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

```
C:\Windows\System32>jps

14212 Jps

C:\Windows\System32>start-dfs.cmd

C:\Windows\System32>jps

12000 DataNode

16488 Jps

24904 NameNode

C:\Windows\System32>start-yarn.cmd

starting yarn daemons

C:\Windows\System32>jps

12000 DataNode

6384 NodeManager

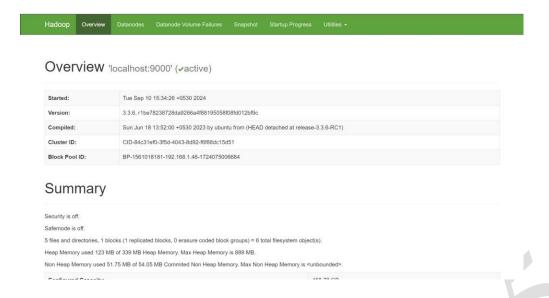
31300 Jps

24904 NameNode

29036 ResourceManager

C:\Windows\System32>
```

Open the browser and go to the URL localhost:9870

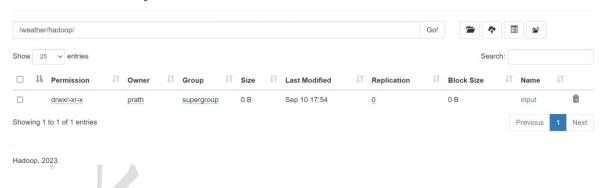


Create a directory in HDFS using the command:

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
24 10.7 24 22.0 28.9 0.00I 999.9 000000
                                                                         943.9 24
690190 13910 20060201_1 54.74 33.0 24 1006
24 10.7 24 22.0 28.9 0.00I 999.9 000000
                                              33.0 24 1006.3 24
                                                                         943.9 24
590190 13910 20060201_2 50.59 33.0 24 1006
24 10.7 24 22.0 28.9 0.00I 999.9 000000
590190 13910 20060201_3 51.67 33.0 24 1006
                                              33.0 24 1006.3 24
                                              33.0 24 1006.3 24
                                                                         943.9 24
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
                  22.0 28.9 0.00I 999.9 000000
24 10.7 24
24 10.7 24 22.0 28.9 0.00I 999.9 000000
                                                                                      15.0
696196 13910 22.0 28.9 0.001 999.9 000000
690190 13910 20060201_6 49.26 33.0 24 1006.3 24
24 10.7 24 22.0 28.9 0.001 999.9 000000
                                                                         943.9 24
590190 13910 20060201_7 55.44 33.0 24 1006
24 10.7 24 22.0 28.9 0.00I 999.9 000000
                                              33.0 24 1006.3 24 943.9 24
24 10.7 24
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:</pre>
            section = "section1"
        elif 10 < hour <= 16:
            section = "section2"
        elif 16 < hour <= 22:</pre>
            section = "section3"
        else:
            section = "section4"
        key out = f"{station}_{date}_{section}"
        value_out = f"{temp} {dew} {wind}"
        print(f"{key out}\t{value out}")
          _ == "__main__":
if __name_
    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}")
me__ == "__main__":
     name
    reduce1()
```

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

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

- -mapper "python C:\Semester7\DataAnalytics\Lab\Ex3\mapper.py" -reducer "python C:\Semester7\DataAnalytics\Lab\Ex3\reducer.py" $^{\land}$
- -input/weather/hadoop/input/sample_weather.txt -output /weather/hadoop/output

```
C:\hadoop-3.3.6\abbinhadoop jar W4000P_HOMEX/share/hadoop/tools/lib/hadoop-streaming-*.jar^-.mapper "python C:/Semester//Datahnalytics/Lab/Ex3/mapper.py" ^ -incut / weather/hadoop/injut/sample_weather.txt -output /weather/hadoop/output package/Dolar: [/c./Users/prath/AppDatalocal/Temp/hadoop-unjarS8271230776608983550] [] C:\Users/prath/AppDatalocal/Temp/hadoop/injut/sample_weather.txt -output /weather/hadoop/output package/Dolar: [/c./Users/prath/AppDatalocal/Temp/hadoop-unjarS8271230776608983550] [] C:\Users/prath/AppDatalocal/Temp/hadoop-unjarS8271230776608983550] [] C:\Users/prath/AppDatalocal/Temp/hadoop-unjarS8271230776608983550] [] C:\Users/prath/AppDatalocal/Temp/hadoop-unjarS827123077660825] [] C:\Users/prath/Prath/AppDatalocal/Temp/hadoop-unjarS827123077660825] [] C:\Users
```

```
Total megabyte-milliseconds taken by all reduce tasks=3352576

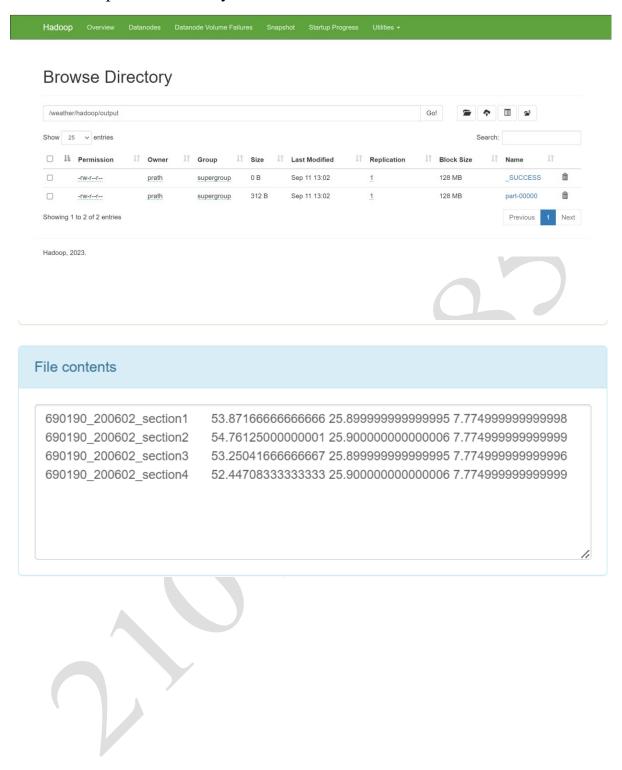
Map-Reduce Framework
Map input records=96
Map output records=96
Map output pytes=3672
Map output materialized bytes=3876
Input split bytes=286
Combine input records=90
Combine input records=90
Combine input records=90
Reduce input groups=4
Reduce input groups=4
Reduce input records=96
Reduce input records=4
Spilled Records=192
Shuffled Maps = 2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=182
CPU time spent (ms)=851
Physical memory (bytes) snapshot=948191744
Virtual memory (bytes) snapshot=984191744
Virtual memory (bytes) snapshot=984819174
Virtual memory (bytes) snapshot=98492412928
Peak Map Virtual memory (bytes)=595987328
Peak Reduce Virtual memory (bytes)=595987328
Peak Reduce Virtual memory (bytes)=277481152
Peak Reduce Virtual memory (bytes)=277481152
Peak Reduce Virtual memory (bytes)=8927481152
Peak Reduce Fhysical memory (bytes)=595987328
Peak Reduce Fhysical m
```

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 54.76125000000001 25.90000000000000 7.7749999999999
690190_200602_section3 53.25041666666667 25.8999999999999 7.7749999999999
690190_200602_section4 52.44708333333333 25.900000000000000 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.