**Instructions for Wikipedia dataset Execution**

CS555: Big Data Computing Lab

horizontal line

# **16th August 2021**

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

**sudo apt-get install maven**

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

**cd /home/iitp**

* 1. Create project directory

**mkdir wiki**

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

**cd wiki**

* 1. Create new folder **src** inside wiki 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 DriverClass.java**
     2. **nano InvertedIndex.java**
     3. **nano LinkGraph.java**
     4. **nano PageRank.java**
     5. **nano PageRankSorting.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 wiki directory

**cd /home/iitp/wiki**

* 1. Create pom.xml file inside pagerank 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/wiki

**mkdir** **inputdata**

1. Copy the input file (input.txt) into inputdata folder on VM from local machine

**$ scp <path-on-local-system> <user-credentials-on-assigned-VM:path-to-copy-into-on-VM>**

**$ scp /home/iitp/CS-555-Lab-2021/Lab-III/wiki/inputdata/input.txt iitp@172.16.27.166:/home/iitp/wiki/inputdata/**

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 directories on HDFS
   1. browse to hadoop installation bin folder

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

* 1. create directory (**wiki**)

**./hadoop fs -mkdir /wiki**

* 1. create sub-directory **inputdata** within wiki directory

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

* 1. create directory (**iprank**)

**./hadoop fs -mkdir /wiki/iprank**

* 1. create directory (**prsort**)

**./hadoop fs -mkdir /wiki/prsort**

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 input file from Local-VM on to HDFS

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

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

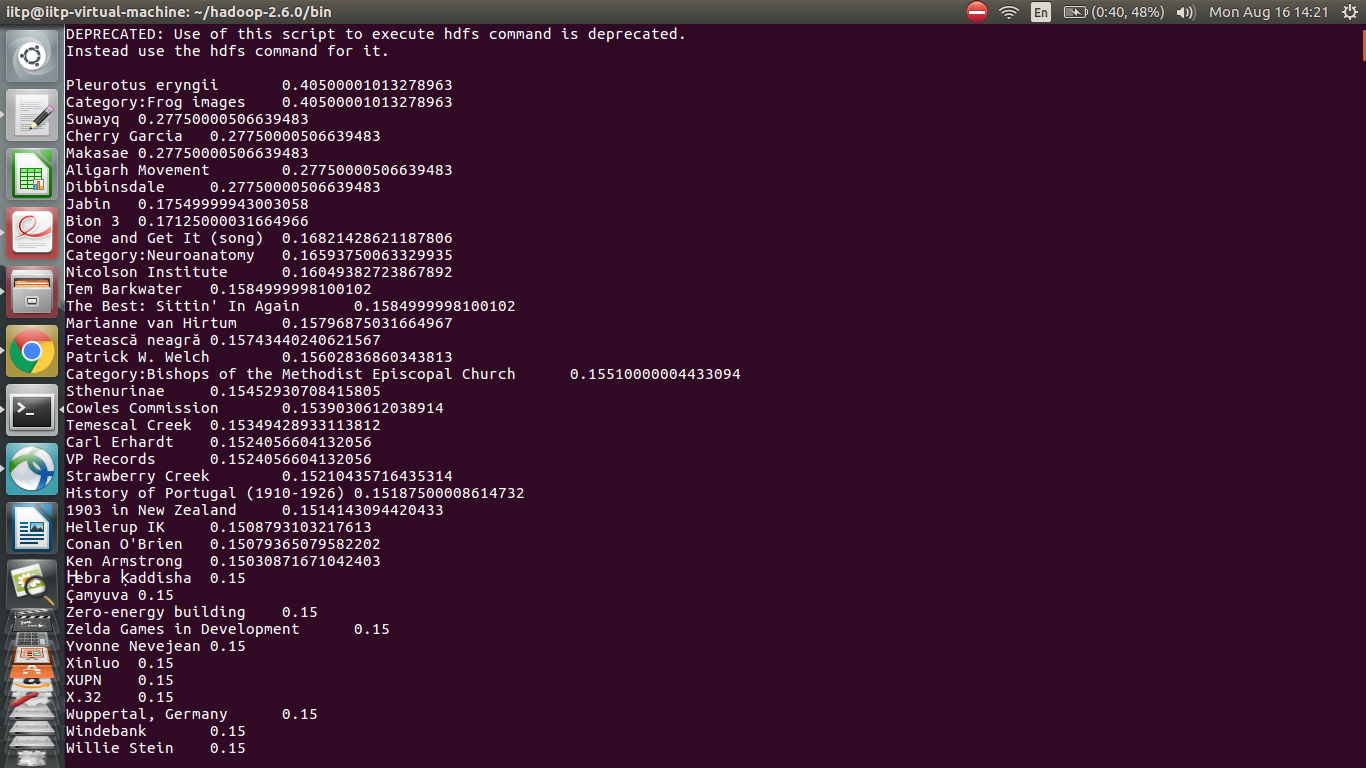
**cd /home/iitp/hadoop-2.6.0/bin**

* 1. Running in terminal

**./hadoop jar /home/iitp/wiki/target/WP-1.0.0.jar DriverClass /wiki/inputdata /wiki/iprank/outputdata /wiki/outputdata /wiki/prsort/outputdata**

* 1. Finding outputs

**./hadoop fs -cat /wiki/prsort/outputdata/part-r-00000**

****