Slurm Recent Developments and Roadmap

Alejandro Sanchez - SchedMD LLC
HPC Knowledge 2017

Slurm Overview

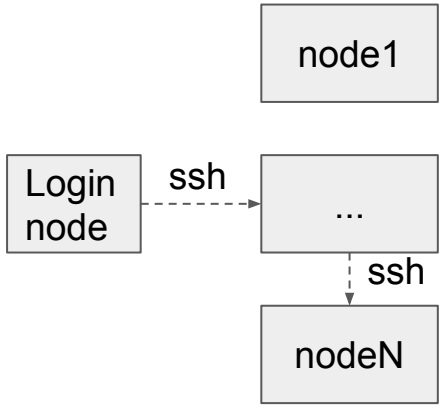
- Workload management system
  - Open source (GPL)
  - Fault tolerant
  - Highly scalable
  - Sophisticated scheduling capabilities (backfill, gang, limits, accounting)
- Used on 6 of top 10 systems from TOP 500 list
  - #1 - Sunway TaihuLight - National Supercomputer Center, China
  - #2 - Tainhe-2 - National Supercomputer Center, China
  - #4 - Sequoia - Lawrence Livermore National Laboratory (LLNL)
  - #5 - Cori - NERSC
  - #8 - Piz Daint - Swiss National Supercomputing Centre (CSCS)
  - #10 - Trinity - Los Alamos National Laboratory (LANL)

Resource Container

- Container for processes spawned outside of Slurm. Some MPI implementations don’t provide Slurm integration (fallback ssh)
- Cgroup container created on compute nodes at job allocation (PrologFlags=contain)
- PAM module puts login shells into that container
  - contribs/pam_slurm_adopt (Author: Ryan Cox, BYU)
- Prevents interference between jobs sharing a node’s resources
- Provides for limits enforcement, accounting, and cleanup.

Resource Container

- Options and use-cases
  - Ignore root (yes | no)
  - No user jobs (deny | ignore)
  - 1 user job. Skip RPC? (yes | no)
  - N>1 user jobs. Which job to adopt?
    - CallerID RPC. Success? Adopt
    - No success (allow | newest | deny)
  - Plugins can use RPC to add pid to extern step



Version 17.02

- Released February 2017
- Many of the changes were for managing federations of clusters
- Relatively few changes visible to users or administrators

Slurmdbd daemon statistics

- sacctmgr show stats - Reports current daemon statistics
- sacctmgr clear stats - Clear daemon statistics
- sacctmgr shutdown - Shutdown the daemon

Slurmdbd daemon statistics

```
$ sacctmgr show stats
Rollup statistics
   Hour      count:8      ave_time:150348 max_time:342905      total_time:1202785
   Day       count:1      ave_time:285012 max_time:285012      total_time:285012
   Month     count:0      ave_time:0          max_time:0           total_time:0

Remote Procedure Call statistics by message type
   DBD_NODE_STATE       ( 1432) count:40  ave_time:979     total_time:39162
   DBD_GET_QOS          ( 1448) count:12  ave_time:949     total_time:11389
   ...

Remote Procedure Call statistics by user
   alex                 (       1001) count:18    ave_time:1342   total_time:24156
   ...
```

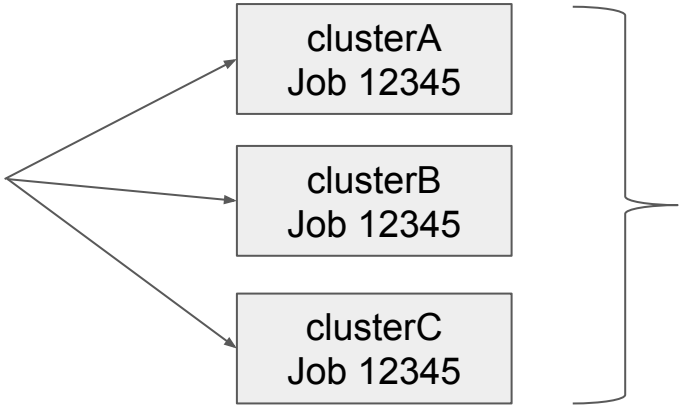
Other 17.02 Changes

- Cgroup containers automatically cleaned up after steps complete
- Added *MailDomain* configuration parameter to qualify email addresses
- Added *PrologFlags=Serial* configuration parameter to prevent Epilog from starting before Prolog completes (even if job cancelled while Prolog is active)
- Added burst buffer support for job arrays
- Memory values changed from 32-bit to 64-bit, increasing maximum supported limit enforcement and schedule for nodes above 2TB
- Removed AIX, BlueGene/L and BlueGene/P support
- Removed sched/wiki and sched/wiki2 plugins (Maui and Moab scheduler support removed)

Federated Clusters Background

- Originally, job IDs were **not unique** across multiple clusters.

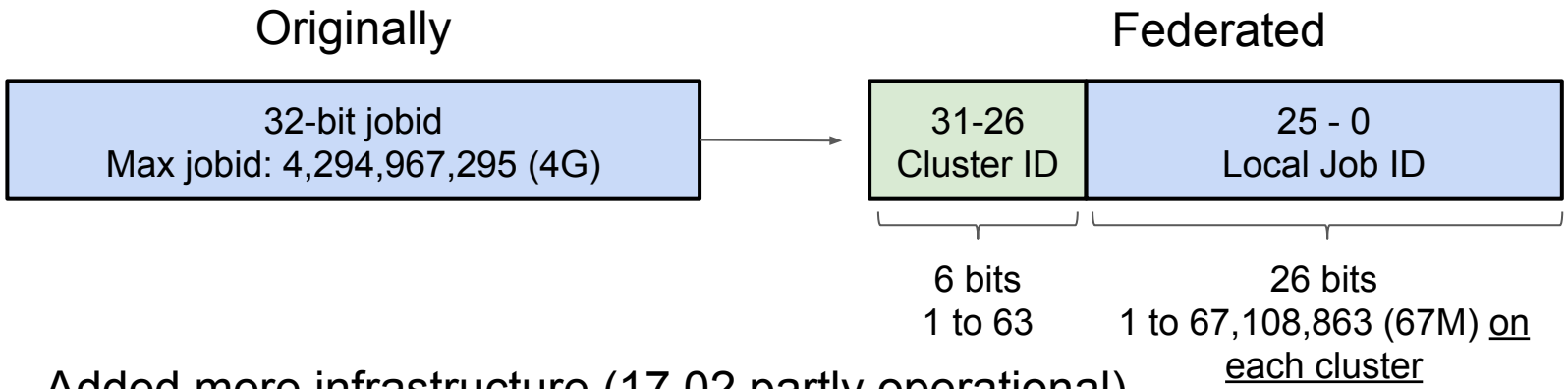
Submission commands have the --cluster (-M) option, which route request to 1 cluster



- Job IDs on each cluster are independent (two jobs can have same ID)
- No Job migration between clusters
- No unified view of system state, each cluster largely independent

Federated Clusters Infrastructure

- Need mechanism to identify the cluster associated with a job ID without using slurmdbd lookup
- Embed cluster ID within the originally 32-bit job ID



- Added more infrastructure (17.02 partly operational)

Version 17.11

- Release November 2017
- More infrastructure for Federated Clusters (fully operational by now)
- Major enhancements in functionality

Persistent Connections

- When a slurmctld is added to a federation of clusters, a persistent connection is created with other slurmctld daemons in the federation + slurmdbd
  - Reduces communication overhead -- only authenticate once
  - Broken connections detected immediately and re-established when needed
  - Controller and SlurmDBD use the same code
- Slurmdbd pushes updates to all clusters in the federation
- A cluster can only be part of one federation at a time

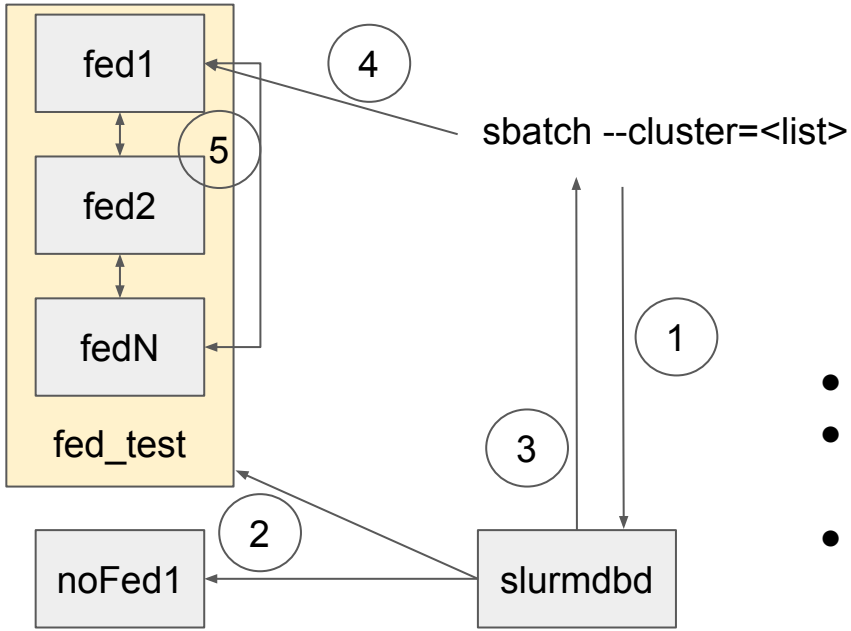
Configuration

- Federations are managed by slurmdbd through sacctmgr
- sacctmgr add federation <fedname> [flags=<list>][clusters=<list>]
  - sacctmgr mod federation <fedname> flags[+|-]=<list>
  - sacctmgr mod federation <fedname> clusters[+|-]=<list>
- sacctmgr mod cluster <cluster> ...
  - set federation=<federation>
  - set features=<features>
    - Features at a cluster level that can be requested by jobs
- FederationParameters=fed_display to show federated view by default
  - --federation, --local, --cluster (-M) used by status commands squeue, sinfo, sprio, etc.

Configuration

- sacctmgr mod cluster <cluster> set fedstate=<state>
  - **ACTIVE** - Cluster will actively accept and schedule federated jobs.
  - **INACTIVE** - Cluster will not schedule or accept any jobs.
  - **DRAIN** - Cluster will not accept any new jobs and will let existing federated jobs complete.
  - **DRAIN+REMOVE** - Cluster will not accept any new jobs and will remove itself from the federation once all federated jobs have completed. When removed from the federation, the cluster will accept jobs as a non-federated cluster.

Job Submission



1. Which cluster/fed has fastest start_time?
2. slurmdbd checks with --test-only
3. slurmdbd returns i.e. fed_test can start soonest
4. sbatch submits job to local cluster fed1
5. Job sibling copies to other clusters in federation

- FedOrigin=fed1 (local cluster)
- FedViableSiblings=fed1,fed2,fed3 (satisfy cluster features)
- FedActiveSiblings=fed2 (cluster whose submission didn't fail)

Scheduling

- Federated jobs contain the locations of all “sibling” jobs
- Each cluster independently schedules each sibling job
- Coordinates with “origin” cluster to start job
  - The origin cluster is determined from the job id
  - Prevents multiple jobs from being started at the same time
  - Policies in place to handle if origin cluster fails
- Once sibling job is started, origin cluster revokes remaining sibling jobs
- Batch jobs can be requeued to federation
- More code added to handle scancel, regression tests, ...

Job Submission

- Interactive Jobs (salloc/srun) example
  1. srun submit the job to a local slurmctld daemon in a federated cluster
  2. Local slurmctld will submit the sibling jobs to slurmctld dameons on other federated clusters
  3. A sibling cluster coordinates with the origin cluster and allocates nodes for the job
  4. The origin cluster removes any pending sibling jobs
  5. The sibling cluster directly notifies to srun passing sibling and allocation information
  6. srun will talk directly to nodes on selected cluster

NOTE: All compute nodes need to be accessible by each submission host!

NOTE: Interactive jobs can not be requeued

Heterogenous Job Allocations

- Almost all job options can be used to specify a resource allocation that heterogeneous resources
  - Different partitions, QOS, node features, memory size, etc.
- Each component is internally managed as a separate job record
- All components are scheduled at the same time
- One or more applications can be launched on the allocation

```
srun --features=haswell --ntasks=1 master : --features=knl,a2a,flat --ntasks=72 slave
```

Other 17.11 Changes

- More flexible advanced reservations (FLEX flag on reservations)
  - Jobs able to use resources inside and outside of the reservation
  - Jobs able to start before and end after the reservation
- Sprio command reports information for every partition associated with a job rather than just one partition
- Support for stacking different interconnect plugins (JobAcctGather)
- Add *scancel --hurry* option
  - Cancel job without staging-out burst buffer files
- See the NEWS and/or RELEASE_NOTES for more

Version 18.08

- Release August 2018
- Eliminate support for Cray/ALPS
  - Must use native Slurm mode

Questions?

Additional Slurm information:

https://slurm.schedmd.com