

General Information about EECS Hadoop Cluster

Since you all have UC email and you are also enrolled in CEAS, you should be able to access the cluster. Cluster can be accessed from your machine if you are connected to UC network. There are 2 ways to do it.

Server: hadoop-gate-0.eecs.uc.edu

UserName: UC 6+2 (for me it is kumarlt)

Password: Your UC password

Hadoop cluster at EECS is configured to be used via command line and the web page I showed yesterday is just for tracking your job status.

Web Site: <http://hadoop-gate-0.eecs.uc.edu>

On this page, you should be able to see following details:

1. your job status
2. how much resources are being used by your job
3. How much time taken by your job to complete.
4. Job Queue status
5. Name of your job which is being executed.
6. Job start time
7. Job finish time.

General HDFS Commands

List directory:

```
$hadoop fs -ls
```

Make directory:

```
$hadoop fs -mkdir <dir_name>
```

Delete directory:

```
$hadoop fs -rm -r <dir_name>
```

Delete File:

\$hadoop fs -rm <file_name>/<file_name>

Compile Java file:

\$hadoop com.sun.tools.javac.Main <filename.java> -d <destination directory>

Create Jar file:

\$jar -cvf <filename.jar> -C <destination directory> <target directory>
if want to use current directory as destination use . (dot)

Putting file to hdfs:

\$hadoop fs -put <filename> <destination directory on hdfs>

Downloading file from hdfs to local machine:

\$hadoop fs -get <source file or directory on hdfs> <destination location>

Running job on hadoop:

hadoop jar <filename.jar> <filename without extension> <input data location on hdfs>
<output data location on hdfs>

Hadoop Streaming

Please have the following link before trying hadoop streaming job:

<http://www.michael-noll.com/tutorials/writing-an-hadoop-mapreduce-program-in-python/>

Streaming Commands:

hadoop jar /usr/hdp/3.0.0.0-1634/hadoop-mapreduce/hadoop-streaming.jar -file /<code directory>/<FileName.py> -mapper /<code directory>/<FileName.py> -input <input_dir>/*.txt -output <output_dir>