Accelerating HPC and AI with Slurm and SchedMD

Nick Ihli, Director - Solutions Engineering and Cloud
nick@schedmd.com
Most people know Slurm!

<table>
<thead>
<tr>
<th>Allocates access</th>
<th>Provides framework</th>
<th>Arbitrates contention</th>
</tr>
</thead>
<tbody>
<tr>
<td>to resources to users for some duration of time for a workload</td>
<td>for starting, executing, and monitoring work on the set of allocated nodes</td>
<td>for resources by managing a queue of pending work</td>
</tr>
</tbody>
</table>

- Policy driven
- Open source
- Fault tolerant
- Highly scalable
- Workload manager
- Job scheduler

Slurm is a workload manager and job scheduler that is policy-driven, open-source, fault-tolerant, and highly scalable.
Slurm on TOP500

7 of the TOP10
And more than 50% of the TOP500 use Slurm

50%+
But what is SchedMD?

Maintainers and supporters of Slurm

- Only organization providing level-3 support
- Training
- Consultation
- Custom Development
Slurm leads in industry trends

**GPU/AI workloads**
- Manages largest GPU/AI systems
- Tight integrations with GPUs

**Hybrid cloud integrations**
- Autoscale cloud resources
- AWS, Google, Microsoft

**Integrations**
- REST API
- Many community tools

**High Scalability**
- Large complex, exascale systems
- 100k+ node/GPUs

**Advanced Container Support**
- Deeper container integration coming in 23.02

**High Throughput**
- 15M jobs/day
- 100M+ jobs/week
Do More with Slurm
Tight GPU integration

- GPUs are a first class citizen like a CPU
- Allows for fine-grained GPU requests
- Bind tasks to GPUs
- MIG support
- Auto-detecting of GPUs
- Constrain jobs to allocated GPUs
- Sets CUDA_VISIBLE_DEVICES environment variable
Slurm is the bridge between on-prem and cloud
Slurm Cloud Autoscaling

- Resume program
  - nodes are "powered on" when required

- Suspend program
  - nodes are "powered off" when no longer required

- Suspend Time
  - Time for node to be IDLE before it is put in power saving mode (deprovisioned)
  - Set globally or per partition

- Timeout settings
  - When to fail if the node has not registered with the controller
Application Integration - Slurm REST API

Client sends a request. (NOT srun, sbatch, salloc)

HTTP Method

GET
POST
PUT
DELETE

HTTP

Server sends a response.
slurmrestd

A tool that runs inside of the Slurm perimeter that will translate JSON/YAML requests into Slurm RPC requests

Diagram:
- slurmctld
- slurmdbd
- slurmrestd
- clients

Connections:
- SLURM RPC
- REST API
Tighter Container Integration - 23.02

- New scrun daemon - goal is to make containers boring for users
- Users have better things to do than learn about the intricacies of containers
- Use Slurm’s existing infrastructure to run containers on compute nodes
- Automatic staging out and in of containers controlled by system administrators
- End requirement that users manually prepare their images on compute nodes.
- Interface directly with OCI runtime clients (Docker or Podman or …)
Do More with SchedMD: A Migration Journey
A Migration Journey

Large Energy Company

• Using their scheduler for many years
  – Can’t just flip a switch and go to production

• Massive scale
  – multiple international sites, nodes, and workloads

• Many integration tools required

3-4 months production
Three Migration Steps

Admin/user education

Training: Help admins identify the commonalities and learn the Slurm way

Wrappers: Use as a bridge to migration not a crutch
- LSF, Grid Engine - command and submission
- PBS - command, submission, environment variables, #PBS scripts

Policy replication

Reevaluate policies
- Are we continuing to produce technical debt due to “doing things how we’ve always done them?”
- Opportunity to take a step back and redefine policies based on Slurm best practices

Optimizing for scale and throughput

Tooling integration

Most time-consuming of the migration journey.
- REST API
- Community integrations
Thank You

schedmd.com  slurmschedmd.com  nick@schedmd.com