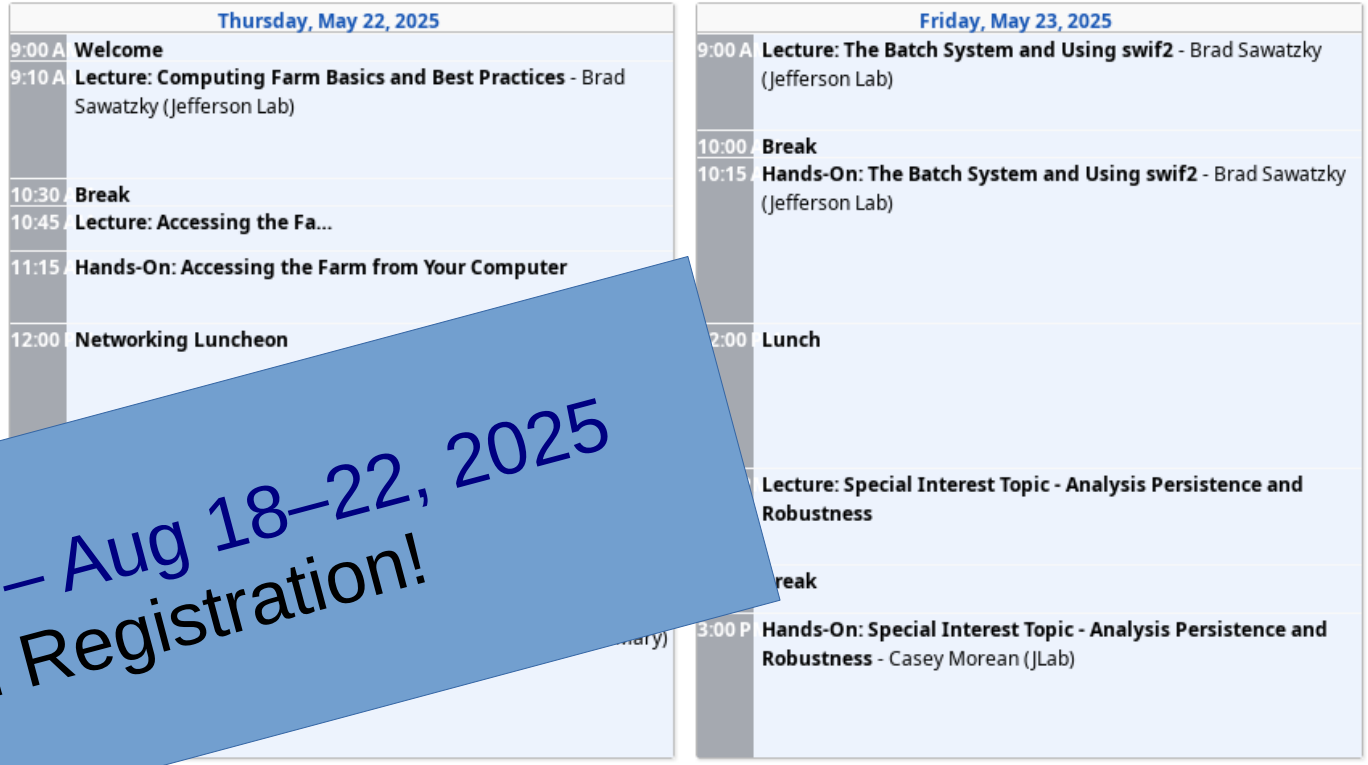
Friday, May 23, 2025

9:00 A	**Lecture: The Batch System and Using swif2** - Brad Sawatzky (Jefferson Lab)
10:00 A	**Break**
10:15 A	**Hands-On: The Batch System and Using swif2** - Brad Sawatzky (Jefferson Lab)
12:00 P	**Lunch**
1:30 P	**Lecture: Special Interest Topic - Analysis Persistence and Robustness**
2:30 P	**Break**
3:00 P	**Hands-On: Special Interest Topic - Analysis Persistence and Robustness** - Casey Morean (JLab)

- Trying to make these at least annual:
  - May 2024 Mini Software Workshop, pt 1
  - Sep 2024 Mini Software Workshop, pt 2
  - May 2025 Computing Bootcamp

GSPDA Mini-Software Workshops

- Thanks to Grad Students PostDoc Assoc. for organizing!
  - Kate Evans (W&M)
  - Cameron Clarke (JLab)
- Targeting new JLab grad students and Farm users
  - ifarm/Farm basics
  - worked examp
    containers, etc
- JSA support c
  - Please advocate unding in FY25!
  - Contributed talks would be great!



Thursday, May 22, 2025

9:00 A	**Welcome**
9:10 A	**Lecture: Computing Farm Basics and Best Practices** - Brad Sawatzky (Jefferson Lab)
10:30 A	**Break**
10:45 A	**Lecture: Accessing the Fa...**
11:15 A	**Hands-On: Accessing the Farm from Your Computer**
12:00 P	**Networking Luncheon**

Friday, May 23, 2025

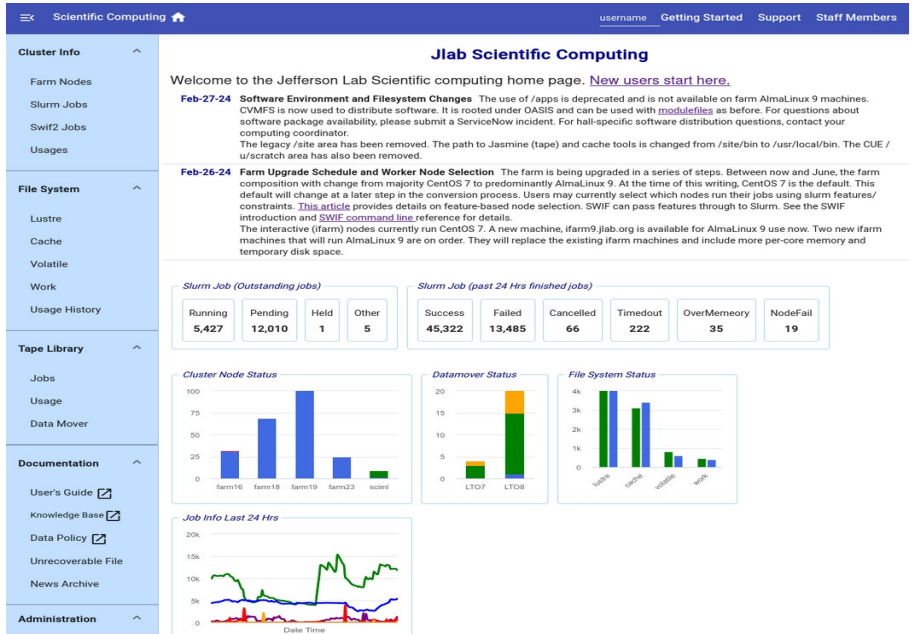
9:00 A	**Lecture: The Batch System and Using swif2** - Brad Sawatzky (Jefferson Lab)
10:00 A	**Break**
10:15 A	**Hands-On: The Batch System and Using swif2** - Brad Sawatzky (Jefferson Lab)
2:00 P	**Lunch**
	Lecture: Special Interest Topic - Analysis Persistence and Robustness
	reak
3:00 P	**Hands-On: Special Interest Topic - Analysis Persistence and Robustness** - Casey Morean (JLab)

G4 Workshop – Aug 18–22, 2025
Limited Registration!

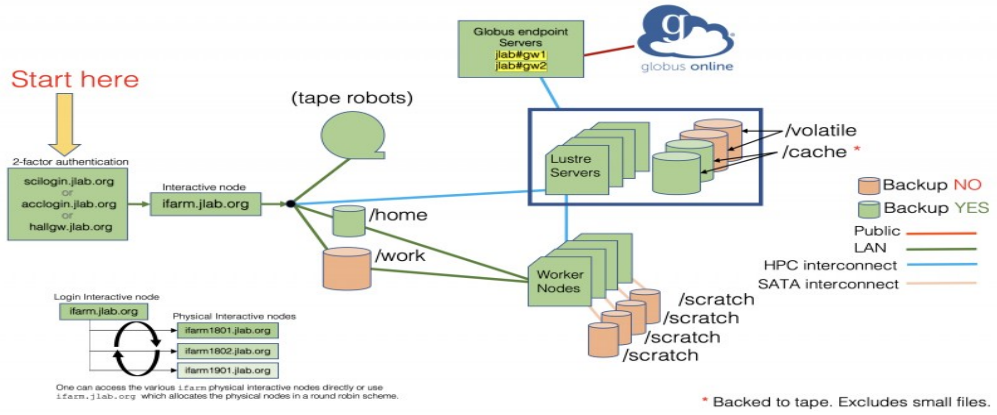
- Trying to make these at least annual:
  - May 2024 Mini Software Workshop, pt 1
  - Sep 2024 Mini Software Workshop, pt 2
  - May 2025 Computing Bootcamp

Information Resources

- scicomp.jlab.org
  - SciComp web page
- scicomp-briefs
  - mailing list for JLab Scientific Computing



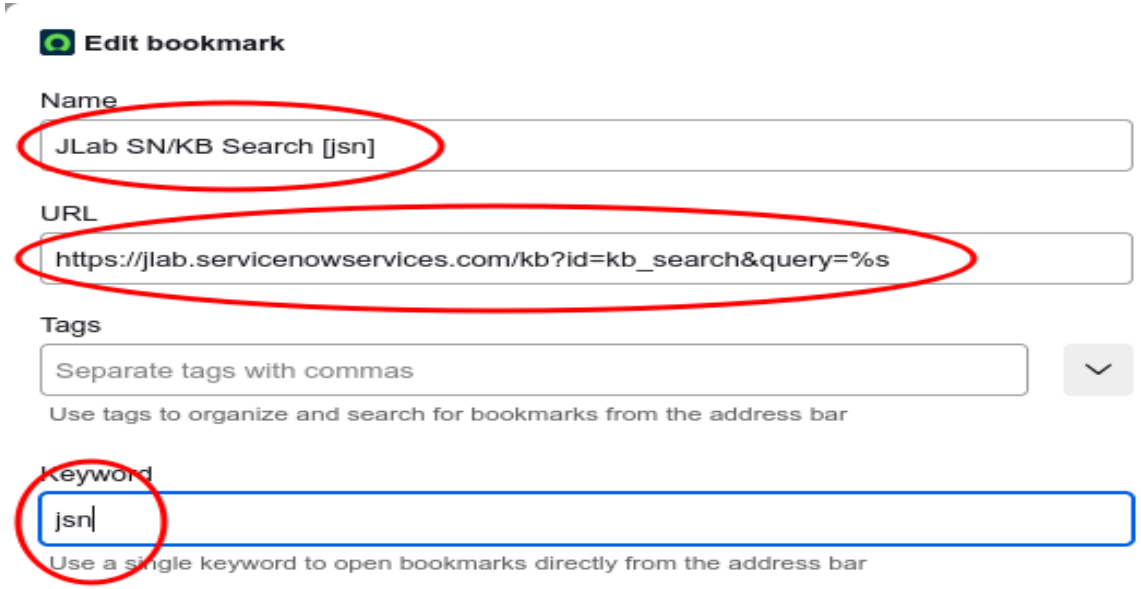
- Documentation links
  - Getting Started
  - SciComp Knowledge Base
  - CST User Portal
  - JLab Helpdesk
    - helpdesk@jlab.org
    - Incident Request



Jefferson Lab

How to find information

- Searching in JLab ServiceNow
  - ServiceNow is where SciComp (and other groups) are putting their documentation.
  - Search all of JLab ServiceNow from within Firefox:
    - Go to https://jlab.servicenowservices.com/scicomp and login (top-right)
    - Bookmark the page
    - Right-click on the bookmark you made and update all 3 fields like so:



- Now you can type 'jsn <keywords>' in the Location bar for instant search

Jefferson Lab

I want to hear from you

- What are the problems / pain-points in your workflows?
- Other question / comments?

Thank you!

Rucio

- Distributed data management system
  - Initially developed for ATLAS
  - Highly scalable & modular
- Features
  - Basic data operations
    - Storage, transfer, deletion
  - Policy based replication (automatic)
  - Designed with distributed storage and ‘protocol agnostic’ data transfer methods in mind
- Can serve as a Replica/File catalog and metadata service
  - Can search its DB and present files matching names, metadata, etc

- JLab Rucio Instance for EIC mature
  - JLab ↔ BNL automated file registration and transport
  - Works well with EIC/OSG simulation workflow
  - More challenging than expected, but worth the work
- Next steps
  - “JLab” workflow integration
    - MOLLER, GlueX top candidates
  - Transparent Jasmine/tape integration
  - Backfill from existing tape library as needed

Containerization Support

- SciComp/PhysDiv working on formal containerization support for Users
  - Documentation
    - ie. Easy ‘on-ramp’ / how-to for common use cases
  - “Official” Infrastructure support
    - Apptainer / Singularity
    - Docker, Podman support for image building and deployment
    - JLab GitLab Container Registry
- Among other benefits, Containers can provide
  - ‘Plug and play’ software configurations
  - SW version / configuration snapshots
  - Ability to run ‘custom’ software frameworks on other datacenters, computers, laptops
- Will also support/streamline workflow capture and data management going forwad

- Apptainer (was Singularity)
  - works on both ifarm and farm
- Podman
  - mostly works on ifarm9 now
  - works on farm
- Kaniko lives
  - Forked and maintained by Chainguard
- Docker
  - not happening on compute clusters
  - but podman == docker (pretty much)
  - Note: docker != dockerhub

Improving Data Management at JLab

- There are a number of ongoing challenges with Data Management at JLab (and elsewhere!)
  - Difficulty capturing the analysis workflow
    - Software toolchain, metadata and calibrations, etc
    - Difficult/impossible to revisit prior ‘working’ code for comparison and cross-checks
  - Raw and Processed data locations may be insufficiently documented
  - Experimental metadata is being scattered more broadly as groups develop distributed and cloud-supported workflows (outside historical lab-provided frameworks)
    - Google Workspace/Groups vs. O365/Teams; Instant Messaging (Slack, Discord, SMS, etc); University/Institution provided wikis, document repos, etc.
  - We must to provide the right combination of Training, Policy, *and* appropriate software tools so the Users/Collaborations *want* to “stay in the fold”.

Improving Data Management at JLab (2)

- Sustainable software and archival support (Containers)
  - Develop infrastructure to make it easier to capture, snapshot, archive, and restore software and workflows.
  - Improve infrastructure support and documentation for Containerizing workflows incl. “How-to’s”, template containers for existing workflows, etc.
  - Provide a Container registry (w/ history), and supporting git-backed code repos
    - Much of this supported by new User-facing GitLab instance managed by JLab (full CI/CD support as well)
    - VM snapshots of Farm environments as 2nd layer of defense on running old code / containers
- Get a handle on what we do have and make it findable
  - New initiative: JLab Experimental Research DB

Building Blocks for the Solution

Improving Data Management at JLab (3)

- "JLab Experimental Research DB" is in the early stages of development
  - Directly supported by recent hire in PhysDiv (Anil Panta, Casey Morean)
  - Provide a '1-stop shop' to store and search information associated with Experiments executed at JLab
    - Provide a searchable database that contains a comprehensive set of information about an experiment.
      - Targets, beam characteristics, kinematics, reaction info
      - Proposals, papers, theses/dissertations, technical documents
        - machine readable "results" databases where available
      - References to raw and processed file locations
      - References to analysis software and workflow Containers
      - References to meta-data sources: RunDBs, Config DBs, Logbooks, Wikis, etc.
- We will take it in bite-sized steps that still provide value to the Lab and User community as a whole. Achievable and still useful will be our guiding principle.
  - Start by cross-referencing existing (but scattered) databases / data sources on and off-site
    - JLab Publication DB, proposal DB, experimental logbooks, wikis, web-sites, MSS URIs/paths (later Rucio datasets), JLab filesystem paths to working environments, etc
    - Reference software snapshots/workflows in the Container Registry
- **This will be a long term project that we will refine as we go.**

Building Blocks for the Solution

Improving Data Management at JLab (4)

- Continue to identify gaps in what we are providing to Users and address them
  - Instant Messaging (Slack, Discord, Teams, SMS, etc) has come to provide very valuable “real-time” support both when an experiment is on the floor and during analysis.
    - Too much “logbook-worthy” information is getting lost here.
    - We’re evaluating our options (Teams for ‘all’, Cloud service licensing, Mattermost, etc.)

Building Blocks for the Solution