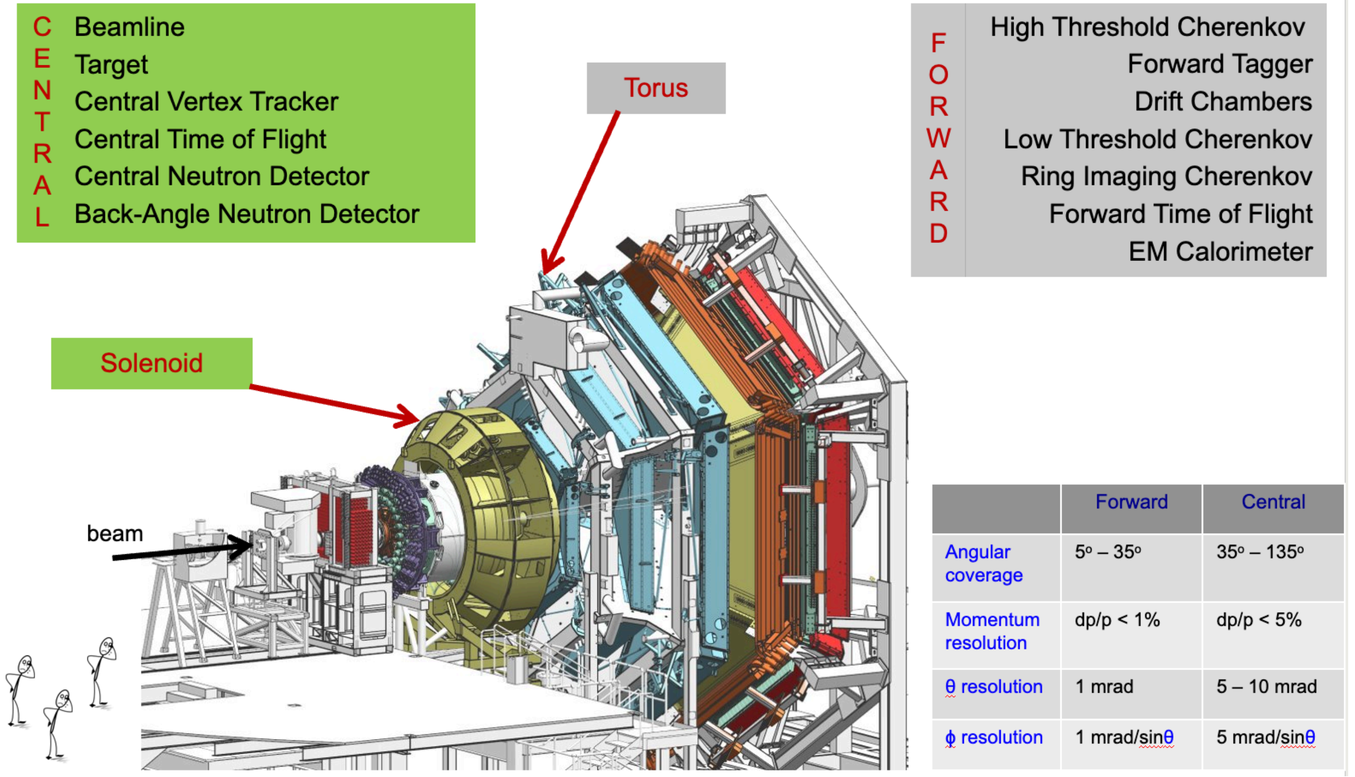
CLAS12 in Hall B at JLab

JLab Software and Computing Workshop 2023

Nathan Baltzell - May 19, 2023

CLAS12 Overview

- Hall B at Jefferson Lab houses a few different experimental setups, but by far the primary one, in terms of beam time, physics diversity, and software and computing requirements, is the CLAS12 experiment
    - **CLAS12** is a double acronym: **C**EBAF **L**arge **A**cceptance **S**pectrometer at **12** GeV, where **CEBAF** is the original acronym for the accelerator facility at JLab
- The science focus of CLAS12 includes the study of 3D nucleon structure, hadron spectroscopy, nuclear effects.
    - Requires polarized beams and targets, and multiple detector systems covering different acceptance and magnetic field regions to measure a variety of final state topologies
    - Different configurations and trigger criteria over the years for different experiments or "run-groups".
- This requires a flexible software system and data representation to support all the different types of physics analyses.
    - One result is that you sometimes have to get your hands a bit more dirty in analyzing data than with a more single-purpose experiment.

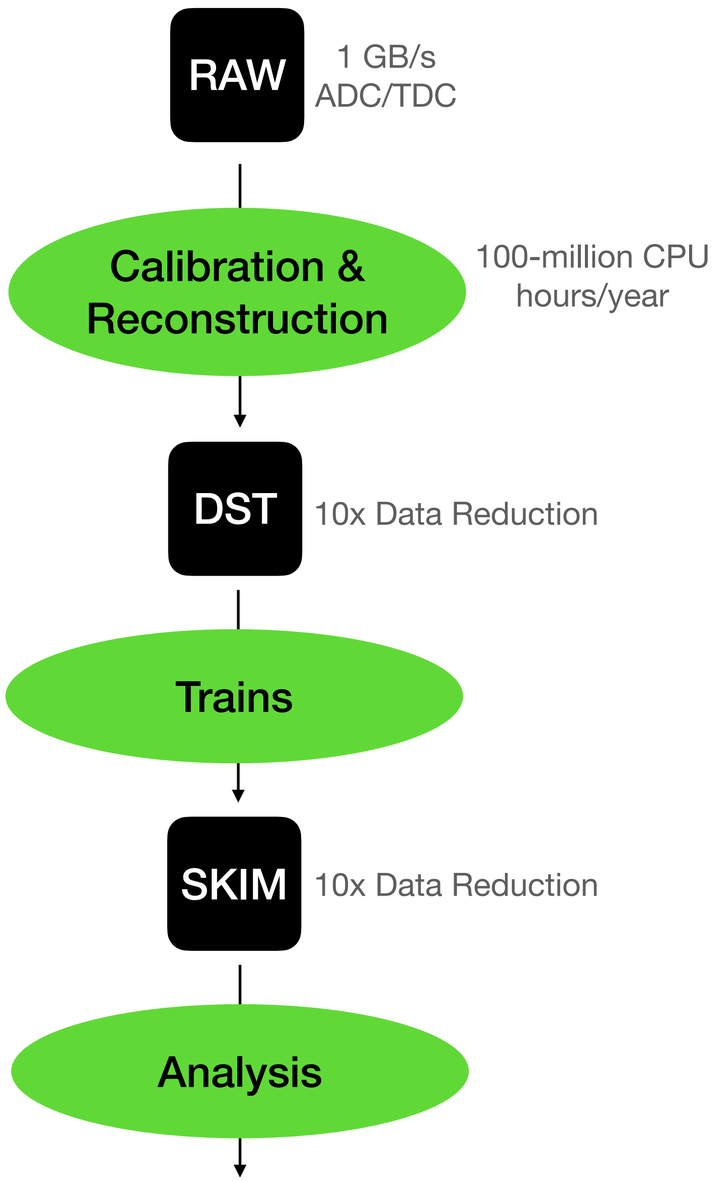


	Forward	Central
Angular coverage	5° – 35°	35° – 135°
Momentum resolution	dp/p < 1%	dp/p < 5%
θ resolution	1 mrad	5 – 10 mrad
ϕ resolution	1 mrad/sinθ	5 mrad/sinθ

Data Workflow, for Perspective

- CLAS12 raw data is on the order of a PB per run-group or ~year
- "DSTs" are on the order of 100 TB for a single run-group, "Data Summary Tapes" is an old term for the high-level quantities for physics analysis, which involves a lot of calibration, conversion, reduction of low-level quantities, ultimately presenting the data as "particles"
- "Trains" are run-group specific, event skims of DSTs for specific reactions or groups of reactions and generally 10-50 times smaller than DSTs
    - Still generally too large to copy and store all data necessary for a full physics analysis on your personal computer
        - although other institutions may have local copies of particular data sets
- We operate pretty "single source", all data originates and is stored at JLab

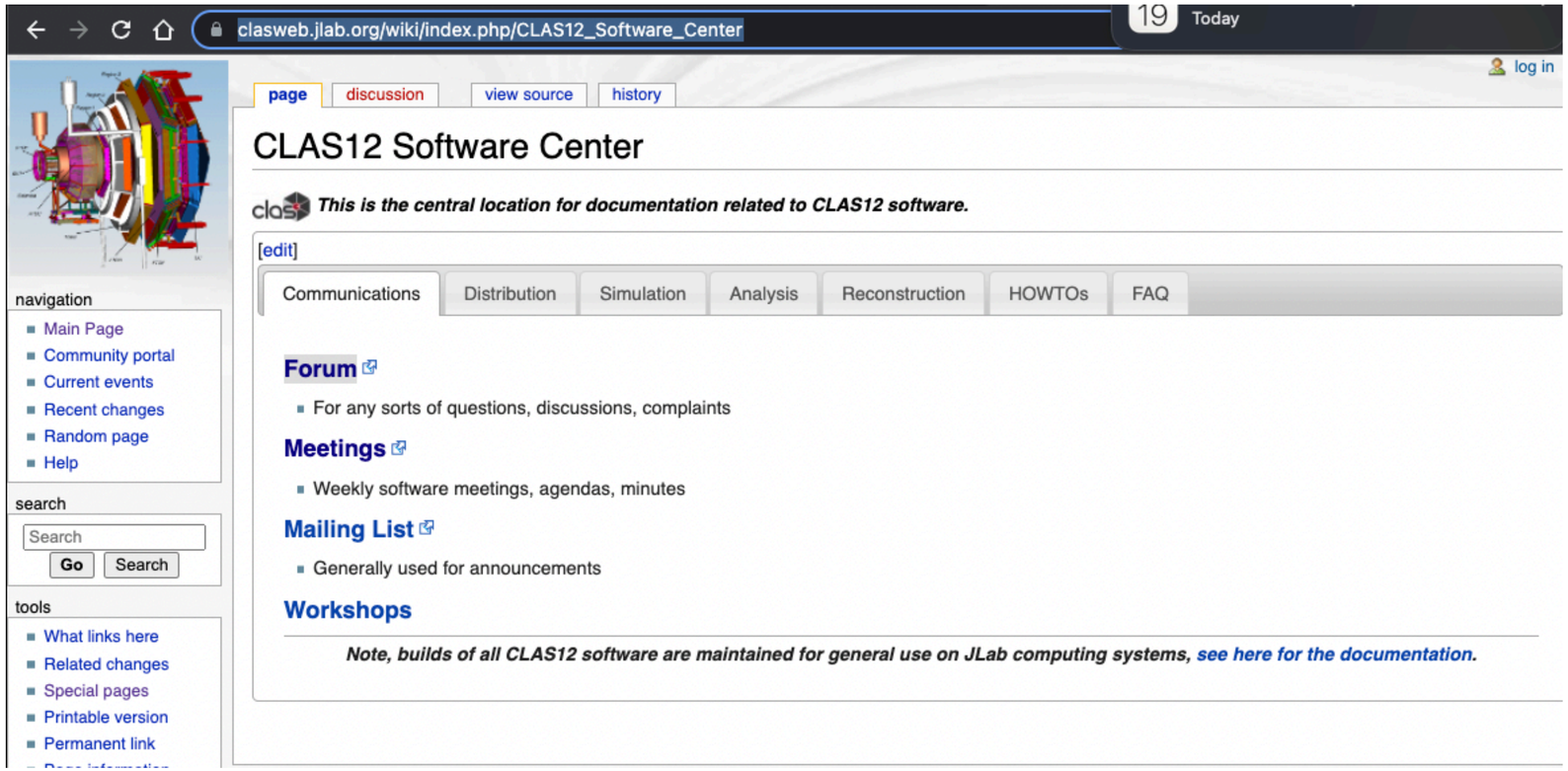
→ using JLab computing resources for final physics analysis is generally required



Software Wiki

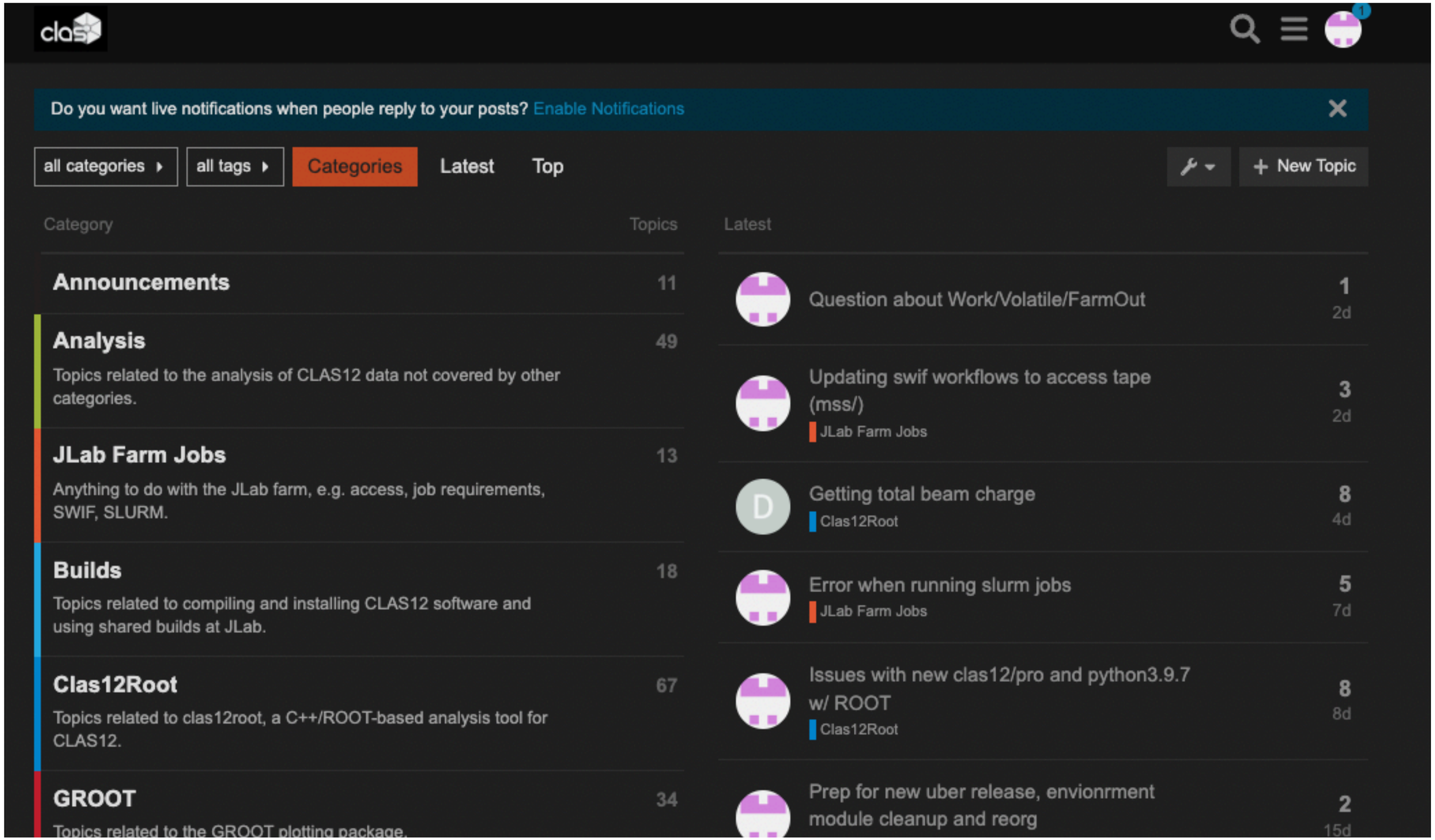
https://clasweb.jlab.org/wiki/index.php/CLAS12_Software_Center

- walk through ...



Discourse Forum

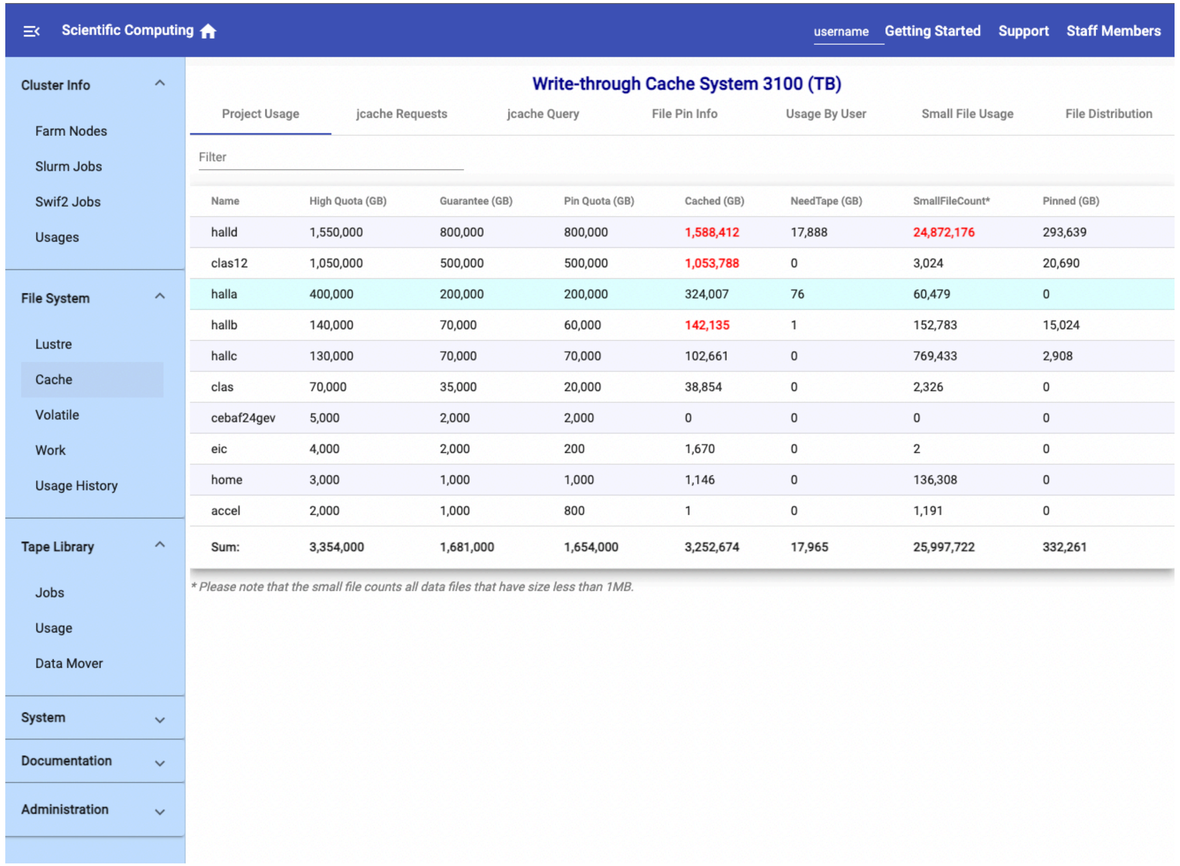
https://clas12.discourse.group/



Computing Access @ JLab

- **Access to computing resources for CLAS12 at JLab requires:**
  - The "clas12-grp" UNIX group for interactive and filesystem access
  - The "clas12" SLURM group for batch farm access
  - Your institution's representative should know who to contact on these

- **Data locations**
  - **/work/clas12** (200 TB, ZFS filesystem w/ backups)
    - free for collaborator's use, manually managed, should be organized within a run group's directory for long-term storage of many TBs
  - **/volatile/clas12** (150/300 TB, Lustre filesystem)
    - free for collaborator's use, temporary storage, with automatic FIFO deletion as more data is written
  - **/cache/clas12** (500/1000 TB, Lustre filesystem)
    - **the final data for publishable physics analysis**
    - a disk-resident cache of data that's really on tape
    - automatic deletion can happen, but only if it's already on tape
  - We try to keep CLAS12 run-groups using a consistent, hierarchical naming scheme within these location, but it's necessary to be in touch with the coordinator of a given run group to know what data should be used



Name	High Quota (GB)	Guarantee (GB)	Pin Quota (GB)	Cached (GB)	NeedTape (GB)	SmallFileCount*	Pinned (GB)
halld	1,550,000	800,000	800,000	1,588,412	17,888	24,872,176	293,639
clas12	1,050,000	500,000	500,000	1,053,788	0	3,024	20,690
halla	400,000	200,000	200,000	324,007	76	60,479	0
hallb	140,000	70,000	60,000	142,135	1	152,783	15,024
hallc	130,000	70,000	70,000	102,661	0	769,433	2,908
clas	70,000	35,000	20,000	38,854	0	2,326	0
cebaf24gev	5,000	2,000	2,000	0	0	0	0
eic	4,000	2,000	200	1,670	0	2	0
home	3,000	1,000	1,000	1,146	0	136,308	0
accel	2,000	1,000	800	1	0	1,191	0
Sum:	3,354,000	1,681,000	1,654,000	3,252,674	17,965	25,997,722	332,261

Software Environment @ JLab

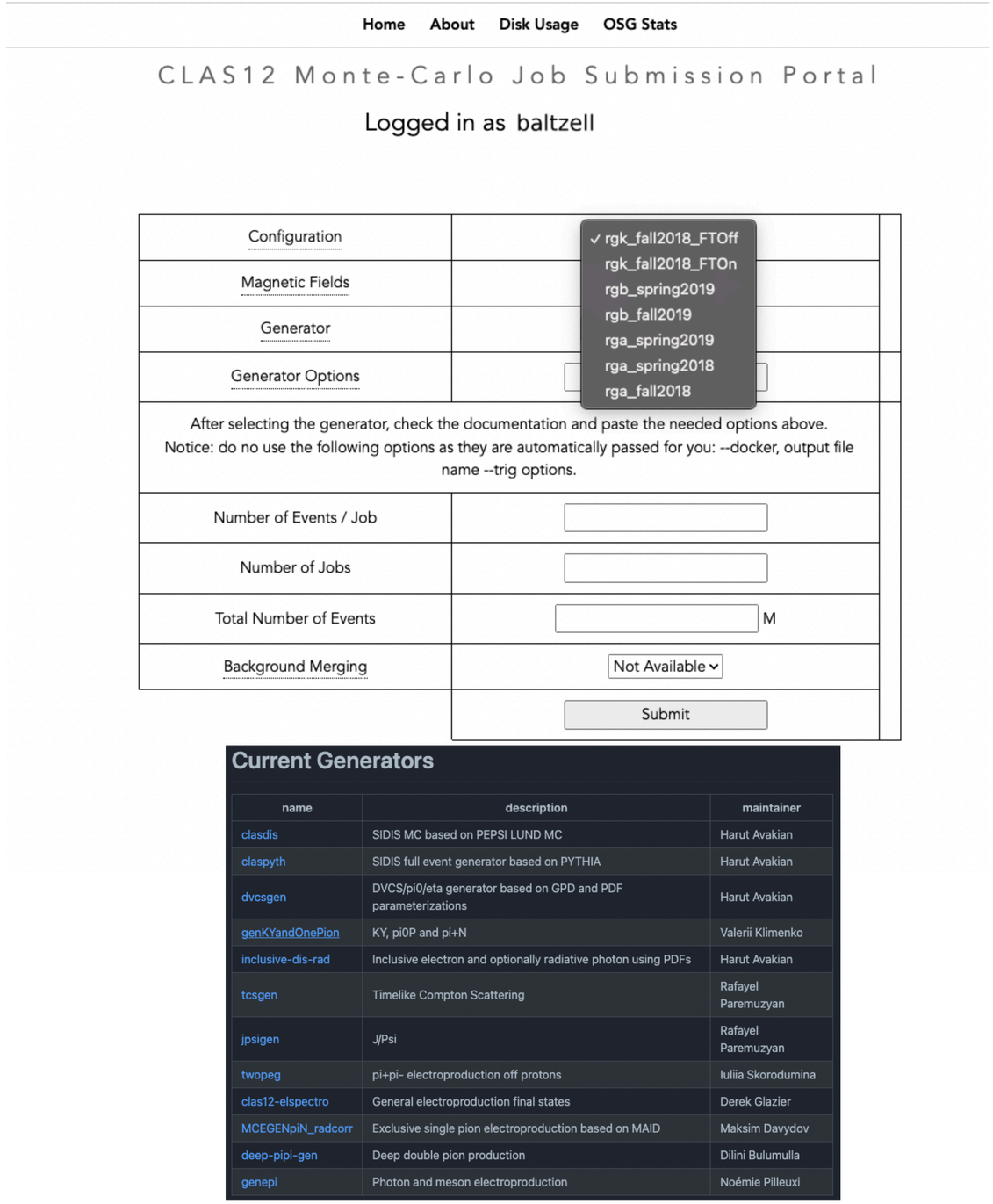
- show it in use ...

```
ifarm1901> module avail
------------------------------------------------- /site/12gev_phys/ceInstall/modulefiles -------------------------------------------------
gemc/2.10  gemc/4.4.2  gemc/5.2  gemc/dev  gemc3/1.0  gemc3/dev  sim/2.5  sim/2.7
gemc/2.11  gemc/5.1    gemc/5.3  gemc/pro  gemc3/1.1  sim/2.4    sim/2.6

------------------------------------------------- /group/clas12/packages/modulefiles -------------------------------------------------
ant/1.10.9    clas12/dev          coatjava/6.5.6.2  evio/5.1          jaw/2.1      maven/3.9.0  rcdb/0.06.00  sqlite/dev
ccdb/1.07.00  clas12/pro          coatjava/6.5.12   graalvm/22.2.0-11 jdk/11.0.2   mcgen/2.14   root/6.20.04  sqlitebrowser/3.12.2
ced/1.4.74    clas12root/1.7.8.c  coatjava/8.2.2    graalvm/22.2.0-17 jdk/14.0.2   mcgen/2.23   root/6.24.06  util/dev
ced/1.5.09    clas12root/1.8.0    coatjava/9.0.0    graalvm/22.3.0-19 jdk/17.0.2   mcgen/2.23b  root/6.26.10  visualvm/2.1.4
clas12/3.4    clas12root/dev      coatjava/9.0.1    gradle/7.5.1      jdk/19.0.1   paw/2005     sqlite/4.3.2  workflow/dev
clas12/4.1    cmake/3.25.0        coatjava/dev      groovy/4.0.3      julia/1.8.5  qadb/1.2.0   sqlite/4.4.0  xrootd/1.0
clas12/4.2    coatjava/6.5.3      coda/3.06         hipo/2.0          maven/3.8.5  qadb/1.2.2   sqlite/4.4.1

------------------------------------------------------------ /apps/modulefiles ------------------------------------------------------------
anaconda2/4.0.5   cmake/3.19.4  cuda/11.4.2  gcc/8.2.0   go/1.15.8     mathematica/11.1    python/3.8.7      singularity/3.8.3  swif2/0.91
anaconda2/4.5.12  cmake/3.19.8  curl/7.59    gcc/8.3.0   go/1.16.12    mathematica/12.0.0  python/3.9.1      singularity/3.8.4
anaconda3/4.5.12  cmake/3.21.1  gcc/4.8.2    gcc/8.4.0   go/1.17.1     maven/3.5.0         python/3.9.5      singularity/3.9.2
anaconda3/4.10.1  cmake/3.22.1  gcc/4.9.2    gcc/9.2.0   go/1.17.5     mpi/openmpi-4.0.1   python3/3.8.7     singularity/3.9.5
ansys/19.2        cmake/3.23.2  gcc/5.1.0    gcc/9.3.0   gpt/3.21      pycharm/2021.1.2    python3/3.9.1     singularity/3.10.0
boost/1.74        cuda/10.0     gcc/5.2.0    gcc/10.2.0  gpt/3.38      python/2.7.18       python3/3.9.5     spack/0.15.4
cmake/3.5.1       cuda/10.1     gcc/5.3.0    gdb/7.11.1  gsl/1.15      python/3.3.1        python3/3.9.7     spack/0.16.0
cmake/3.13.4      cuda/11.2.0   gcc/6.4.0    git/2.31.1  java/1.7      python/3.4.3        scons/4.4.0       spark/2.4.4
cmake/3.18.4      cuda/11.2.1   gcc/7.2.0    go/1.11.4   java/1.8      python/3.6.8        singularity/3.7.1 swif/1.2.2
cmake/3.18.6      cuda/11.2.2   gcc/8.1.0    go/1.15.4   jlabapps/1.0  python/3.7.3        singularity/3.7.4 swif2/0.8
```

Simulations

https://gemc.jlab.org/web_interface/index.php
https://github.com/JeffersonLab/clas12-mcgen

- Simulations are a critical part of the majority of physics analyses in CLAS12, with its complex geometric acceptance and varying detector configurations across run-groups
- CLAS12 leverages the Open Science Grid for 90+% of its simulations
    - moves our simulation workload to the OSG and off of JLab
    - utilizes a web portal to streamline complex job configurations for the user
    - runs a physics event generator (more can be added) plus GEANT4 simulation of the CLAS12 detector, plus full reconstruction of particles through magnetic fields with Kalman filtering
    - It's about a 50/50 split between dedicated resources and truly "open" ones on OSG
        - The CLAS12 collaboration has about 5 institutions, in Europe and USA, that provide dedicated CPU resources that we access via the support of OSG



Analysis

https://clasweb.jlab.org/wiki/index.php/CLAS12_Software_Center#tab=Analysis

- CLAS12 analyses are done in C++ and Java, about 50/50, but also Python and, and Fortran(!)
    - No single, standard framework, but all share the CLAS12 DST format
- Walk through the documentation ...

Questions?