Ex.No.10

IMPLEMENT THE MAPREDUCE PROGRAM TO PERFORM PROCESSING ON TEMPERATURE DATA COLLECTED FROM SENSORS

AIM:

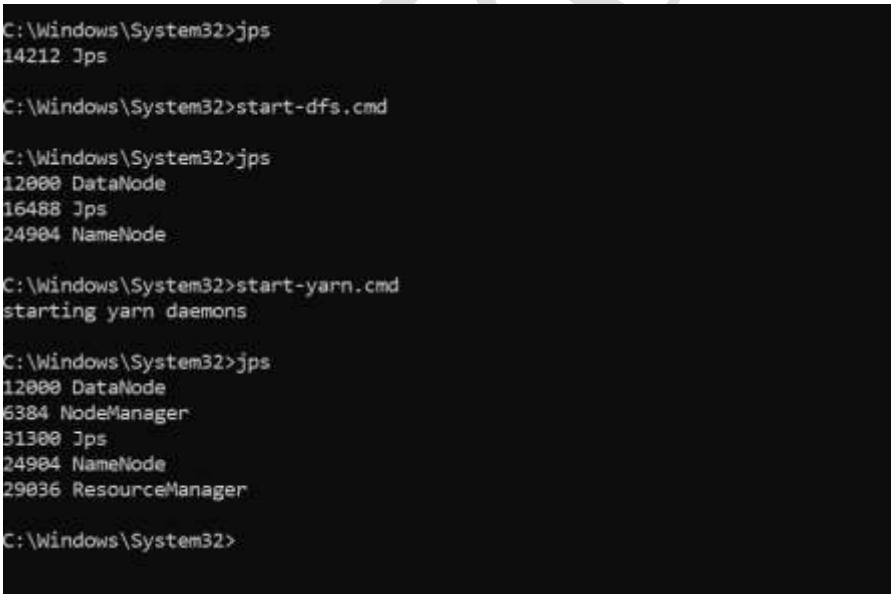
To implement a Map Reduce program to perform processing on the temperature data collected from sensors

PROCEDURE:

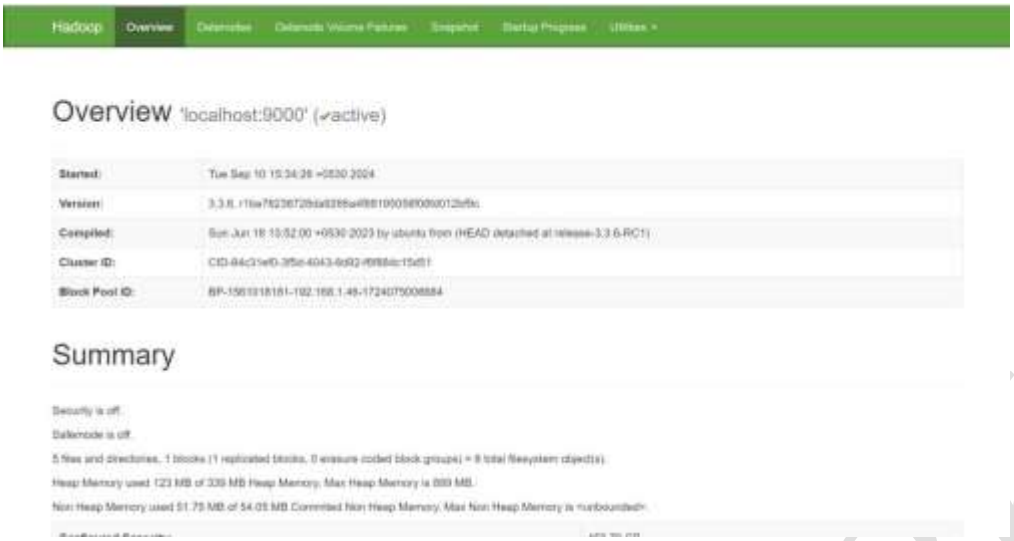
Open command prompt and run as administrator

Start Hadoop services by typing in the following commands:

- start-dfs.cmd
- start-yarn.cmd



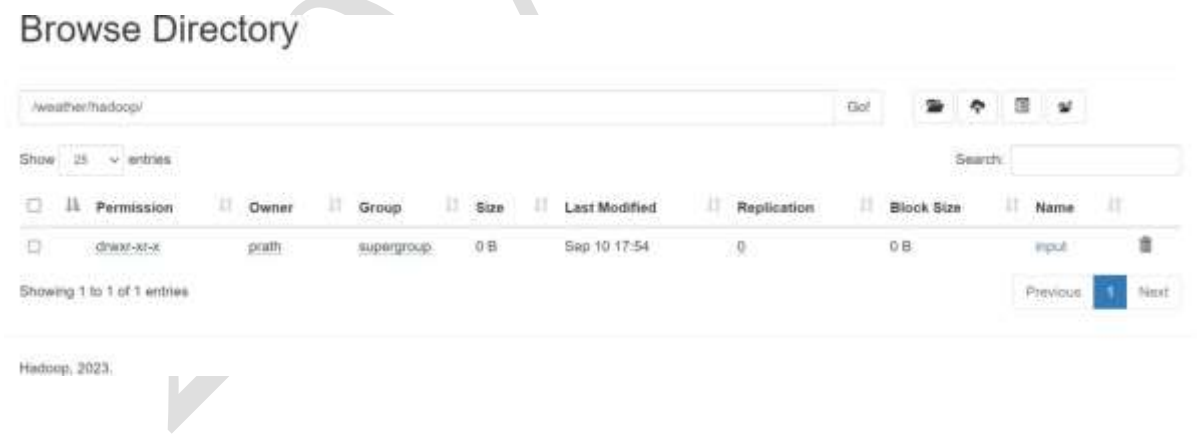
Open the browser and go to the URL localhost:9870



Create a directory in HDFS using the command:

hdfs dfs -mkdir -p /weather/hadoop/input

```
C:\hadoop-3.3.6\sbin>hdfs dfs -mkdir -p /weather/hadoop/input

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



Copy the input file to HDFS using the command:

hdfs dfs -put C:/Semester7/DataAnalytics/Lab/Ex3/sample_weather.txt /weather/hadoop/input

```
C:\hadoop-3.3.6\sbin>hdfs dfs -put C:/Semester7/DataAnalytics/Lab/Ex3/sample_weather.txt /weather/hadoop/input
```

Display the contents of the file using this command:

hdfs dfs -cat /weather/hadoop/input/sample_weather.txt

```
C:\hadoop-3.3.6\sbin>hdfs dfs -cat /weather/hadoop/input/sample_weather.txt
690190 13910  20060201_0  51.75     33.0 24  1006.3 24   943.9 24   15.0
24   10.7 24   22.0   28.9    0.00I 999.9  000000
690190 13910  20060201_1  54.74     33.0 24  1006.3 24   943.9 24   15.0
24   10.7 24   22.0   28.9    0.00I 999.9  000000
690190 13910  20060201_2  50.59     33.0 24  1006.3 24   943.9 24   15.0
24   10.7 24   22.0   28.9    0.00I 999.9  000000
690190 13910  20060201_3  51.67     33.0 24  1006.3 24   943.9 24   15.0
24   10.7 24   22.0   28.9    0.00I 999.9  000000
690190 13910  20060201_4  65.67     33.0 24  1006.3 24   943.9 24   15.0
24   10.7 24   22.0   28.9    0.00I 999.9  000000
690190 13910  20060201_5  55.37     33.0 24  1006.3 24   943.9 24   15.0
24   10.7 24   22.0   28.9    0.00I 999.9  000000
690190 13910  20060201_6  49.26     33.0 24  1006.3 24   943.9 24   15.0
24   10.7 24   22.0   28.9    0.00I 999.9  000000
690190 13910  20060201_7  55.44     33.0 24  1006.3 24   943.9 24   15.0
24   10.7 24   22.0   28.9    0.00I 999.9  000000
690190 13910  20060201_8  64.05     33.0 24  1006.3 24   943.9 24   15.0
```

Create mapper.py and reducer.py files

mapper.py

```python
import sys

def map1():
    for line in sys.stdin:
        tokens = line.strip().split()
        if len(tokens) < 13:
            continue

        station = tokens[0]
        if "STN" in station:
            continue

        date_hour = tokens[2]
        temp = tokens[3]
        dew = tokens[4]
        wind = tokens[12]

        if temp == "9999.9" or dew == "9999.9" or wind == "999.9":
            continue
        hour = int(date_hour.split("_")[-1])
        date = date_hour[:date_hour.rfind("_")-2]

        if 4 < hour <= 10:
            section = "section1"
        elif 10 < hour <= 16:
            section = "section2"
        elif 16 < hour <= 22:
            section = "section3"
        else:
            section = "section4"

        key_out = f"{station}_{date}_{section}"
        value_out = f"{temp} {dew} {wind}"
        print(f"{key_out}\t{value_out}")

if __name__ == "__main__":
    map1()
```

reducer.py

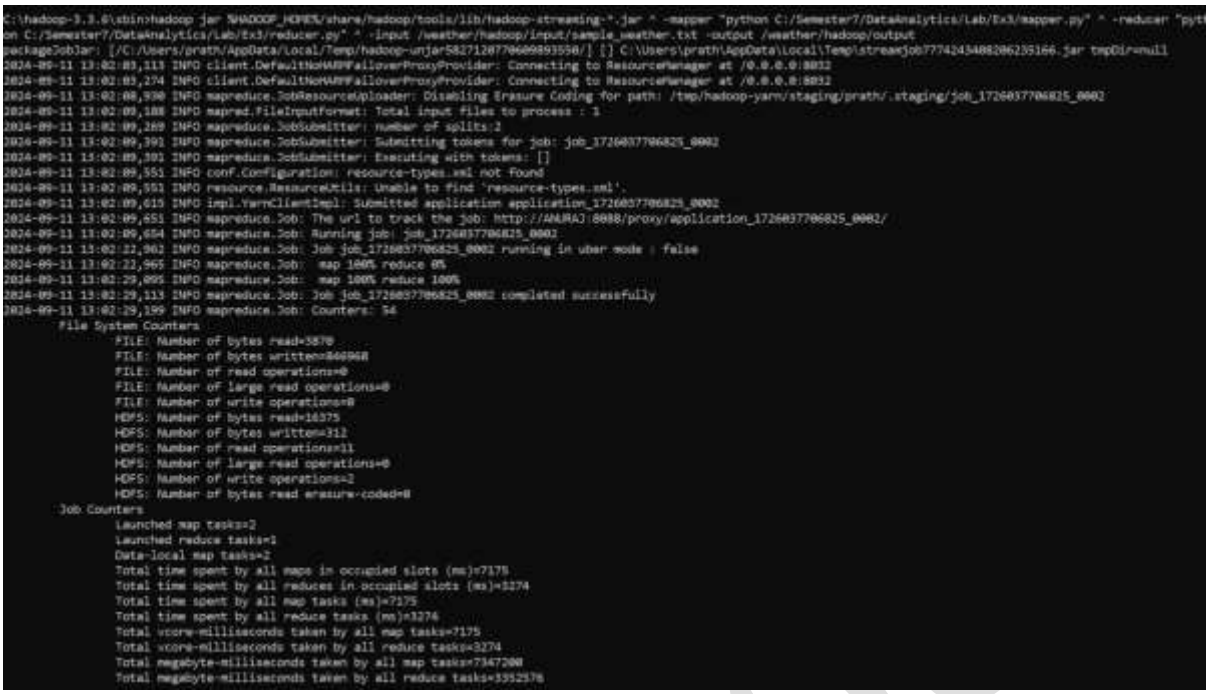
```
import sys

def reduce1():
    current_key = None
    sum_temp, sum_dew, sum_wind = 0, 0, 0
    count = 0

    for line in sys.stdin:
        key, value = line.strip().split("\t")
        temp, dew, wind = map(float, value.split())

        if current_key is None:
            current_key = key

        if key == current_key:
            sum_temp += temp
            sum_dew += dew
            sum_wind += wind
            count += 1
        else:
            avg_temp = sum_temp / count
            avg_dew = sum_dew / count
            avg_wind = sum_wind / count
            print(f"{current_key}\t{avg_temp} {avg_dew} {avg_wind}")
            current_key = key
            sum_temp, sum_dew, sum_wind = temp, dew, wind
            count = 1

    if current_key is not None:
        avg_temp = sum_temp / count
        avg_dew = sum_dew / count
        avg_wind = sum_wind / count
        print(f"{current_key}\t{avg_temp} {avg_dew} {avg_wind}")
if __name__ == "__main__":
    reduce1()
```

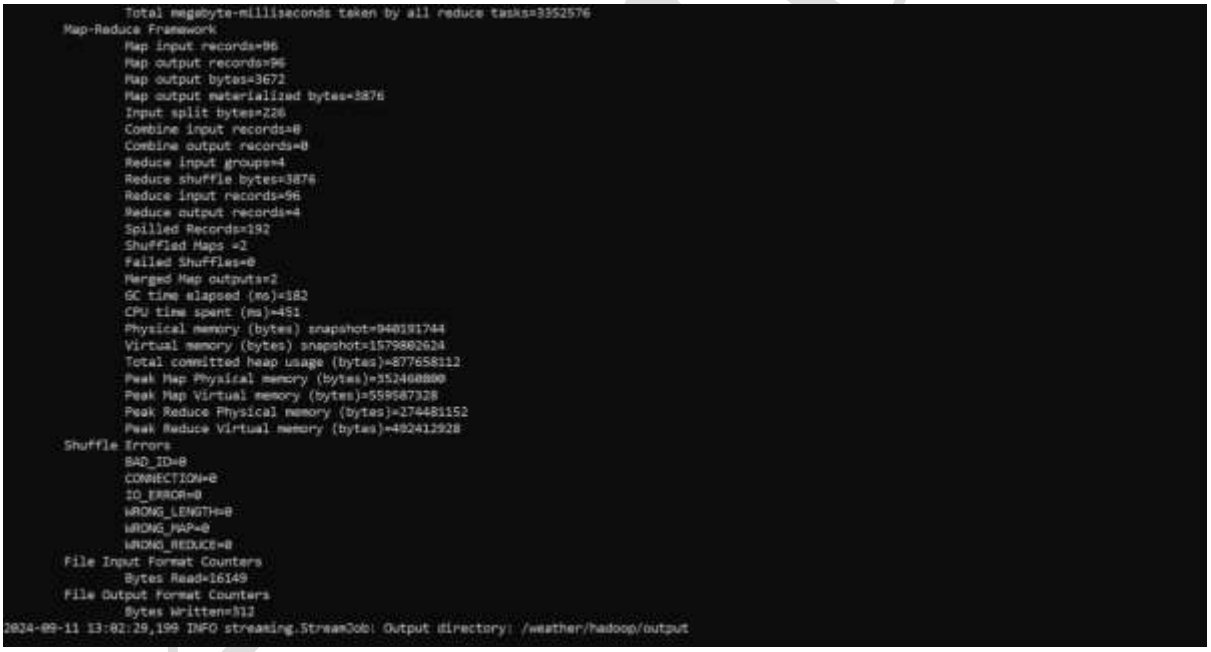
Run the Hadoop Streaming Job and give the file paths to the input, mapper and reducer using the following command:

hadoop jar %HADOOP_HOME%\share\hadoop\tools\lib\hadoop-streaming-*.jar^

-mapper "python C:\Semester7\DataAnalytics\Lab\Ex3\mapper.py" -reducer "python C:\Semester7\DataAnalytics\Lab\Ex3\reducer.py"^

-input/weather/hadoop/input/sample_weather.txt -output /weather/hadoop/output



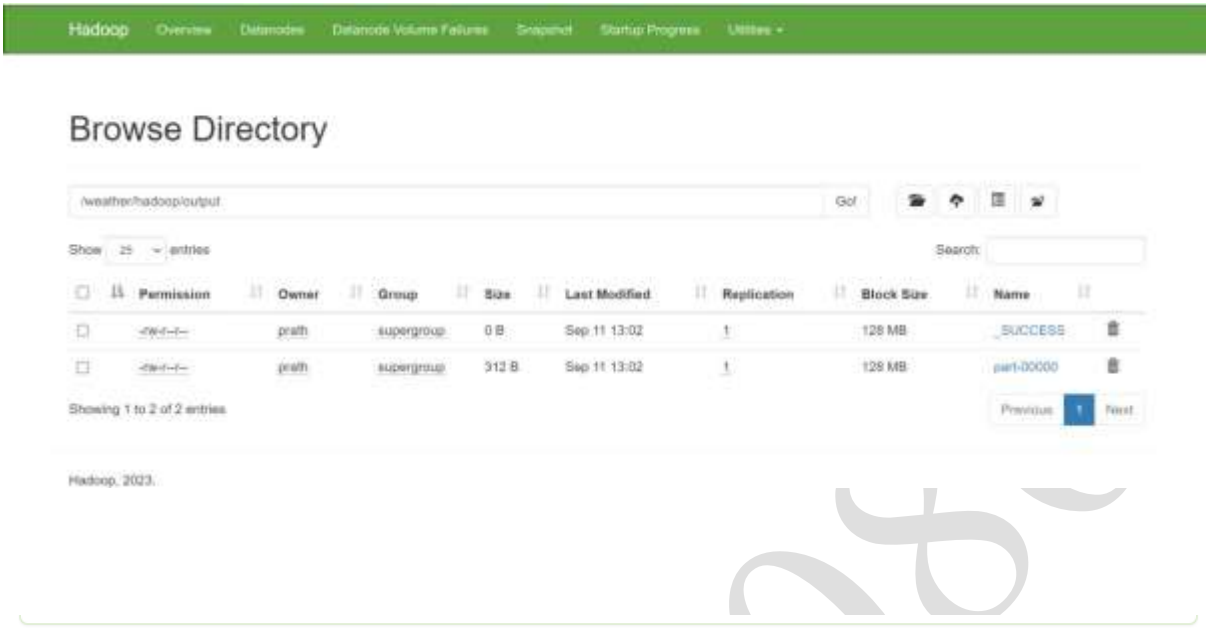


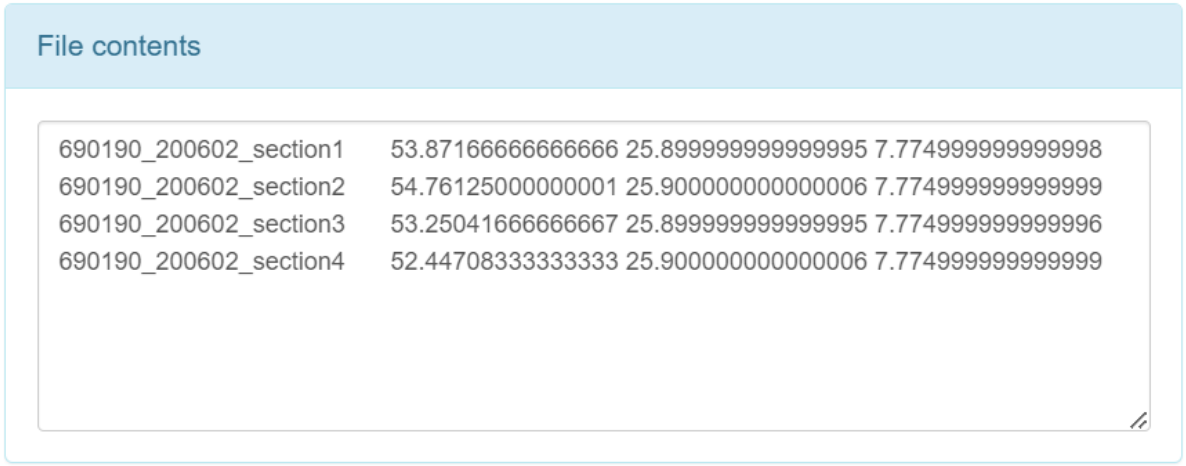
View the output using the command:

hdfs dfs -cat /weather/hadoop/output/part-00000

```
C:\hadoop-3.3.6\sbin>hdfs dfs -cat /weather/hadoop/output/part-00000
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
```

View the output on the file system in browser





RESULT:

Thus, to implement the Map Reduce program to perform processing on the temperature data collected from sensors was completed successfully.