

# Instructions Reduce Side Join Execution

CS555: Big Data Computing Lab

---

7<sup>th</sup> September 2021

1. Install Maven

- a. Open the terminal and type:

**sudo apt-get install maven**

2. create project directory : (ReduceSideJoin)

- a. Browse to home directory

**cd /home/iitp**

- b. Create project directory

**mkdir ReduceSideJoin**

3. Copy files with directory structure from local system to VM:

- a. Copy the source code files directory

**scp -r /home/iitp/CS-555-Lab-2021/Lab-VI/ReduceSideJoin/src/  
iitp@172.16.27.166:/home/iitp/ReduceSideJoin**

- b. Copy the Inputdata directory with input file

**scp -r /home/iitp/CS-555-Lab-2021/Lab-VI/ReduceSideJoin/Inputdata/  
iitp@172.16.27.166:/home/iitp/ReduceSideJoin**

- c. Copy the buildfile pom.xml

**scp -r /home/iitp/CS-555-Lab-2021/Lab-VI/ReduceSideJoin/pom.xml  
iitp@172.16.27.166:/home/iitp/ReduceSideJoin**

4. Compile the java source files

- a. Browse to the ReduceSideJoin directory

**cd ReduceSideJoin**

- b. Compile source files and create jar file

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

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 (**ReduceSideJoin**)

**./hadoop fs -mkdir ReduceSideJoin**

- c. create directory (**Inputdata**) inside ReduceSideJoin directory on HDFS

**./hadoop fs -mkdir ReduceSideJoin/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 the multiple input files from Local

**./hadoop dfs -put /home/iitp/ReduceSideJoin/Inputdata/customers  
ReduceSideJoin/Inputdata/**

and

**./hadoop dfs -put /home/iitp/ReduceSideJoin/Inputdata/sales  
ReduceSideJoin/Inputdata/**

8. Check if the input files have been correctly moved onto HDFS

**./hadoop dfs -ls -R ReduceSideJoin/Inputdata/**

9. Running the program

- a. browse to the bin directory of hadoop installation

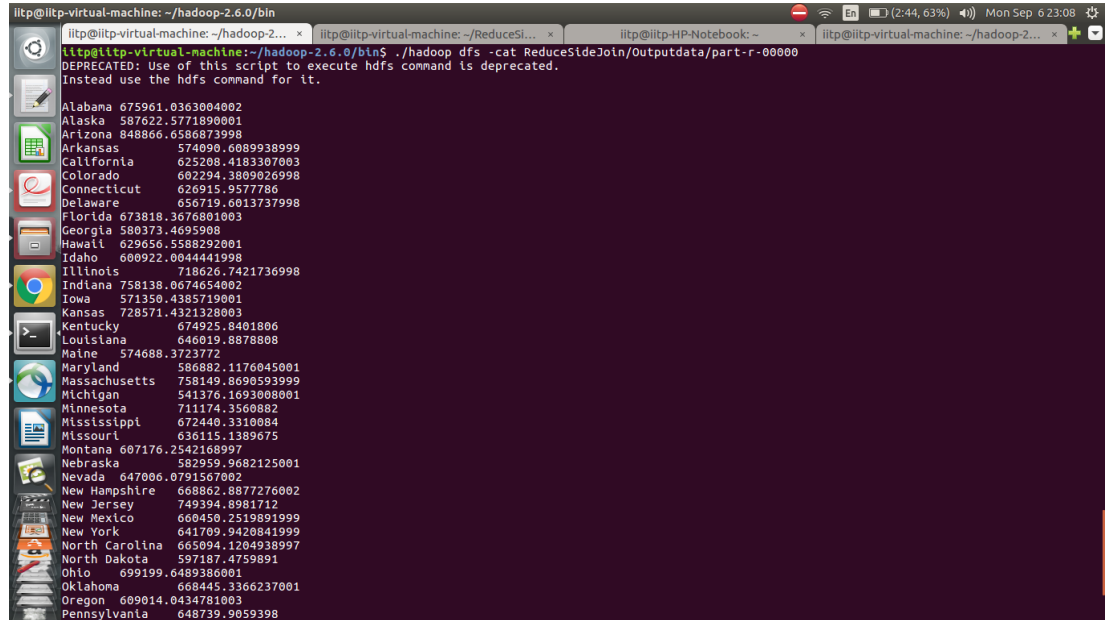
**cd /home/iitp/hadoop-2.6.0/bin**

- b. Running in terminal

```
./hadoop jar /home/iitp/ReduceSideJoin/target/ReduceSide-1.0.0.jar  
CalculateTotalSales ReduceSideJoin/Inputdata/customers  
ReduceSideJoin/Inputdata/sales ReduceSideJoin/Outputdata
```

c. Finding outputs

```
./hadoop dfs -cat ReduceSideJoin/Outputdata/part-r-00000
```



```
iitp@iitp-virtual-machine: ~/hadoop-2.6.0/bin
iitp@iitp-virtual-machine: ~/hadoop-2.6.0/bin$ ./hadoop dfs -cat ReduceSideJoin/Outputdata/part-r-00000
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
Alabama 675961.0363004002
Alaska 587622.5771890001
Arizona 848866.6586873998
Arkansas 574090.6089938999
California 625208.4183307003
Colorado 602294.3809026998
Connecticut 626915.9577780
Delaware 656719.6013737998
Florida 673818.3676801003
Georgia 580373.4695908
Hawaii 629656.5588292001
Idaho 680922.0044441998
Illinois 718626.7421736998
Indiana 758138.0674654002
Iowa 571350.4385719001
Kansas 728571.4321328003
Kentucky 674925.8401806
Louisiana 646019.8878808
Maine 574688.3723772
Maryland 586802.1176045001
Massachusetts 758149.8690593999
Michigan 541376.1693008001
Minnesota 711174.3560882
Mississippi 672440.3310084
Missouri 636115.1389675
Montana 687176.2542168997
Nebraska 582959.9682125001
Nevada 647006.0791567002
New Hampshire 668862.8877276002
New Jersey 749394.8981712
New Mexico 660450.2519891999
New York 641709.9420841999
North Carolina 665094.1204938997
North Dakota 597187.4759891
Ohio 699199.6489386001
Oklahoma 668445.3366237001
Oregon 609014.0434781003
Pennsylvania 648739.9059398
```