



703651 PS Parallel Systems

A Crash Course in Clusters and Job Submission

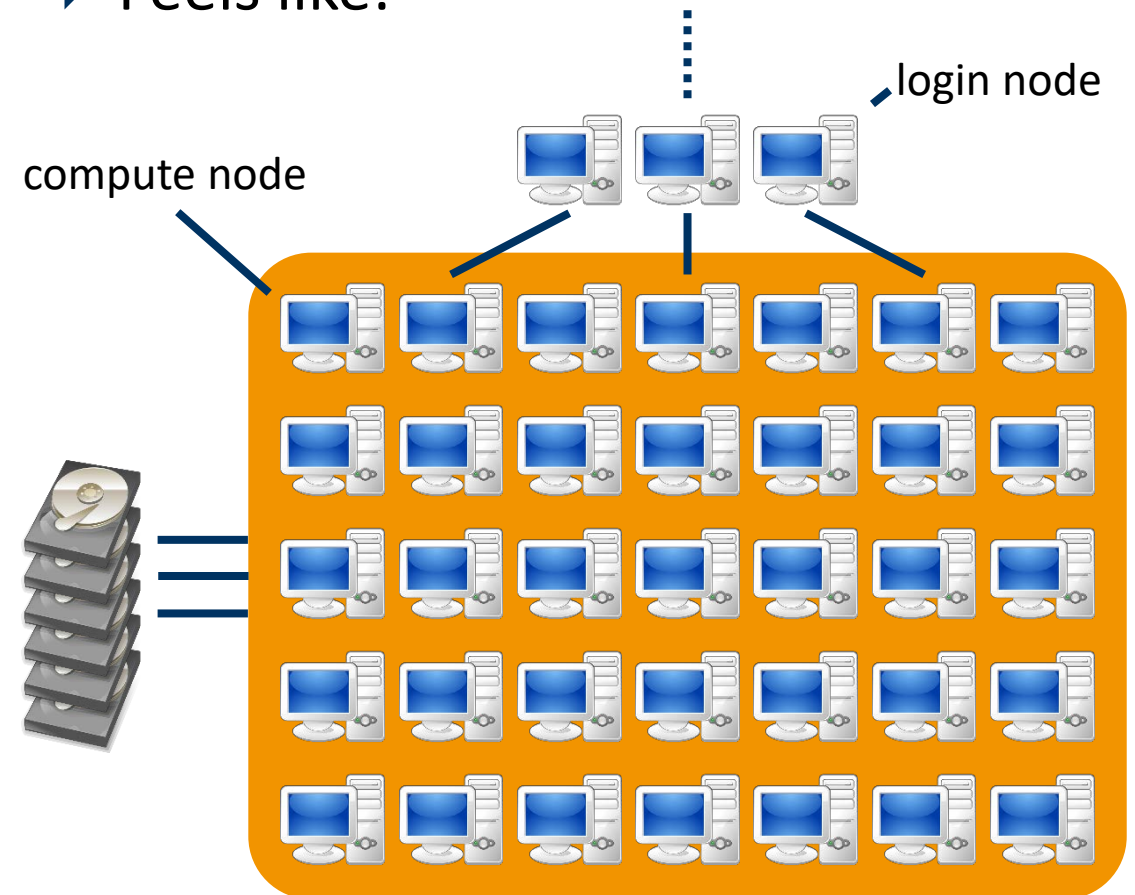
Philipp Gschwandtner

Clusters and Supercomputers

► Looks like:



► Feels like:



Get User Credentials, Log in and Change Your Password!

- ▶ `ssh cbxxxxxx@lcc2.uibk.ac.at`
- ▶ Change password with `passwd`
- ▶ don't use these credentials for anything other than this course
 - ▶ coin mining isn't worth it anyways...

Submission Systems

- ▶ Responsible for resource management and job orchestration
 - ▶ used to submit or cancel “jobs”, query their status, get information about cluster, ...
- ▶ Very popular: SLURM
 - ▶ modern, complex but very capable
 - ▶ de-facto standard on most systems these days
- ▶ On LCC2: Sun Grid Engine (SGE)
 - ▶ older and deprecated ☹️
 - ▶ switch to SLURM currently in progress



Jobs: Submission, Deletion, Status

▶ `qsub name_of_script`

- ▶ allocates resources
- ▶ sets up environment
- ▶ executes application
- ▶ frees allocation

▶ `qdel job_id_list`

- ▶ terminates application
- ▶ frees up resources

▶ `qstat`

- ▶ queries for job status

```
[cb761011@login.lcc2 ~]$ qsub job.script
Your job 30369 ("my_test_job") has been submitted
[cb761011@login.lcc2 ~]$ qdel 30369
cb761011 has deleted job 30369
[cb761011@login.lcc2 ~]$ qsub job.script
Your job 30370 ("my_test_job") has been submitted
[cb761011@login.lcc2 ~]$ qstat
```

job-ID	prior	name	user	state	submit/start at	queue	slots	ja-task-ID
30370	0.00000	my_test_jo	cb761011	qw	09/30/2019 15:02:01		8	

SGE Job Scripts

```
#!/bin/bash

# Execute job in the queue "std.q" unless you have special requirements.
#$ -q std.q
# The batch system should use the current directory as working directory.
#$ -cwd
# Name your job. Unless you use the -o and -e options, output will
# go to a unique file name.ojob_id for each job.
#$ -N my_test_job
# Redirect output stream to this file.
#$ -o output.dat
# Join the error stream to the output stream.
#$ -j yes
# Specify parallel environment (list available ones with qconf -spl)
#$ -pe openmpi-2perhost 8

/bin/hostname
```

Parallel Environments

- ▶ Specifies number of CPU/core slots to allocate per node and in total
 - ▶ `openmpi-Xperhost Y` X slots per node, Y slots in total
- ▶ List available environments with `qconf -spl`
 - ▶ `openmpi-1perhost`
 - ▶ `openmpi-2perhost`
 - ▶ `openmpi-4perhost`
 - ▶ `openmpi-8perhost`
- ▶ Be careful with non-exclusive allocations!

Action!

- ▶ Submit the job, wait for it to finish, check the output
 - ▶ What happened and what did you expect?

Modules System

- ▶ Used to modify the user environment
(environment variables, most notably PATH & LD_LIBRARY_PATH)
- ▶ `module avail`
- ▶ `module load`
- ▶ `module list`
- ▶ `module unload`

Fix Job Script

- ▶ add a line to load the required MPI module (e.g. just before program execution)
 - ▶ `module load openmpi/4.0.1`
- ▶ fix the program execution line in the jobscript
 - ▶ `mpiexec -n 8 /bin/hostname`
- ▶ Re-submit and check the output
- ▶ Happy now?

Compiling and Running MPI programs

- ▶ compiler wrappers for C/C++ (set all required flags and directories)
 - ▶ `mpicc`
 - ▶ `mpic++`
- ▶ execution wrapper for MPI programs
 - ▶ `mpiexec -n [num_processes] /path/to/application`

Further Information on SGE, Job Scripts and LCC2

- ▶ Refer to ZID's help pages
 - ▶ LCC2 Status: <http://login.lcc2.uibk.ac.at/cgi-bin/state.pl>
 - ▶ LCC2: <https://www.uibk.ac.at/zid/systeme/hpc-systeme/lcc2/>
 - ▶ SGE: <https://www.uibk.ac.at/zid/systeme/hpc-systeme/common/tutorials/sge-howto.html>
- ▶ Consult manpages or “The Internet”
- ▶ Ask me

Image Sources

- ▶ Cluster Photo: https://forschungsinfrastruktur.bmbwf.gv.at/de/fi/hpc-compute-cluster-leo3-leo3e_513
- ▶ SLURM: https://justjimsthoughts.blogspot.com/2017/07/trivia_24.html