

### **Program 7**

**From the following link extract the weather data <https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/all>**

**Create a Map Reduce program to**

**a) find average temperature for each year from NCDC data set.**

**AverageDriver**

```
package temp;

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 the input and output parameters");
            System.exit(-1);
        }

        Job job = new Job();

        job.setJarByClass(AverageDriver.class);

        job.setJobName("Max temperature");

        FileInputFormat.addInputPath(job, new Path(args[0]));

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

```

job.setMapperClass(AverageMapper.class);
job.setReducerClass(AverageReducer.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}

```

### **AverageMapper**

```

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;

    public void map(LongWritable key, Text value, Mapper<LongWritable, Text, Text,
IntWritable>.Context context) throws IOException, InterruptedException {

        int temperature;

        String line = value.toString();

        String year = line.substring(15,
19); if (line.charAt(87) == '+') {

            temperature = Integer.parseInt(line.substring(88, 92));

        } else {

```

```

temperature = Integer.parseInt(line.substring(87, 92));
}
String quality = line.substring(92, 93);
if (temperature != 9999 && quality.matches("[01459]"))
context.write(new Text(year), new IntWritable(temperature));
}
}

```

AverageReducer

```

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> {

    public void reduce(Text key, Iterable<IntWritable> values, Reducer<Text, IntWritable,
Text, IntWritable>.Context context) throws IOException, InterruptedException {

        int max_temp = 0;

        int count = 0;

        for (IntWritable value : values) {

            max_temp += value.get();

            count++;

        }

        context.write(key, new IntWritable(max_temp / count));

    }
}

```

```

C:\hadoop-3.3.0\sbin>hadoop jar C:\avgtemp.jar temp.AverageDriver /input_dir/temp.txt /avgtemp_outputdir
2021-05-15 14:52:58,835 INFO client.DefaultHadoopWallerProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2021-05-15 14:52:59,005 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-05-15 14:52:59,111 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-user/staging/warehouse/staging/job_1621860230696_0005
2021-05-15 14:52:59,795 INFO Input.FileInputFormat: Total input files to process : 1
2021-05-15 14:52:59,795 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-15 14:52:59,873 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1621860230696_0005
2021-05-15 14:52:59,873 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-15 14:52:59,937 INFO conf.Configuration: resource-types.xml not found
2021-05-15 14:52:59,938 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-15 14:52:59,938 INFO impl.YarnClientImpl: Submitted application application_1621860230696_0005
2021-05-15 14:52:59,952 INFO mapreduce.Job: The url to track the job: http://LAPTOP-3G2H5D-0000/pspy/application_1621860230696_0005/
2021-05-15 14:52:59,953 INFO mapreduce.Job: Running job: job_1621860230696_0005
2021-05-15 14:53:00,048 INFO mapreduce.Job: Job job_1621860230696_0005 running in uber mode : false
2021-05-15 14:53:00,048 INFO mapreduce.Job: map 0% reduce 0%
2021-05-15 14:53:02,754 INFO mapreduce.Job: map 100% reduce 0%
2021-05-15 14:53:03,368 INFO mapreduce.Job: map 100% reduce 100%
2021-05-15 14:53:05,967 INFO mapreduce.Job: Job job_1621860230696_0005 completed successfully
2021-05-15 14:53:06,006 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=72250
        FILE: Number of bytes written=674341
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=894880
        HDFS: Number of bytes written=8
        HDFS: Number of read operations=4
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=5702

```

```

C:\hadoop-3.3.0\sbin>hdfs dfs -ls /avgtemp_outputdir
Found 2 items
-rw-r--r-- 1 Anusree supergroup 0 2021-05-15 14:53 /avgtemp_outputdir/_SUCCESS
-rw-r--r-- 1 Anusree supergroup 8 2021-05-15 14:53 /avgtemp_outputdir/part-r-000000
C:\hadoop-3.3.0\sbin>hdfs dfs -cat /avgtemp_outputdir/part-r-000000
1901 46
C:\hadoop-3.3.0\sbin>

```

**b) find the mean max temperature for every month MeanMaxDriver.class**

```
package meanmax;
```

```
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 MeanMaxDriver {
public static void main(String[] args) throws Exception {
if (args.length != 2) {
System.err.println("Please Enter the input and output parameters");
System.exit(-1);
}
Job job = new Job();
job.setJarByClass(MeanMaxDriver.class);
job.setJobName("Max temperature");
FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
job.setMapperClass(MeanMaxMapper.class);
job.setReducerClass(MeanMaxReducer.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}

```

### **MeanMaxMapper.class**

```

package meanmax;
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;

```

```

import org.apache.hadoop.mapreduce.Reducer;

public class MeanMaxReducer extends Reducer<Text, IntWritable, Text, IntWritable> {

    public void reduce(Text key, Iterable<IntWritable> values, Reducer<Text, IntWritable,
    Text, IntWritable>.Context context) throws IOException, InterruptedException {

        int max_temp = 0;

        int total_temp = 0;

        int count = 0;

        int days = 0;

        for (IntWritable value : values) {

            int temp = value.get();

            if (temp > max_temp)

                max_temp = temp;

            count++;

            if (count == 3) {

                total_temp += max_temp;

                max_temp = 0;

                count = 0;

                days++;

            }

        }

        context.write(key, new IntWritable(total_temp / days));

    }

}

```

```

C:\hadoop-3.3.0\bin>hadoop jar C:\meanmax.jar meanmax.HadoopDriver /input_dir/tmp.txt /meanmax_output
2021-05-21 20:28:05,254 INFO client.DefaultHadoopFileSystemProvider: Connecting to ResourceManager at /0.0.0.0:8032
2021-05-21 20:28:06,062 INFO mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-05-21 20:28:06,934 INFO mapreduce.JobResourceUploader: Disabling Insecure Coding for path: /tmp/hadoop-yarn/staging/meanmax/staging/job_162168943895_0001
2021-05-21 20:28:08,438 INFO input.FileInputFormat: Total input files to process : 1
2021-05-21 20:28:09,187 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-21 20:28:09,741 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_162168943895_0001
2021-05-21 20:28:09,741 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-21 20:28:10,629 INFO conf.Configuration: resource-types.xml not found
2021-05-21 20:28:10,638 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-21 20:28:10,674 INFO impl.YarnClientImpl: Submitted application application_162168943895_0001
2021-05-21 20:28:11,085 INFO mapreduce.Job: The url to track the job: http://hadoop-36329530-0000/primary/application_162168943895_0001/
2021-05-21 20:28:11,088 INFO mapreduce.Job: Running job: job_162168943895_0001
2021-05-21 20:28:25,345 INFO mapreduce.Job: Job job_162168943895_0001 running in uber mode : false
2021-05-21 20:28:25,389 INFO mapreduce.Job:  map 0% reduce 0%
2021-05-21 20:28:40,664 INFO mapreduce.Job:  map 100% reduce 0%
2021-05-21 20:28:54,812 INFO mapreduce.Job:  map 100% reduce 100%
2021-05-21 20:28:54,963 INFO mapreduce.Job: Job job_162168943895_0001 completed successfully
2021-05-21 20:28:55,171 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=5983
        FILE: Number of bytes written=648891
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=65488
        HDFS: Number of bytes written=74
        HDFS: Number of read operations=0
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=4077
        Total time spent by all reduces in occupied slots (ms)=7511
        Total time spent by all map tasks (ms)=4077
        Total time spent by all reduce tasks (ms)=7511
        Total score-millisecons taken by all map tasks=4077
        Total score-millisecons taken by all reduce tasks=7511
        Total mapbyte-millisecons taken by all map tasks=4270648
        Total mapbyte-millisecons taken by all reduce tasks=7591264

```

```

C:\hadoop-3.3.0\sbin>hdfs dfs -cat /meanmax_output/*

```

```

01      4
02      0
03      7
04     44
05    100
06    168
07    219
08    198
09    141
10    100
11     19
12      3

```

```

C:\hadoop-3.3.0\sbin>

```