Instructions for Uber Pickup Data Analysis

CS555: Big Data Computing Lab

10 August 2021

Dispatching Base Number: The NYC Taxi & Limousine Commission (TLC) base company code of the base that dispatched the Uber

Active Vehicles: shows the number of active Uber vehicles for a particular date and company (base)

Trips: is the number of trips for a particular base and date.

The base codes are for the following Uber bases:

B02512 : Unter, **B02598** : Hinter, **B02617** : Weiter, **B02682** : Schmecken, **B02764** : Danach-NY , **B02765** : Grun, **B02835** : Dreist, **B02836** : Drinnen

Problem Statement: **Finds the days on which each basement has more trips.**

1. create project directory : (uber)
    a. Browse to home directory

        cd /home/iitp

    b. Create project directory

        mkdir uber

2. Create source file: (**Basement_trips.java**)
    a. Browse into newly created uber directory

        cd /home/iitp/uber

    b. Create file

        nano Basement_trips.java

        paste the lines from the source code provided

        Note:- **To save file: Press- CTRL + o followed by Enter button**

To Exit Press:- CTRL + x from nano editor

3. Create input directory (inputdata) for input files

   cd /home/iitp/uber

   mkdir inputdata

4. Copy the input file (uberdata.txt) into inputdata folder

   cd /home/iitp/uber/inputdata

   nano uberdata.txt

   paste the lines from the provided input file

5. Start all hadoop services
   a. Browse to hadoop installation sbin sub-directory

      cd /home/iitp/hadoop-2.6.0/sbin

   b. start all services

      ./start-all.sh

      Note:- Enter password when prompted

6. Create input directory on HDFS
   a. browse to hadoop installation bin folder

      cd /home/iitp/hadoop-2.6.0/bin

   b. create directory (**uber**)

      ./hadoop fs -mkdir /uber

   c. create subdirectory (inputdata) inside uber on HDFS

      ./hadoop fs -mkdir /uber/inputdata

7. Copy the input text file from local directory to HDFS
   a. browse to hadoop installation bin folder

      cd /home/iitp/hadoop-2.6.0/bin

   b. Copy from Local

./hadoop dfs -put /home/iitp/uber/inputdata/uberdata.txt /uber/inputdata

8. Compile the Source Code
    a. export the Hadoop classpath

    export
    HADOOP_CLASSPATH=/usr/lib/jvm/java-1.8.0-openjdk-amd64/lib/tools.jar

    b. browse to bin folder of hadoop installation

    cd /home/iitp/hadoop-2.6.0/bin

    c. Compile

    ./hadoop com.sun.tools.javac.Main /home/iitp/uber/Basement_trips.java

9. Create Jar file
    a. Browse to uber directory on your VM

    cd /home/iitp/uber

    jar cf basement_trips.jar Basement_trips*.class

10. Running the program
    a. browse to the bin directory of hadoop installation

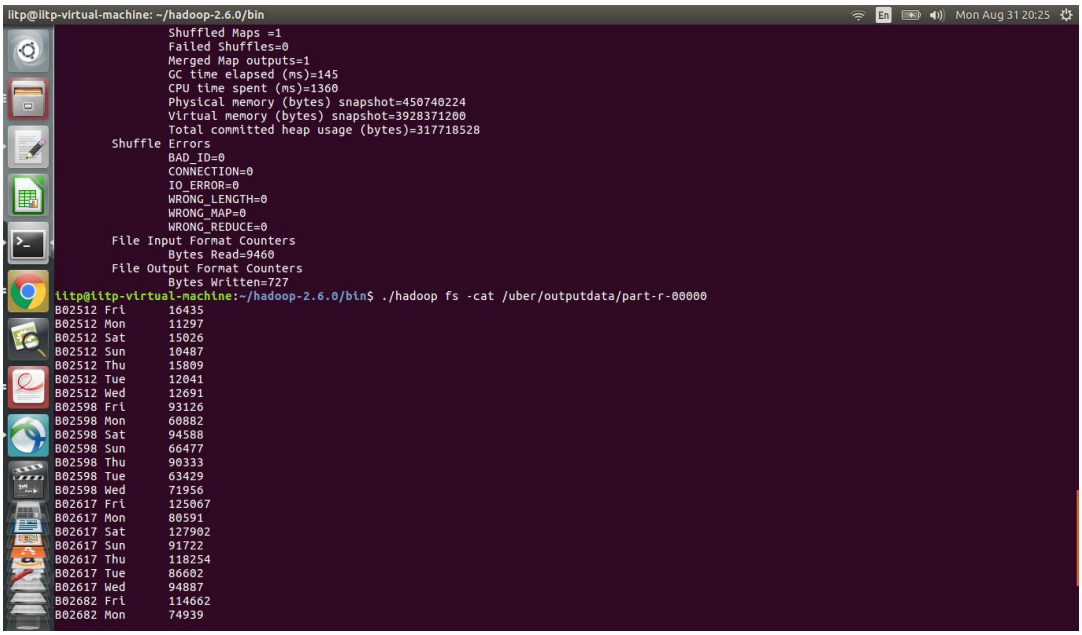
    cd /home/iitp/hadoop-2.6.0/bin

    b. Running in terminal

    ./hadoop jar /home/iitp/uber/basement_trips.jar Basement_trips
    /uber/inputdata/ /uber/outputdata

    c. Finding outputs

    ./hadoop fs -cat /uber/outputdata/part-r-00000


iitp@iitp-virtual-machine: ~/hadoop-2.6.0/bin
Mon Aug 31 20:25
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=145
CPU time spent (ms)=1360
Physical memory (bytes) snapshot=450740224
Virtual memory (bytes) snapshot=3928371200
Total committed heap usage (bytes)=317718528
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=9460
File Output Format Counters
Bytes Written=727
iitp@iitp-virtual-machine:~/hadoop-2.6.0/bin$ ./hadoop fs -cat /uber/outputdata/part-r-00000
B02512 Fri 16435
B02512 Mon 11297
B02512 Sat 15026
B02512 Sun 10487
B02512 Thu 15809
B02512 Tue 12041
B02512 Wed 12691
B02598 Fri 93126
B02598 Mon 60882
B02598 Sat 94588
B02598 Sun 66477
B02598 Thu 90333
B02598 Tue 63429
B02598 Wed 71956
B02617 Fri 125067
B02617 Mon 80591
B02617 Sat 127902
B02617 Sun 91722
B02617 Thu 118254
B02617 Tue 86602
B02617 Wed 94887
B02682 Fri 114662
B02682 Mon 74939