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:

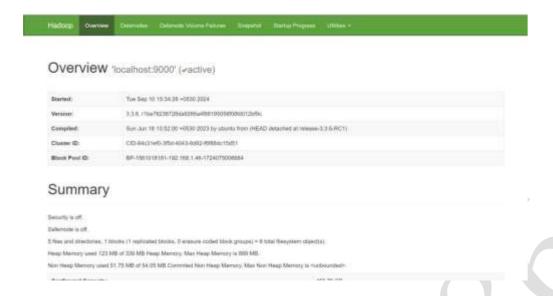
Open command prompt and run as administrator

Start Hadoop services by typing in the following commands:

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

```
:\Windows\System32>jps
14212 Jps
:\Windows\System32>start-dfs.cmd
::\Windows\System32>jps
12000 DataNode
16488 Jps
4904 NameNode
:\Windows\System32>start-yarn.cmd
tarting yarn daemons
:\Windows\System32>jps
12000 DataNode
6384 NodeManager
31300 Jps
24984 NameNode
19036 ResourceManager
::\Windows\System32>
```

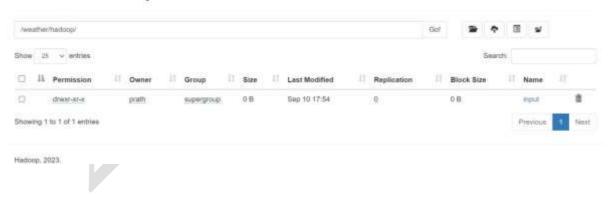
Open the browser and go to the URL localhost:9870



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>_

Browse Directory



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 8.001 999.9 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 0.001 999.9 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 8.001 999.9 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 8.001 999.9 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 8.001 999.9 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 8.001 999.9 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 8.001 999.9 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 8.001 999.9 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 8.001 999.9 000000
690190 13910 20060201 5 55.44 33.0 24 1006.3 24 943.9 24 15.0
24 10.7 24 22.0 28.9 8.001 999.9 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 8.001 999.9 000000
690190 13910 20060201 8 64.05 33.0 24 1006.3 24 943.9 24 15.0
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

```
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
        tf 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)")
me__ == "__main__":
                 main
     name
    reduce1()
```

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

 $hadoop \quad jar \quad \% HADOOP_HOME\% \\ share \\ hadoop \\ tools \\ lib \\ hadoop - streaming \\ *.jar^{\land}$

- -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

```
C. Chadoso 9.8. Outbin Houdoop jer $400000 jerSk/shara/hadoop/tools/10/hadoop-streaming-*jer - masser "pytton C./Semester/Ostakaslytis/Lab/Ex/masper.py" - reducer "pytton C./Semester/Ostakaslytis/Lab/Ex/masper.python C./Semester/Ostakasly
```

```
Total magasyte-milliseconds taken by all reduce tasks=3552676

Map-induce framework

Map input records=86

Map output records=86

Map output bytes=3672

Map output sterialized bytes=8278

Input split bytes=367

Map output interialized bytes=8288

Combine input records=80

Reduce input records=80

Reduce input groups=4

Reduce input groups=4

Reduce input records=86

Maduse upput records=86

Maduse input records=86

Marged Haps = 2

Falled Shuffles=8

Marged Haps = 2

Falled Shuffles=8

Marged Haps = 2

Falled Shuffles=8

Marged Hap outputs=2

GV time slapsed (no)=182

GV time slapsed (no)=182

GV time lapsed (no)=182

GV time lapsed (no)=182

File Input format (bytes) srapshot=157988203

Total convicted heap usage (bytes)=5785812

Maw Nap Mysical memory (bytes)=558587728

Pauk Maduse Virtual memory (bytes)=558587728

Pauk Maduse Virtual memory (bytes)=558587728

Fauk Maduse Virtual memory (bytes)=558587728

Fauk Maduse Virtual memory (bytes)=558587728

File Input format Counters

Bytes Resel=6149

File Input format Counters

Bytes Resel=6149

File Input format Counters

Bytes Resel=6149

File Output format Counters

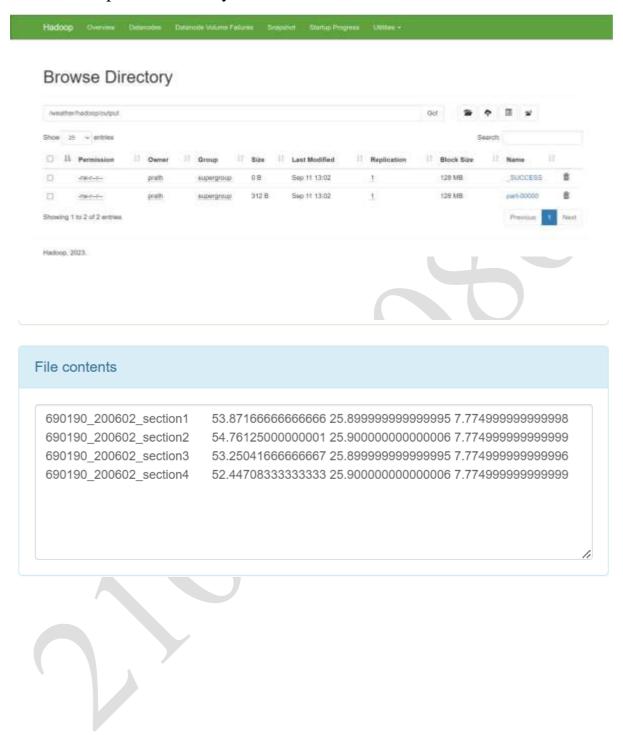
Bytes Resel=6149
```

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 25.89999999999995 7.7749999999998
690190_200602_section2 54.761250000000001 25.90000000000000 7.7749999999999
690190_200602_section3 53.25041666666667 25.8999999999999 7.77499999999999
690190_200602_section4 52.44708333333333 25.9000000000000000 7.77499999999999
```

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.