Perlmuter

What was, what is, what will be

Sergiu Weisz
sergiu.weisz@upb.ro
18/09/2024

Perlmutter

- Runs HPE Cray OS

- Only whole-node scheduling

- 3072 nodes running 2x AMD EPYC 7763, 64 core, 512 GB memory

- Jobs run in Shifter containers

- Outside connection from nodes

- Full CVMFS support

- SFAPI submission

Issues when last speaking

- VO running under my user
    - Moved to aliprod user on VM

- VO running "by hand"
    - Moved to a systemd user unit

- CE running from JAR, agents running from CVMFS

- Key generated by hand
    - Still a limitation for automation
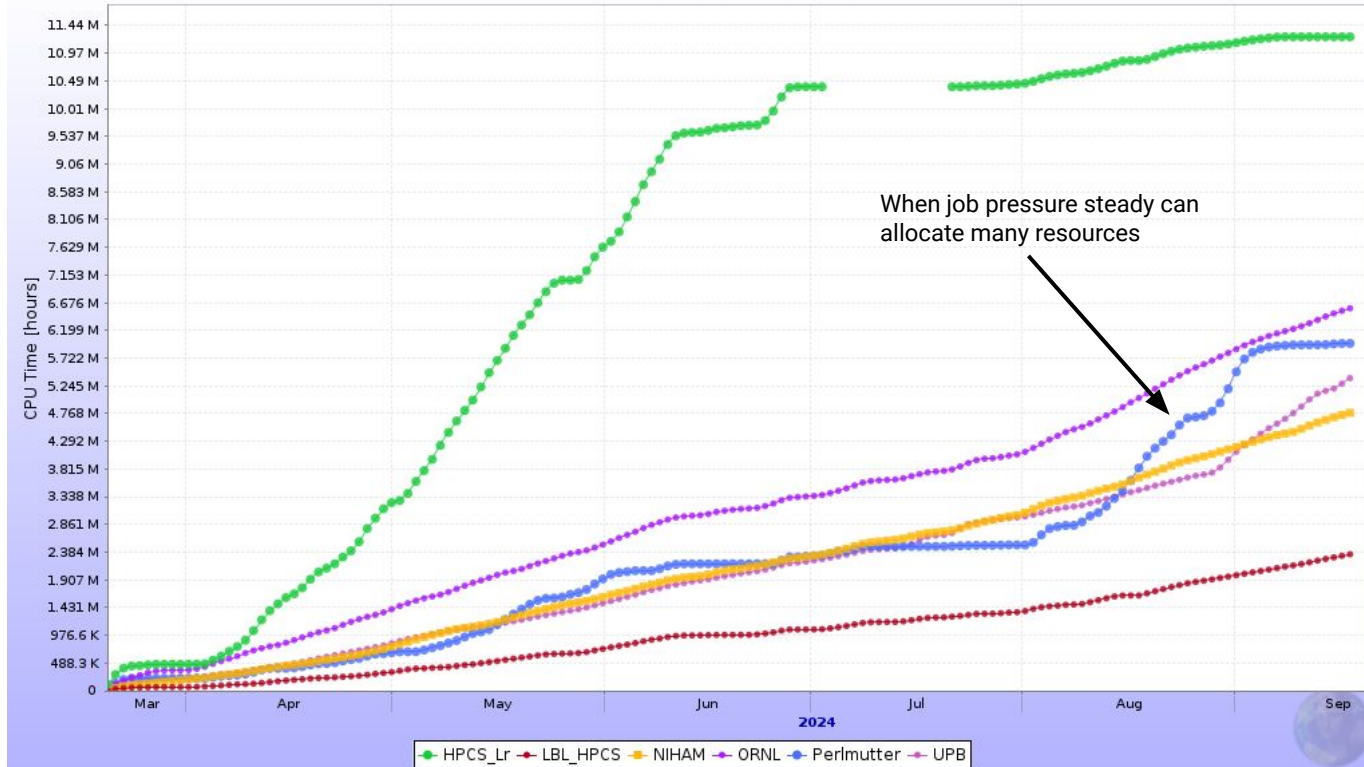   - Irakli can setup the keys now
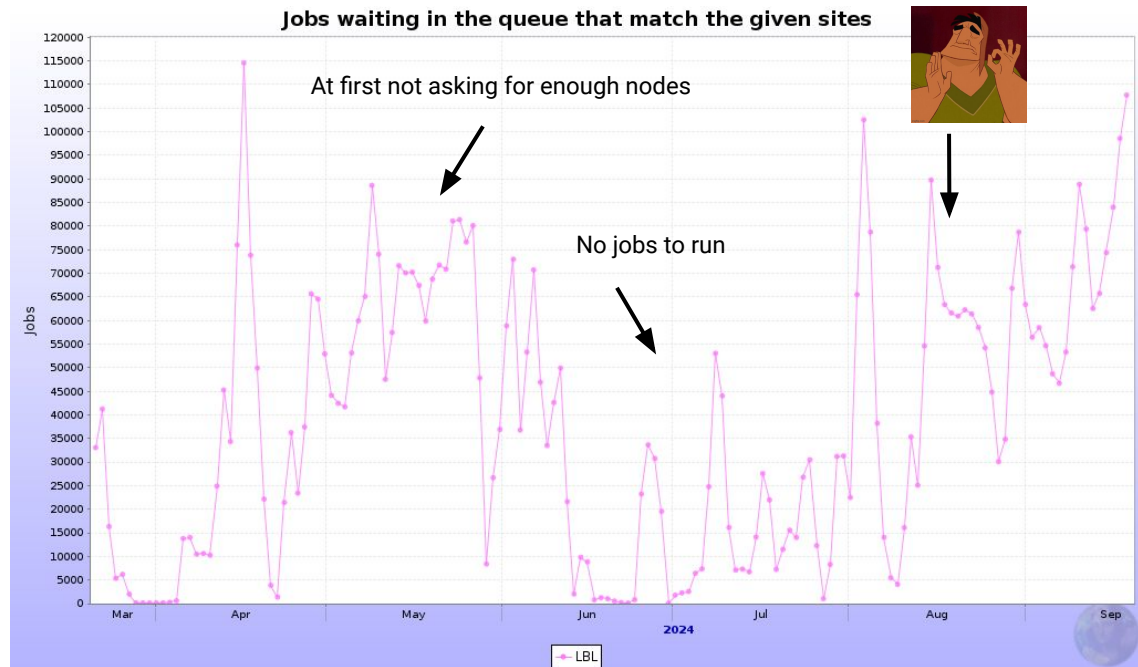
Story so far

Active jobs in Perlmutter

5 queued
20 running jobs

10 queued
50 running jobs

20 queued
100 running jobs
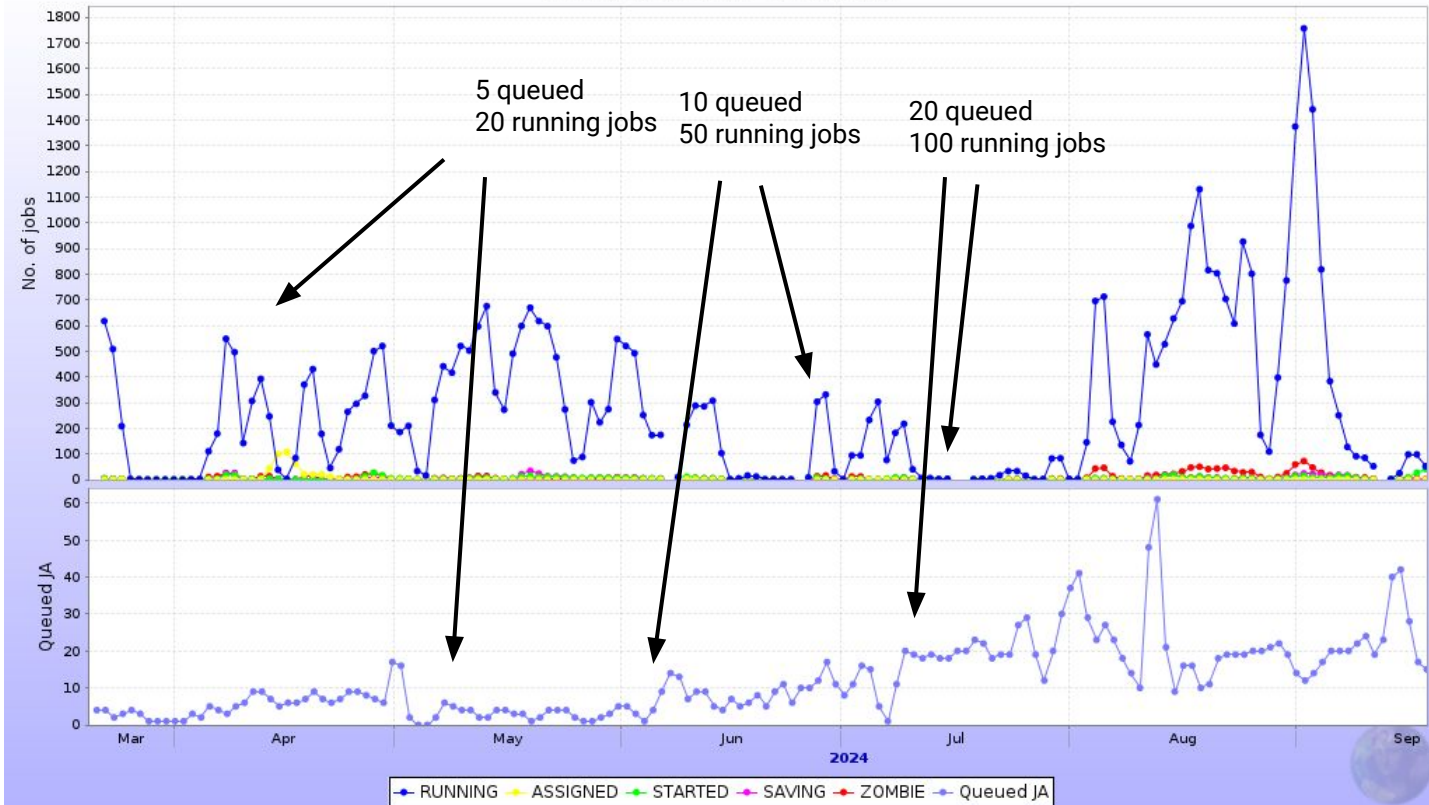
CPU hours offered

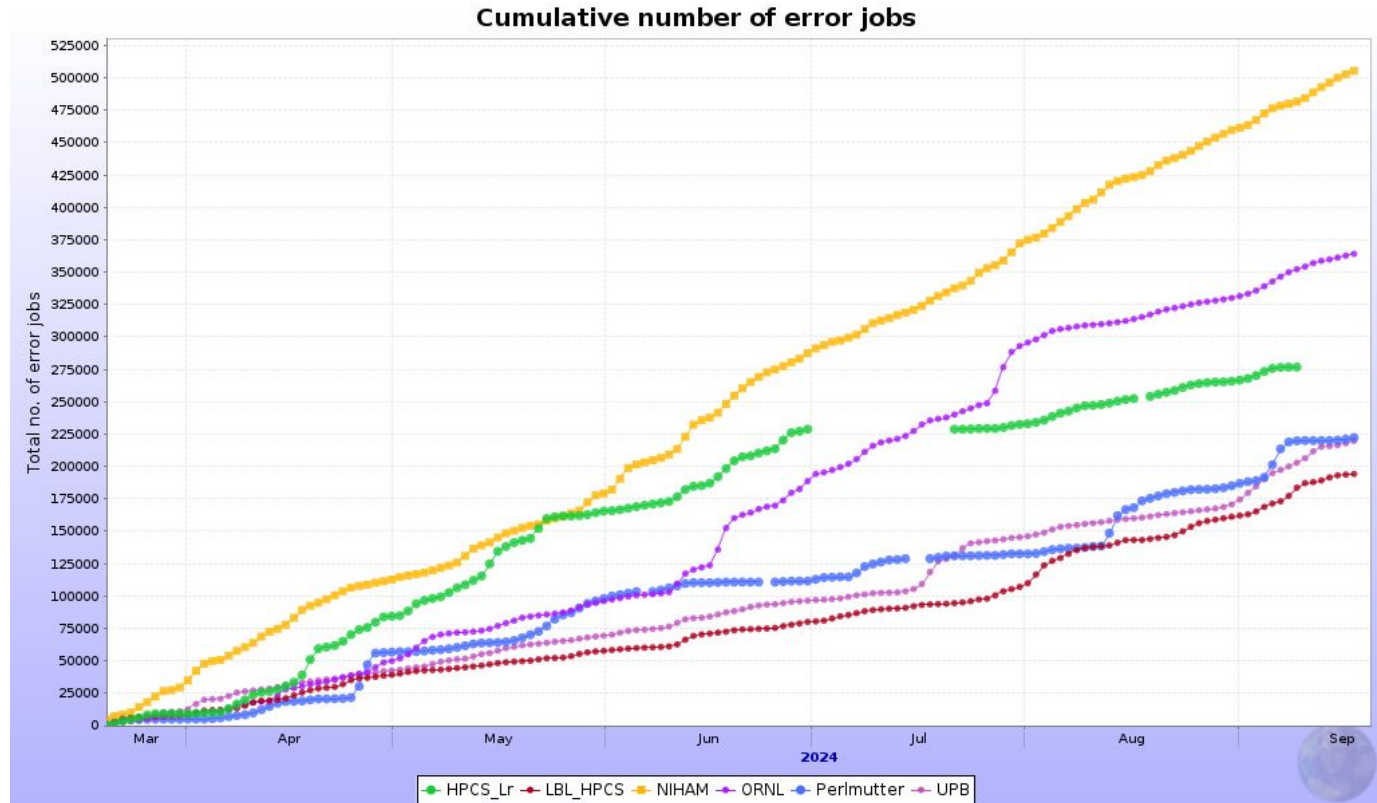Total CPU time for ALICE jobs

When job pressure steady can allocate many resources

Why is job allocation an issue

- Only running MC and DAQ jobs

- Depend on MC job pressure and Perlmutter queue

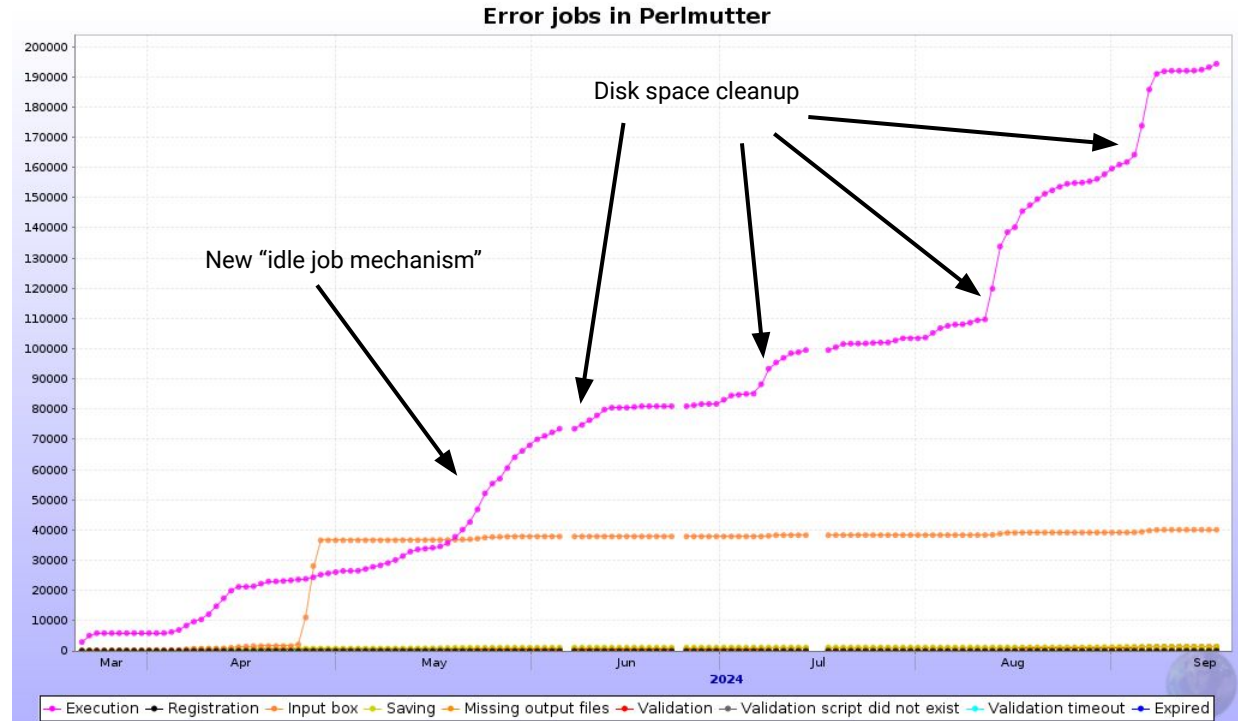**Jobs waiting in the queue that match the given sites**

At first not asking for enough nodes

No jobs to run

Perlmutter errors

Same as similarly
sized sites

Cumulative number of error jobs

Deeper dive into errors

- Sudden fails leave uncleaned workdirs

- Running hyperloop jobs lead to ERROR_E jobs

Error jobs in Perlmutter

Disk space cleanup

New "idle job mechanism"

Legend: Execution · Registration · Input box · Saving · Missing output files · Validation · Validation script did not exist · Validation timeout · Expired

Current issues

- Patches needed to be able to run CE out of CVMFS
    - Hardcoded paths for SFAPI key

- Need to automate job workdir cleanup

- Perlmutter queue contention may be high

- No OSG reporting

- No automated or longterm key management

Perlmutter queue contention

Active jobs in Perlmutter