

Notes on Supercomputing with Hyak

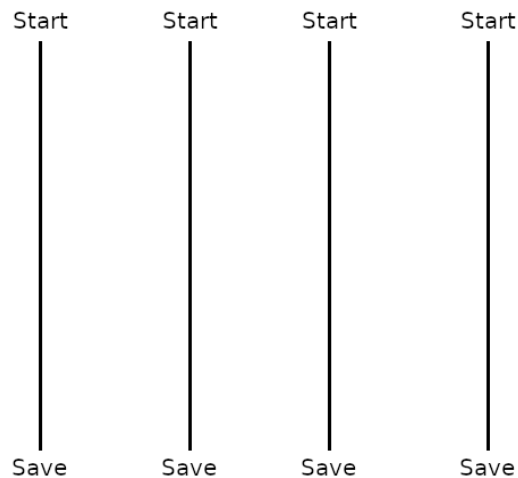
M.P.Ross

January 2, 2019

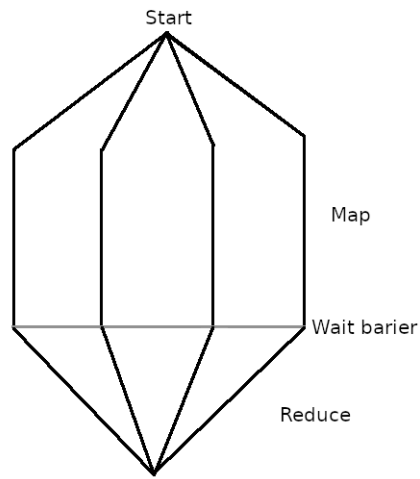
1 Parallel Computing

There are two main types of parallel computing:

1. Embarrassingly parallel: completely independent processes
 - speed-up = number of cores
 - can be done easily with many nodes and many cores



2. Map-Reduce: data shared between processes
 - speed-up = $\frac{1}{(1-p)+p/n}$ where p is the fraction of code which is parallelizable and n is the number of cores
 - harder to do across nodes but not impossible
 - can be stacked multiple times
 - must have a wait barrier to wait for all processes to be done before having them communicate to each other



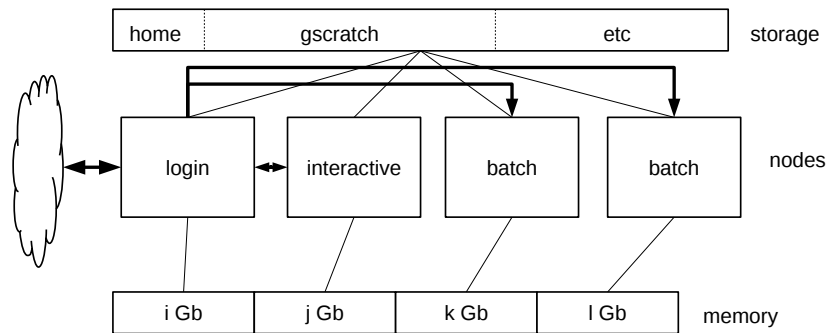
2 Hyak Structure

Supercomputers are essentially a collection of high end computers connected via a high speed network that share resources. Each computer is called a node and all nodes share memory and storage.

Two types of nodes designed to run code:

1. Interactive: user has full control using command line
2. Batch: user submits job and node runs without user control

There's also a login node that is designed to only start jobs and manipulate files. All nodes have 28 cores, have access to the entire storage, and have a set partition of memory (can be up to 150 GB). Nodes are allocated using the slurm scheduler which will give priority based on many parameters including amount of time requested, memory usage, number of nodes, user's history, etc.



3 Useful Commands

Logging in: gets you into the login node:

```
ssh username@mox.hyak.uw.edu
```

Get interactive node:

```
srun -p group --time=time limit --mem=amount of memory
      -N number of nodes --pty what to do
```

```
srun -p stf --time=1:00:00 --mem=10GB -N 2 --pty /bin/bash
```

stf = student technology fee

To load software:

```
module load (software name)
```

Submit batch job:

First make script using google or examples then

```
sbatch myscript.slurm
```

See status of job:

```
scontrol show job (job #)
```