BIG DATA ANALYTICS

22AIE312 – LABSHEET 4

Implement a MapReduce program to process weather data and find the highest temperature recorded by a weather station.

- Implement the Mapper class. The mapper should read the input text and extract the temperature from each record. Output the station ID as the key and the temperature as the value.

- Implement the Reducer class. The reducer should receive the key-value pairs from the mapper and find the maximum temperature for each station.

- Write the driver class to configure and run the MapReduce job.

Start Hadoop

```
hadoop@DESKTOP-O7BKI3U:~$ $HADOOP_HOME/sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [DESKTOP-O7BKI3U]
Starting resourcemanager
Starting nodemanagers
hadoop@DESKTOP-O7BKI3U:~$ jps
1154 NameNode
2307 Jps
1572 SecondaryNameNode
1306 DataNode
1787 ResourceManager
1933 NodeManager
hadoop@DESKTOP-O7BKI3U:~$
```

Create and Upload Input Data to HDFS

```
hadoop@DESKTOP-O7BKI3U:~$ nano weather_data.txt
hadoop@DESKTOP-O7BKI3U:~$ hdfs dfs -mkdir -p /weather/input
hadoop@DESKTOP-O7BKI3U:~$ hdfs dfs -put weather_data.txt /weather/input/
put: `/weather/input/weather_data.txt': File exists
hadoop@DESKTOP-O7BKI3U:~$ hdfs dfs -ls /weather/input
Found 1 items
-rw-r--r--   1 hadoop supergroup        162 2025-03-12 09:49 /weather/input/weather_data.txt
hadoop@DESKTOP-O7BKI3U:~$
```

```
  GNU nano 4.8                                          weather_data.txt
USW00094889 2024-01-01  12
USW00094889 2024-01-02  15
USW00094889 2024-01-03  10
USC00123456 2024-01-01  22
USC00123456 2024-01-02  24
USC00123456 2024-01-03  21
```

MapReduce Program

```
hadoop@DESKTOP-O7BKI3U:~$ nano MaxTemperature.java
hadoop@DESKTOP-O7BKI3U:~$ hadoop com.sun.tools.javac.Main MaxTemperature.java
hadoop@DESKTOP-O7BKI3U:~$ jar cf maxtemperature.jar MaxTemperature*.class
hadoop@DESKTOP-O7BKI3U:~$ hadoop jar maxtemperature.jar MaxTemperature /weather/input /weather/output
2025-03-12 10:03:37,482 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2025-03-12 10:03:37,766 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-12 10:03:37,792 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/
hadoop/.staging/job_1741753748851_0001
2025-03-12 10:03:38,535 INFO input.FileInputFormat: Total input files to process : 1
2025-03-12 10:03:38,600 INFO mapreduce.JobSubmitter: number of splits:1
2025-03-12 10:03:38,790 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1741753748851_0001
2025-03-12 10:03:38,790 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-12 10:03:38,935 INFO conf.Configuration: resource-types.xml not found
```

Results

```
hadoop@DESKTOP-O7BKI3U:~$ hdfs dfs -ls /weather/output
Found 2 items
-rw-r--r--   1 hadoop supergroup          0 2025-03-12 10:03 /weather/output/_SUCCESS
-rw-r--r--   1 hadoop supergroup         30 2025-03-12 10:03 /weather/output/part-r-00000
hadoop@DESKTOP-O7BKI3U:~$ hdfs dfs -cat /weather/output/part-r-00000
USC00123456     24
USW00094889     15
hadoop@DESKTOP-O7BKI3U:~$
```