CLAS12 Software & Computing: Miscellaneous!

N. Baltzell

CLAS Collaboration Meeting, June 2024

Possible Topics for Discussion/Demonstration

- authentication, login nodes, ifarm nodes

- documentation,
  forum, https://scicomp.jlab.org, https://clasweb.jlab.org/wiki/index.php/CLAS12_Software_Center

- batch jobs, SLURM/SWIF details

- disk usage, filesystems, quotas

- data transfer, clas12 data discovery

- system utilities

- clas12 software

- github/gitlab@jlab, RUCIO, jupyterhub, XRootD, CVMFS

FYI: el9 @ JLab

- See `/etc/redhat-release` to know for sure what operating system you're on
  - the default on the JLab batch farm is long el9, but `ssh ifarm` still goes to old el7

- System default umask is more restrictive, so permissions for newly created files/directories do not include group (much less world) which means nobody can read them except you and su!
  - To restore the previous default, `umask 022`., e.g., in your shell's login files
  - To restore group read permissions on previously created files, `find . –exec chmod g+r {} \;`
    - and directories also need the executable bit, so `find . -type d -exec chmod g+x {} \;`
    - but may also need to set some executable bits for particular files …

- `/tmp` filesystem is mounted "noexec", and some software can't handle it, e.g., Apptainer, Java SQLite
  - But environment variables can be used to define a /tmp alternative, see `module show tmpfs`, and coatjava prior to 10.0.1 requires that!

- `/apps` and `/site` are gone, replaced by either system installations or CVMFS
  - so your login environment may be broken if relying on them, e.g. sourcing `/site/env/syscshrc` without checking for its existence (or the operating system)

- el9 requires x86-64-v2 instruction set, so won't build/run on 15+ year-old CPUs (and not a good container choice for OSG)

CLAS12 software environment

- Changes in recent months

  - version controlled at https://github.com/jeffersonlab/clas12-env

  - only one installation location at JLab, new and cleaned up with only necessary software versions installed, and on CVMFS

  - environment setup.*sh scripts are deprecated, to be removed by the end of the year

    - during the transition, `/group/clas12/packages/setup.*sh` were maintained to work on both el7 and el9 and automatically do the right thing

    - but now the stuff that necessitated shell script/tricks is either gone with el9 or absorbed into our modulefiles Tcl, so all the shell script does at this point is module use /path/to/modulefiles (and this stops proliferation of multiple advertised ways to setup the environment …)

  - builds for el9, included updating various packages for newer dependency versions (mostly python stuff) and some system changes (e.g., noexec /tmp)

- Demo ...

Data Transfers - demo?

- The traditional: scp/sftp
  - Note, this requires 2-factor for access to scicomp filesystems, but it's super easy with ssh ControlMaster and ProxyJump if you use it regularly

- A better option: rsync
  - It's got looots of options, can tunnel through ssh, generally requires a documentation refresher every time you use it

- Globus (Personal Connect)
  - Like rsync in that it supports resuming partial trasnfers, e.g. if you lose network mid-transfer, or if new files have appeared it'll only get the new ones
  - Supported by faster network at JLab than for scp/sftp/ssh
  - Web browser interface: https://www.globus.org/globus-connect-personal
  - Command line interface
    - pip3 install globus-cli
    - globus login
    - globus whoami
    - export GLOBUS_LOCAL=$(globus endpoint local-id)
    - globus endpoint search jlab
    - export GLOBUS_JLAB=#####
    - globus ls $GLOBUS_JLAB:/expphy/work/clas12
    - globus transfer $GLOBUS_JLAB:/expphy/work/clas12/users/baltzell/alma9.sif $GLOBUS_LOCAL:$HOME/alma9.sif
    - globus task list
    - globus task show 1234567890

Batch jobs – demo?

- Run a simple SLURM job, demonstrating
  - Making a script
    - The shebang
    - The SLURM directives
      - CPU/disk/memory/time requests
      - clean/inherit environment
      - log file location
      - job arrays
      - Set –x –e
  - Jobs
    - Printing environment
    - Accessing log files
    - Module use, load
    - which hipo-utils
  - Need to read data from tape? SWIF

System utilities – demo?

- awk|sed|grep|diff | tail –f | ps | htop
- find
  - -name '*.txt'
  - -mtime +7
  - -empty
  - -type f/d
  - -exec {} \;
- xargs
  - instead of writing a for loop
- grep –e asdf –e qwer
  - sometimes you need an "OR"
- sed –i
  - if you need this, you're probably fixing something that'd ideally be fixed differently, but sometimes you must …
  - note, it's different, -i' ', on MacOS for in-place editing
- tmux/screen

Oh, and CLAS6 – demo?

apptainer run /group/clas/builds/centos79-clas6.sif