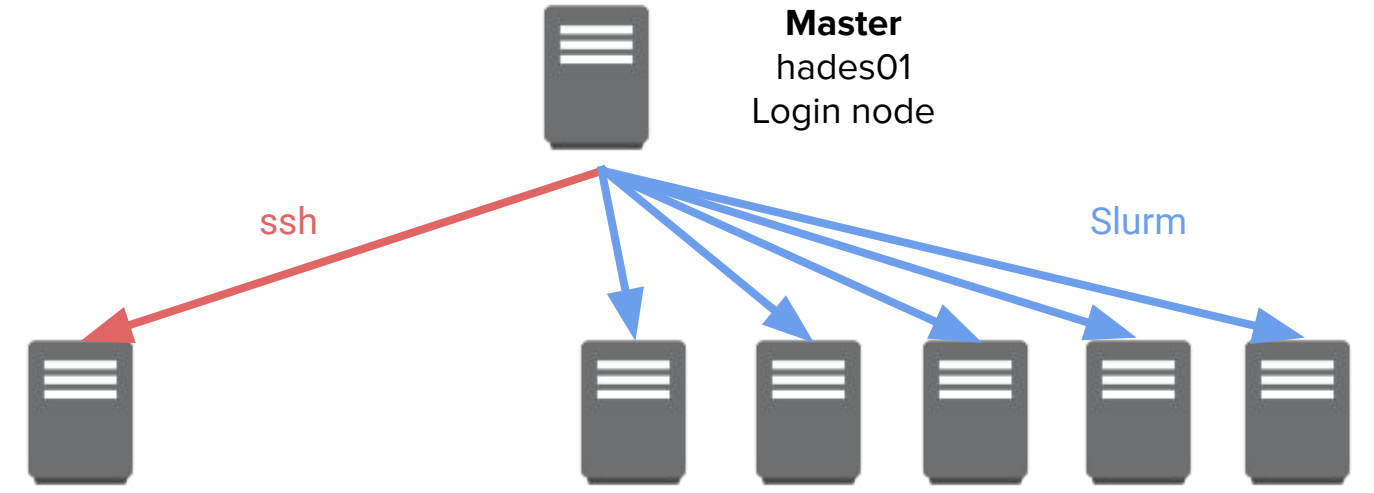
Lab3
GPU Cluster

Introduction to Parallel Computing
2022/03/22

Hardware

- hades01
  - CPU: Intel Xeon X5670 (2.93GHz) * 2
  - Memory: 96GB
  - No GPU
- hades02 - 07
  - CPU: ---
  - Memory: 64GB
  - GPU: GTX 1080*2

Architechure



You cannot use GPU on hades03-07 directly !

Software

- Linux 3.10
- Compilers
  - GCC 10.1.0
  - Clang 10.0.1
  - NVCC 11.0.194 (CUDA 11)
- Wordload Scheduler
  - Slurm 20.02.5

Available Resources

- Login node (200% CPU max)
- hades02
  - 2 GPUs, use by `export CUDA_VISIBLE_DEVICES` to env
- The remaining are compute nodes, available via job submission
  - Max nodes: 2
  - Max GPU: 2
  - Max wall time: 5 minutes
  - Max jobs run at any time: 2
- Priority
  - Favor short running jobs(based on wall time)
  - Favor less resource demanding jobs (based on number of GPUs)
  - Favor jobs which are queued for longer

SSH Credentials

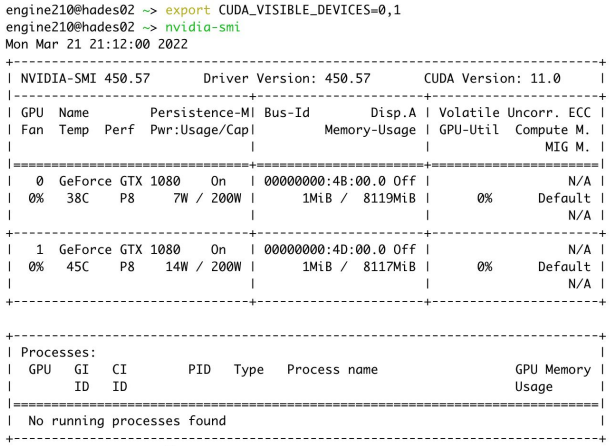
- Hostname
  - hades.cs.nthu.edu.tw
- Username
  - Same as apollo
- Password
  - Same as apollo
- Check the slides of Lab1 if you forget how to login to the server
  - If you don’t want to re-type your password when ssh into hades02
    => Generate the public/private key pair on hades, and ssh-copy-id into hades!

Compile A CUDA Program

- Compile a CUDA program
  - `nvcc` [options] [exe] [target]
  - e.g. `nvcc -o exe src.cu`
- Most of usages are same as GNU compiler
- Check `nvcc --help` for more options

hades02

- ssh to hades02
- export CUDA_VISIBLE_DEVICES=0 for using the first GPU
- export CUDA_VISIBLE_DEVICES=1 for using the second GPU
- export CUDA_VISIBLE_DEVICES=0,1 for using the two GPUs
- Run program on hades02 directly



Slurm Scheduler (For hades[03-07])

- Most of options are same as the apollo scheduler

-p <PARTITION>	**<PARTITION> should be ipc22**
-N <NODES>	<NODES> should be the number of nodes to run the job
-n <PROCESSES>	<PROCESSES> are the number of processes to launch
-c <CORES>	Each process can use up to <CORES> CPU cores
-t <TIME>	Time limit with format “mm:ss”
-j <NAME>	The name of job, will be displayed on squeue
-w <NODE_LIST>	The specific <NODE_LIST> you want to run
--gres=gpu:<GPUS>	**<GPUS> is the numbers of GPUs to run the job**

Example

```
[ipc22sxx@hades01 ~]$ srun -p ipc22 --gres=gpu:1 -N 1 -n 1 -w hades04 ./deviceQuery
```

Demo

- CUDA-MEMCHECK
  `cuda-memcheck /home/ipc22/share/lab3/memcheck_demo`
- CUDA-GDB
  Not available on hades, but you can try it on your own computer.

Practice

- Google Form: https://forms.gle/DHteQJ44vQh943mW8
- Please finish the form before 3/29 23:59

Note for VSCode Terminal users. If command not found, try:

- `bash -l`
- MPI on apollo (mpicc, mpirun, etc.):
  `source /opt/intel/mpi/intel64/bin/mpivars.sh`
- CUDA on hades (nvcc, etc.):
  `source /etc/profile.d/cuda.sh`