STAT 4410/8416 Homework 6

VanSteenbergen Nicolaas

Due on Dec 9, 2018

1.

a)

INPUT
[cloudera@quickstart ~]$ hadoop fs -mkdir wordcount
[cloudera@quickstart ~]$ hadoop fs -mkdir wordcount/input
[cloudera@quickstart ~]$ hadoop fs -ls

OUTPUT
Found 1 items drwxr-xr-x - cloudera cloudera 0 2018-12-07 10:32 wordcount

b)

INPUT
hadoop fs -put words.txt wordcount/input

OUTPUT
hadoop fs -ls wordcount/input
Found 1 items -rw-r–r– 1 cloudera cloudera 54601 2018-12-07 10:10 wordcount/input/words.txt

c)

INPUT
[cloudera@quickstart ~]$ mkdir wordcount_classes

[cloudera@quickstart ~]$ ls

OUTPUT
[cloudera@quickstart ~]$ ls cloudera-manager kerberos Videos cm_api.py lib wordcount_classes Desktop Music WordCount.java Documents parcels words.txt Downloads Pictures words.txt~ eclipse pig_1544127357145.log workspace enterprise-deployment.json Public express-deployment.json Templates

d)

INPUT
javac -cp /usr/lib/hadoop/client-0.20/* -d wordcount_classes WordCount.java

jar -cvf wordcount.jar -C wordcount_classes/ .

hadoop jar wordcount.jar org.myorg.WordCount wordcount/input wordcount/output

OUTPUT
[cloudera@quickstart ~]$ hadoop fs -ls wordcount/output

Found 2 items -rw-r–r– 1 cloudera cloudera 0 2018-12-07 10:50 wordcount/output/_SUCCESS -rw-r–r– 1 cloudera cloudera 28935 2018-12-07 10:50 wordcount/output/part-00000

2.

a)

[cloudera@quickstart ~]$ pig

INPUT
grunt> ls wordcount/output

OUTPUT
hdfs://quickstart.cloudera:8020/user/cloudera/wordcount/output/_SUCCESS 0 hdfs://quickstart.cloudera:8020/user/cloudera
00000 28935

b)

INPUT

dat = LOAD 'wordcount/output/part-00000' AS (words:chararray, count:int);

datLimit10 = LIMIT dat 10;

DUMP datLimit10;

OUTPUT
("Be,1) (”Are,1) ("Cut,1) (”Don't,2) ("Aren't,1) (”Dell,",1) (”Except,1) ("Blessed,1) (”Coffee¿',1) (”Dilling-
ham",1)

c)

INPUT
dat = LOAD 'wordcount/output/part-00000' AS (words:chararray, count:int);

sorted = ORDER dat BY count DESC;

limitDat = LIMIT sorted 10;

DUMP limitDat;

OUTPUT
(the,434) (and,325) (to,280) (of,265) (a,209) (in,173) (I,153) (that,150) (is,103) (for,94)