Field Notes From the Frontlines of Slurm Support

Tim Wickberg
SchedMD

Slurm User Group Meeting 2017

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Field Notes - Overview

- Random notes, observations, and configuration preferences.
- My opinions, not those of my employer.
  - Although if you file a support case, you'll see a lot of the same opinions repeated back. :)
- … and I can be somewhat opinionated.
  - They're not strict requirements, you may have other valid approaches.
  - Slurm has a lot of flexibility in setup and configuration to suit disparate environments, and not all or any of this may apply to you. YMMV.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Field Notes - Overview

- Feel free to ask questions throughout.
  - Although I may defer more involved discussions until afterwards in the interest of time.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Log Rotation

- Common request and a best practice
- Frequently mishandled though
- If using logrotate, make sure that the correct daemons are being signalled at the appropriate time.
- slurmdbd needs to be SIGHUP'd when slurmdbd.log is rotated, several sites have had this misconfigured and lost logs.
- All the daemons will restart on SIGHUP.
  - Make sure the config files are in a usable state, or things may crash.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Log Rotation

- The 17.11 release adds SIGUSR2 to {slurmctld, slurmdbd, slurmd} to drop + reopen the log files only. Avoids reconfiguration. Strongly preferred.
- One other mistake: at least one cluster manager has shipped logrotate scripts that call *scontrol reconfigure* when rotating local log files.
  - Causes O(N) reconfigurations within the cluster in quick succession, leading to some temporary performance issues.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Fixed limit on an account's usage

- QOS setting of Flags=NoDecay
  - Allows you to establish a fixed-size allocation of resource for a group.
  - A bank account, if you will.
  - GrpTRESMins, GrpWall, UsageRaw will not decay even when using PriorityDecayHalfLife
  - Jobs running against the QOS will eventually hit a limit.
  - Reset the RawUsage value to zero to reset, or change the limits.
  - Thanks to Josh Samuelson (U. Nebraska Lincoln) for the patch.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Allocating Memory

- If you use CR_{CORE,CPU,SOCKET}_MEMORY, make sure to set DefMemPerCPU or DefMemPerNode.
  - Choose one or the other. DefMemPerCPU is usually preferable.
  - Set DefMemPerCPU roughly equal to Memory / CPUs.
    - But preferably round it down a bit. Usually to the closest GB.
    - Helps fill the node up better
  - With no setting, the first job allocated that did not specific --mem will receive all memory on the node.
    - Leaving 0 MB free for other jobs.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Node configuration

- Must have all settings less than or equal to the available hardware in the node.
- If a node comes up with more sockets/threads/cores/memory that's okay, although Slurm will still schedule based on the configured values.
- I encourage sites to round their Memory values down to the nearest GB though.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Node configuration - continued

- Rounding down the memory value does two things:
  - Avoids the node being marked down if the BIOS reports slightly less memory.
    - This can happen due to a firmware change, DIMM replacement, or even kernel upgrade. If the node has 4MB less than configured, it'll stay offline.
    - Gives the OS some breathing room when you're using CR_*_MEMORY.
      - If Slurm can't allocate that space, it'll never be divvied up and assigned. Helps avoid the dreaded OOM.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Node configuration - continued

- Alternate approach:
  - The MemSpec limit setting on a Node can set aside an amount of space for the slurmd process and OS.
    - The slurmd processes cgroup will be limited to that amount of space, so make sure it’s not too small though.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Node configuration

- You cannot add or remove nodes on the fly.
- Internal data structures related to scheduling and communication are designed around fixed-size bit field for performance reasons, and cannot be changed without a restart.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Partition Priority

- Show of hands - how many people use the Priority setting on a partition in conjunction with the PriorityWeightPartition setting to adjust user job priorities?

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Partition Priority

- Yes - this is a trick question. It's not doing what you may think it is. The change to user priorities is inconsequential.
- When scheduling jobs, higher Priority partitions are considered as a whole separate tier to service.
- If any job in a higher Priority partition is pending, no jobs will be scheduled (or backfilled) on node that is shared with a higher tiered partition.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Partition Priority

- So the only effect of the PriorityWeightPartition has been to change the numbers around.
- But priority numbers only matter *when compared to other jobs at the same Tier.*

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Partition Priority

- The 16.05 release split the Priority setting on a Partition into two parts:
  - PriorityTier - only sets this preemption tier, no effect on job priority values.
  - PriorityJobFactor - normalized, then used with PriorityWeightPartition to change job priority values. Does not set the Tier.
- For backwards compatibility, if Priority is set, it's applied to both PriorityTier and PriorityJobFactor.
  - Although the PriorityJobFactor setting doesn't accomplish much.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Backfill tuning problems

- Most common backfill problem: bf_window set too short.
- Should be equal to the largest time limit on any partition, within reason. Max of ~ 2 weeks is my preference.
  - Larger values lead to higher memory consumption, and slower backfill performance.
- Too small of a value will starve large jobs indefinitely.
- If the highest priority job won't start within the current backfill window, smaller jobs will backfill in ahead of it.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Backfill tuning problems

- Then the main scheduler will run, decide that large job can't start when originally planned, and reschedule it into the future again.
- This cycle can then repeat indefinitely.
- We're looking into a new tuning flag to prevent this behavior, but do not have a solution just yet.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

On performance, time, and such matters

- The time() and gettimeofday() syscalls need to be fast for good slurmctld performance.
- slurmctld calls then **very** frequently.
- Certain VM platforms do not handle this well, as they disable the Linux vDSO that is usually in use.
  - The vDSO avoids a context switch to the kernel and back and makes time() a direct memory reference. slurmctld effectively assumes this is in use.
  - Xen especially guilty of this.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

On performance, time, and such matters

- Hardware:
  - Fewer faster cores on the slurmctld host is preferred.
  - Fast path to StateSaveLocation.
    - IOPS this filesystem can sustain is a major bottleneck to job throughput.
      - At least 1 directory and two files created per job.
      - And the corresponding unlink() calls will add to the load.
    - Although using job arrays instead of individual job records will help significantly, as only one job script and environment file is saved per entire array.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

On performance, time, and such matters

- slurmctld is highly threaded.
  - 10 - 100’s of threads in use at any time.
- Unfortunately, slurmctld is not highly concurrent.
  - Internal datastructure locking will limit the concurrent thread count in a lot of circumstances to… one thread.
  - We’re working on this, and have made some minor improvements lately.
- Hardware:
  - Fewer, faster cores on the slurmctld host is vastly preferred.
  - Dual 16-core Xeons are overkill, the higher clock rate (and cheaper!) 6-cores will perform

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Ways to keep your configs tidy

- For both Nodes and Partitions there is a reserved keyword of DEFAULT.
- Used to set default options that will apply to all successive lines.
  - Also why you can't have a partition named "default".
- Next DEFAULT line will override those previously set values.
  - But not reset all of them. E.g., if the first sets Features=foo,bar, and the second sets Weight=100, all values after the second line with get both settings by default.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Ways to keep your configs tidy

- Before

```
NodeName=node01 CoresPerSocket=4 Sockets=2 ThreadsPerCore=2 RealMemory=252928 Gres=gpu:k80:4 Weight=2
NodeName=node03 CoresPerSocket=12 Sockets=2 ThreadsPerCore=2 RealMemory=252928 Weight=2
NodeName=node04 CoresPerSocket=12 Sockets=2 ThreadsPerCore=2 RealMemory=252928 Weight=2
NodeName=node05 CoresPerSocket=12 Sockets=2 ThreadsPerCore=2 RealMemory=252928 Weight=2
NodeName=node06 CoresPerSocket=12 Sockets=2 ThreadsPerCore=2 RealMemory=252928 Weight=2
NodeName=node11 CoresPerSocket=8 Sockets=2 ThreadsPerCore=2 RealMemory=123904 Weight=4
NodeName=node12 CoresPerSocket=8 Sockets=2 ThreadsPerCore=2 RealMemory=123904 Weight=4
NodeName=node13 CoresPerSocket=8 Sockets=2 ThreadsPerCore=2 RealMemory=123904 Weight=4
NodeName=node14 CoresPerSocket=8 Sockets=2 ThreadsPerCore=2 RealMemory=123904 Weight=4
NodeName=node15 CoresPerSocket=11 Sockets=2 ThreadsPerCore=2 RealMemory=245760 Weight=2
NodeName=node16 CoresPerSocket=11 Sockets=2 ThreadsPerCore=2 RealMemory=245760 Weight=2
NodeName=node17 CoresPerSocket=16 Sockets=2 ThreadsPerCore=2 RealMemory=381952 Weight=1
NodeName=node18 CoresPerSocket=16 Sockets=2 ThreadsPerCore=2 RealMemory=381952 Weight=1
NodeName=node19 CoresPerSocket=10 Sockets=2 ThreadsPerCore=2 RealMemory=187392 Weight=3
NodeName=node20 CoresPerSocket=10 Sockets=2 ThreadsPerCore=2 RealMemory=187392 Weight=3
```

Ways to keep your configs tidy

- After

```
NodeName=DEFAULT Sockets=2 ThreadsPerCore=2
NodeName=node01 CoresPerSocket=4 RealMemory=252928 Gres=gpu:k80:4 Weight=2
NodeName=node[03-06] CoresPerSocket=12 RealMemory=252928 Weight=2
NodeName=node[11-14] CoresPerSocket=8 RealMemory=123904 Weight=4
NodeName=node[15-16] CoresPerSocket=11 RealMemory=245760 Weight=2
NodeName=node[17-18] CoresPerSocket=16 RealMemory=381952 Weight=1
NodeName=node[19-20] CoresPerSocket=10 RealMemory=187392 Weight=3
```

Job Submit Plugins

- Allow you to add arbitrary business logic to filter/deny/modify user jobs at submission time.
- Run within the slurmctld process, and have access to all internal data structures.
  - One caveat: they can also potentially corrupt anything and everything if misdesigned… be careful.
- Optional job_submit.lua plugin uses an interpreted and “safe” script provided by the site, rather than needing to code in C.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Job Submit Plugins

- Lua (or a new C plugin) can do whatever you want.
- Potentially call out to external APIs.
- Enforce whatever arbitrary logic you like.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Job Submit Plugins

- Demo

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Job Submit Plugins

- Caveats
  - Remember that I mentioned earlier that slurmctld is highly threaded, but not highly concurrent?
  - For safety reasons, the job_submit plugin operates with write locks held on the two main internal data structures.
  - So no other threads are able to run concurrently.
  - Make the scripts fast, or pay the price in system throughput and responsiveness.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Job Submit Plugins

- Demo

Copyright 2017 SchedMD LLC
https://www.schedmd.com

SlurmDBD is hung

- Most common source of problems with slurmdbd is from overly-aggressive sacct requests.
  - Caused by mistake, or bad scripting.
- Depending on your system, ‘sacct -S 2010-01-01’ may need to return millions of rows.
  - May push the slurmdbd host out of memory, or just take an unreasonably long time to complete.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

SlurmDBD is hung

- New slurmdbd.conf option of MaxQueryTimeRange allows admins to restrict the maximum time range users can query in a single command.

```
bbaggins@zoidberg:/home/tim$ sacct -S 2007-01-01
       JobID    JobName  Partition    Account  AllocCPUS      State ExitCode
------------ ---------- ---------- ---------- ---------- ---------- --------
sacct: error: slurmdbd: Too wide of a date range in query
```

Copyright 2017 SchedMD LLC
https://www.schedmd.com

High-Availability

- My preferences:
  - Setup the primary and backup slurmctld hosts.
    - Should be directly attached to a fast shared filesystem for the StateSaveLocation
  - Co-locate the slurmdbd with its own MariaDB instance. Do not worry about running a backup slurmdbd.
  - If short on machines, host the backup slurmctld on the same machine as slurmdbd. But keep them split by default.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

High-Availability

- Common issues:
  - StateSaveLocation needs to be available (and fast!) on both primary and backup at the same time.
  - If the StateSaveLocation filesystem crashes, your cluster will be unavailable. Choose setup carefully, and avoid introducing unneeded complexity.
  - E.g., I would prefer a single slurmctld with a local SSD in it for a lot of environments, over a more elaborate deployment involving DRDB + GFS.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

High-Availability

- Similarly, for slurmdbd the real failure domain is the MySQL/MariaDB installation.
  - In my experience, HA approaches to MySQL/MariaDB are more likely to introduce outages than the hardware failures they're trying to prevent.
  - Just to be clear - you should still keep backups.
  - slurmctld can survive and keep the cluster running while spooling messages destined for the slurmdbd for a while.
    - Although this depends on your environment, and the outstanding messages in the spool are currently capped at 3xNodes + MaxJobCount.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

High-Availability

- Triggers within slurmctld can help supplement your monitoring.
- And yes, you should probably be doing some sort of cluster monitoring.
  - Waiting for grad students to flood your inbox when something breaks doesn’t count.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

FairsharePriority normalization

- For fairshare priority calculation, all of the component values are normalized from 0 to 1.
- This includes, and may be counter-intuitive, values such as the Partition and QOS Priority settings.
- So… with a PartitionPriority=2, and a PriorityWeightPartition=1000, the fairshare contribution from this is… anywhere between 0 and 1000, depending on the highest value for ParititionPriority in the cluster.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

FairsharePriority normalization

- So… either keep this in mind when adjusting the priority values
- Or add in dummy Partition and QOS settings to fix the maximum range, and account for this in your PriorityWeight{Partition,QOS} options.
- E.g., add a partition definition with no nodes:

```
Partition=normalizer PartitionPriority=100 Hidden=yes
```

- Credit to Doug Jacobsen for the idea.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Email

- Users can request email notifications be sent.
- By default, slurmctld will assume the local MTA will know how to send a message to “username”.
- In 17.02, the MailDomain option was added.
- Note that your users can request these status messages be sent to arbitrary destinations. E.g., their personal Gmail account.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Email - Story Time!

- On a system at a previous employer, a user had set their batch scripts to automatically submit a new script on completion.
  - This is a bad idea in general.
- And requested email notifications.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Email - Story Time!

- And then the filesystem that their commands needed to run on went down. But their script didn't bother checking that their simulation commands had actually run before submitting the next job.
- Slurm ran several hundred thousand of these jobs in under an hour.
- And the local MTA handled sending the load without issue.
- But the campus Exchange cluster… not so much.

Copyright 2017 SchedMD LLC
https://www.schedmd.com

Email

- So… you may want to institute some form of rate limiting in your local MTA.
- If you'd like to disable email entirely you can:
  - Use a job_submit plugin to strip the option off all submissions.
    - Or maybe just from certain accounts.
  - Set MailProg to a command that discards the request.
    - E.g.: /bin/true

Copyright 2017 SchedMD LLC
https://www.schedmd.com