

## **LABORATORY PROGRAM – 6**

### **Weather Program in HADOOP**

#### **QUESTIONS**

From the following link extract the weather data

<https://github.com/tomwhite/hadoopbook/tree/master/input/ncdc/all>

a) Create a MapReduce program to find average temperature for each year from NCDC data set.

b) find the mean max temperature for every month

#### **Driver Code**

```
package temp;

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 AverageDriver {

    public static void main(String[] args) throws Exception {
        if (args.length != 2) {
            System.err.println("Please enter both input and output parameters.");
            System.exit(-1);
        }

        // Creating a configuration and job instance
        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "Average Calculation");
        job.setJarByClass(AverageDriver.class);

        // Input and output paths
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
```

```

// Setting mapper and reducer classes
job.setMapperClass(AverageMapper.class);
job.setReducerClass(AverageReducer.class);

// Output key and value types
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);

// Submitting the job and waiting for it to complete
System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}

```

### Mapper Code

```

package temp;

import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class AverageMapper extends Mapper<LongWritable, Text, Text, IntWritable> {

    public static final int MISSING = 9999;

    @Override

    public void map(LongWritable key, Text value, Context context)
        throws IOException, InterruptedException {

        String line = value.toString();

        // Extract year from fixed position
        String year = line.substring(15, 19);

        int temperature;

        // Determine if there's a '+' sign
        if (line.charAt(87) == '+') {
            temperature = Integer.parseInt(line.substring(88, 92));
        } else {

```

```

temperature = Integer.parseInt(line.substring(87, 92));
}
// Quality check character
String quality = line.substring(92, 93);
// Only emit if data is valid
if (temperature != MISSING && quality.matches("[01459]")) {
context.write(new Text(year), new IntWritable(temperature));
}
}
}

```

### Reducer Code

```

package temp;

import java.io.IOException;

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

public class AverageReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
    @Override
    public void reduce(Text key, Iterable<IntWritable> values,
        Context context) throws IOException, InterruptedException {
        int sumTemp = 0;
        int count = 0;
        for (IntWritable value : values) {
            sumTemp += value.get();
            count++;
        }
        if (count > 0) {
            int average = sumTemp / count;
            context.write(key, new IntWritable(average));
        }
    }
}

```

```
}  
  
}
```

## OBSERVATION

```
hadoop@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ 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 [bmscscse-HP-Elite-Tower-800-G9-Desktop-PC]  
Starting resourcemanager  
Starting nodemanagers  
hadoop@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ jps  
7056 DataNode  
7332 SecondaryNameNode  
7638 ResourceManager  
8231 Jps  
5883 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar  
7804 NodeManager  
6877 NameNode  
hadoop@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /\n> ^C  
hadoop@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /\nFound 4 items  
drwxr-xr-x - hadoop supergroup          0 2025-04-15 15:00 /FFF  
drwxr-xr-x - hadoop supergroup          0 2025-04-15 15:34 /LLL  
drwxr-xr-x - hadoop supergroup          0 2024-05-13 14:46 /file  
drwxr-xr-x - hadoop supergroup          0 2024-05-13 15:18 /newDataFlair  
hadoop@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /weather  
ls: '/weather': No such file or directory  
hadoop@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -mkdir /weather  
hadoop@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -copyFromLocal /home/hadoop/Desktop/1901.txt /weather/test.txt
```

```
hadoop@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/AverageTemperature.jar AverageDriver /weather/test.txt /weather/output  
2025-05-06 14:59:23,239 INFO Impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties  
2025-05-06 14:59:23,279 INFO Impl.MetricsSystemImpl: Scheduled metric snapshot period at 10 second(s).  
2025-05-06 14:59:23,279 INFO Impl.MetricsSystemImpl: JobTracker metrics system started  
2025-05-06 14:59:23,340 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
2025-05-06 14:59:23,393 INFO Input.FileInputFormat: Total input files to process : 1  
2025-05-06 14:59:23,422 INFO mapreduce.JobSubmitter: number of splits:1  
2025-05-06 14:59:23,487 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local91822813_0001  
2025-05-06 14:59:23,488 INFO mapreduce.JobSubmitter: Executing with tokens: []  
2025-05-06 14:59:23,560 INFO mapreduce.Job: The url to track the job: http://localhost:8080/  
2025-05-06 14:59:23,560 INFO mapreduce.Job: Running job: job_local91822813_0001  
2025-05-06 14:59:23,561 INFO mapred.LocalJobRunner: OutputCommitter set in config null  
2025-05-06 14:59:23,564 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory  
2025-05-06 14:59:23,565 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2  
2025-05-06 14:59:23,565 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false  
2025-05-06 14:59:23,602 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter  
2025-05-06 14:59:23,603 INFO mapred.LocalJobRunner: Starting task: attempt_local91822813_0001_m_000000_0  
2025-05-06 14:59:23,615 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory  
2025-05-06 14:59:23,615 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2  
2025-05-06 14:59:23,615 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false  
2025-05-06 14:59:23,622 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]  
2025-05-06 14:59:23,624 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/weather/test.txt:0+888190  
2025-05-06 14:59:23,658 INFO mapred.MapTask: (EQUATOR) 0 kvl 26214396(104857584)  
2025-05-06 14:59:23,658 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100  
2025-05-06 14:59:23,658 INFO mapred.MapTask: sort limit at 038860800  
2025-05-06 14:59:23,658 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600  
2025-05-06 14:59:23,658 INFO mapred.MapTask: kvstart = 26214396; length = 6553600  
2025-05-06 14:59:23,660 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
```

```

2025-05-06 14:59:24,581 INFO mapreduce.Job: Counters: 36
  File System Counters
    FILE: Number of bytes read=153118
    FILE: Number of bytes written=1493804
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1776380
    HDFS: Number of bytes written=8
    HDFS: Number of read operations=15
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
    HDFS: Number of bytes read erasure-coded=0
  Map-Reduce Framework
    Map input records=6565
    Map output records=6564
    Map output bytes=59076
    Map output materialized bytes=72210
    Input split bytes=103
    Combine input records=0
    Combine output records=0
    Reduce input groups=1
    Reduce shuffle bytes=72210
    Reduce input records=6564
    Reduce output records=1
    Spilled Records=13128
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=0
    Total committed heap usage (bytes)=1266679808
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=888190
  File Output Format Counters
    Bytes Written=8

```

```

    Bytes Written=8
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /weather
Found 2 items
drwxr-xr-x  - hadoop supergroup          0 2025-05-06 14:59 /weather/output
-rw-r--r--  1 hadoop supergroup    888190 2025-05-06 14:50 /weather/test.txt
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /weather/output
Found 2 items
-rw-r--r--  1 hadoop supergroup          0 2025-05-06 14:59 /weather/output/_SUCCESS
-rw-r--r--  1 hadoop supergroup          8 2025-05-06 14:59 /weather/output/part-r-00000
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /weather/output/part-r-00000
1901    46
hadoop@bmscece-HP-Elite-Tower-800-G9-Desktop-PC:~$ █

```