Instructions for Netflix data Execution

CS555: Big Data Computing Lab

24th August 2021

1. Install Maven
    a. Open the terminal and type:

       sudo apt-get install maven

2. create project directory : (netflix)
    a. Browse to home directory

       cd /home/iitp

    b. Create project directory

       mkdir netflix

3. Create source files:
    a. Browse into netflix directory

       cd netflix

    b. Create new folder **src** inside netflix directory

       mkdir src

    c. Browse into src directory

       cd src

    d. Create new folder **main** inside src directory

       mkdir main

    e. Browse into **main** directory

       cd main

    f. Create new folder **java** inside main directory

       mkdir java

    g. Browse into java directory

cd java

h. Create source files within java directory
   i. **nano CooccurrenceMatrixGenerator.java**
   ii. **nano CooccurrenceMatrixNormalization.java**
   iii. **nano Driver.java**
   iv. **nano MatrixCellMultiplication.java**
   v. **nano MatrixCellSum.java**
   vi. **nano RatingDataReader.java**
   vii. **nano UserRatingAveraging.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

i. Browse into netflix directory

   cd /home/iitp/netflix

j. Create pom.xml file inside netflix directory

   nano pom.xml

   paste the lines from the xml file provided

k. Compile the java source files

   mvn clean && mvn compile && maven package

4. Create input directory (inputdata) for input files

   cd /home/iitp/netflix

   mkdir inputdata

5. Copy the input file (input.txt) into inputdata folder

   cd /home/iitp/netflix/inputdata

   nano input.txt

   paste the lines from the provided input file

6. 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

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

        cd /home/iitp/hadoop-2.6.0/bin

    b. create directory (**netflix**)

        ./hadoop fs -mkdir /netflix

    c. create directory (**inputdata**) inside netflix directory on HDFS

        ./hadoop fs -mkdir /netflix/inputdata

8. 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/netflix/inputdata/input.txt /netflix/inputdata

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

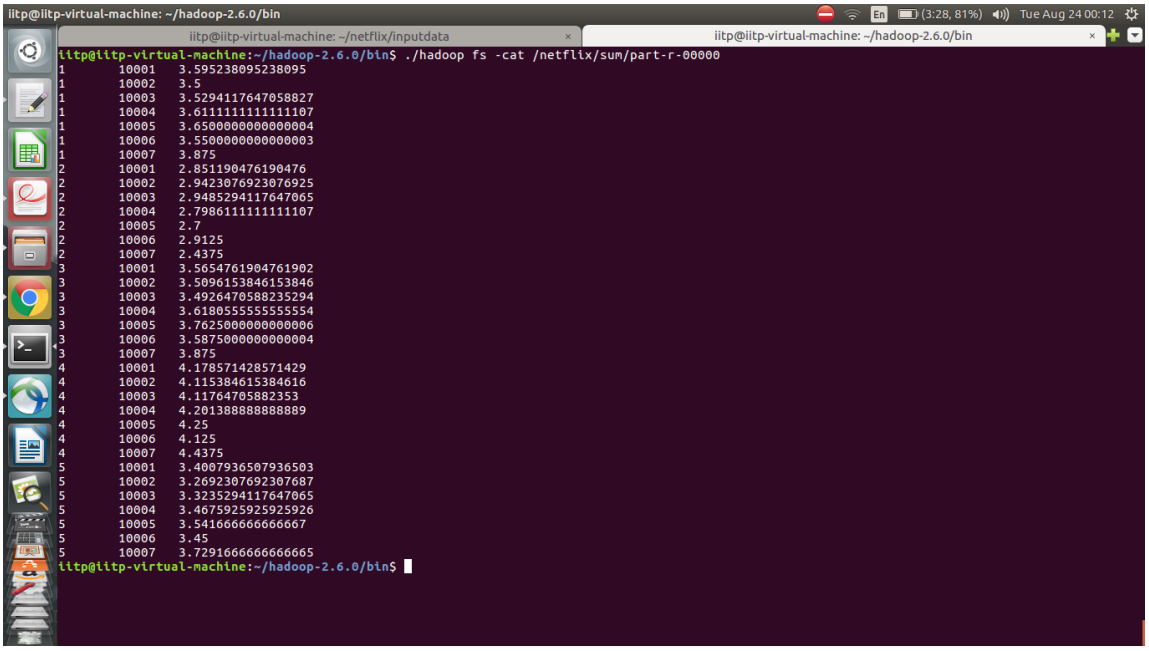
        cd /home/iitp/hadoop-2.6.0/bin

    b. Running in terminal

        ./hadoop jar /home/iitp/netflix/target/NetflixRec-1.0.0.jar Driver /netflix/inputdata /netflix/userrating /netflix/cooccurencegenerator /netflix/cooccurencenormalize /netflix/useraveragerating /netflix/multiplication /netflix/sum

    c. Finding outputs

        ./hadoop fs -cat /netflix/sum/part-r-00000


iitp@iitp-virtual-machine: ~/hadoop-2.6.0/bin
Tue Aug 24 00:12
iitp@iitp-virtual-machine: ~/netflix/inputdata
iitp@iitp-virtual-machine: ~/hadoop-2.6.0/bin
iitp@iitp-virtual-machine:~/hadoop-2.6.0/bin$ ./hadoop fs -cat /netflix/sum/part-r-00000
1 10001 3.595238095238095
1 10002 3.5
1 10003 3.5294117647058827
1 10004 3.6111111111111107
1 10005 3.6500000000000004
1 10006 3.5500000000000003
1 10007 3.875
2 10001 2.851190476190476
2 10002 2.9423076923076925
2 10003 2.9485294117647065
2 10004 2.7986111111111107
2 10005 2.7
2 10006 2.9125
2 10007 2.4375
3 10001 3.5654761904761902
3 10002 3.5096153846153846
3 10003 3.4926470588235294
3 10004 3.6180555555555554
3 10005 3.7625000000000006
3 10006 3.5875000000000004
3 10007 3.875
4 10001 4.178571428571429
4 10002 4.115384615384616
4 10003 4.11764705882353
4 10004 4.201388888888889
4 10005 4.25
4 10006 4.125
4 10007 4.4375
5 10001 3.4007936507936503
5 10002 3.2692307692307687
5 10003 3.3235294117647065
5 10004 3.4675925925925926
5 10005 3.541666666666667
5 10006 3.45
5 10007 3.7291666666666665
iitp@iitp-virtual-machine:~/hadoop-2.6.0/bin$