Introduction to SLURM & SLURM batch scripts

Anita Orendt
Assistant Director
Research Consulting & Faculty Engagement
anita.orendt@utah.edu

4 February 2016
Overview of Talk

• Basic SLURM commands
• SLURM batch directives
• Accounts and Partitions
• SLURM Environment Variables
• Running an Interactive Batch job
• SLURM Batch scripts
• Where to get more Information

2/4/2016
http://www.chpc.utah.edu
Basic SLURM commands

• `sinfo` - shows partition/node state
• `sbatch <scriptname>` - launches a batch script
• `squeue` - shows all jobs in the queue
  – `squeue -u <username>` - shows only your jobs
• `scancel <jobid>` - cancels a job
SLURM Batch Directives

#SBATCH --time 1:00:00  ← wall time of a job (or -t)
#SBATCH --partition=name  ← partition to use (or -p)
#SBATCH --account=name  ← account to use (or -A)
#SBATCH --nodes=2  ← number of nodes (or -N)
#SBATCH --ntasks 12  ← total number of tasks (or -n)
#SBATCH --ntasks-per-node=12  ← OR can use tasks per node
#SBATCH --mail-type=FAIL,BEGIN,END  ← events on which to send email
#SBATCH --mail-user=name@example.com  ← email address to use
#SBATCH -o slurm-%j.out-%N  ← name for stdout; %j is job#, %N node
#SBATCH -e slurm-%j.err-%N  ← name for stderr; %j is job#, %N node
#SBATCH --constraint “C20”  ← can use features given for nodes (or –C)
Accounts and Partitions

- You need to specify an account and partition to run jobs
- You can see a list of partitions using the `sinfo` command
- For general allocation usage the partition is the cluster name
- If no allocation (or out of allocation) use `clusternamelfreecycle` for partition
- Your account is typically your PI’s name (e.g., if my PI is Baggins, I use the "baggins" account) – there are a few exceptions!
- Private node accounts and partition have the same name – PI last name with cluster abbreviation, e.g., baggins-kp, baggins-em, etc
- Private nodes can be used as a guest using the "owner-guest" account and the cluster-guest partition
Useful Slurm Aliases

• Bash version

  alias si="sinfo -o \"%20P %5D %14F %8z %10m %10d %11l %16f %N\""
  alias si2="sinfo -o \"%20P %5D %6t %8z %10m %10d %11l %16f %N\"
  alias sq="squeue -o \"%8i %12j %4t %10u %20q %20a %10g %20P %10Q %5D %11l %11L %R\""

• Tcsh version

  alias si 'sinfo -o "%20P %5D %14F %8z %10m %11l %16f %N"
  alias si2 'sinfo -o "%20P %5D %6t %8z %10m %10d %11l %N"
  alias sq 'squeue -o "%8i %12j %4t %10u %20q %20a %10g %20P %10Q %5D %11l %11L %R""
SLURM Environment variables

- Depend on SLURM Batch Directives used, especially choices of ntasks, ntasks-per-per node
- Can get them for a given set of directives by using the env command inside a script (or in a srun session).
- Some useful env variables:
  - $SLURM_JOBID
  - $SLURM_SUBMIT_DIR
  - $SLURM_NNODES
  - $SLURM_NTASKS
Running interactive batch jobs

- An interactive command is launched through the `srun` command
  
  ```
  srun --time=1:00:00 --nodes=1 --account=chpc
    --partition=ember --pty /bin/tcsh -l
  ```

- Launching an interactive job automatically forwards environment information, including X11 forwarding.

- `"--pty"` must be set to shell preferred for the session (either `/bin/tcsh` or `/bin/bash`)

- `-l` (lower case “L”) at the end required
Basic SLURM script flow

1. Set up the #SBATCH directives for the scheduler to request resources for job
2. Set up the working environment, by loading appropriate modules
3. If necessary, add any additional libraries or programs to $PATH and $LD_LIBRARY_PATH, or set other environment needs
4. Set up temporary/scratch directories if needed
5. Switch to the working directory
6. Run the program
7. Clean up any temporary files or directories
Basic SLURM script - bash

#!/bin/bash
#SBATCH --time=02:00
#SBATCH --nodes=1
#SBATCH -o slurmjob-%j
#SBATCH --ntasks-per-node=16
#SBATCH --account=owner-guest
#SBATCH --partition=kingspeak-guest

#Set up whatever package we need to run with
module load somemodule

#set up the temporary directory
TMPDIR=/scratch/local/u0123456/data
mkdir -p $TMPDIR

#Set up the path to the working directory
WORKDIR=/uufs/chpc.utah.edu/common/home/u0123456/data
cd $WORKDIR

#Run the program with our input
BINDIR=/uufs/chpc.utah.edu/sys/pkg/mypackage/bin
$BINDIR/myprogram < $WORKDIR/input > $WORKDIR/output
rm -rf $TMPDIR
Basic SLURM script - tcsh

#!/bin/tcsh
#SBATCH --time=02:00
#SBATCH --nodes=1
#SBATCH -o slurmjob-%j
#SBATCH --ntasks-per-node=16
#SBATCH --account=owner-guest
#SBATCH --partition=kingspeak-guest

#Set up whatever package we need to run with
module load somemodule

#set up the temporary directory
set TMPDIR /scratch/local/u0123456/data
mkdir -P $TMPDIR

#Set up the path to the working directory
set WORKDIR /uufs/chpc.utah.edu/common/home/u0123456/data
cd $WORKDIR

#Run the program with our input
set BINDIR /uufs/chpc.utah.edu/sys/pkg/mypackage/bin
BINDIR/myprogram < $WORKDIR/input > $WORKDIR/output
rm -rf $TMPDIR
Parallel Execution

• If needed, create the node list:
  - `srun hostname | sort -u > nodefile.$SLURM_JOBID`
  - `srun hostname | sort > nodefile.$SLURM_JOBID`

• MPI installations at CHPC are SLURM aware, so `mpirun` will work without a machinefile (unless you are manipulating the machinefile in your scripts)

• Alternatively, you can use the `srun` command instead, but you need to compile with a more recently compiled MPI

• Mileage may vary, and for different MPI distributions, `srun` or `mpirun` may be preferred (check our slurm page on the CHPC website for more info or email us)
Slurm Documentation at CHPC

- https://www.chpc.utah.edu/documentation/software/slurm.php

Other good documentation sources

- http://slurm.schedmd.com/documentation.html
Getting Help

- CHPC website and wiki
  - [www.chpc.utah.edu](http://www.chpc.utah.edu)
    - Getting started guide, cluster usage guides, software manual pages, CHPC policies
- Jira Ticketing System
  - Email: [issues@chpc.utah.edu](mailto:issues@chpc.utah.edu)
- Help Desk: 405 INSCC, 581-6440 (9-5 M-F)
- We use [chpc-hpc-users@lists.utah.edu](mailto:chpc-hpc-users@lists.utah.edu) for sending messages to users; also have Twitter accounts for announcements -- @CHPCOutages & @CHPCUpdates