Exp No: 10   **Date: 15.09.2024**

IMPLEMENT THE MAX TEMPERATURE MAPREDUCE PROGRAM TO IDENTIFY THE YEAR WISE MAXIMUM TEMPERATURE FROM SENSOR DATA

AIM:

To implement the Max temperature Map Reduce program to identify the year-wise maximum temperature from the sensor data.

PROCEDURE:

1. Open command prompt and run as administrator

Go to hadoop sbin directory

```
C:\Windows\system32>cd C:\Hadoop\sbin

C:\Hadoop\sbin>_
```
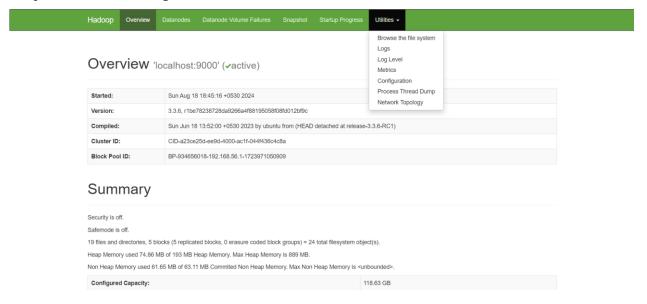
Note:

1. Check hadoop/data/datanode and hadoop/data/namenode and if both folders are empty, type "hdfs namenode -format".
2. Check python version with "python --version".
3. Check "C:\Python39\" is added in Environment variables > System variables > Path, if not add your python path.
4. Check Environment variables > System variables > HADOOP_HOME is set as "C:\Hadoop".

2. Start Hadoop Services

start-dfs.cmd

start-yarn.cmd

```
C:\Hadoop\sbin>start-dfs.cmd

C:\Hadoop\sbin>start-yarn.cmd
starting yarn daemons

C:\Hadoop\sbin>jps
13120 NameNode
2384 NodeManager
4100 DataNode
7956 ResourceManager
9124 Jps
```

3. Open the browser and go to the URL "localhost:9870"

4. Create a Directory in HDFS

hdfs dfs -mkdir -p /maxweather/hadoop/input


```
C:\Hadoop\sbin>hdfs dfs -mkdir -p /maxweather/hadoop/input
```

5. Copy the Input File to HDFS

hdfs dfs -put C:/Users/Admin/max_weather_cc/weather_data.txt /maxweather/hadoop/input/


```
C:\Hadoop\sbin>hdfs dfs -put C:/Users/Admin/max_weather_cc/weather_data.txt /maxweather/hadoop/input/

C:\Hadoop\sbin>hdfs dfs -ls /maxweather/hadoop/input/
Found 1 items
-rw-r--r--   1 Admin supergroup       3264 2024-09-15 14:49 /maxweather/hadoop/input/weather_data.txt

C:\Hadoop\sbin>hdfs dfs -cat /maxweather/hadoop/input/weather_data.txt
0029029070999992021010720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00311+99999098351ADDGF102991999999999999999999
0029029070999992021020720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00121+99999098351ADDGF102991999999999999999999
0029029070999992021030720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00231+99999098351ADDGF102991999999999999999999
0029029070999992021040720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00411+99999098351ADDGF102991999999999999999999
0029029070999992021050720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00501+99999098351ADDGF102991999999999999999999
0029029070999992021060720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00211+99999098351ADDGF102991999999999999999999
0029029070999992022010720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00221+99999098351ADDGF102991999999999999999999
0029029070999992022020720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00311+99999098351ADDGF102991999999999999999999
0029029070999992022030720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00401+99999098351ADDGF102991999999999999999999
0029029070999992022040720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00501+99999098351ADDGF102991999999999999999999
0029029070999992022050720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00311+99999098351ADDGF102991999999999999999999
0029029070999992022060720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00241+99999098351ADDGF102991999999999999999999
0029029070999992023010720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00301+99999098351ADDGF102991999999999999999999
0029029070999992023020720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00421+99999098351ADDGF102991999999999999999999
0029029070999992023030720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00531+99999098351ADDGF102991999999999999999999
0029029070999992023040720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00211+99999098351ADDGF102991999999999999999999
0029029070999992023050720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00301+99999098351ADDGF102991999999999999999999
0029029070999992023060720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00311+99999098351ADDGF102991999999999999999999
0029029070999992024010720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00111+99999098351ADDGF102991999999999999999999
0029029070999992024020720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00231+99999098351ADDGF102991999999999999999999
0029029070999992024030720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00301+99999098351ADDGF102991999999999999999999
0029029070999992024040720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00411+99999098351ADDGF102991999999999999999999
0029029070999992024050720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00501+99999098351ADDGF102991999999999999999999
0029029070999992024060720004+64333+023450FM-12+000599999V0202501N027819999999N0000001N9+00211+99999098351ADDGF102991999999999999999999
```

Note:

mapper.py:

```python
#!/usr/bin/env python3
import sys
for line in sys.stdin:
    line = line.strip()
    if not line:
        continue
    try:
        # Extract relevant fields from the raw data
        year = line[15:19]  # Extract year from the line
        temp_str = line[90:92]  # Extract temperature from the line
        quality = line[92:93]  # Extract quality indicator

        # Check if the temperature is valid and the quality is acceptable
        if temp_str != "+9999" and quality in ['0', '1', '4', '5', '9']:
            temp = int(temp_str)
            print(f"{year}\t{temp}")
    except Exception as e:
        sys.stderr.write(f"Error processing line: {line}\nException: {str(e)}\n")
```

combiner.py:

```python
#!/usr/bin/env python
import sys
from collections import defaultdict
current_year = None
temp_set = set()
for line in sys.stdin:
    line = line.strip()
    if not line:
        continue
    try:
        year, temp_str = line.split('\t')
        temp = int(temp_str)
        if current_year == year:
            temp_set.add(temp)
```

```python
        else:
            if current_year:
                # Print the year and the set of temperatures
                for t in temp_set:
                    print(f"{current_year}\t{t}")
            current_year = year
            temp_set = {temp}
    except Exception as e:
        sys.stderr.write(f"Error processing line: {line}\nException: {str(e)}\n")
# Output the set of temperatures for the last year
if current_year:
    for t in temp_set:
        print(f"{current_year}\t{t}")
```

reducer.py:

```python
#!/usr/bin/env python
import sys
current_year = None
max_temp = None
for line in sys.stdin:
    line = line.strip()
    if not line:
        continue
    try:
        year, temp_str = line.split('\t')
        temp = int(temp_str)
        if current_year == year:
            if temp > max_temp:
                max_temp = temp
        else:
            if current_year:
                # Print the maximum temperature for the previous year
                print(f"{current_year}\t{max_temp}")
            current_year = year
            max_temp = temp
    except Exception as e:
        sys.stderr.write(f"Error processing line: {line}\nException: {str(e)}\n")
```

Output the maximum temperature for the last year

if current_year:

   print(f"{current_year}\t{max_temp}")

6. Run the Hadoop Streaming Job

hadoop jar %HADOOP_HOME%\share\hadoop\tools\lib\hadoop-streaming-*.jar ^

 -input /maxweather/hadoop/input/weather_data.txt ^

 -output /maxweather/hadoop/output ^

 -mapper "python C:\Users\Admin\max_weather_cc\mapper.py" ^

 -combiner "python C:\Users\Admin\max_weather_cc\combiner.py" ^

 -reducer "python C:\Users\Admin\max_weather_cc\reducer.py" ^

 -file C:\Users\Admin\max_weather_cc\mapper.py ^

 -file C:\Users\Admin\max_weather_cc\combiner.py ^

 -file C:\Users\Admin\max_weather_cc\reducer.py

```
C:\Hadoop\sbin>hadoop jar %HADOOP_HOME%\share\hadoop\tools\lib\hadoop-streaming-*.jar ^
More?    -input /maxweather/hadoop/input/weather_data.txt ^
More?    -output /maxweather/hadoop/output ^
More?    -mapper "python C:\Users\Admin\max_weather_cc\mapper.py" ^
More?    -combiner "python C:\Users\Admin\max_weather_cc\combiner.py" ^
More?    -reducer "python C:\Users\Admin\max_weather_cc\reducer.py" ^
More?    -file C:\Users\Admin\max_weather_cc\mapper.py ^
More?    -file C:\Users\Admin\max_weather_cc\combiner.py ^
More?    -file C:\Users\Admin\max_weather_cc\reducer.py
2024-09-15 16:05:06,874 WARN streaming.StreamJob: -file option is deprecated, please use generic option
packageJobJar: [C:\Users\Admin\max_weather_cc\mapper.py, C:\Users\Admin\max_weather_cc\combiner.py, C:\U
[] C:\Users\Admin\AppData\Local\Temp\streamjob4190031441542396910.jar tmpDir=null
2024-09-15 16:05:09,340 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at
2024-09-15 16:05:09,758 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at
2024-09-15 16:05:11,009 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hado
2024-09-15 16:05:12,204 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-15 16:05:12,612 INFO mapreduce.JobSubmitter: number of splits:2
2024-09-15 16:05:12,963 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1726391798653_0008
2024-09-15 16:05:12,963 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-09-15 16:05:13,260 INFO conf.Configuration: resource-types.xml not found
2024-09-15 16:05:13,262 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-09-15 16:05:13,471 INFO impl.YarnClientImpl: Submitted application application_1726391798653_0008
2024-09-15 16:05:13,575 INFO mapreduce.Job: The url to track the job: http://DESKTOP-TF65P79:8088/proxy,
2024-09-15 16:05:13,583 INFO mapreduce.Job: Running job: job_1726391798653_0008
2024-09-15 16:05:33,469 INFO mapreduce.Job: Job job_1726391798653_0008 running in uber mode : false
2024-09-15 16:05:33,472 INFO mapreduce.Job:  map 0% reduce 0%
2024-09-15 16:05:54,357 INFO mapreduce.Job:  map 100% reduce 0%
2024-09-15 16:06:06,579 INFO mapreduce.Job:  map 100% reduce 100%
2024-09-15 16:06:08,617 INFO mapreduce.Job: Job job_1726391798653_0008 completed successfully
2024-09-15 16:06:08,873 INFO mapreduce.Job: Counters: 55
```

```
        File Input Format Counters
                Bytes Read=4896
        File Output Format Counters
                Bytes Written=32
2024-09-15 16:06:08,910 INFO streaming.StreamJob: Output directory: /maxweather/hadoop/output
```
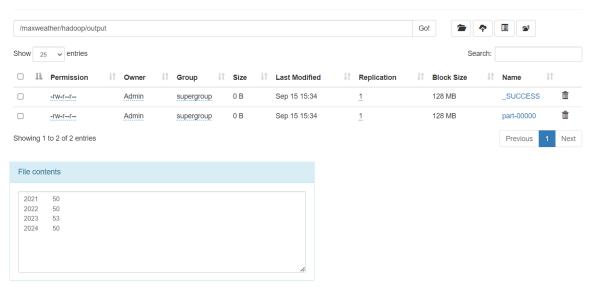
7. View the Output

hdfs dfs -cat /maxweather/hadoop/output/part-00000

```
C:\Hadoop\sbin>hdfs dfs -cat /maxweather/hadoop/output/part-00000
2021    50
2022    50
2023    53
2024    50
```

8. Once the map reduce operations are performed successfully, the output will be present in the specified directory.

"/maxweather/hadoop/output/part-00000"

Browse Directory

| | | /maxweather/hadoop/output | | | | | Go! | | | | |

Show 25 entries   Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	-rw-r--r--	Admin	supergroup	0 B	Sep 15 15:34	1	128 MB	_SUCCESS	🗑
☐	-rw-r--r--	Admin	supergroup	0 B	Sep 15 15:34	1	128 MB	part-00000	🗑

Showing 1 to 2 of 2 entries   Previous 1 Next

File contents
2021   50
2022   50
2023   53
2024   50

9. Stop Hadoop Services

stop-dfs.cmd

stop-yarn.cmd

```
C:\Hadoop\sbin>stop-dfs.cmd
SUCCESS: Sent termination signal to the process with PID 3636.
SUCCESS: Sent termination signal to the process with PID 14040.

C:\Hadoop\sbin>stop-yarn.cmd
stopping yarn daemons
SUCCESS: Sent termination signal to the process with PID 10644.
SUCCESS: Sent termination signal to the process with PID 8316.

INFO: No tasks running with the specified criteria.

C:\Hadoop\sbin>
```

RESULT:

Thus, to implement the Max temperature Map Reduce program to identify the year-wise maximum temperature from the sensor data was completed successfully.