

Instructions for PageRank Execution

CS555: Big Data Computing Lab

9th August 2021

1. Install Maven
 - a. Open the terminal and type:
sudo apt-get install maven
2. create project directory : (pagerank)
 - a. Browse to home directory
cd /home/iitp
 - b. Create project directory
mkdir pagerank
3. Create source files:
 - a. Browse into pagerank directory
cd pagerank
 - b. Create new folder **src** inside pagerank 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 PRNode.java**
 - ii. **nano PageRank.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 pagerank directory

cd /home/iitp/pagerank

- j. Create pom.xml file inside pagerank 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/pagerank

mkdir inputdata

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

cd /home/iitp/pagerank/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 (page)

```
./hadoop fs -mkdir /page
```

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/pagerank/inputdata/input.txt /page
```

9. Running the program

- a. browse to the bin directory of hadoop installation

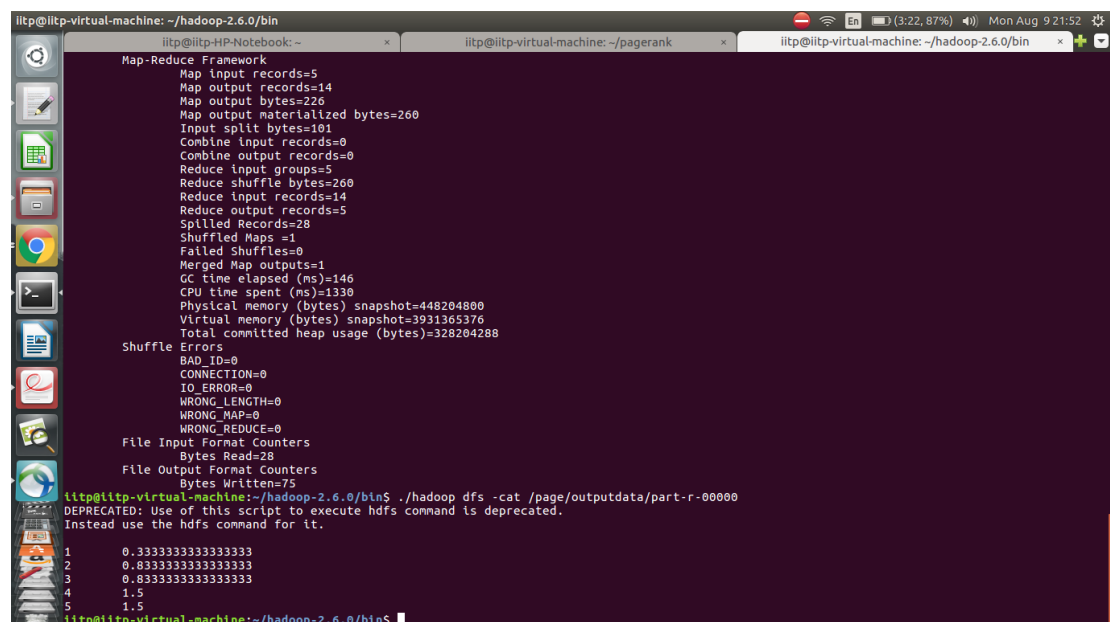
```
cd /home/iitp/hadoop-2.6.0/bin
```

- b. Running in terminal

```
./hadoop jar /home/iitp/pagerank/target/PR-1.0.0.jar PageRank /page/  
/page/outputdata
```

- c. Finding outputs

```
./hadoop fs -cat /page/outputdata/part-r-000000
```



```
iitp@iitp-virtual-machine: ~/hadoop-2.6.0/bin
iitp@iitp-HP-Notebook: ~
iitp@iitp-virtual-machine: ~/pagerank
iitp@iitp-virtual-machine: ~/hadoop-2.6.0/bin

Map-Reduce Framework
  Map input records=5
  Map output records=14
  Map output bytes=226
  Map output materialized bytes=260
  Input split bytes=101
  Combine input records=0
  Combine output records=0
  Reduce input groups=5
  Reduce shuffle bytes=260
  Reduce input records=14
  Reduce output records=5
  Spilled Records=28
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=146
  CPU time spent (ms)=1330
  Physical memory (bytes) snapshot=448204800
  Virtual memory (bytes) snapshot=3931365376
  Total committed heap usage (bytes)=328204288

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=28
File Output Format Counters
  Bytes Written=75

iitp@iitp-virtual-machine:~/hadoop-2.6.0/bin$ ./hadoop dfs -cat /page/outputdata/part-r-000000
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
1 0.3333333333333333
2 0.8333333333333333
3 0.8333333333333333
4 1.5
5 1.5
iitp@iitp-virtual-machine:~/hadoop-2.6.0/bin$
```