**Instructions Reduce Side Join Execution**

CS555: Big Data Computing Lab

horizontal line

# **7th September 2021**

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

**sudo apt-get install maven**

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

**cd /home/iitp**

* 1. Create project directory

**mkdir ReduceSideJoin**

1. Copy files with directory structure from local system to VM:
   1. 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**

* 1. 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**

* 1. 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**

1. Compile the java source files
   1. Browse to the ReduceSideJoin directory

**cd ReduceSideJoin**

* 1. Compile source files and create jar file

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

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 (**ReduceSideJoin**)

**./hadoop fs -mkdir ReduceSideJoin**

* 1. create directory (**Inputdata**) inside ReduceSideJoin directory on HDFS

**./hadoop fs -mkdir ReduceSideJoin/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 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/**

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

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

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

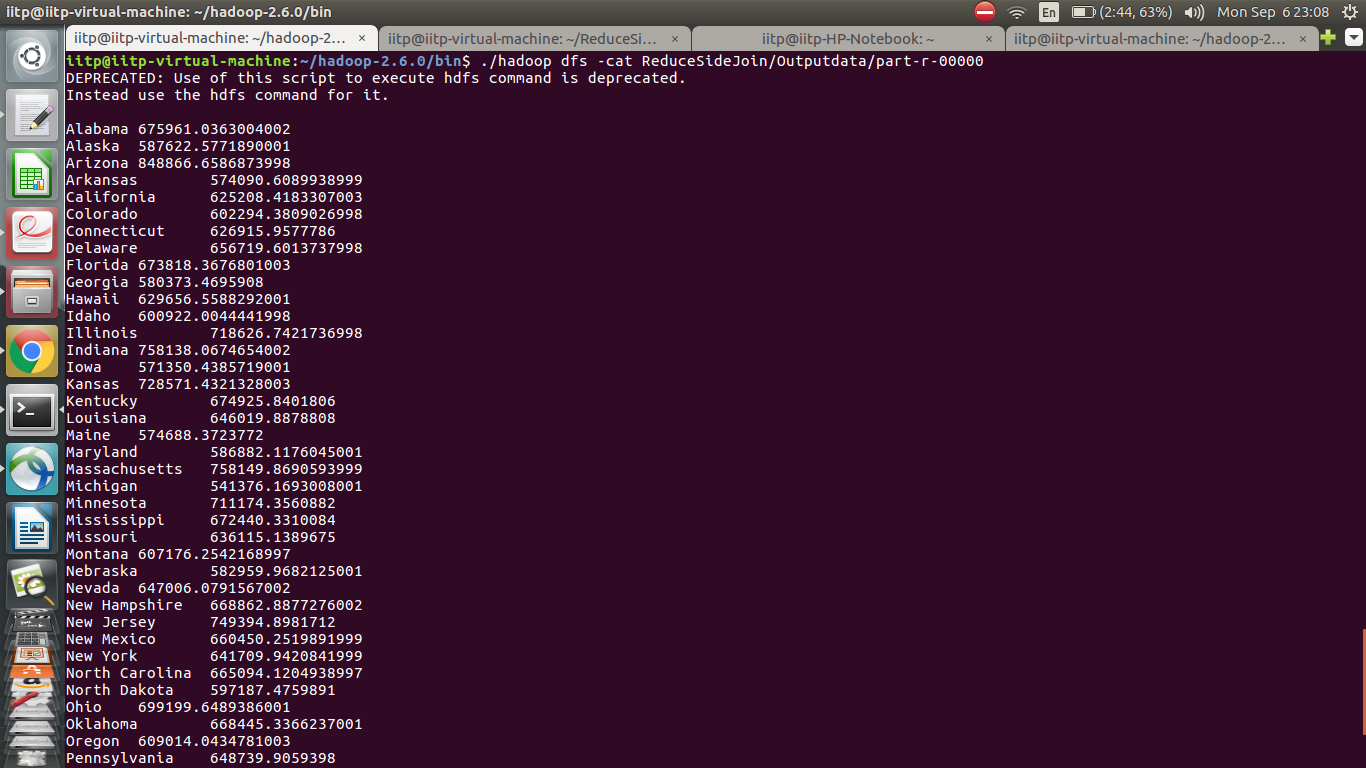
**cd /home/iitp/hadoop-2.6.0/bin**

* 1. Running in terminal

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

* 1. Finding outputs

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

****