**Steps for MapReduce Program for Weather dataset**

1. Open Eclipse
2. Erase the old program and close the program windows present.
3. File New Java Project

Give the project name as **MaxTemp**

1. Click on Next Button
2. Click on Libraries and select Add External Jars
3. File system usr Lib hadoop . (select all the jar files and click on OK Button)
4. Once again in Libraries Tab select **Add external Jar**
5. Click client and select all JAR files. Click on OK and click on Finish Button
6. Select Max Temp from left panel u can see three dropdowns
7. Creating **Mapper File**

Right click on src.

New class

Name field : **MaxTempMapper**

**Click on Finish**

1. Creating **Reducer File**

Right click on src.

New class

Name field : **MaxTempReducer**

**Click on Finish**

1. Creating **driver File**

Right click on src.

New class

Name field : **MaxTemp**

**Click on Finish**

1. Copy the mapper, reducer and driver code into the java files
2. Change the class name of mapper, reducer and reducer and change the conf name to **MaxTemp**
3. Change the **arg value** of input and output path
4. Right click on **MaxTemp** in the left panel select **export**
5. Select java JAR file.
6. Click on **Browse button**
7. Select Desktop. Give the name as **MaxTemp.jar**
8. Click on **OK** Button. *U can see the jar file created in the desktop*
9. Keep the dataset in the desktop tempinput.txt
10. Open terminal

hadoop dfs –mkdir /maxoutput

1. We have to go to desktop as we have dataset

cd Desktop

1. Copy the dataset to hadoop

Hadoop dfs –copyFromLocal tempinput.txt /maxinput

1. Hadoop jar MaxTemp.jar MaxTemp /maxinput /max output

**Output screen:**

**cloudera@quickstart ~]$ hadoop dfs -mkdir /maxinput**

**DEPRECATED: Use of this script to execute hdfs command is deprecated.**

**Instead use the hdfs command for it.**

**[cloudera@quickstart ~]$ cd Desktop**

**[cloudera@quickstart Desktop]$ cd Desktop/**

**bash: cd: Desktop/: No such file or directory**

**[cloudera@quickstart Desktop]$ hadoop dfs -copyFromLocal tempinput.txt /maxinput**

**DEPRECATED: Use of this script to execute hdfs command is deprecated.**

**Instead use the hdfs command for it.**

**[cloudera@quickstart Desktop]$ hadoop jar MaxTemp.jar MaxTemp /maxinput /maxoutput**

**23/04/24 22:07:42 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032**

**23/04/24 22:07:46 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.**

**23/04/24 22:07:48 INFO input.FileInputFormat: Total input paths to process : 1**

**23/04/24 22:07:49 INFO mapreduce.JobSubmitter: number of splits:1**

**23/04/24 22:07:50 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1682395925952\_0001**

**23/04/24 22:07:52 INFO impl.YarnClientImpl: Submitted application application\_1682395925952\_0001**

**23/04/24 22:07:52 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application\_1682395925952\_0001/**

**23/04/24 22:07:52 INFO mapreduce.Job: Running job: job\_1682395925952\_0001**

**23/04/24 22:08:50 INFO mapreduce.Job: Job job\_1682395925952\_0001 running in uber mode : false**

**23/04/24 22:08:50 INFO mapreduce.Job: map 0% reduce 0%**

**23/04/24 22:09:15 INFO mapreduce.Job: map 100% reduce 0%**

**23/04/24 22:09:36 INFO mapreduce.Job: map 100% reduce 100%**

**23/04/24 22:09:37 INFO mapreduce.Job: Job job\_1682395925952\_0001 completed successfully**

**23/04/24 22:09:38 INFO mapreduce.Job: Counters: 49**

**File System Counters**

**FILE: Number of bytes read=61**

**FILE: Number of bytes written=286393**

**FILE: Number of read operations=0**

**FILE: Number of large read operations=0**

**FILE: Number of write operations=0**

**HDFS: Number of bytes read=635**

**HDFS: Number of bytes written=17**

**HDFS: Number of read operations=6**

**HDFS: Number of large read operations=0**

**HDFS: Number of write operations=2**

**Job Counters**

**Launched map tasks=1**

**Launched reduce tasks=1**

**Data-local map tasks=1**

**Total time spent by all maps in occupied slots (ms)=21659**

**Total time spent by all reduces in occupied slots (ms)=19101**

**Total time spent by all map tasks (ms)=21659**

**Total time spent by all reduce tasks (ms)=19101**

**Total vcore-milliseconds taken by all map tasks=21659**

**Total vcore-milliseconds taken by all reduce tasks=19101**

**Total megabyte-milliseconds taken by all map tasks=22178816**

**Total megabyte-milliseconds taken by all reduce tasks=19559424**

**Map-Reduce Framework**

**Map input records=5**

**Map output records=5**

**Map output bytes=45**

**Map output materialized bytes=61**

**Input split bytes=105**

**Combine input records=0**

**Combine output records=0**

**Reduce input groups=2**

**Reduce shuffle bytes=61**

**Reduce input records=5**

**Reduce output records=2**

**Spilled Records=10**

**Shuffled Maps =1**

**Failed Shuffles=0**

**Merged Map outputs=1**

**GC time elapsed (ms)=497**

**CPU time spent (ms)=3040**

**Physical memory (bytes) snapshot=347107328**

**Virtual memory (bytes) snapshot=3015774208**

**Total committed heap usage (bytes)=226365440**

**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=530**

**File Output Format Counters**

**Bytes Written=17**

**[cloudera@quickstart Desktop]$ hadoop dfs -cat /maxoutput/part-r-00000**

**DEPRECATED: Use of this script to execute hdfs command is deprecated.**

**Instead use the hdfs command for it.**

**1949 111**

**1950 22**

**[cloudera@quickstart Desktop]$**

1. To see the output

hadoop dfs -cat /maxoutput/part-r-00000

1. We can also see the output from the browser

**Localhost:50070**

In utilities tab select browse file system and we can see the output name Maxoutput

click on it and download.