Labsheet 4

22AIE312 Big Data Analytics Labsheet 4

Name: Aniketh Vijesh
Roll No: AM.EN.U4AIE22009

Question

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.

Programs

WeatherMapper.java

```
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

import java.io.IOException;

public class WeatherMapper extends Mapper<Object, Text, Text, IntWritable>
{
    private Text stationID = new Text();
    private IntWritable temperature = new IntWritable();

    public void map(Object key, Text value, Context context) throws
IOException, InterruptedException {
        String[] parts = value.toString().split(",");
        if (parts.length == 5) {
            stationID.set(parts[0]);  // First column is the station ID
            temperature.set(Integer.parseInt(parts[4]));  // Last column
is the temperature
            context.write(stationID, temperature);
        }
```

```
    }
}
```

WeatherReducer.java

```
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

import java.io.IOException;

public class WeatherReducer extends Reducer<Text, IntWritable, Text,
IntWritable> {
    public void reduce(Text key, Iterable<IntWritable> values, Context
context) throws IOException, InterruptedException {
        int maxTemp = Integer.MIN_VALUE;

        for (IntWritable temp : values) {
            maxTemp = Math.max(maxTemp, temp.get());
        }

        context.write(key, new IntWritable(maxTemp));
    }
}
```

WeatherDriver.java

```
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WeatherDriver {
    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "Weather Max Temperature");

        job.setJarByClass(WeatherDriver.class);
        job.setMapperClass(WeatherMapper.class);
        job.setReducerClass(WeatherReducer.class);

        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
```

```
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
}
```

Input

```
ST003,2024,10,09,19

ST002,2023,11,24,24

ST003,2024,09,18,8

ST003,2023,11,12,35

ST004,2024,04,08,11

ST001,2023,07,16,30

ST002,2024,09,28,43

ST003,2023,03,01,27

ST001,2024,01,21,14

ST001,2023,11,08,40

ST002,2023,05,23,48

ST001,2023,02,15,11

ST002,2023,05,04,26

ST001,2023,08,16,9

ST004,2023,09,04,21

ST002,2024,01,17,17

ST003,2024,10,01,45

ST003,2024,04,09,3

ST002,2023,09,06,2

ST002,2023,02,01,43
```

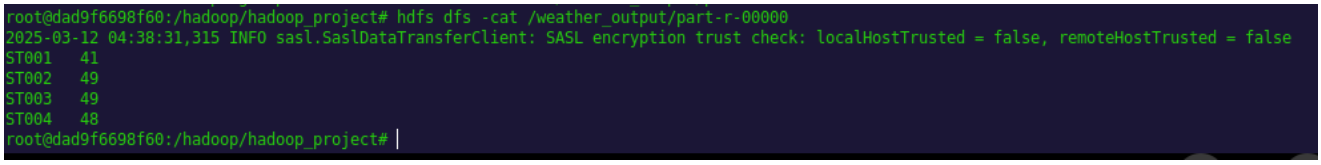
```
ST003,2023,01,11,36

ST001,2024,05,08,-9

ST004,2023,07,15,2

ST003,2024,05,25,20

ST001,2024,05,16,16

ST004,2024,07,09,9

ST004,2024,06,06,11

ST003,2024,05,15,15

ST004,2024,11,17,32

ST001,2023,07,09,17

ST003,2024,10,09,-3

ST002,2023,11,28,49

ST003,2024,04,15,1

ST003,2023,02,08,33

ST003,2024,09,26,7

ST001,2024,04,03,38

ST003,2023,11,15,42

ST002,2024,04,23,3

ST002,2023,05,23,-3

ST003,2024,01,09,43

ST003,2024,09,13,10

ST003,2024,06,15,-8

ST004,2023,11,13,10

ST004,2024,08,07,13
```

```
ST003,2023,06,03,29

ST002,2024,07,19,25

ST002,2024,06,25,41

ST001,2023,10,05,22

ST004,2024,07,21,29

ST004,2024,11,05,20

ST003,2024,02,25,49

ST003,2023,11,25,3

ST003,2024,10,16,-1

ST003,2024,07,14,21

ST003,2024,05,03,15

ST003,2024,08,11,34

ST001,2023,01,20,35

ST003,2023,09,06,12

ST003,2023,04,05,8

ST004,2023,08,06,8

ST001,2023,06,15,41

ST004,2023,08,18,44

ST001,2024,04,03,13

ST001,2023,10,06,10

ST001,2024,07,08,12

ST003,2024,04,25,49

ST002,2024,05,28,15

ST003,2023,05,22,32

ST002,2023,02,17,14
```

```
ST002,2024,05,03,49
ST004,2024,03,15,17
ST002,2024,07,03,10
ST001,2024,01,17,28
ST001,2024,02,25,0
ST001,2024,11,28,39
ST002,2023,01,24,-3
ST003,2024,10,25,17
ST003,2024,02,18,21
ST002,2023,10,20,-6
ST002,2024,03,22,18
ST001,2023,07,22,-4
ST001,2024,09,12,37
ST002,2023,03,22,14
ST003,2024,12,20,47
ST002,2024,03,03,7
ST004,2024,02,06,25
ST001,2023,10,26,0
ST003,2023,11,12,-10
ST001,2023,04,16,6
ST001,2024,06,04,20
ST003,2023,12,21,25
ST001,2024,06,04,37
ST004,2023,03,18,48
```

```
ST004,2023,09,22,39

ST004,2024,10,12,4

ST002,2024,01,13,4

ST004,2023,09,22,-4

ST004,2024,04,10,2

ST001,2023,12,20,11

ST004,2023,05,11,30
```

Output



Commands

- compilation: `javac -classpath $(hadoop classpath) -d . WeatherMapper.java WeatherReducer.java WeatherDriver.java`
- generating jar files: `jar cf weather.jar WeatherMapper.class WeatherReducer.class WeatherDriver.clas`
- running the hadoop job: `hadoop jar weather.jar WeatherDriver /weather_input /weather_output`