Experiment No: 10 Roll No: 210701147

Implement a MapReduce Program to Process a Weather Dataset

AIM:

To implement a MapReduce Program to Process a Weather dataset using Hadoop.

PROCEDURE:

1. Open command prompt and run as administrator. Next start Hadoop using the **start-dfs.cmd** and **start-yarn.cmd** command.
2. Open the browser and go to the URL localhost:9870.
3. In command prompt create a directory called weather and upload the weather dataset using the **hdfs dfs -put** command.
4. Inorder to verify that the dataset was uploaded successfully, use the command **hdfs dfs -ls /weather** and it will show the dataset.
5. Create the mapper and reducer python files and save it.
6. Execute the mapreduce program using the following command

```
C:\Users\Sajjad\OneDrive\Documents\DataAnalytics>hadoop jar C:/hadoop-3.3.6/share/hadoop/tools/lib/hadoop-streaming-3.3.6.jar ^
-file C:/Users/Sajjad/OneDrive/Documents/DataAnalytics/mapper3.py ^
-file C:/Users/Sajjad/OneDrive/Documents/DataAnalytics/reducer3.py ^
-input /weather/weather_dataset.txt ^
 -output /weather/output ^
-mapper "python mapper3.py" ^
-reducer "python reducer3.py"
```

7. To verify that program has been successfully executed, The output will be present in the localhost directory.

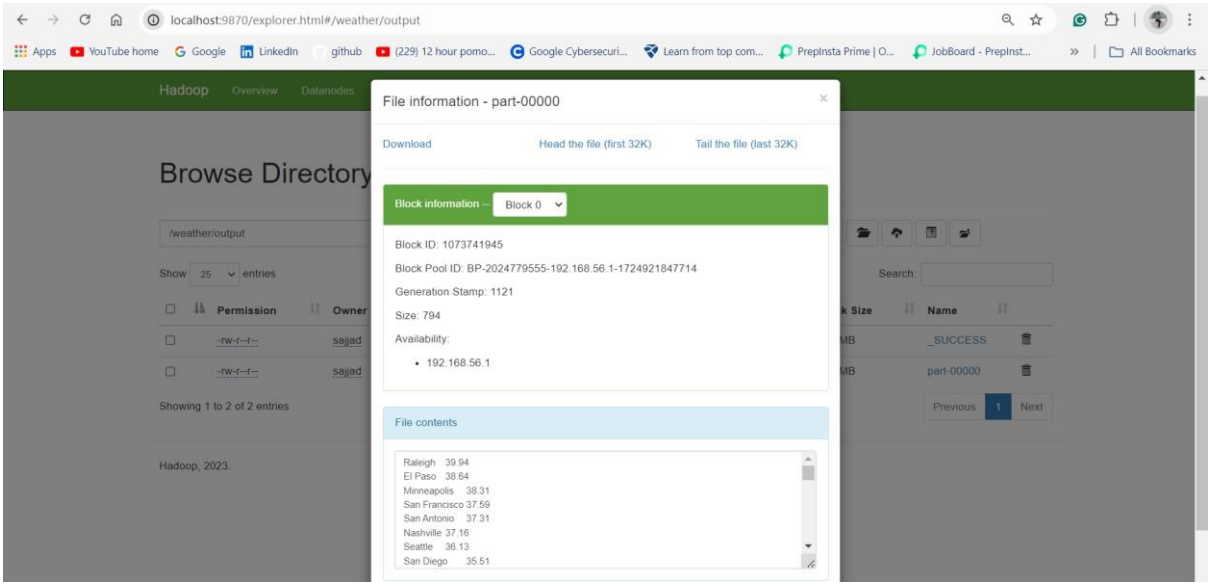
OUTPUT:

1.
```
C:\>cd C:\hadoop-3.3.6\sbin

C:\hadoop-3.3.6\sbin>start-dfs.cmd

C:\hadoop-3.3.6\sbin>start-yarn.cmd
starting yarn daemons

C:\hadoop-3.3.6\sbin>jps
15968 NodeManager
33264 Jps
23876 NameNode
20728 ResourceManager
17500 DataNode
```

2.
```
C:\hadoop-3.3.6\bin>hdfs dfs -mkdir -p /weather

C:\hadoop-3.3.6\bin>hdfs dfs -put C:\Users\Sajjad\OneDrive\Documents\DataAnalytics\weather_dataset.txt /weather

C:\hadoop-3.3.6\bin>
```

3.
```
C:\Users\Sajjad\OneDrive\Documents\DataAnalytics>hadoop jar C:/hadoop-3.3.6/share/hadoop/tools/lib/hadoop-streaming-3.3.6.jar ^
More? -file C:/Users/Sajjad/OneDrive/Documents/DataAnalytics/mapper3.py ^
More? -file C:/Users/Sajjad/OneDrive/Documents/DataAnalytics/reducer3.py ^
More? -input /weather/weather_dataset.txt ^
More? -output /weather/output ^
More? -mapper "python mapper3.py" ^
More? -reducer "python reducer3.py"
2024-09-08 13:32:19,614 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [C:/Users/Sajjad/OneDrive/Documents/DataAnalytics/mapper3.py, C:/Users/Sajjad/OneDrive/Documents/DataAnalytics/reducer3.py, /C:/Users/Sajjad/AppData/Local/Temp/hadoop-unjar2564494122007576O5/] []
C:\Users\Sajjad\AppData\Local\Temp\streamjob1626265326661482232.jar tmpDir=null
2024-09-08 13:32:20,861 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-08 13:32:21,185 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-08 13:32:21,947 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/sajjad/.staging/job_1725781366746_0002
2024-09-08 13:32:22,433 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-08 13:32:22,530 INFO mapreduce.JobSubmitter: number of splits:2
2024-09-08 13:32:23,189 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1725781366746_0002
2024-09-08 13:32:23,189 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-09-08 13:32:23,389 INFO conf.Configuration: resource-types.xml not found
2024-09-08 13:32:23,389 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-09-08 13:32:23,528 INFO impl.YarnClientImpl: Submitted application application_1725781366746_0002
2024-09-08 13:32:23,609 INFO mapreduce.Job: The url to track the job: http://envy24:8088/proxy/application_1725781366746_0002/
2024-09-08 13:32:23,612 INFO mapreduce.Job: Running job: job_1725781366746_0002
2024-09-08 13:32:33,788 INFO mapreduce.Job: Job job_1725781366746_0002 running in uber mode : false
2024-09-08 13:32:33,790 INFO mapreduce.Job:  map 0% reduce 0%
2024-09-08 13:32:42,965 INFO mapreduce.Job:  map 50% reduce 0%
2024-09-08 13:32:43,976 INFO mapreduce.Job:  map 100% reduce 0%
2024-09-08 13:32:51,061 INFO mapreduce.Job:  map 100% reduce 100%
2024-09-08 13:32:51,070 INFO mapreduce.Job: Job job_1725781366746_0002 completed successfully
                Bytes Written=794
```

4.
```
2024-09-08 13:32:51,199 INFO streaming.StreamJob: Output directory: /weather/output

C:\Users\Sajjad\OneDrive\Documents\DataAnalytics>
```

5. 

RESULT:

Thus the above Implement a MapReduce Program to Process a Weather Dataset has been executed successfully.