Cluster Usage Notes and Guidelines
CSE 490h Winter 08

Below are a few notes and guidelines to help you use the Hadoop cluster and be considerate of others in the class:

- As project deadlines approach, the cluster will likely incur heavy load. This means that the longer you wait, the longer your projects will take to complete. So start early.

- None of the projects (with the exception of perhaps the final project) will require Hadoop jobs that take longer than about 30 minutes to run. If your job is taking up the entire cluster and is taking longer than 30 minutes, you are probably doing something wrong. Please kill it so that others can run their jobs. If we see jobs running on the cluster for exceedingly long times, we will kill them.

- The DFS (Distributed File System) that we are using has no file permissions. That means you can delete the entire contents of the file system with one command – so think before you type, and when using the eclipse plug-in, think before you hit OK! (Yes, this actually happened last quarter.)

- Problems with your mapper/reducer code will manifest themselves in unexpected ways. If your mapper fails because of a bug in your code, you will not see this in the driver output. The likeliest scenario is that your driver will report problems reading mapper/reducer output from the map/reduce nodes. You can then use the job tracker page to look at the failures on the nodes.

- The job tracker page can be found at http://hadoop.cs.washington.edu:50030/