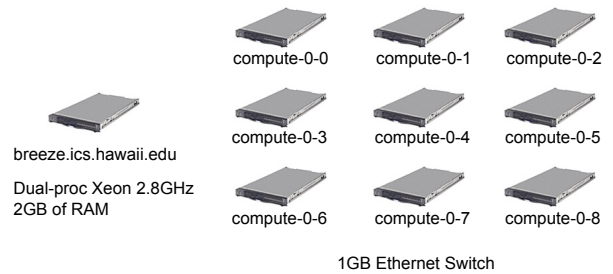


Principles of High Performance Computing (ICS 632)

How to use the cluster

Our Cluster

- You now all have an account on our cluster
- The cluster is called breeze:



Our Cluster

- Question: once I am logged in to breeze, what do I do?
- Clusters are always organized as
 - A front end node
 - To compile code (and do **minimal** testing)
 - To **submit jobs**
 - Compute nodes
 - To run the code
 - You don't ssh to these directly
 - In most production clusters it's disallowed
 - There is a file system mounted over all nodes
 - Can be fast, can be slow, depending
 - Each node has a local storage as well
 - For our programming assignment we won't have I/O issues, but perhaps for your projects

Batch Schedulers

- Most production clusters are managed via a **batch scheduler**:
 - You ask the batch scheduler to give you X nodes for Y minutes to run program Z
 - At some point, the program will be started.
 - Later on you can look at the program output
- This is really different from what you're used to, and honestly is sort of painful
 - No interactive execution
- Necessary because:
 - Since most applications are in this for high performance, they'd better be alone on their compute nodes
 - There are not enough compute nodes for everybody at all times
- The batch scheduler on the cluster is called Torque/Maui

How to use Torque/Maui?

- You need to learn how to do three things
 - Check the status of the platform (optional)
 - Submit a job
 - Check on job status
 - Cancel a job
- All can be done from the command line
- Let's go through some typical examples

Checking the status of the platform

- There is a low-level command to check the status of individual nodes: pbsnodes
- It simply returns the list of available nodes
 - Includes status
 - Includes physical characteristics
- Let's try it...

Checking the status of the platform

- A higher-level command is: showq
- Showq shows the status of the **normal** queue
 - Which jobs are running
 - Which jobs are idle: could be running, but just not enough space on the machine
 - Which jobs are blocked: can't be running on the machine, but perhaps later
 - E.g., too many running/idle jobs from the current user
- Let's try it

Submitting a 1-node Job

- Say I want to submit a job that does a simple command, to the default queue
 - In this class we'll all submit to the normal queue
- Say we want to do "echo hello; sleep 20"
- I can simply do:
% echo "echo hello; sleep 20" | qsub
- Let's try it and look at the status ...

Stdout and Stderr

- In the previous example 2 files were created:
 - STDIN.o1
 - STDIN.e1
- The name of the file corresponds to where the job came from
 - In this case Stdin
- The number at the end is the ID of the job
- The .o means: here is the stdout produced by the job
- The .e means: here is the stderr produced by the job

Job Scripts

- To control a bit more what happens, one has to write a job script
- Here is a simple script

```
#PBS -l nodes=1:ppn=2          very important!!
#PBS -l walltime=5:00:00
#PBS -o myprogram.out
#PBS -e myprogram.err
```

```
cd $PBS_O_WORKDIR
```

```
./myprogram arg1 arg2
```

- Let's try it with simply: qsub my_script ...

Environment variables

- The batch scheduler exports environment variables to the script
- In the previous example we saw \$PBS_O_WORKDIR
- There are others
 - http://www.clusterresources.com/wiki/doku.php?id=torque:2.1_job_submission
- An important one is: \$PBS_NODEFILE
 - The list of hosts allocated to the job
 - In our case it's just one host
- Let's try it ...

Canceling a job

- This is done with the qdel command
- Let's submit a long job and then delete it...



That's pretty much it

- we'll talk about multi-node jobs later
- we'll talk about how batch schedulers work, and/or how they should work
 - A lot of theory (which we'll gloss over)
 - A lot of engineering/practice
 - The two are not very connected
 - A bunch of interesting new issues
 - Essentially, we still don't know how to share and play nice



Sample Batch Script

- There is a sample one-node batch script in /home/casanova/public on the cluster
- You must take it and modify it for your needs, according to the comments therein
 - very little modification
- Let's look at it right now...