LIU,HONGYANG

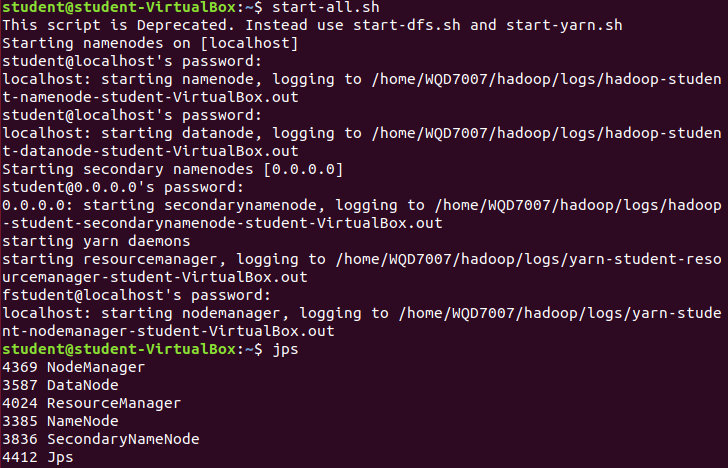
Author: LIU,HONGYANG

Matrix Number: 17201091/1

### Before the Lab Test:

##### Start the Hadoop and check the daemons:

start-all.sh  
jps



##### DownLoad dabasets:



### LabTest:

##### Part 1:

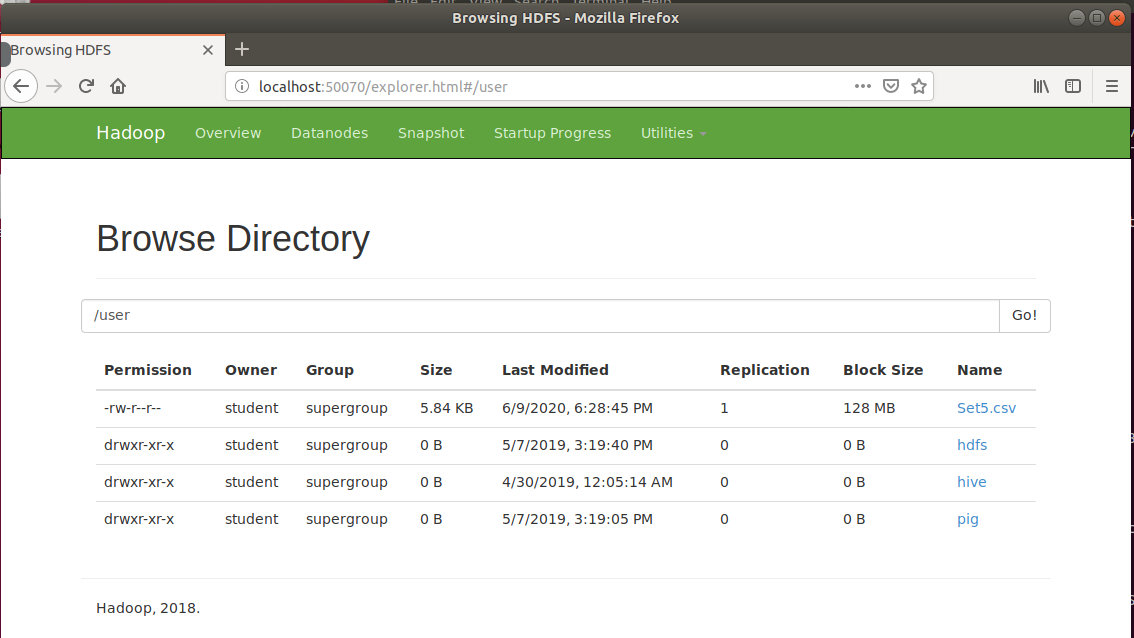
1.Import the downloaded dataset to HDFS

code:

hdfs dfs -put ~/Desktop/Set5.csv /user/Set5.csv



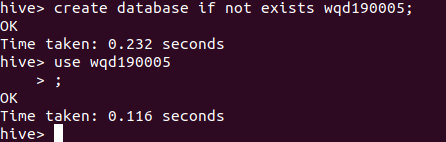
Results:



2.By using Hive or Pig, identify 5 rows of data that have the

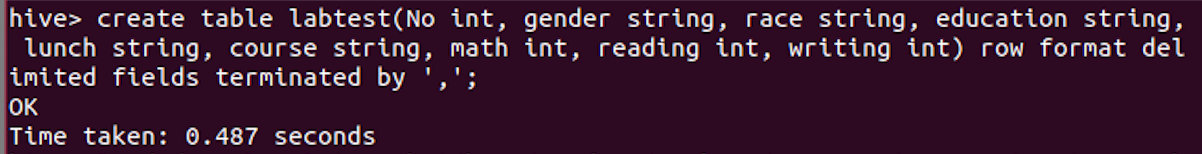
1. highest reading score.
2. lowest CGPA.

Create database:

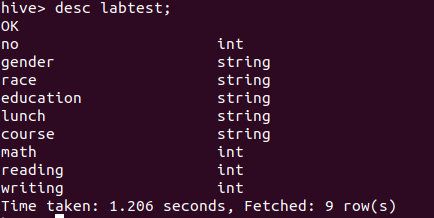


Create table:

create table labtest(No int, gender string, race string, education string, lunch string, course string, math int, reading int, writing int) row format delimited fields terminated by ',';

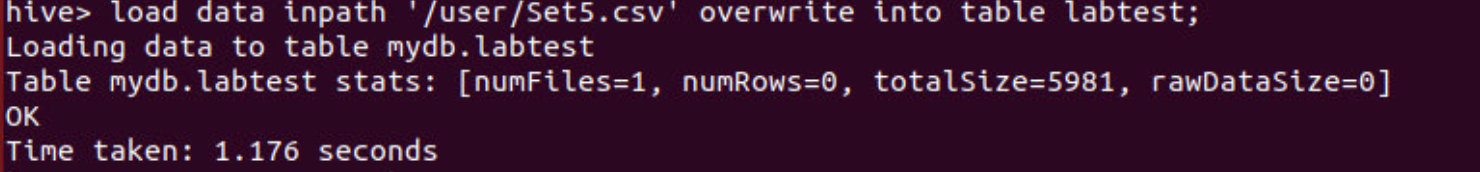


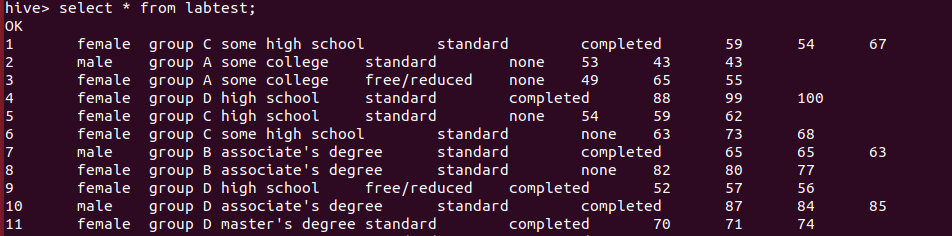
desc labtest;



load data from hdfs to hive:

load data inpath '/user/Set5.csv' overwrite into table labtest;

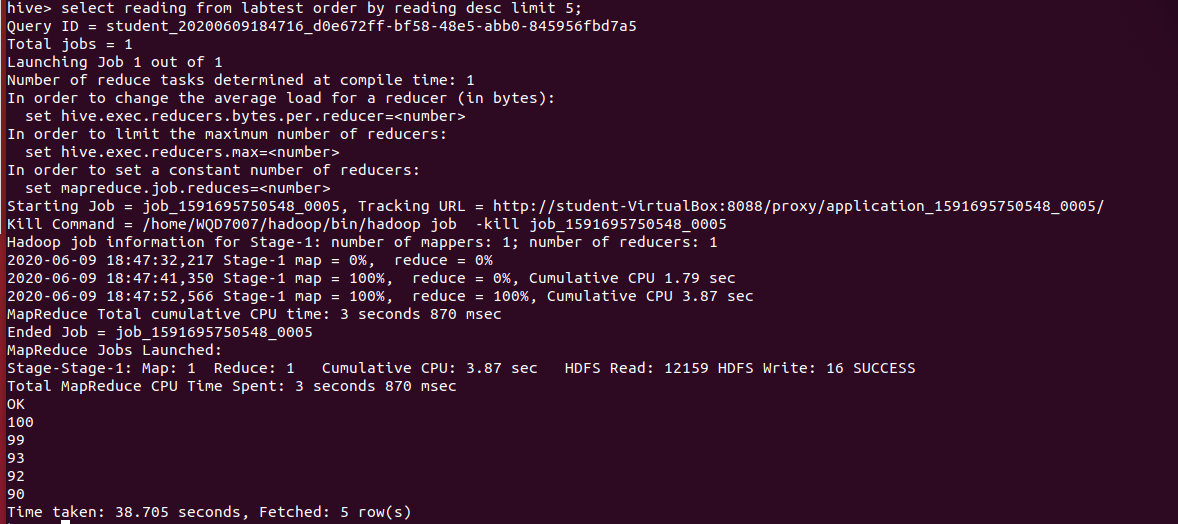




###### Answer 1:

**highest reading score(5 rows):**

select reading from labtest order by reading desc limit 5;

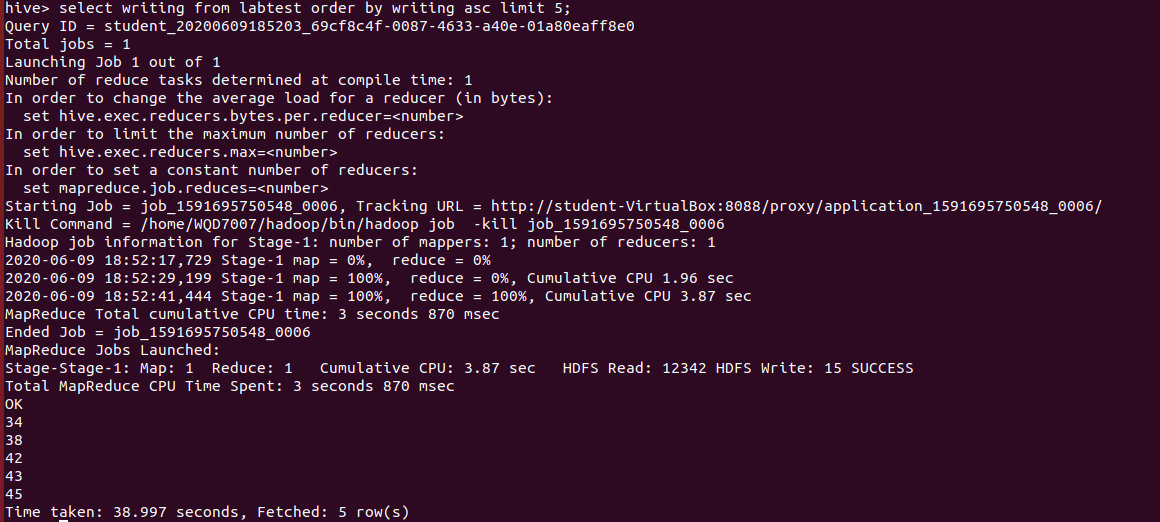


They are 100, 99, 93, 92, 90

###### Answer 2:

**lowest writing score(5 rows):**

select writing from labtest order by writing asc limit 5;

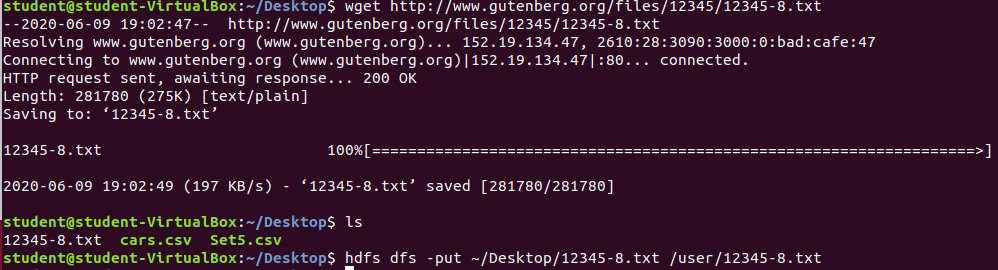


They are 34, 38, 42,43,45

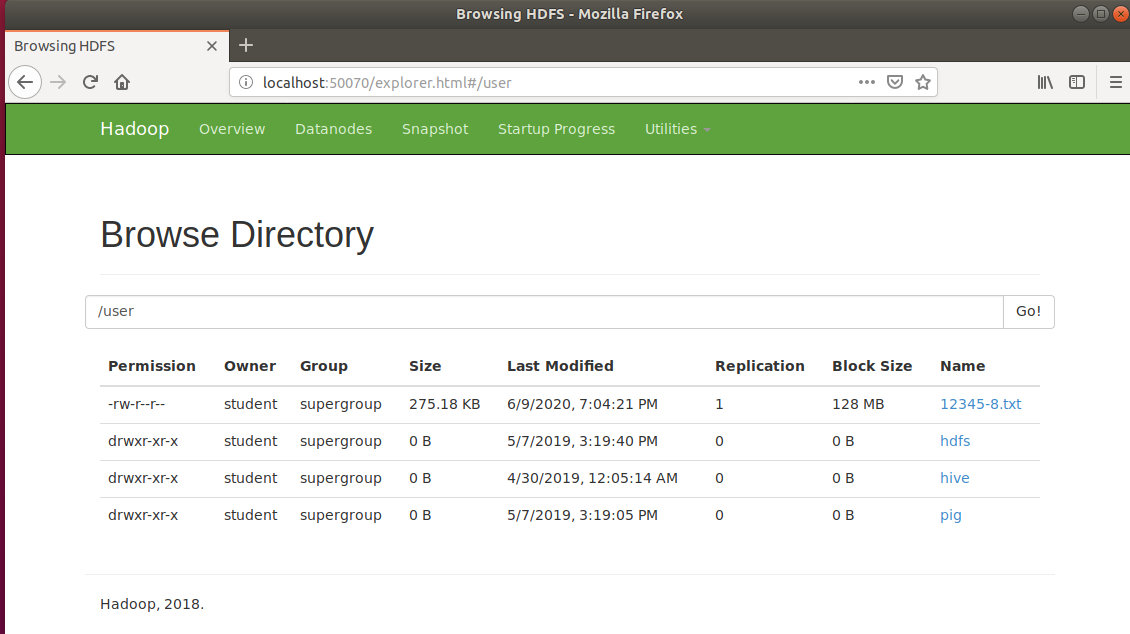
##### Part 2:

download file and upload it to hdfs

wget http://www.gutenberg.org/files/12345/12345-8.txt  
hdfs dfs -put ~/Desktop/12345-8.txt /user/12345-8.txt

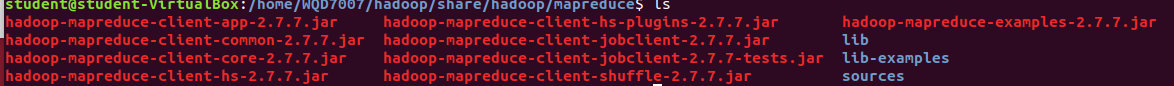


###### Answer 1:



##### Answer 2:

enter the file example file: hadoop-mapreduce-examples-2.7.7.jar

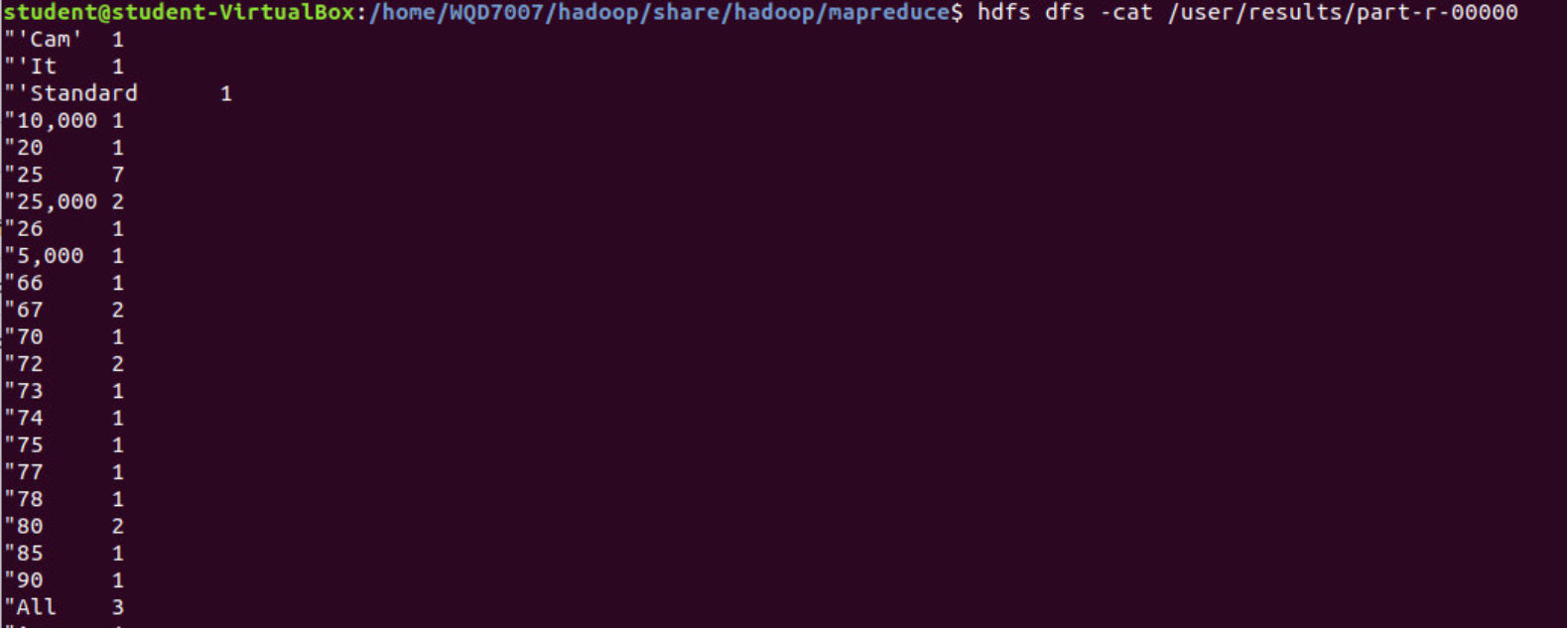


run shell:

hadoop jar hadoop-mapreduce-examples-2.7.7.jar wordcount /user/12345-8.txt /user/results

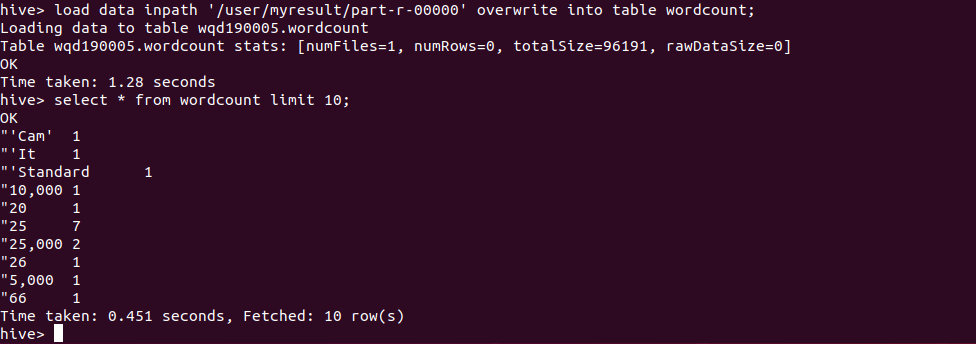
check results:

hdfs dfs -cat /user/results/part-r-0000

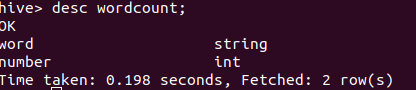


##### Answer 3:

create table wordcount(word string, number int) row format delimited fields terminated by '\t';

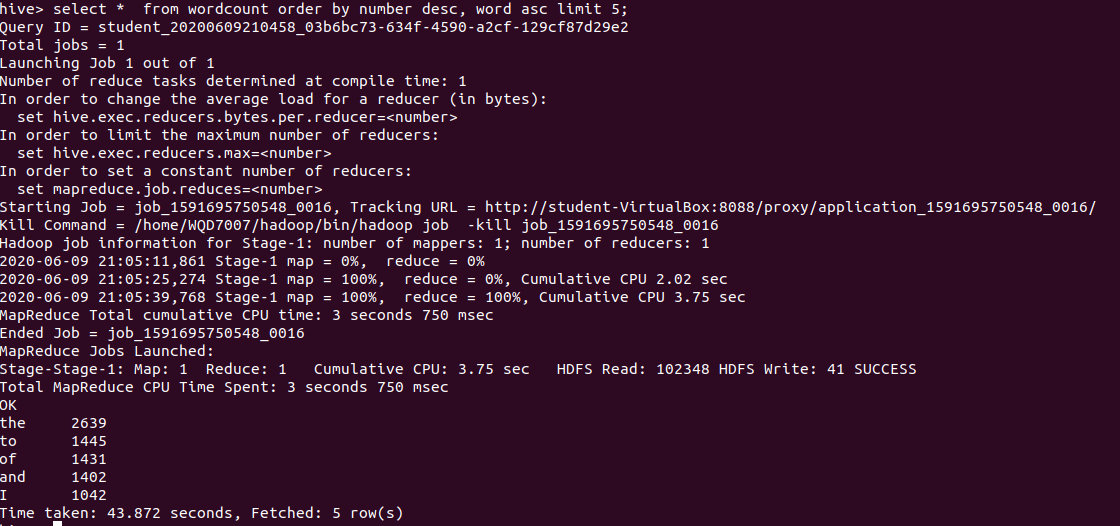


desc wordcount;



a

select \* from wordcount order by number desc, word asc limit 5;



b

select \* from wordcount where number=5 order by word desc limit 5;

