

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

```
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

Overview 'localhost:9000' (active)

| | |
|----------------|--|
| Started: | Tue Sep 10 15:34:26 +0530 2024 |
| Version: | 3.3.6-116a762367295a0285a498100098000012ef5c |
| Compiled: | Sun Jun 18 15:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1) |
| Cluster ID: | CID-84c25e0-3f5e-4043-8d92-49884c15d51 |
| Block Pool ID: | BP-1561318181-192.168.1.48-1724075008884 |

Summary

Security is off.
Balancer is off.

5 files and directories, 1 blocks (1 replicated blocks, 0 erasure coded block groups) = 8 total filesystem object(s).

Heap Memory used 123 MB of 339 MB Heap Memory. Max Heap Memory is 800 MB.
Non-Heap Memory used 51.75 MB of 54.05 MB Committed Non-Heap Memory. Max Non-Heap Memory is unbounded.

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

/weather/hadoop/

Show 25 entries

| Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|------------|-------|------------|------|---------------|-------------|------------|-------|
| drwxr-xr-x | prath | supergroup | 0 B | Sep 10 17:54 | 0 | 0 B | input |

Showing 1 to 1 of 1 entries

Previous Next

Hadoop, 2023.

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

```
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
```

```

C:\hadoop-3.3.6\sbin>hadoop jar $HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-*.jar ^ -mapper "python C:/Semester7/DatAnalytics/Lab/Ex3/mapper.py" ^ -reducer "python C:/Semester7/DatAnalytics/Lab/Ex3/reducer.py" ^ -input /weather/hadoop/input/sample_weather.txt -output /weather/hadoop/output
packageJobJar: [/C:/Users/prath/AppData/Local/Temp/hadoop-unjar5827128779668250002/] [] C:/Users/prath/AppData/Local/Temp/hadoop-unjar5827128779668250002.jar tmpDir=null
2024-09-11 13:02:09,113 INFO client.DefaultHadoopFallbackProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-11 13:02:09,124 INFO client.DefaultHadoopFallbackProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-11 13:02:09,188 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-11 13:02:09,269 INFO mapreduce.JobSubmitter: number of splits:2
2024-09-11 13:02:09,393 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1726037706825_0002
2024-09-11 13:02:09,393 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-09-11 13:02:09,553 INFO conf.Configuration: resource-types.xml not found
2024-09-11 13:02:09,553 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-09-11 13:02:09,619 INFO impl.YarnClientImpl: Submitted application application_1726037706825_0002
2024-09-11 13:02:09,654 INFO mapreduce.Job: The url to track the job: http://HAMA01:8088/proxy/application_1726037706825_0002/
2024-09-11 13:02:09,654 INFO mapreduce.Job: Running job: job_1726037706825_0002
2024-09-11 13:02:11,862 INFO mapreduce.Job: Job job_1726037706825_0002 running in uber mode : false
2024-09-11 13:02:12,965 INFO mapreduce.Job: map 100% reduce 0%
2024-09-11 13:02:20,895 INFO mapreduce.Job: map 100% reduce 100%
2024-09-11 13:02:20,113 INFO mapreduce.Job: Job job_1726037706825_0002 completed successfully
2024-09-11 13:02:20,195 INFO mapreduce.Job: Counters: 54

    File System Counters
      FILE: Number of bytes read=3876
      FILE: Number of bytes written=86964
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=16375
      HDFS: Number of bytes written=312
      HDFS: Number of read operations=11
      HDFS: Number of large read operations=0
      HDFS: Number of write operations=2
      HDFS: Number of bytes read erasure-coded=0

    Job Counters
      Launched map tasks=2
      Launched reduce tasks=1
      Data-local map tasks=2
      Total time spent by all maps in occupied slots (ms)=7175
      Total time spent by all reduces in occupied slots (ms)=3274
      Total time spent by all map tasks (ms)=7175
      Total time spent by all reduce tasks (ms)=3274
      Total vcore-milliseconds taken by all map tasks=7175
      Total vcore-milliseconds taken by all reduce tasks=3274
      Total megabyte-milliseconds taken by all map tasks=7367200
      Total megabyte-milliseconds taken by all reduce tasks=3352576

```

```

    Total megabyte-milliseconds taken by all reduce tasks=3352576
Map-Reduce Framework
  Map input records=86
  Map output records=96
  Map output bytes=3672
  Map output materialized bytes=3876
  Input split bytes=226
  Combine input records=0
  Combine output records=0
  Reduce input groups=4
  Reduce shuffle bytes=3876
  Reduce input records=96
  Reduce output records=4
  Spilled Records=192
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=182
  CPU time spent (ms)=451
  Physical memory (bytes) snapshot=9440191744
  Virtual memory (bytes) snapshot=1579802624
  Total committed heap usage (bytes)=877658112
  Peak Map Physical memory (bytes)=352460800
  Peak Map Virtual memory (bytes)=559567328
  Peak Reduce Physical memory (bytes)=274481152
  Peak Reduce Virtual memory (bytes)=492412928

  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0

  File Input Format Counters
    Bytes Read=16149
  File Output Format Counters
    Bytes Written=812
2024-09-11 13:02:29,199 INFO streaming.StreamJob: Output directory: /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 7.774999999999999
690190_200602_section2 54.761250000000001 25.900000000000006 7.774999999999999
690190_200602_section3 53.250416666666667 25.899999999999995 7.774999999999999
690190_200602_section4 52.447083333333333 25.900000000000006 7.774999999999999

```

View the output on the file system in browser

Navigation: Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

Path: /weather/hadoop/output

Show: 25 entries Search:

| Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|------------|-------|------------|-------|---------------|-------------|------------|------------|
| -rw-r--r-- | prath | supergroup | 0 B | Sep 11 13:02 | 1 | 128 MB | SUCCESS |
| -rw-r--r-- | prath | supergroup | 312 B | Sep 11 13:02 | 1 | 128 MB | part-00000 |

Showing 1 to 2 of 2 entries

Hadoop, 2023.

File contents

```
690190_200602_section1 53.87166666666666 25.899999999999995 7.774999999999998
690190_200602_section2 54.761250000000001 25.900000000000006 7.774999999999999
690190_200602_section3 53.25041666666667 25.899999999999995 7.774999999999996
690190_200602_section4 52.44708333333333 25.900000000000006 7.774999999999999
```

RESULT:

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