

## **Ex No: 4 Weather Report POC using Hadoop Streaming**

### **AIM:**

To write a Hadoop Streaming MapReduce program in Python to analyze weather data and generate a report containing maximum and minimum temperatures per day.

Algorithm:

### **Mapper Algorithm**

1. Read a line from input. 2. Split the line into datetime and temperature. 3. Extract date from datetime. 4. Emit (date, temperature).

### **Reducer Algorithm**

1. Receive (date, list of temperatures) from all mappers.  
2. Track current date and temperature values.  
3. When date changes, output max and min temperature for previous date.  
4. After loop ends, output max and min for the last date.

### **Python Implementation**

#### **Mapper (mapper.py)**

```
#!/usr/bin/env python3
import sys
#Inputformat: "datetime,temp"
#Example:"2025-09-01 14:00,35"
for line in sys.stdin:

    try:
        line = line.strip()
        datetime, temp = line.split(",")
        date = datetime.split(" ")[0] # extract only date
        print(f"{date}\t{temp}")
    except:
        continue # skip malformed lines
```

#### **Reducer (reducer.py)**

```
#!/usr/bin/env python3
import sys
current_date = None
temps = []
for line in sys.stdin:

    line = line.strip()
```

```

ifnotline:
    continue
date,temp = line.split("\t")
temp=float(temp)
ifcurrent_date == date:

    temps.append(temp)
else:
    ifcurrent_date:
        #output result for the previous date
        print(f"{current_date}\tmax={max(temps)}\tmin={min(temps)}")
    current_date = date
    temps= [temp]

#Finaloutput for the last date
ifcurrent_date:
    print(f"{current_date}\tmax={max(temps)}\tmin={min(temps)}")

```

#### Sample Input (weather\_data.txt)

2025-09-01 14:00,35 2025-09-01 15:00,33 2025-09-01  
 16:00,37 2025-09-02 14:00,32 2025-09-02 15:00,34

#### Running the Program in Hadoop Streaming

hadoop jar /path/to/hadoop-streaming.jar \

```

-input /user/hadoop/weather_data.txt \
-output /user/hadoop/weather_output \
-mapper mapper.py \
-reducer reducer.py \
-file mapper.py \
-file reducer.py

```

#### Sample Output

2025-09-01 max=37.0 min=33.0  
 2025-09-02 max=34.0 min=32.0

#### Result

TheHadoop Streaming MapReduce program was successfully executed to generate a daily weather report showing maximum and minimum temperatures.