

## NCDC dataset

Create a Map Reduce program to

- find average temperature for each year from NCDC data set.
- find the mean max temperature for every month.

### 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 14:52:50,635 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2021-05-15 14:52:51,005 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 14:52:51,111 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Anusree/staging/job_1621060230696_0005
2021-05-15 14:52:51,735 INFO input.FileInputFormat: Total input files to process : 1
2021-05-15 14:52:52,751 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-15 14:52:53,073 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1621060230696_0005
2021-05-15 14:52:53,073 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-15 14:52:53,237 INFO conf.Configuration: resource-types.xml not found
2021-05-15 14:52:53,238 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-15 14:52:53,312 INFO impl.YarnClientImpl: Submitted application application_1621060230696_0005
2021-05-15 14:52:53,352 INFO mapreduce.Job: The url to track the job: http://LAPTOP-JG329ESD:8088/proxy/application_1621060230696_0005/
2021-05-15 14:52:53,353 INFO mapreduce.Job: Running job: job_1621060230696_0005
2021-05-15 14:53:06,640 INFO mapreduce.Job: Job job_1621060230696_0005 running in uber mode : false
2021-05-15 14:53:06,643 INFO mapreduce.Job: map 0% reduce 0%
2021-05-15 14:53:12,750 INFO mapreduce.Job: map 100% reduce 0%
2021-05-15 14:53:19,000 INFO mapreduce.Job: map 100% reduce 100%
2021-05-15 14:53:25,967 INFO mapreduce.Job: Job job_1621060230696_0005 completed successfully
2021-05-15 14:53:26,096 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=72210
    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=894860
    HDFS: Number of bytes written=8
    HDFS: Number of read operations=8
    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)=3782

```

```

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>

```

## 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.io.Text;

import org.apache.hadoop.mapreduce.Mapper;

public class MeanMaxMapper 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 month = line.substring(19, 21);

    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(month), new IntWritable(temperature));

    }

}

```

### **MeanMaxReducer.class**

```

package meanmax;

import java.io.IOException;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.Text;

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));

}
```

```

[hadoop-3.3.0]$ bin/hadoop jar C:\Programs\Hadoop\hadoop\hadoop.jar /input/dfs/dfs.txt /meanmax_output
2023-05-22 20:28:05,258 INFO client.DefaultHadoopFileSystemHadoopReader: Connecting to NameNode at 192.168.0.8:8022
2023-05-22 20:28:06,062 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-05-22 20:28:06,063 INFO mapreduce.JobResourceUploader: Building Resource Listing for path: /mapreduce-servicing/output/staging/job_302108043095_0001
2023-05-22 20:28:06,428 INFO InputFileConfiguration: Total input files to process: 1
2023-05-22 20:28:06,387 INFO mapreduce.JobSubmitter: number of splits: 1
2023-05-22 20:28:06,742 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_302108043095_0001
2023-05-22 20:28:06,743 INFO mapreduce.JobSubmitter: Enclosing with tokens: {}
2023-05-22 20:28:08,829 INFO conf.Configuration: resource-types.xml not found
2023-05-22 20:28:08,830 INFO resource.ResourceStatistics: unable to find 'resource-types.xml'
2023-05-22 20:28:18,076 INFO impl.VersionClientImpl: Submitted application application_302108043095_0001
2023-05-22 20:28:21,065 INFO mapreduce.Job: The url to track the job: http://192.168.0.8:8022/track/application_302108043095_0001/
2023-05-22 20:28:21,066 INFO mapreduce.Job: Launch job: job_302108043095_0001
2023-05-22 20:28:25,385 INFO mapreduce.Job: job_302108043095_0001 running is over mode: false
2023-05-22 20:28:25,387 INFO mapreduce.Job: map 06 reduce 06
2023-05-22 20:28:06,663 INFO mapreduce.Job: map 00S reduce 0S
2023-05-22 20:28:58,012 INFO mapreduce.Job: map 00S reduce 00S
2023-05-22 20:28:58,013 INFO mapreduce.Job: job_302108043095_0001 completed successfully
2023-05-22 20:28:59,178 INFO mapreduce.Job: (tasks): 54

File System Counters
  FILE: Number of bytes read:5882
  FILE: Number of bytes written:44801
  FILE: Number of read operations:0
  FILE: Number of large read operations:0
  FILE: Number of write operations:0
  HFS: Number of bytes read:60484
  HFS: Number of bytes written:24
  HFS: Number of read operations:0
  HFS: Number of large read operations:0
  HFS: Number of write operations:0
  HFS: Number of bytes read uncacheable:0

Job Counters
  Launched map tasks:1
  Launched reduce tasks:1
  Map-Local map tasks:1
  Total time spent by all maps in occupied slots (ms):4077
  Total time spent by all reducers in occupied slots (ms):7511
  Total time spent by all map tasks (ms):4077
  Total time spent by all reduce tasks (ms):7511
  Total mapreduce time taken by all map tasks:4077
  Total mapreduce time taken by all reduce tasks:7511
  Total mapreduce time taken by all map tasks:4077
  Total mapreduce time taken by all reduce tasks:7511
  Total mapreduce time taken by all map tasks:4077
  Total mapreduce time taken by all reduce tasks:7511

```

```

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>

```