Best Practices for Supporting Diverse HPC User Storage Needs with Finite Storage Resources

Chris Fuson,   OLCF

Bilel Hadri,   KAUST

Frank Indiviglio,   NOAA

CUG 2018

Wednesday, May 23

05:10 – 06:20

BOF Format

Open floor discussion

Three topics - 20 minutes per topic:

1. Managing storage levels, storage distribution, and ensuring availability

2. Dealing with exceptions to policies

3. Tools and methods used to report and monitor storage usage and availability

Oak Ridge Leadership Computing Facility

- Mission: Provide the computational and data science resources required to solve the world's most impactful scientific & engineering problems.

- New projects and users are added to system throughout year

- Average 250 projects per year and over 1,000 active users per year

- Users cover multiple science domains and experience levels

- Data storage and workflow requirements vary by project

- OLCF users are spread out all over the globe and come from academia, national laboratories, other government agencies, and industry
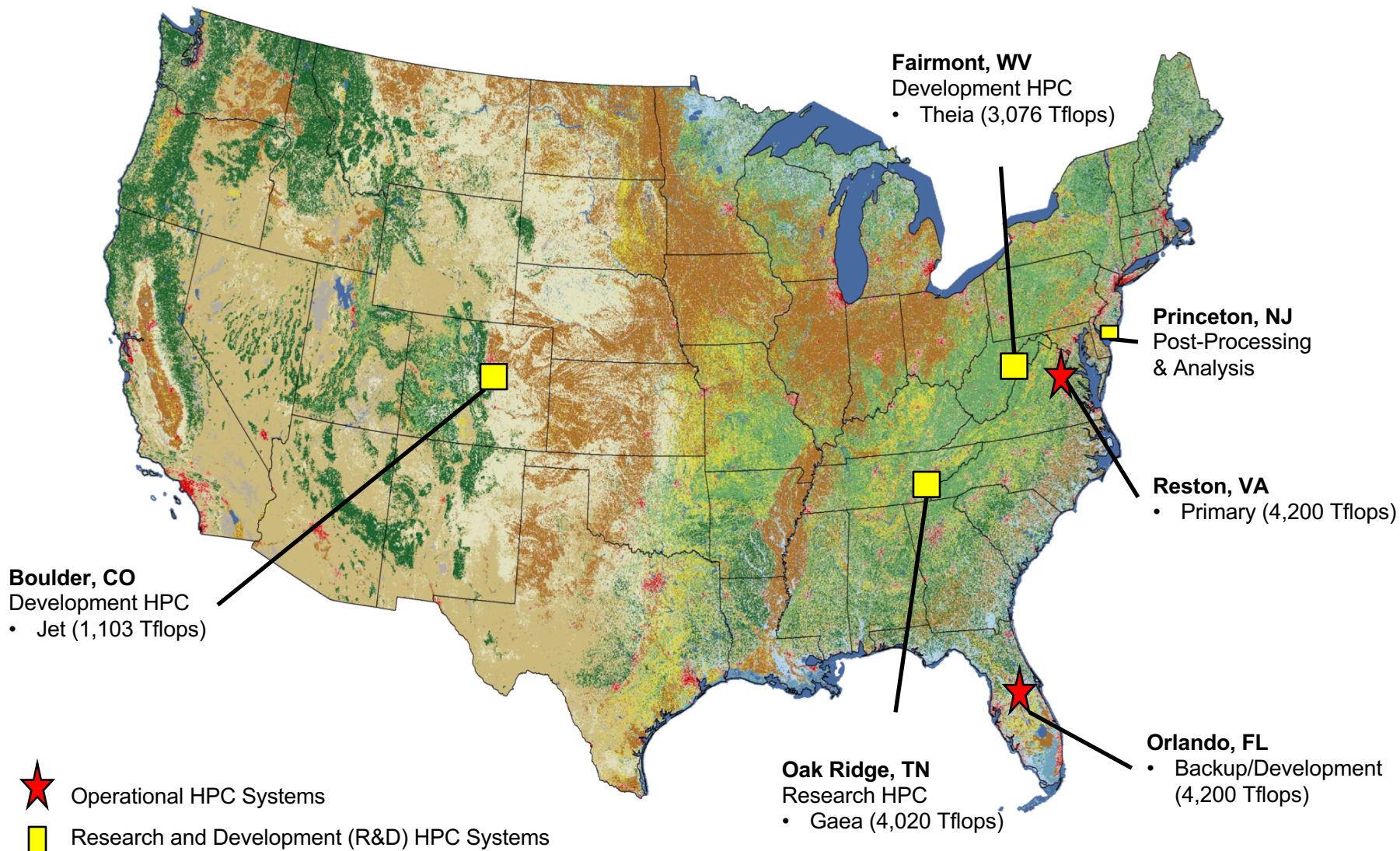
Spider

Center-Wide Lustre File System

32 Petabytes

26,000+ clients

2,016 OSTs

1 TB/s

NOAA's High Performance Computing Locations and Systems

Fairmont, WV
Development HPC
- Theia (3,076 Tflops)

Princeton, NJ
Post-Processing
& Analysis

Reston, VA
- Primary (4,200 Tflops)

Boulder, CO
Development HPC
- Jet (1,103 Tflops)

Orlando, FL
- Backup/Development
  (4,200 Tflops)

Oak Ridge, TN
Research HPC
- Gaea (4,020 Tflops)

★ Operational HPC Systems

▢ Research and Development (R&D) HPC Systems

CUG 2018 – BoF Best Practices for Supporting Diverse HPC User Storage Needs with Finite Storage Resources

Bilel Hadri, **Maciej Olchowik**
KAUST Supercomputing Laboratory

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Shaheen
Ecosystem/Users

Storage	Sonexion 2000 Lustre appliance	17.6 Peta Bytes of usable storage Over 500 GB/s bandwidth
Burst Buffer	DataWarp	Intel Solid Sate Devices (SSD) fast data cache. 1.5 Petabytes of capacity Over 1.5 TB/s bandwidth
Archive	Tiered Adaptive Storage (TAS)	Hierarchical storage with 200 TB disk cache and 20 PB of tape storage, using a spectra logic tape library (Upgradable to 100 PB)

- 7.2 PF Cray XC40 (Haswell) supporting over 700 users, 300 projects.
- PFS reached 75% after 2 years of production.

Field of Science	CPU hours	% overall
CFD/CSM	766,887,001	31.59%
ErSE-GS	479,933,780	19.77%
Material Science	442,328,890	18.22%
ErSE-AS	277,852,890	11.44%
AMCS	255,868,987	10.54%
Bioscience	141,445,255	5.83%
Physics	52,611,730	2.17%
Others	10,870,534	0.45%
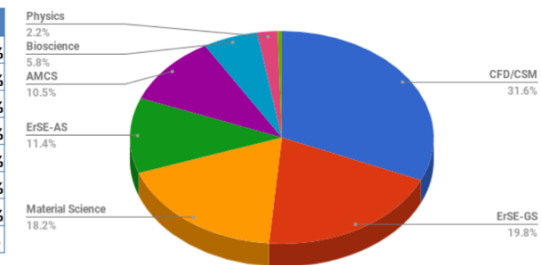
More than 2.4 Billion Core hours in the last 32 months.

# Managing storage levels, storage distribution, and ensuring availability

1. How is your center organized to provide long-term and short-term data storage?
2. Do you purge data on regular intervals or only as needed?
3. Do you notify users before purging, or provide users lists of purged or to-be-purged data?
4. How are quotas enforced; through a hard write stop or another method?
5. Do you notify users when they are close to their quota or have reached the limit?
6. Is your user base generally happy with the layout and policies used at your center?
7. Is there a process for determining initial purge periods and quota limits?
8. Does your center have areas that are not limited by quotas or purge?

Dealing with exemptions to policies

- ***Managing exemptions is managing customer priorities, which makes it a labor intensive activity.***

1. How do users request exemptions to quotas, purge, or other policies?

2. What is the exemption review process used at your center?

3. How do you monitor exemptions?

4. Does your center have procedures that include notifying users before removing exemptions?

Tools and methods used to report and monitor storage usage and availability

1. What types of data does your center monitor to track storage usage and availability?
   - Volume, size and traffic
2. Does your center use commercial tools or center-developed tools to monitor storage?
   - Mc Gyver solutions using different tools to correlate data
     - Cray Sonnexion system monitoring
     - Darshan/xalt tools
     - Scripts on the storage usage
3. How are reports provided to staff for review and monitoring?
   - Weekly usage reports shared to CS (PoC with PIs)
   - Follow up with users and PI
4. Is usage, availability, and storage policies reviewed regularly?
   - Occasionally, following high usage, performance issues.
5. How do users monitor their usage and compare against center policies?
   - When needed following a warning email from the staff or requesting larger quota
   - Dedicated workshop on data management
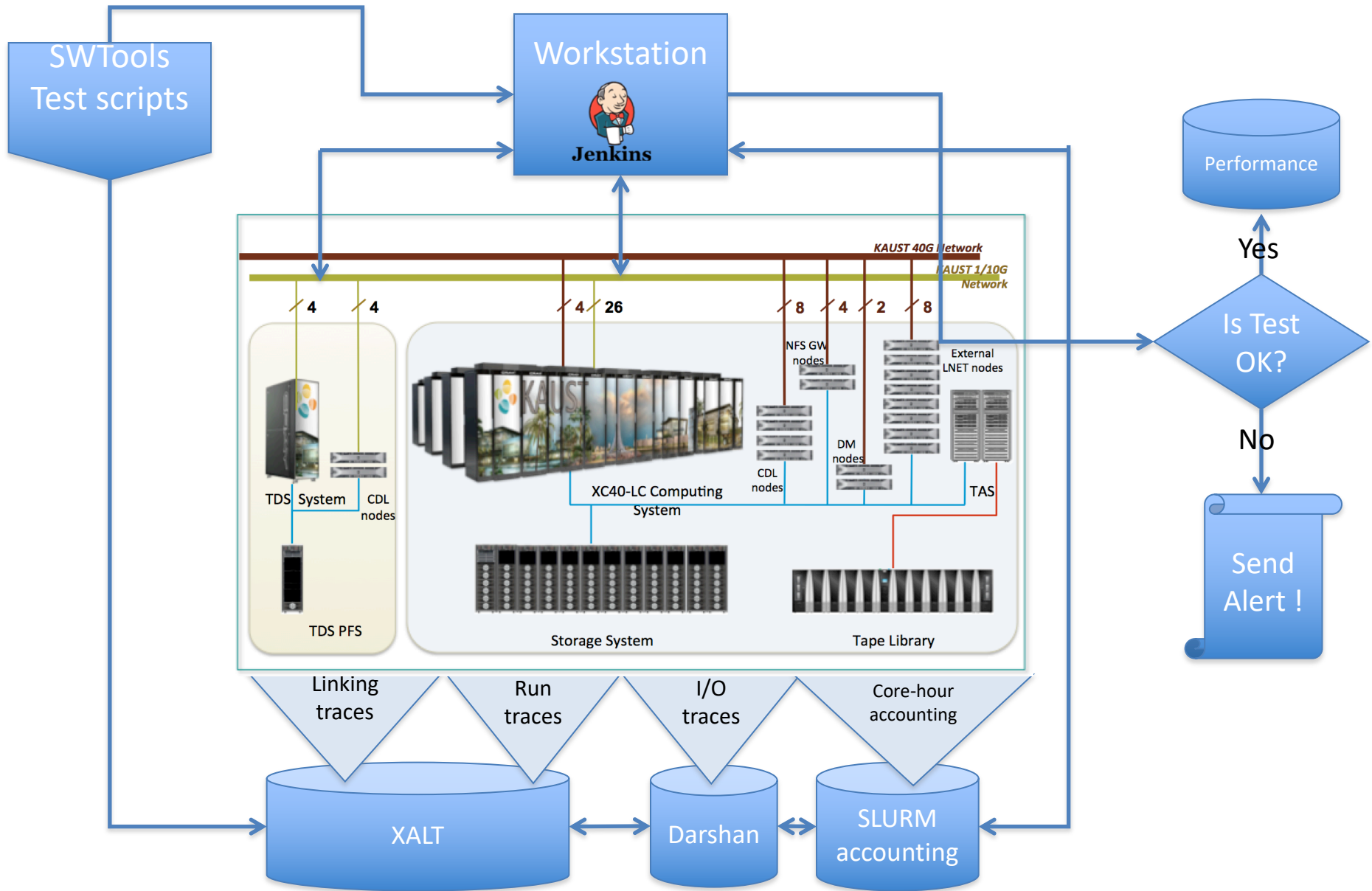
Contact Us

- Chris Fuson,   fusoncb@ornl.gov
- Bilel Hadri,   bilel.hadri@kaust.edu.sa
- Frank Indiviglio,   frank.indiviglio@noaa.gov