**Instructions for Netflix data Execution**

CS555: Big Data Computing Lab

horizontal line

# **24th August 2021**

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

**sudo apt-get install maven**

1. create project directory : (netflix)
   1. Browse to home directory

**cd /home/iitp**

* 1. Create project directory

**mkdir netflix**

1. Create source files:
   1. Browse into netflix directory

**cd netflix**

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

**mkdir src**

* 1. Browse into src directory

**cd src**

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

**mkdir main**

* 1. Browse into **main** directory

**cd main**

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

**mkdir java**

* 1. Browse into java directory

**cd java**

* 1. Create source files within java directory
     1. **nano CooccurrenceMatrixGenerator.java**
     2. **nano CooccurrenceMatrixNormalization.java**
     3. **nano Driver.java**
     4. **nano MatrixCellMultiplication.java**
     5. **nano MatrixCellSum.java**
     6. **nano RatingDataReader.java**
     7. **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**

* 1. Browse into netflix directory

**cd /home/iitp/netflix**

* 1. Create pom.xml file inside netflix directory

**nano pom.xml**

paste the lines from the xml file provided

* 1. Compile the java source files

**mvn clean && mvn compile && maven package**

1. Create input directory (inputdata) for input files

cd /home/iitp/netflix

**mkdir** **inputdata**

1. 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

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

**cd /home/iitp/hadoop-2.6.0/sbin**

* 1. start all services

**./start-all.sh**

Note:- Enter password when prompted

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

**cd /home/iitp/hadoop-2.6.0/bin**

* 1. create directory (**netflix**)

**./hadoop fs -mkdir /netflix**

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

**./hadoop fs -mkdir /netflix/inputdata**

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

**cd /home/iitp/hadoop-2.6.0/bin**

* 1. Copy from Local

**./hadoop dfs -put /home/iitp/netflix/inputdata/input.txt /netflix/inputdata**

1. Running the program
   1. browse to the bin directory of hadoop installation

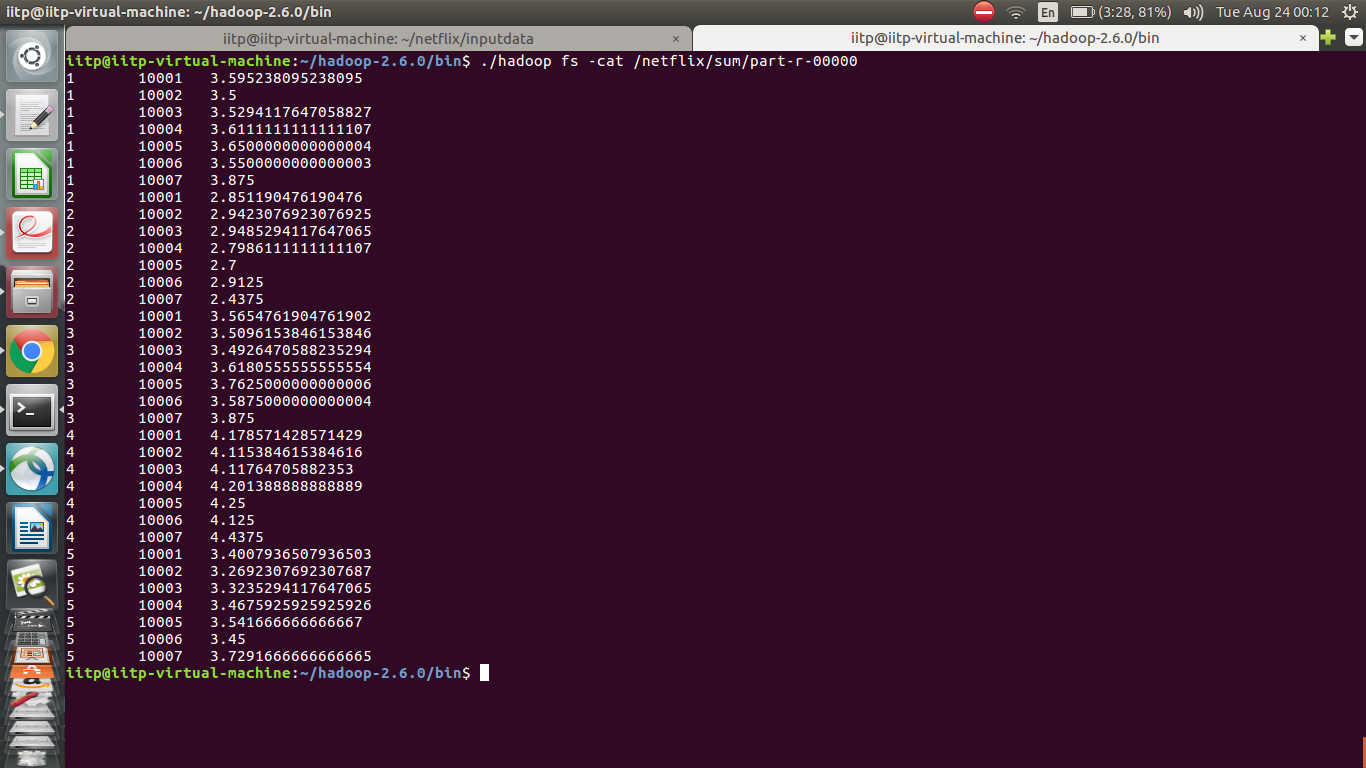
**cd /home/iitp/hadoop-2.6.0/bin**

* 1. 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**

* 1. Finding outputs

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

****