Note: We provide a summary below, but the official instructions on getting a Flux account, which is necessary to access the CAEN Hadoop cluster are also here: http://cac.engin.umich.edu/getting-started

Summary:

- 0. Get access
 - a. Get a cluster account by <u>filling this form</u> (see Appendix 1) and waiting for the confirmation from the administrator. Put my name in the advisor fields. https://www.engin.umich.edu/form/cacaccountapplication
 - b. Get MToken
 - i. Go to the MCard center in the Student Activity Building.515 East Jefferson. Hours: M-F, 8 am 5pm
 - ii. Don't sign in, but simply ask for an MToken.
 - iii. You will get an instruction to activate the MToken.
 - iv. You might choose to get a Software Token.

Copied and modified from the tutorial by CAEN lab

1. Login to Hadoop cluster

(Please treat \$ as commands in terminal and {xyz} as variables)

- a. Connect to UM VPN
- b. \$ ssh {uniqname}@flux-login.engin.umich.edu
 - i. Use your MToken (6 digits w/o space)
 - ii. User your UM password
- 2. Transfer data from your personal folder to the Hadoop HDFS
 - a. \$ hdfs dfs -copyFromLocal /home/{uniqname}/input.txt input.txt
- 3. Count words by Pig (online tutorial)

- a. Test locally first:
- b. \$ pig -x local wordcount.pig
- c. Once tested, run on the cluster\$ pig -Dmapreduce.job.queuename=si601f14 wordcount.pig
- d. See something like this:

```
2013-09-06 10:15:27,427 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLay er.MapReduceLauncher - Success!
```

- 4. show directory (a) and file contents (b)
 - a. \$ hdfs dfs -ls
 - b. \$ hdfs dfs -cat wordcount/part-r-00000
- 5. clean files in HDFS
 - a. \$ hdfs dfs -rm input.txt
 - b. \$ hdfs dfs -rmr wordcount

C.

- 6. Upload your dataset to Hadoop cluster
 - a. Open FTP client (ex: FileZilla, see Appendix 2)
 - b. URL: flux-xfer.engin.umich.edu

Protocol: SFTP

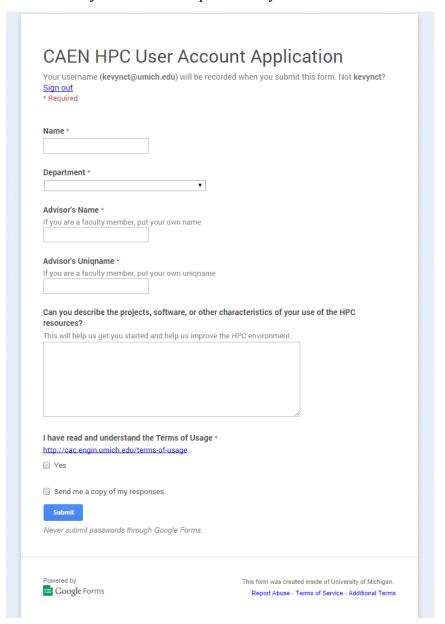
User: {uniqname}

Password: {password}

- c. Upload the file "input.txt" to your personal folder
- d. Upload the script "wordcount.pig" to your personal folder

Appendix 1

Note: Put my name and uniqname kevynct in the advisor fields.



Appendix 2

