EXP 3: Map Reduce program to process a weather dataset.

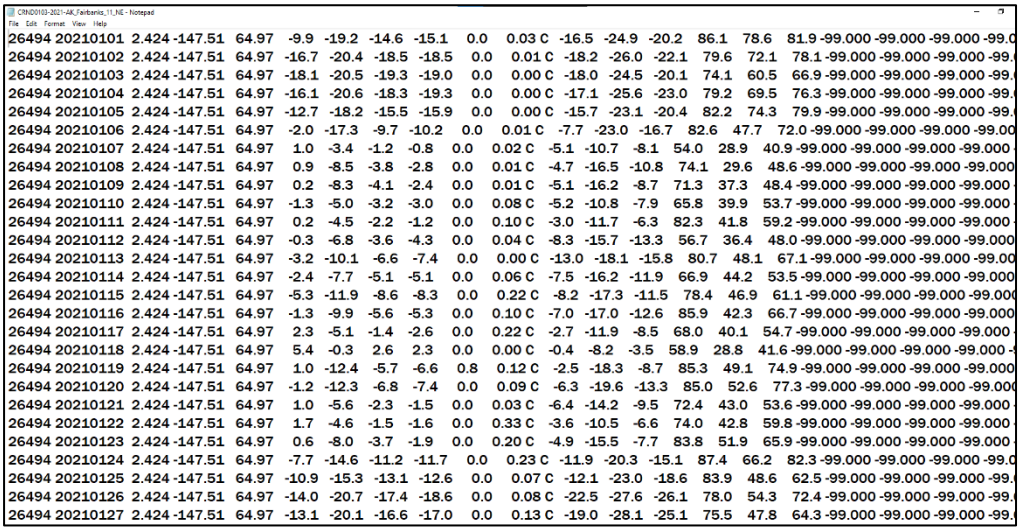
AIM:

To implement MapReduce program to process a weather dataset.

PROCEDURE:

Step 1: Create Data File:

Create a file named "sample_weather.txt" and populate it with text data that you wish to analyse.



CRND0103-2021-AK_Fairbanks_11_NE - Notepad

File Edit Format View Help

```
26494 20210101 2.424 -147.51 64.97  -9.9 -19.2 -14.6 -15.1  0.0  0.03 C -16.5 -24.9 -20.2  86.1  78.6  81.9 -99.000 -99.000 -99.000 -99.0
26494 20210102 2.424 -147.51 64.97 -16.7 -20.4 -18.5 -18.5  0.0  0.01 C -18.2 -26.0 -22.1  79.6  72.1  78.1 -99.000 -99.000 -99.000 -99.
26494 20210103 2.424 -147.51 64.97 -18.1 -20.5 -19.3 -19.0  0.0  0.00 C -18.0 -24.5 -20.1  74.1  60.5  66.9 -99.000 -99.000 -99.000 -99.
26494 20210104 2.424 -147.51 64.97 -16.1 -20.6 -18.3 -19.3  0.0  0.00 C -17.1 -25.6 -23.0  79.2  69.5  76.3 -99.000 -99.000 -99.000 -99.
26494 20210105 2.424 -147.51 64.97 -12.7 -18.2 -15.5 -15.9  0.0  0.00 C -15.7 -23.1 -20.4  82.2  74.3  79.9 -99.000 -99.000 -99.000 -99.
26494 20210106 2.424 -147.51 64.97  -2.0 -17.3  -9.7 -10.2  0.0  0.01 C  -7.7 -23.0 -16.7  82.6  47.7  72.0 -99.000 -99.000 -99.000 -99.00
26494 20210107 2.424 -147.51 64.97   1.0  -3.4  -1.2  -0.8  0.0  0.02 C  -5.1 -10.7  -8.1  54.0  28.9  40.9 -99.000 -99.000 -99.000 -99.000
26494 20210108 2.424 -147.51 64.97   0.9  -8.5  -3.8  -2.8  0.0  0.01 C  -4.7 -16.5 -10.8  74.1  29.6  48.6 -99.000 -99.000 -99.000 -99.000
26494 20210109 2.424 -147.51 64.97   0.2  -8.3  -4.1  -2.4  0.0  0.01 C  -5.1 -16.2  -8.7  71.3  37.3  48.4 -99.000 -99.000 -99.000 -99.000
26494 20210110 2.424 -147.51 64.97  -1.3  -5.0  -3.2  -3.0  0.0  0.08 C  -5.2 -10.8  -7.9  65.8  39.9  53.7 -99.000 -99.000 -99.000 -99.000
26494 20210111 2.424 -147.51 64.97   0.2  -4.5  -2.2  -1.2  0.0  0.10 C  -3.0 -11.7  -6.3  82.3  41.8  59.2 -99.000 -99.000 -99.000 -99.000
26494 20210112 2.424 -147.51 64.97  -0.3  -6.8  -3.6  -4.3  0.0  0.04 C  -8.3 -15.7 -13.3  56.7  36.4  48.0 -99.000 -99.000 -99.000 -99.000
26494 20210113 2.424 -147.51 64.97  -3.2 -10.1  -6.6  -7.4  0.0  0.00 C -13.0 -18.1 -15.8  80.7  48.1  67.1 -99.000 -99.000 -99.000 -99.00
26494 20210114 2.424 -147.51 64.97  -2.4  -7.7  -5.1  -5.1  0.0  0.06 C  -7.5 -16.2 -11.9  66.9  44.2  53.5 -99.000 -99.000 -99.000 -99.000
26494 20210115 2.424 -147.51 64.97  -5.3 -11.9  -8.6  -8.3  0.0  0.22 C  -8.2 -17.3 -11.5  78.4  46.9  61.1 -99.000 -99.000 -99.000 -99.00
26494 20210116 2.424 -147.51 64.97  -1.3  -9.9  -5.6  -5.3  0.0  0.10 C  -7.0 -17.0 -12.6  85.9  42.3  66.7 -99.000 -99.000 -99.000 -99.000
26494 20210117 2.424 -147.51 64.97   2.3  -5.1  -1.4  -2.6  0.0  0.22 C  -2.7 -11.9  -8.5  68.0  40.1  54.7 -99.000 -99.000 -99.000 -99.000
26494 20210118 2.424 -147.51 64.97   5.4  -0.3   2.6   2.3  0.0  0.00 C  -0.4  -8.2  -3.5  58.9  28.8  41.6 -99.000 -99.000 -99.000 -99.000 -
26494 20210119 2.424 -147.51 64.97   1.0 -12.4  -5.7  -6.6  0.8  0.12 C  -2.5 -18.3  -8.7  85.3  49.1  74.9 -99.000 -99.000 -99.000 -99.000
26494 20210120 2.424 -147.51 64.97  -1.2 -12.3  -6.8  -7.4  0.0  0.09 C  -6.3 -19.6 -13.3  85.0  52.6  77.3 -99.000 -99.000 -99.000 -99.00
26494 20210121 2.424 -147.51 64.97   1.0  -5.6  -2.3  -1.5  0.0  0.03 C  -6.4 -14.2  -9.5  72.4  43.0  53.6 -99.000 -99.000 -99.000 -99.000
26494 20210122 2.424 -147.51 64.97   1.7  -4.6  -1.5  -1.6  0.0  0.33 C  -3.6 -10.5  -6.6  74.0  42.8  59.8 -99.000 -99.000 -99.000 -99.000
26494 20210123 2.424 -147.51 64.97   0.6  -8.0  -3.7  -1.9  0.0  0.20 C  -4.9 -15.5  -7.7  83.8  51.9  65.9 -99.000 -99.000 -99.000 -99.000
26494 20210124 2.424 -147.51 64.97  -7.7 -14.6 -11.2 -11.7  0.0  0.23 C -11.9 -20.3 -15.1  87.4  66.2  82.3 -99.000 -99.000 -99.000 -99.0
26494 20210125 2.424 -147.51 64.97 -10.9 -15.3 -13.1 -12.6  0.0  0.07 C -12.1 -23.0 -18.6  83.9  48.6  62.5 -99.000 -99.000 -99.000 -99.
26494 20210126 2.424 -147.51 64.97 -14.0 -20.7 -17.4 -18.6  0.0  0.08 C -22.5 -27.6 -26.1  78.0  54.3  72.4 -99.000 -99.000 -99.000 -99.
26494 20210127 2.424 -147.51 64.97 -13.1 -20.1 -16.6 -17.0  0.0  0.13 C -19.0 -28.1 -25.1  75.5  47.8  64.3 -99.000 -99.000 -99.000 -99.
```

Step 2: Mapper Logic - mapper.py:

Create a file named "mapper.py" to implement the logic for the mapper. The mapper will read input data from STDIN, split lines into words, and output each word with its count.

mapper.py:

```
#!/usr/bin/env python3
import sys

THRESHOLD_HOT = 30  # You can adjust this threshold
THRESHOLD_COLD = 0  # You can adjust this threshold

for line in sys.stdin:
    line = line.strip()
    parts = line.split(',')

    if len(parts) >= 4:
        date = parts[0]
        max_temp = float(parts[3])
        min_temp = float(parts[4])

        if max_temp > THRESHOLD_HOT:
            print(f"hot\t{date}\t{max_temp}")

        if min_temp < THRESHOLD_COLD:
            print(f"cold\t{date}\t{min_temp}")
```

Step 3: Reducer Logic - reducer.py:
Create a file named "reducer.py" to implement the logic for the reducer. The reducer will aggregate the occurrences of each word and generate the final output.

reducer.py:

```
#!/usr/bin/env python3
import sys

current_marker = None
current_max = None
current_min = None
hot_days = []
cold_days = []

for line in sys.stdin:
    line = line.strip()
    marker, date, temperature = line.split('\t')

    temperature = float(temperature)

    if marker == "hot":
        hot_days.append((date, temperature))
    elif marker == "cold":
        cold_days.append((date, temperature))

# Output the results
print("Hot Days:")
for date, temp in hot_days:
    print(f"{date}\t{temp}")

print("\nCold Days:")
for date, temp in cold_days:
    print(f"{date}\t{temp}")
```

Step 4: Prepare Hadoop Environment:
Start the Hadoop daemons and create a directory in HDFS to store your data. Run the following commands to store the data in the WeatherData Directory.

```
start-all.cmd
cd C:/Hadoop/sbin
hdfs dfs -mkdir /WeatherData
hdfs dfs -put C:/Users/user/Documents/DataAnalytics2/input.txt /WeatherData
hadoop jar C:\hadoop\share\hadoop\tools\lib\hadoop-streaming-3.3.6.jar ^
-input /user/input/sample_weather.txt ^
-output /user/output ^
-mapper "C:\Users\Navneeth\Desktop\SEM-7\HadoopLab\exp3\mapper.py" ^
-reducer "C:\Users\Navneeth\Desktop\SEM-7\HadoopLab\exp3\reducer.py"
```

Step 5: Check Output:

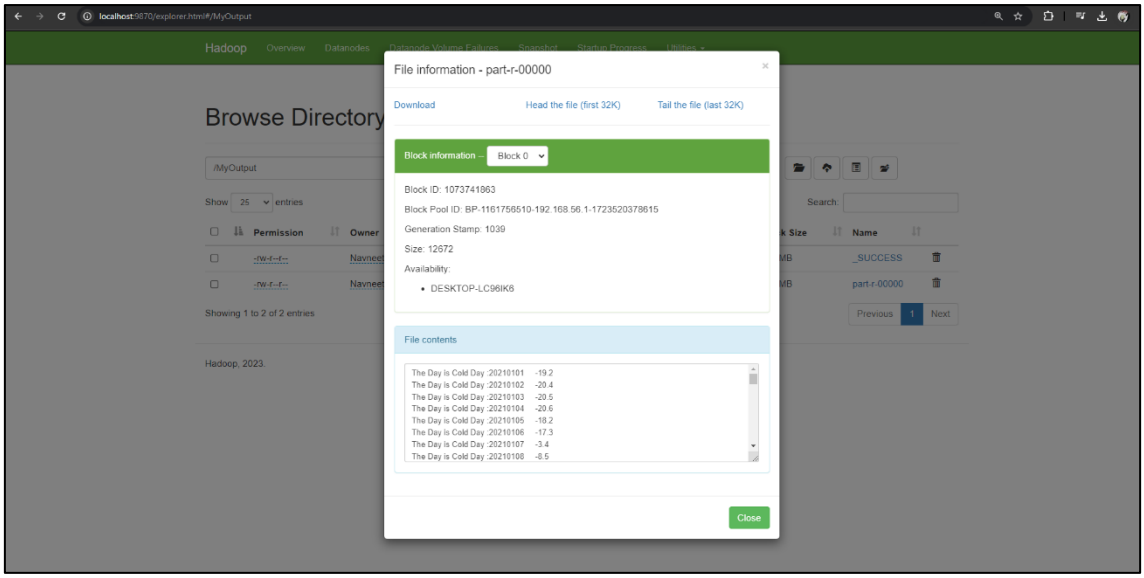
Check the output of the Word Count program in the specified HDFS output directory.

hdfs dfs -cat /WeatherData/output/part-00000

OUTPUT:

```
C:\Windows\system32>hadoop jar "C:\Users\Navneeth\Desktop\SEM-7\HadoopLab\Hadoop-Exp-3-WeatherDataset\out\artifacts\Hadoop_Exp
set.jar" /CRND0103-2021-AK_Fairbanks_11_NE.txt /MyOutput
2024-09-02 20:41:30,771 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-02 20:41:31,367 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the To
 ToolRunner to remedy this.
2024-09-02 20:41:31,412 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Navnee
2024-09-02 20:41:32,300 INFO input.FileInputFormat: Total input files to process : 1
2024-09-02 20:41:32,389 INFO mapreduce.JobSubmitter: number of splits:1
2024-09-02 20:41:32,532 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1725289772902_0001
2024-09-02 20:41:32,532 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-09-02 20:41:32,733 INFO conf.Configuration: resource-types.xml not found
2024-09-02 20:41:32,734 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-09-02 20:41:33,186 INFO impl.YarnClientImpl: Submitted application application_1725289772902_0001
2024-09-02 20:41:33,237 INFO mapreduce.Job: The url to track the job: http://DESKTOP-LC96IK6:8088/proxy/application_1725289772
2024-09-02 20:41:33,238 INFO mapreduce.Job: Running job: job_1725289772902_0001
2024-09-02 20:41:42,446 INFO mapreduce.Job: Job job_1725289772902_0001 running in uber mode : false
```

The Day is Cold Day :20210101	-19.2
The Day is Cold Day :20210102	-20.4
The Day is Cold Day :20210103	-20.5
The Day is Cold Day :20210104	-20.6
The Day is Cold Day :20210105	-18.2
The Day is Cold Day :20210106	-17.3
The Day is Cold Day :20210107	-3.4
The Day is Cold Day :20210108	-8.5
The Day is Cold Day :20210109	-8.3
The Day is Cold Day :20210110	-5.0
The Day is Cold Day :20210111	-4.5
The Day is Cold Day :20210112	-6.8
The Day is Cold Day :20210113	-10.1
The Day is Cold Day :20210114	-7.7
The Day is Cold Day :20210115	-11.9
The Day is Cold Day :20210116	-9.9
The Day is Cold Day :20210117	-5.1
The Day is Cold Day :20210118	-0.3



RESULT:

Thus, the program for weather dataset using Map Reduce has been executed successfully.