

How to access the UM CAEN Hadoop cluster

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 [filling this form](#) (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 {username}@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/{username}/input.txt input.txt](#)

3. Count words by Pig ([online tutorial](#))

How to access the UM CAEN Hadoop cluster

- 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.mapReduceLayer.
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: [{username}](#)
Password: [{password}](#)
 - c. Upload the file “input.txt” to your personal folder
 - d. Upload the script “wordcount.pig” to your personal folder

How to access the UM CAEN Hadoop cluster

Appendix 1

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

CAEN HPC User Account Application

Your username (kevynct@umich.edu) will be recorded when you submit this form. Not kevynct?
[Sign out](#)
* Required

Name *

Department *

Advisor's Name *
If you are a faculty member, put your own name

Advisor's Unickname *
If you are a faculty member, put your own unickname


Can you describe the projects, software, or other characteristics of your use of the HPC resources?
This will help us get you started and help us improve the HPC environment.

I have read and understand the Terms of Usage *
<http://cac.engin.umich.edu/terms-of-usage>

☐ Yes

☐ Send me a copy of my responses.

Never submit passwords through Google Forms.

Powered by
 Google Forms

This form was created inside of University of Michigan.
[Report Abuse](#) - [Terms of Service](#) - [Additional Terms](#)

How to access the UM CAEN Hadoop cluster

Appendix 2

