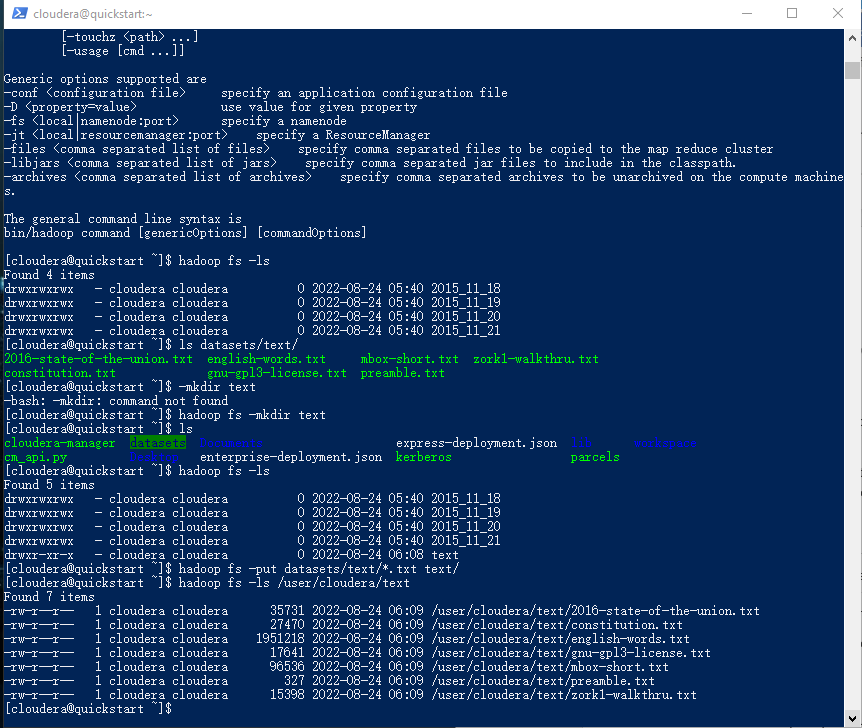
Name: Suihin Wong

SU email: swong18@syr.edu

1.

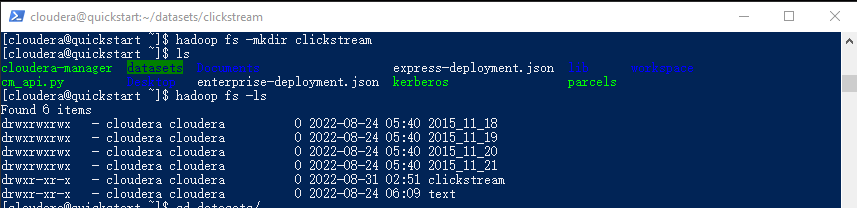
Hadoop fs -mkdir text

Hadoop fs -put datasets/text/\*.txt text/



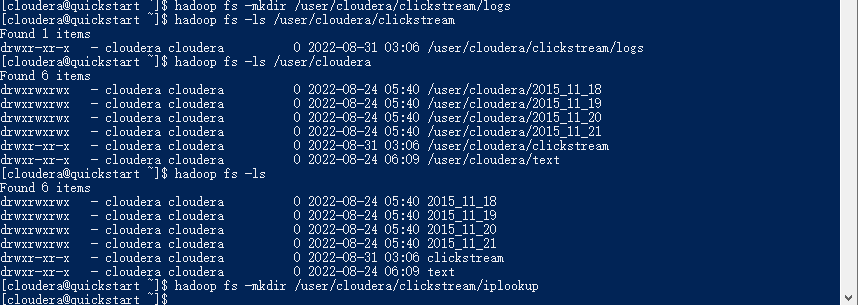
2.

Hadoop fs -mkdir clickstream



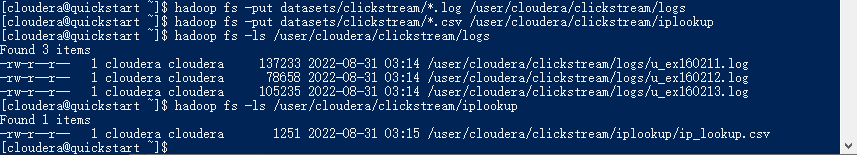
Hadoop fs -mkdir /user/cloudera/clickstream/logs

Hadoop fs -mkdir /user/cloudera/clickstream/iplookup



Hadoop fs -put datasets/clickstream/\*.log /user/cloudera/clickstream/logs

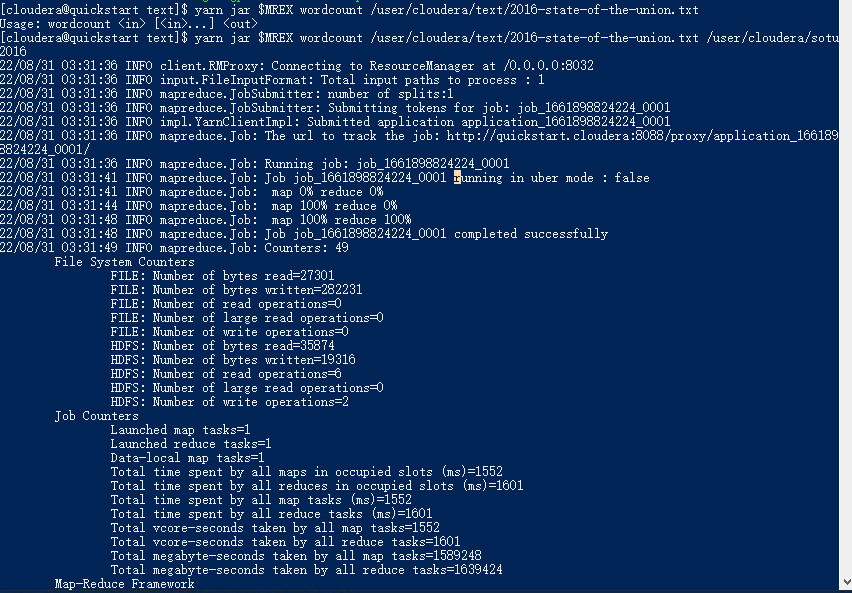
Hadoop fs -put datasets/clickstream/\*.csv /user/cloudera/clickstream/iplookup

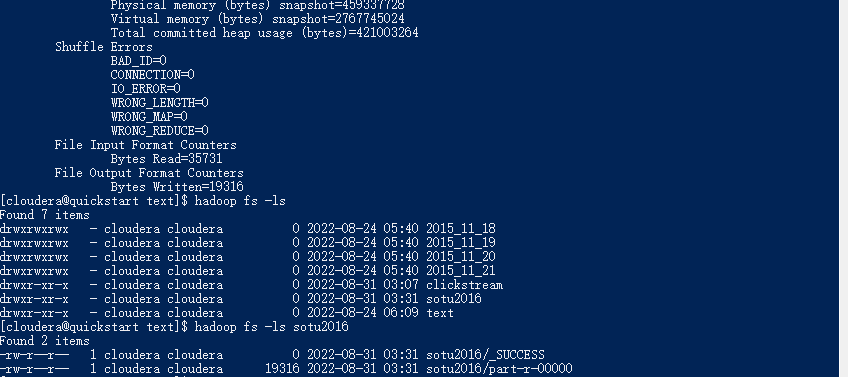


3.

yarn jar $MREX wordcount /user/cloudera/text/2016-state-of-the-union.txt /user/cloudera/sotu2016

hadoop fs -cat sotu2016/\*







“are” appears 27 times.

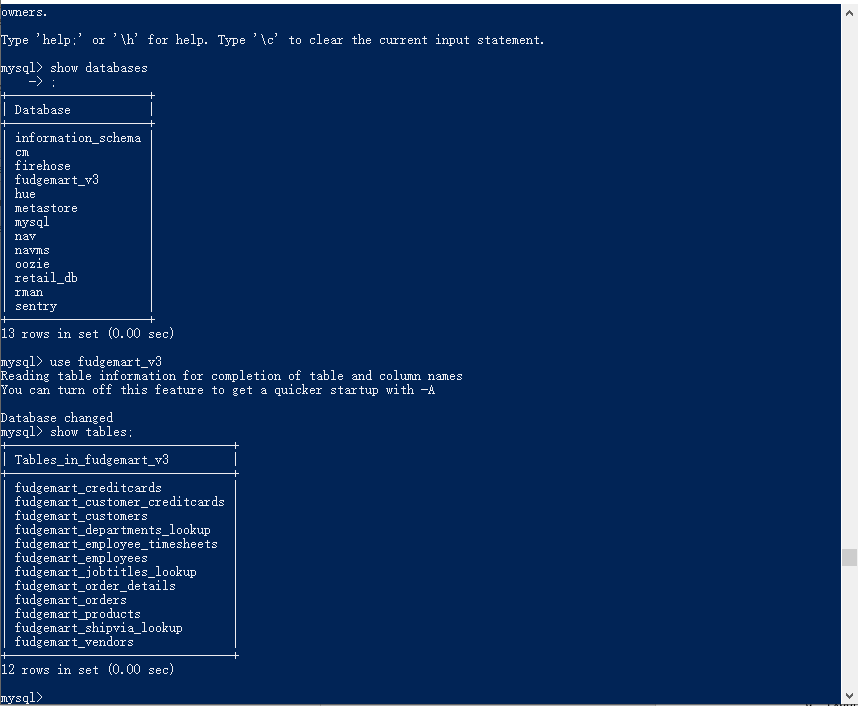
4.

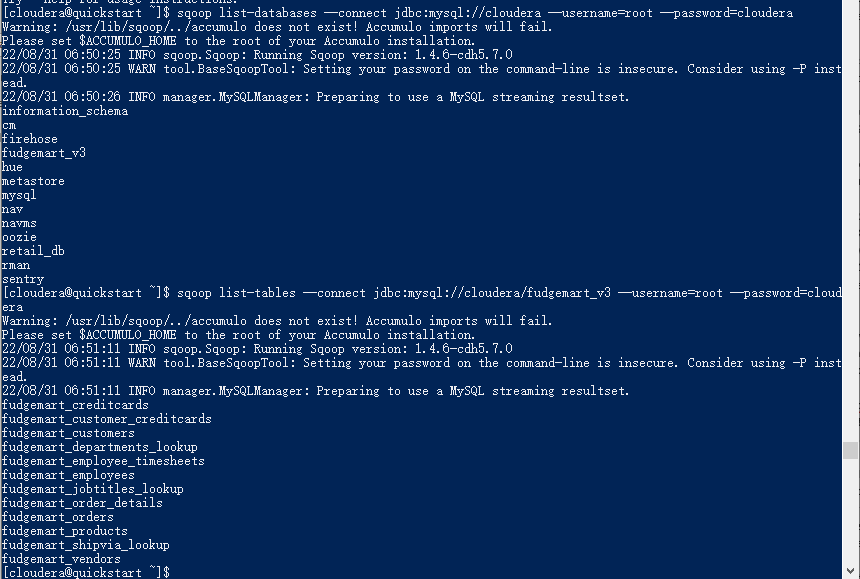
Use fudgemart\_v3

Show tables;

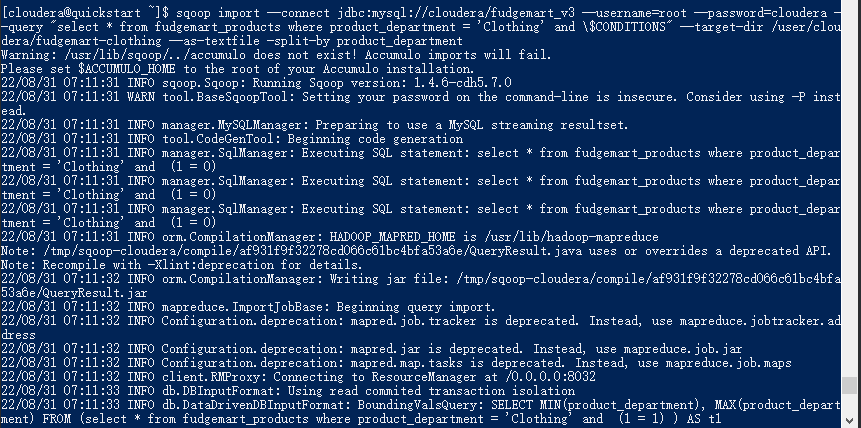
sqoop list-databases --connect jdbc:mysql://cloudera --username=root --password=cloudera

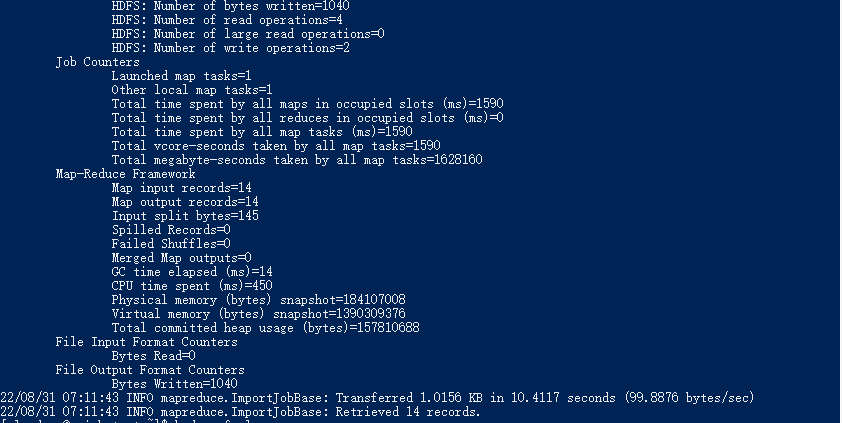
sqoop list-tables --connect jdbc:mysql://cloudera/fudgemart\_v3 --username=root --password=cloudera

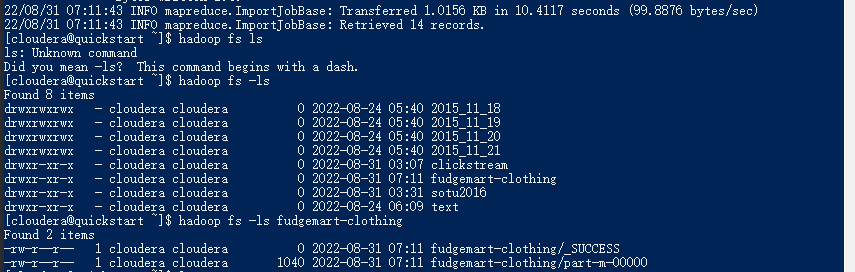




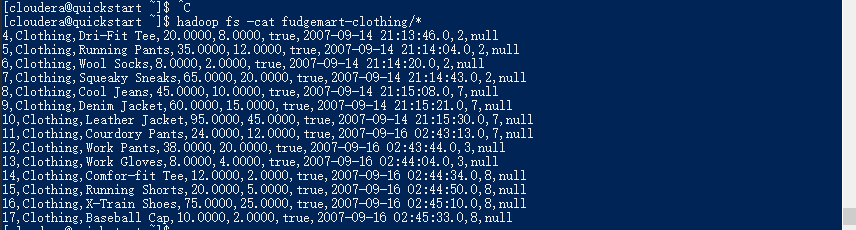
sqoop import --connect jdbc:mysql://cloudera/fudgemart\_v3 --username=root --password=cloudera --query "select \* from fudgemart\_products where product\_department = 'Clothing' and \$CONDITIONS" --target-dir /user/cloudera/fudgemart-clothing --as-textfile -split-by product\_department







Hadoop fs -cat fudgemart-clothing/\*



5.

Hadoop fs -mkdir tweets

Hadoop fs -ls

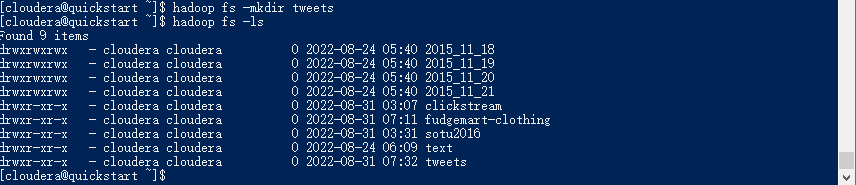
hadoop fs -put datasets/tweets/sample-tweet-stream.psv tweets/

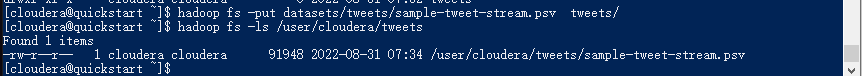
hadoop fs -ls /user/cloudera/tweets

create databse twitter

Use twitter

create table tweets (id int, timestamp TIMESTAMP, date\_time DATETIME, username VARCHAR(20), tweet\_text VARCHAR(50))





sqoop export --connect jdbc:mysql://cloudera/twitter --username=root --password=cloudera --table tweets --export-dir /user/cloudera/tweets --input-fields-terminated-by '\t'

