

EXERCISE NO.: 04

WEATHER REPORT POC USING MAPREDUCE

AIM:

To implement a MapReduce program that analyzes weather data and generates a report containing maximum and minimum temperatures for each day or time interval, thereby demonstrating the use of Hadoop MapReduce for statistical analysis.

SCRIPT:

weather.csv

```
Date,Time,Temperature,Humidity,Pressure
2025-11-13,06:00,22,85,1012
2025-11-13,12:00,28,70,1010
2025-11-13,18:00,24,75,1008
2025-11-14,06:00,21,80,1011
2025-11-14,12:00,30,65,1009
2025-11-14,18:00,25,70,1007
```

Upload the file to HDFS

```
!hdfs dfs -mkdir -p /user/bdt/weather/input
!hdfs dfs -put weather.csv /user/bdt/weather/input/
```

mapper.py

```
import sys
for line in sys.stdin:
    if line.strip() and not line.startswith("Date"):
        date, time, temp, *rest = line.strip().split(',')
        print(f"{date}\\t{temp}")
```

reducer.py

```
import sys
from collections import defaultdict
temps = defaultdict(list)
for line in sys.stdin:
```

```
date, temp = line.strip().split("\t")
temps[date].append(float(temp))
for date, values in temps.items():
    print(f"{date}\tMin:{min(values)}\tMax:{max(values)}")
```

Hadoop Streaming

```
!chmod +x mapper.py reducer.py
```

```
!hadoop jar $HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-*.jar \
-input /user/bdt/weather/input/ \
-output /user/bdt/weather/output/ \
-mapper mapper.py \
-reducer reducer.py \
-file mapper.py \
-file reducer.py
```

OUTPUT:

2025-11-13	Min:22.0	Max:28.0
2025-11-14	Min:21.0	Max:30.0

RESULT:

Thus, the MapReduce program successfully generated a daily weather report with max/min temperatures.