Ex.No – 3 Roll No – 210701149

Implement a MapReduce Program to Process a Weather Dataset

AIM:

To implement a MapReduce program to process a weather dataset using Hadoop.

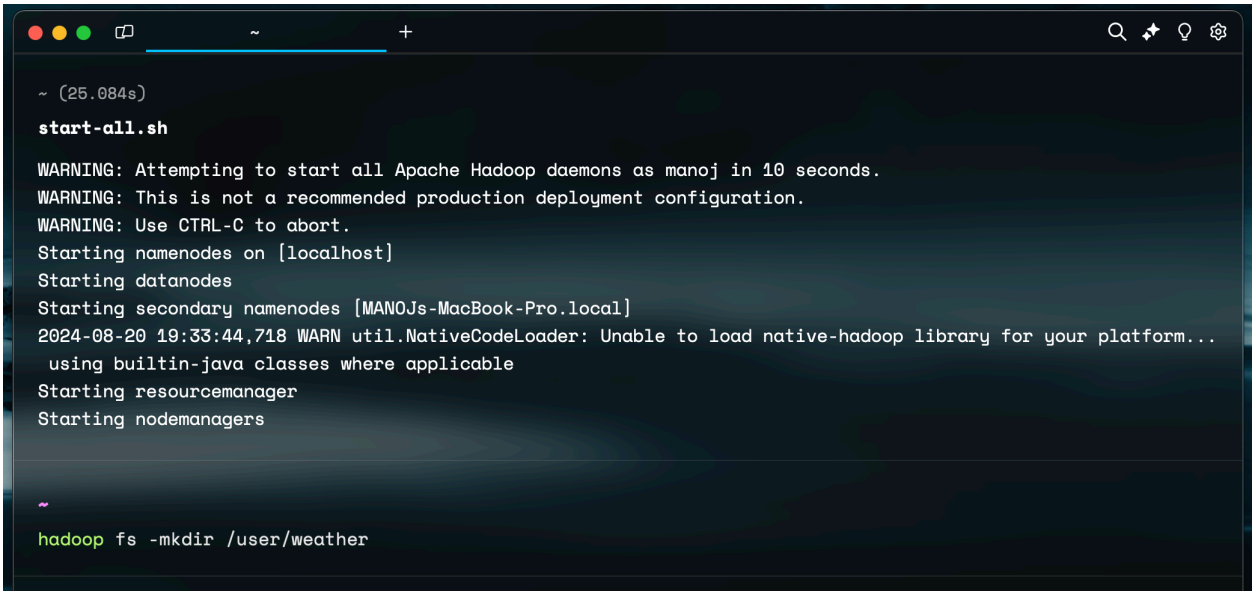
PROCEDURES:

1. Open the terminal and start Hadoop using start-all.sh command
2. Open the browser and go to the URL localhost:9870.
3. In the terminal using the command hadoop fs -mkdir /user create a directory called user.
4. Upload the sample_weather.txt file to hdfs using the command hadoop fs -put sample_weather.txt /user.

   Then perform the mapreduce operation using the command

   ```
   hadoop jar /path/to/hadoop-streaming.jar \
   -files /path/to/mapper.py, /path/to/reducer.py \
   -input /path/to/input \
   -output /path/to/output \
   -mapper mapper.py \
   -reducer reducer.py
   ```
5. Check the output using the command hadoop fs -cat /user/output/part-00000.

OUTPUT:

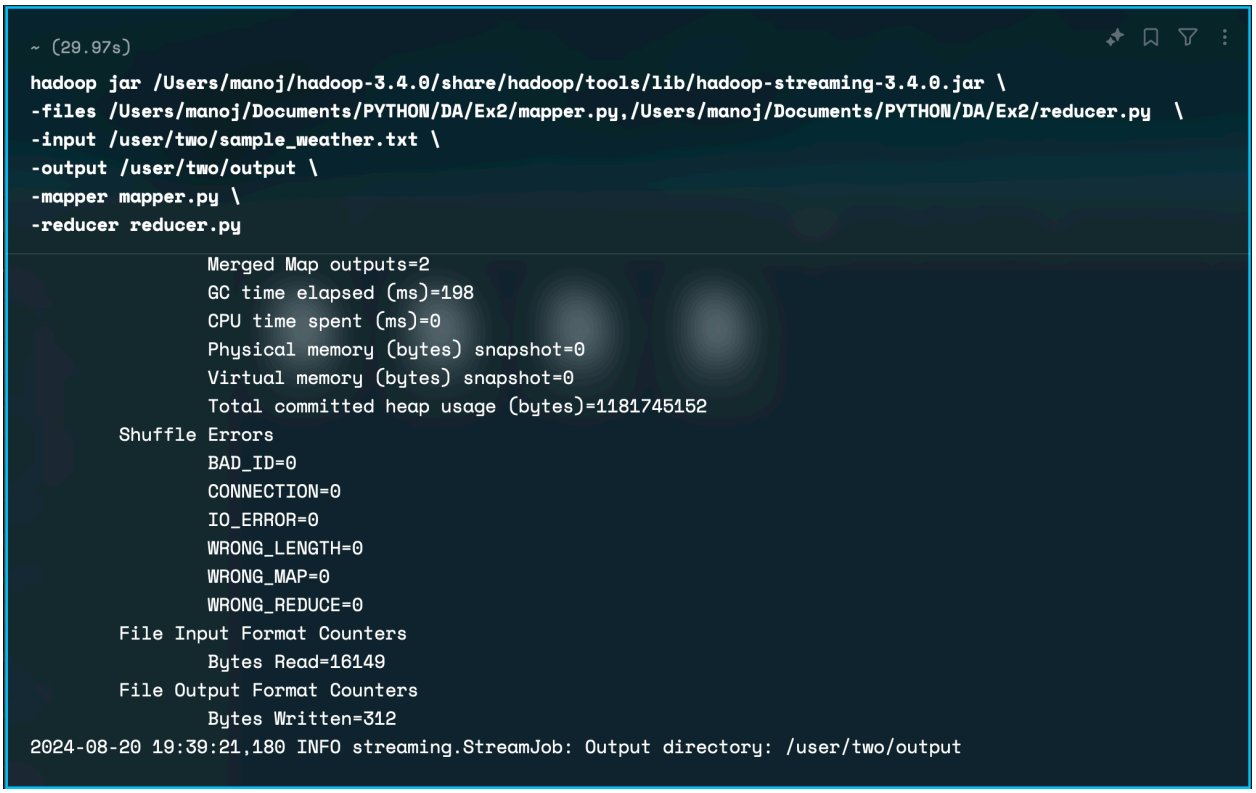


```
~ (25.084s)
start-all.sh

WARNING: Attempting to start all Apache Hadoop daemons as manoj in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [MANOJs-MacBook-Pro.local]
2024-08-20 19:33:44,718 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
 using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers


~
hadoop fs -mkdir /user/weather
```

```
~
hadoop fs -put /Users/manoj/Downloads/sample_weather.txt /user/weather
```



```
~ (29.97s)
hadoop jar /Users/manoj/hadoop-3.4.0/share/hadoop/tools/lib/hadoop-streaming-3.4.0.jar \
-files /Users/manoj/Documents/PYTHON/DA/Ex2/mapper.py,/Users/manoj/Documents/PYTHON/DA/Ex2/reducer.py  \
-input /user/two/sample_weather.txt \
-output /user/two/output \
-mapper mapper.py \
-reducer reducer.py

                Merged Map outputs=2
                GC time elapsed (ms)=198
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=1181745152
        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=16149
        File Output Format Counters
                Bytes Written=312
2024-08-20 19:39:21,180 INFO streaming.StreamJob: Output directory: /user/two/output
```

```
hadoop fs -cat /user/two/output/part-00000

2024-08-20 19:41:14,090 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
 using builtin-java classes where applicable
690190_200602_section1  53.87166666666666 25.899999999999995 7.774999999999998
690190_200602_section2  54.76125000000001 25.900000000000006 7.774999999999999
690190_200602_section3  53.25041666666667 25.899999999999995 7.774999999999996
690190_200602_section4  52.44708333333333 25.900000000000006 7.774999999999999
```

RESULT:

Thus, to implement a MapReduce program to process a weather dataset has been successfully.