Slurm at the George Washington University

Tim Wickberg - wickberg@gwu.edu

Slurm User Group Meeting 2015

September 16, 2015

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Colonial One

What's new?

- Only major change was switch to FairTree
  - Thanks to BYU for that feature
  - Less complaints of anomalous priority behavior
- Looking forward to TRES and expanded QOS options in 15.08
- One email per job array is also a nice incidental benefit
  - One user ran through 100x 1000-array jobs in one night
  - Took their Gmail account two days to recover

What's new?

- slurmdbd migration yesterday was painless.
  - 14 minutes for 1.6 million job records
  - MySQL database on a single pair of SAS disks
  - Accounting db is ~500MB uncompressed as a mysql_dump file
- 15.08 migration Monday during maintenance outage
  - Not doing that step live, although I don't expect any issues

Colonial One - Current System

- Dell C8220 cluster, ~200 nodes
  - 50x GPU nodes, each with dual NVIDIA K20 GPUs
  - 150x CPU nodes, each with dual 2.6GHz 8-core Intel Xeon CPUs, and 64/128/256GB of RAM
  - 2TB, 48-core ‘hugemem’ node
- FDR Infiniband, 2-to-1 oversubscription
- Dell NSS NFS server, ~120TB
- Dell / Terascala HSS Lustre storage system, ~250TB

Priority modeling

- Complicated priority model due to funding model
- `Using priority/multifactor` plugin
- Scheduler will try to balance groups runtime over a longer period, so they match their priority targets
- Top-level: contributing schools, colleges, research centers
  - Second-level: research groups, departments
    - Third: separate research groups in department
      - Fourth: individual users in group

Fairshare model - top tier

```
# sacctmgr show assoc tree format="account,fairshare" | editing
-------------------- ---------
root                         1
 c1                     parent
 cbi                      1880
 ccas                    10080
 other                    1960
 seas                     4320
 sphhs                     720
```

- c1 account - support folks, give them top ranking
- “other” - 10% open to anyone (10points per node)
- Everyone else - 90 points per node

Fairshare model - secondary tier

```
# sacctmgr show assoc tree format="account,fairshare" | editing
-------------------- ---------
root                         1
 ccas                    10080
  astrophysics              17
  ccas-other                31
  economics                  1
  qcd                       16
```

- at the second tier, set 1 point per node
- fairshare is calculated separately at each level, ratio between accounts at each level is all that matters

Fairshare

- Useful commands:
  - sshare
  - sprio
  - sreport

- Force users to specify timelimit:

```
JobSubmitPlugins="job_submit/require_timelimit"
```

Support

- Level 3 support contract with SchedMD
- Filed zero bugs in the past year - code just works
- As demonstrated yesterday, I trust Slurm more than any other component in the cluster.

Feature requests

- Independent fairshare hierarchies for specific partitions
  - Our current model implies GPU and CPU nodes cycles are interchangeable
    - No one has complained about this… yet
  - Likely fix: use Federated Clusters in next release
- sreport support for partitions
  - same rationale as above
  - using ugly/slow bash scripts to aggregate statistics
  - Likely fix, again: split into clusters and Federate

Feature requests

- Ancillary commands for users to figure out what the problem with their job submission is.
  - “swtf” command?
    - Succinct description of current priority level, other issues that may be delaying launch
    - Exposing users to sprio and sreport hasn’t been effective, too much info to parse. Need something simple: “You run too much. Your recent use is 500% of your fairshare value. Buy more hardware for your group.”
  - Some addressed by new descriptions in 15.08

Feature requests

- Better status in sinfo
  - “idle” nodes aren’t necessarily idle, likely waiting for additional nodes to free to start larger jobs
  - Introduce new “earmark” state?

Feature requests

- “Speculative Fairshare Accounting”
  - Account for the full value of active allocations in fairshare immediately
    - Refund unused cycles after completion
  - Current behavior seems to be to wait until after job completion to book usage
  - Causes odd priority swings, exacerbated by long run times
  - (BYU also wants to see this… but Ryan’s slides were already fixed)

Novel (ab)uses of Slurm

- Two non-traditional uses of the Slurm scheduler
  - Fastscratch - dynamic SSD scratch space allocation
  - Backups

- “If all you have is a hammer, everything looks like a nail.”
  - Slurm is a pretty good hammer...

Fastscratch

- Motivation
  - Genomic sequencing apps
  - “Big data”
- Random I/O with mixed read+writes
  - Our NFS and Lustre fileservers hate this
  - SSDs handle this much better than disks, but...
    - We can’t afford to install large SSDs into all nodes
    - And don’t want a “special” set of nodes just for them
    - And this wouldn’t deal with shared access across nodes
- Possible solution - build fileserver with ~3 TB of SSDs
  - And allocate space to jobs on demand

Fastscratch

- Need to manage space and assign to active jobs only
- First approach… some way to use GRES?
  - GRES works with devices in nodes, this is a separate system
  - Jobs shouldn't have access to the fileserver
    - Let alone launch jobs on it

- There's a second mechanism in Slurm that tracks resources: Licenses

Fastscratch

- Need to manage space and assign to active jobs only
- First approach – some way to use GRES?
  - GRES wo… devic…odes, this is a separate system
  - Jobs shouldn t… to the fileserver
    - Let alone…
- There's a second mechanism in Slurm that tracks resources: Licenses

Burst Buffers

Backups

- Disk backup servers
  - Located in a separate datacenter
  - ~100TB usable in 4RU.
- ZFS on FreeBSD, uses zfs snapshots
- Dedicated transfer node located within cluster - “syncbox”
- Goal: run rsync on separate user directories in parallel.
- Bottlenecks are
  - (1) SSH encryption speed (limited by core speed),
  - (2) TCP throughput between locations, then
  - (3) disk I/O.
- Running in parallel lets us get past (1) and (2).

Backups

- Create special partition:

```
PartitionName=syncbox Nodes=syncbox MaxTime=7-0
RootOnly=YES Hidden=YES Shared=YES:8
```

- Run each rsync as separate job:

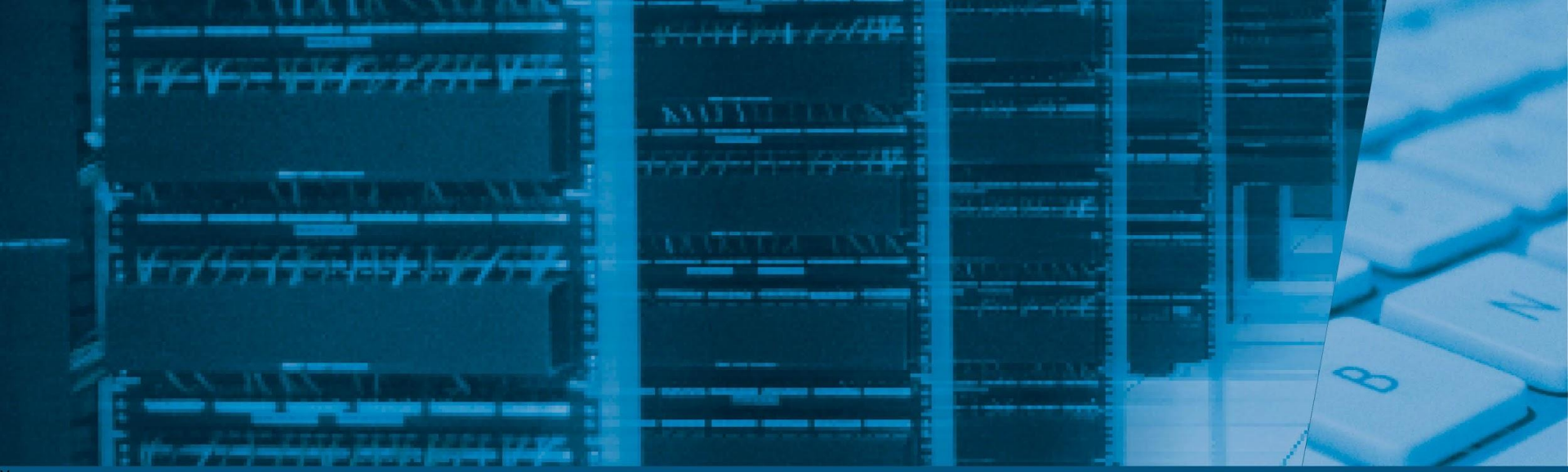
```
#!/bin/bash
#SBATCH -p syncbox --share -t 1-0

# sync
rsync -avd /home/$1 backup1:/storage/home

# snapshot
ssh backup1 zfs snapshot storage/home/$i@$(date +%Y%m%
d)
```

Backups

- Running for over a year, once weekly.
- Works perfectly, full backup pass takes ~6 hours.
- No tape libraries or tape-like software to wrangle.



Thank You

Documentation: http://colonialone.gwu.edu
Twitter: @GWColonialOne
Support Email: hpchelp@gwu.edu

Tim Wickberg
wickberg@gwu.edu