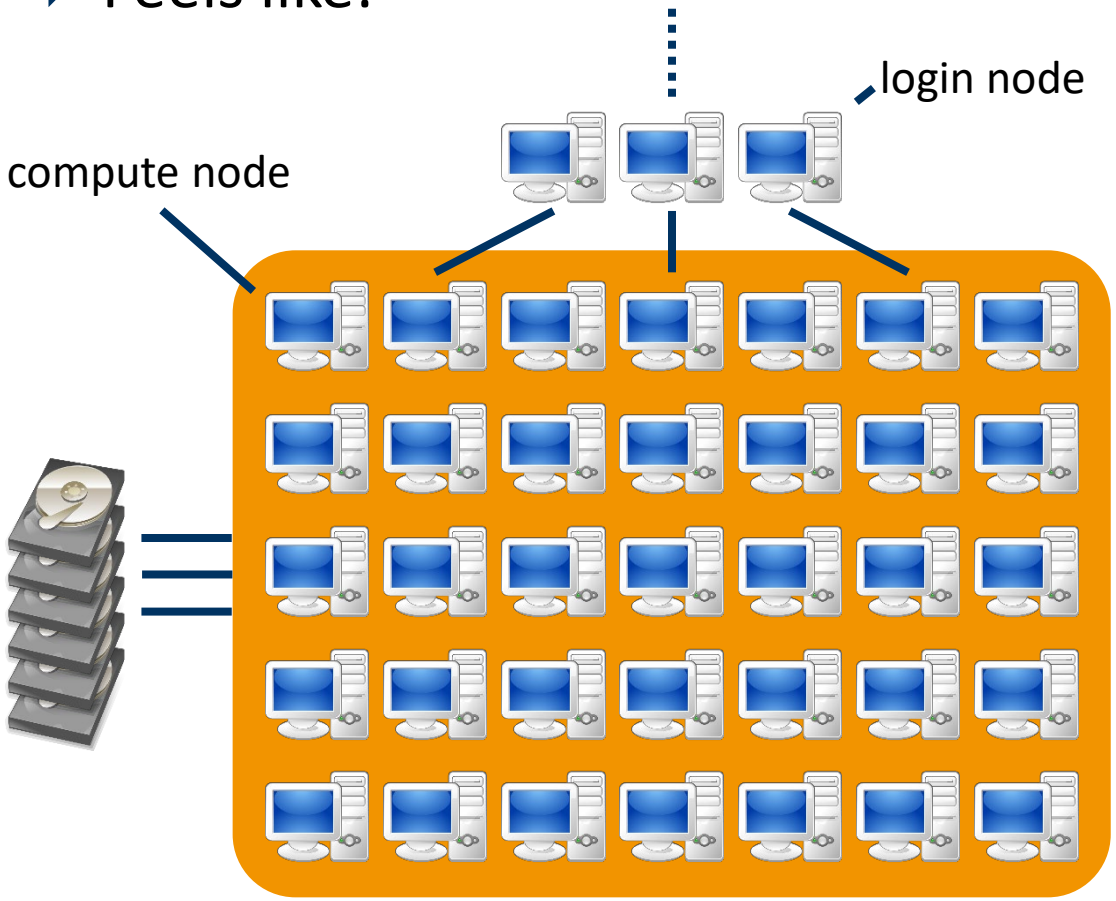
703309 PS High-Performance Computing
A Crash Course in Clusters and Job Submission

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~~

SLURM!!11!!

Jobs: Submission, Deletion, Status

- `sbatch name_of_script`
  - allocates resources
  - sets up environment
  - executes application
  - frees allocation
- `scancel job_id_list`
  - terminates application
  - frees up resources
- `squ (squeue -u $USER)`
  - queries for job status
  - `squeue` for all users

```
[cb761011@login.lcc2 ~]$ sbatch job.sh
Submitted batch job 184
[cb761011@login.lcc2 ~]$ scancel 184
[cb761011@login.lcc2 ~]$ sbatch job.sh
Submitted batch job 185
[cb761011@login.lcc2 ~]$ squ
JOBID PRIORITY PARTITION NAME USER     STATE   NODES CPUS TIME_LIMIT NODELIST(REASON)
185   504      lva       test cb761011 RUNNING 1     8    30:00      n002
```

SLURM Job Scripts

```
#!/bin/bash

# Execute job in the partition "lva" unless you have special requirements.
#SBATCH --partition=lva
# Name your job to be able to identify it later
#SBATCH --job-name test
# Redirect output stream to this file
#SBATCH --output=output.log
# Maximum number of tasks (=processes) to start in total
#SBATCH --ntasks=8
# Maximum number of tasks (=processes) to start per node
#SBATCH --ntasks-per-node=8
# Enforce exclusive node allocation, do not share with other jobs
#SBATCH --exclusive

/bin/hostname
```

Additional Useful SLURM Tools

- `sinfo`
  - list information on partitions, nodes, etc.
- `scontrol`
  - alter properties of pending jobs
- `sacct`
  - show accounting information

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/4.0.3`
- fix the program execution line in the jobscript
  - `mpiexec -n $SLURM_NTASKS /bin/hostname`
- Re-submit and check the output
- Happy now?

Compiling and Running MPI programs

- MPI is an inter-process communication library
  - provides a header + library files (*.so/*.a)
  - more information in the lecture

- 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`

Storage and Compilers on LCC2

- LCC2 has two main storage mount points
  - `/home/cb76/<username>` limited to 500 MB
  - `/scratch/<username>` limited to 100 GB
  - Both are network-mounted on all login and compute nodes
- LCC2's system gcc is ancient (4.8.5)
  - use modules to find and load newer versions (e.g. 10.0.3)

Further Information on SLURM, Job Scripts and LCC2

- Refer to ZID's help pages
  - LCC2 Status: https://login.lcc2.uibk.ac.at/cgi-bin/slurm.pl
  - LCC2: https://www.uibk.ac.at/zid/systeme/hpc-systeme/lcc2/
  - SLURM: https://www.uibk.ac.at/zid/systeme/hpc-systeme/common/tutorials/slurm-tutorial.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