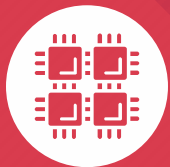




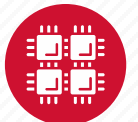
PBS to Slurm Migration

September 23, 2020



Ohio Supercomputer Center

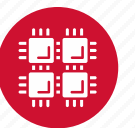
An OH·TECH Consortium Member





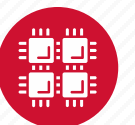
Summer Wang

Senior Client Engineer



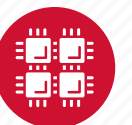
Background

- What is Slurm
 - Open source package for job scheduling and resource management
 - Originated at LLNL in 2001
 - Commercial support and services from SchedMD starting 2010
 - Very widely used
- OSC has been using Torque/Moab for many years
 - Support for Torque/Moab ends December 31st



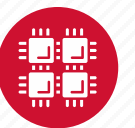
OSC Slurm Migration Timeline

- Pitzer
 - September 22, 2020 (Completed)
 - With deployment of new Pitzer hardware
- Owens
 - Early- to mid-November 2020
 - Planning work just started
- Ruby
 - No migration scheduled
 - Planned decommission: End of 2020



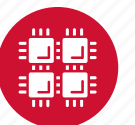
A Big Challenge for OSC Users?

- Do NOT Panic
- We try to reduce challenges caused by the migration:
 - PBS Compatibility layer enabled
 - Most PBS syntax has Slurm equivalence
 - Early user program (open to all) to test Slurm



Early User Program

- Between Aug 31 and Sept 21
- 102 individual users submitted jobs in Slurm
 - ~27% of total active Pitzer users
- 55 Slurm-related tickets
 - Helped updating user documentations
 - Deployed patches to fix bugs
 - Some were added to Slurm migration issues:
https://www.osc.edu/supercomputing/knowledge-base/slurm_migration/slurm_migration_issues

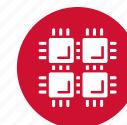


Job Submission Options

Torque/Moab directive	Slurm directive
#PBS -N myjob	#SBATCH --job-name=myjob
#PBS -l walltime=1:00:00	#SBATCH --time=1:00:00
#PBS -l nodes=N:ppn=M	#SBATCH --nodes=N --ntasks-per-node=M
#PBS -l nodes=N:ppn=M:gpus=G	#SBATCH --nodes=N --ntasks-per-node=M #SBATCH --gpus-per-node=G
#PBS -l mem=Xgb	#SBATCH --mem=Xgb
#PBS -l software=pkg1+1%pkg2+4	#SBATCH --licenses=pkg1:1,pkg2:4
#PBS -A account	#SBATCH -account=account

Exhaustive list:

https://www.osc.edu/supercomputing/knowledge-base/slurm_migration/how_to_prepare_slurm_job_scripts

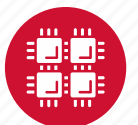


Job Environment Variables

Torque/Moab environment variable	Slurm environment variable
<code>\$PBS_JOBID</code>	<code>\$_SLURM_JOB_ID</code>
<code>\$PBS_JOBNAME</code>	<code>\$_SLURM_JOB_NAME</code>
<code>\$PBS_QUEUE</code>	<code>\$_SLURM_JOB_PARTITION</code>
<code>\$PBS_O_WORKDIR</code>	<code>\$_SLURM_SUBMIT_DIR</code>
<code>\$PBS_NODEFILE</code>	<code>\$_SLURM_JOB_NODELIST</code>
<code>\$PBS_NUM_NODES</code>	<code>\$_SLURM_JOB_NUM_NODES</code>
<code>\$PBS_NP</code>	<code>\$_SLURM_NTASKS</code>
<code>\$PBS_NUM_PPN</code>	<code>\$_SLURM_TASKS_PER_NODE</code>
<code>\$PBS_WALLTIME</code>	<code>\$_SLURM_TIME_LIMIT</code>

Exhaustive list:

https://www.osc.edu/supercomputing/knowledge-base/slurm_migration/how_to_prepare_slurm_job_scripts

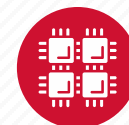


Submit & Manage Batch Jobs

Torque/Moab	Slurm
qsub <jobscript>	sbatch <jobscript>
qdel <jobid>	scancel <jobid>
qhold <jobid>	scontrol hold <jobid>
qrls <jobid>	scontrol release <jobid>
qstat -u <user>	squeue -u <user>

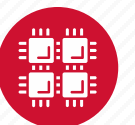
More information:

https://www.osc.edu/supercomputing/knowledge-base/slurm_migration/how_to_monitor_and_manage_jobs



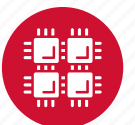
Notable Differences between Torque/Moab and Slurm

- First line of job script in Slurm must be '#!<shell>'
 - However, scripts missing '#!<shell>' still work because we patch the qsub/sbatch wrapper
- Slurm jobs start in the submission directory rather than \$HOME
- Slurm jobs have stdout and stderr output log files combined by default
- Slurm can send email when your job reaches certain percentage of walltime limit:
 - `sbatch --mail-type=TIME_LIMIT_90 myjob.txt`



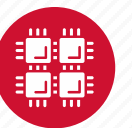
A Simple Example

Explanations	Torque	Slurm
Line that specifies the shell	No need	<code>#!/bin/bash</code>
Resource specification	<code>#PBS -l walltime=1:00:00</code> <code>#PBS -l nodes=2:ppn=40</code> <code>#PBS -N hello</code> <code>#PBS -A PZS0712</code>	<code>#SBATCH --time=1:00:00</code> <code>#SBATCH --nodes=2</code> <code>#SBATCH --ntasks-per-node=40</code> <code>#SBATCH --job-name=hello</code> <code>#SBATCH --account=PZS0712</code>
Variables, path, and modules	<code>cd \$PBS_O_WORKDIR</code> <code>module load intel</code>	<code>cd \$SLURM_SUBMIT_DIR</code> <code>module load intel</code>
Launch and run application	<code>mpicc -O2 hello.c -o hello</code> <code>mpiexec ./hello > hello_results</code>	<code>mpicc -O2 hello.c -o hello</code> <code>srun ./hello > hello_results</code>



Demo

- Interactive jobs
 - `qsub -I`
 - `salloc`
 - `sinteractive`
- Non-interactive jobs: examples are available in `/users/oscggen/xwang/slurm/workshop`
 - `pr -m -t pbs_job.txt slurm_job.txt`
 - `qsub/sbatch pbs_job.txt`
 - `qsub/sbatch slurm_job.txt`
 - `qsub/sbatch mix_job.txt`
 - `qstat/squeue -u userid`

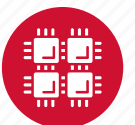


Using --constraint

- Pitzer is a heterogeneous cluster. Jobs may land on different types of CPUs if the user submits the same job script repeatedly
- `#SBATCH --ntasks=1` (or `#PBS -l nodes=1:ppn=1`) can land on nodes in the following table

CPU type	Per node	Usable memory/core
Dual Intel Xeon 6148s Skylake @2.4GHz	40 cores	4556 MB
Dual Intel Xeon 8268s Cascade Lakes @2.9GHz	48 cores	3797 MB

- Add `#SBATCH --constraint=40core/48core`
- More information:
https://www.osc.edu/resources/technical_support/supercomputers/pitzer/guidance_on_requesting_resources_on_pitzer

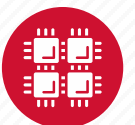


Job Information

- Job statistic information like below not available in output by default

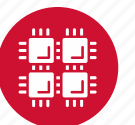
```
-----  
Resources requested:  
nodes=1:ppn=4  
mem=17260mb  
-----  
Resources used:  
cput=00:00:01  
walltime=00:00:13  
mem=0.004GB  
vmem=0.226GB  
-----
```

- Use 'sacct' command: <https://slurm.schedmd.com/sacct.html>
- More information will be provided on our webpage soon



Take-home Messages

- Slurm has been implemented on Pitzer on Sept 22
- If you use Pitzer but have not submitted any job in Slurm environment. Simply submit your PBS job script the same way as you did before
 - If it works, great! No immediate action is needed. However, please start to convert your script into Slurm
 - If it does not work, do NOT debug the PBS script in Slurm environment. Please convert your script to Slurm and contact oschelp@osc.edu for assistance, if needed





Questions?

