

Expt-3

Map Reduce program to process a weather dataset.

AIM:

To implement MapReduce program to process a weather dataset.

PROCEDURE:

1. Create Weather Dataset:

`nano weather_data.txt` **Example**

content:

20220101 30.5

20220102 29.8

2. Mapper Program (mapper.py):

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

```
for line in sys.stdin: line =
```

```
line.strip()
```

```
month = line[4:6] # Extracting month temp
```

```
= line[7:11] # Extracting temperature
```

```
print(f'{month}\t{temp}')
```

3. Reducer Program (reducer.py):

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

```
current_month = None current_max_temp
```

```
= -float('inf')
```

```
for line in sys.stdin: line = line.strip()
```

```
month, temp = line.split('\t')
```

```
try:
```

```
temp = float(temp)
```

```
except ValueError:
    continue

if current_month == month:
    current_max_temp = max(current_max_temp, temp) else:

if current_month:
    print(f'{current_month}\t{current_max_temp}')
    current_month = month
    current_max_temp
    = temp

if current_month == month:
    print(f'{current_month}\t{current_max_temp}')
```

4. Run the Program:

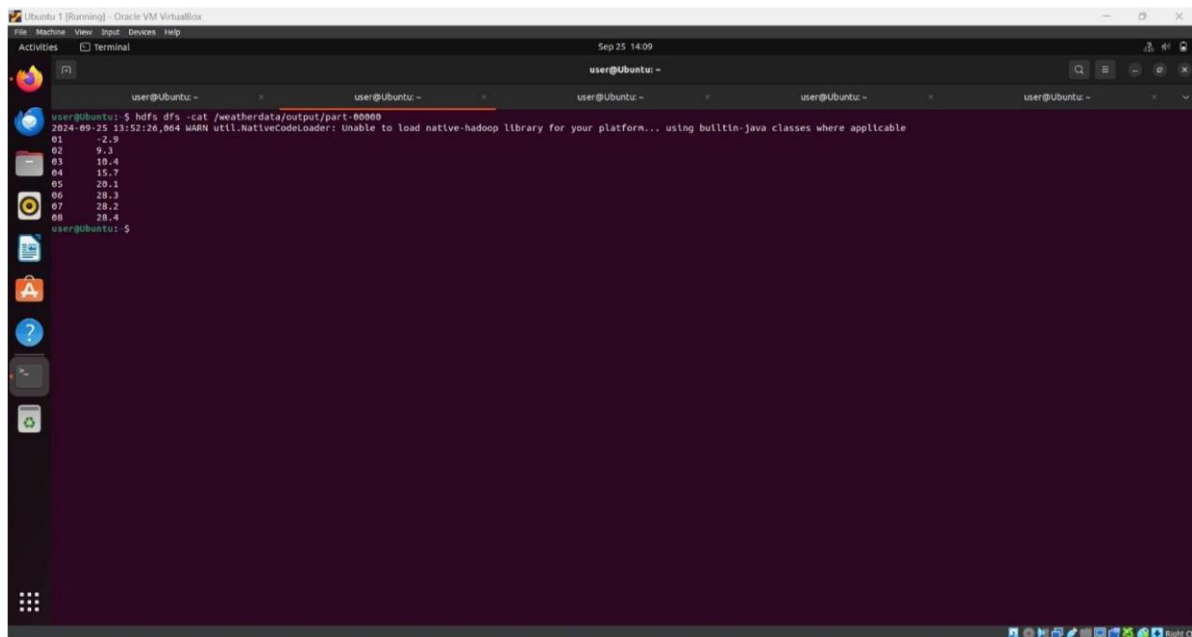
```
hdfs dfs -mkdir /weatherdata
hdfs dfs -copyFromLocal weather_data.txt /weatherdata

hadoop jar $HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-*.jar \
-input /weatherdata/weather_data.txt \
-output /weatherdata/output \
-mapper mapper.py \
-reducer reducer.py
```

5. Check Output:

```
hdfs dfs -cat /weatherdata/output/part-00000
```

OUTPUT:



The screenshot shows a terminal window titled "Ubuntu 1 [Running] - Oracle VM VirtualBox". The terminal displays the output of a MapReduce job. The command executed is `user@ubuntu:~$ hdfs dfs -cat /weatherdata/output/part-00000`. The output shows a warning message: `2024-09-25 13:52:26,864 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable`. This is followed by a list of 8 lines of weather data, each starting with a line number (01 to 08) and a temperature value. The terminal window also shows the Ubuntu desktop environment with various icons on the left and bottom.

```
user@ubuntu:~$ hdfs dfs -cat /weatherdata/output/part-00000
2024-09-25 13:52:26,864 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
01 -2.9
02 9.3
03 10.4
04 15.7
05 29.1
06 28.3
07 28.2
08 28.4
user@ubuntu:~$
```

RESULT:

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