Mock DSBDA
(ELE min max)

Name : Sahil Bhoye
rn : 33305

Program :

```
package PackageDemo;

import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.util.GenericOptionsParser;

public class MaxElectricityConsumption {

   public static void main(String[] args) throws Exception {
      Configuration c = new Configuration();
      String[] files = new GenericOptionsParser(c, args).getRemainingArgs();
      Path input = new Path(files[0]);
      Path output = new Path(files[1]);
      Job j = Job.getInstance(c, "maxMinElectricityConsumption");
      j.setJarByClass(MaxElectricityConsumption.class);
      j.setMapperClass(MapForMaxMinConsumption.class);
      j.setReducerClass(ReduceForMaxMinConsumption.class);
      j.setMapOutputKeyClass(Text.class);
      j.setMapOutputValueClass(Text.class);
      j.setOutputKeyClass(Text.class);
      j.setOutputValueClass(Text.class);
      FileInputFormat.addInputPath(j, input);
      FileOutputFormat.setOutputPath(j, output);
      System.exit(j.waitForCompletion(true) ? 0 : 1);
   }

   public static class MapForMaxMinConsumption extends Mapper<LongWritable, Text, Text,
Text> {
      public void map(LongWritable key, Text value, Context context) throws IOException,
InterruptedException {
         String line = value.toString();
         String[] values = line.split(",");

         String year = values[0].trim();
         int maxConsumption = Integer.MIN_VALUE;
```

```
        int minConsumption = Integer.MAX_VALUE;

        for (int i = 1; i < values.length; i++) {
            try {
                int consumption = Integer.parseInt(values[i].trim());
                if (consumption > maxConsumption) {
                    maxConsumption = consumption;
                }
                if (consumption < minConsumption) {
                    minConsumption = consumption;
                }
            } catch (NumberFormatException e) {
                continue; // Ignore invalid data
            }
        }

        context.write(new Text(year), new Text(maxConsumption + "," + minConsumption));
    }
}

public static class ReduceForMaxMinConsumption extends Reducer<Text, Text, Text, Text> {
    public void reduce(Text year, Iterable<Text> values, Context context) throws IOException,
InterruptedException {
        int maxConsumption = Integer.MIN_VALUE;
        int minConsumption = Integer.MAX_VALUE;

        for (Text value : values) {
            String[] parts = value.toString().split(",");
            int maxVal = Integer.parseInt(parts[0]);
            int minVal = Integer.parseInt(parts[1]);

            if (maxVal > maxConsumption) {
                maxConsumption = maxVal;
            }
            if (minVal < minConsumption) {
                minConsumption = minVal;
            }
        }
        context.write(year, new Text("Max: " + maxConsumption + ", Min: " + minConsumption));
    }
}
}
```

DATA :

2000,23,23,24,32,42,52,62,62,62,62,52,62
2001,26,27,28,28,28,30,31,31,31,30,30,30
2002,31,32,32,32,33,34,35,36,36,34,34,34
2003,39,38,39,39,39,41,42,43,40,39,38,38
2004,38,39,39,39,39,41,41,41,00,40,39,39
2005,26,26,21,38,42,52,62,62,62,62,52,62

2006,26,27,28,28,28,36,34,38,34,31,39,30
2007,31,85,32,32,33,34,35,36,36,34,34,34
2008,39,38,39,39,39,41,42,20,40,39,38,38
2009,38,39,65,39,39,41,41,41,00,40,39,39
2010,38,39,39,39,55,41,41,41,00,40,39,39
2011,23,23,66,32,42,52,62,62,62,62,52,62
2012,26,27,28,28,99,30,31,31,31,30,30,30
2013,31,32,121,32,33,34,35,36,36,34,34,34
2014,39,38,39,39,53,41,42,43,40,39,38,38
2015,38,39,64,39,39,41,41,41,00,40,39,39

Output :

```
[cloudera@quickstart Desktop]$ hadoop fs -put Ele elee3
[cloudera@quickstart Desktop]$ hadoop jar bbb.jar
PackageDemo.MaxElectricityConsumption elee3 MRDir6
25/03/25 23:56:26 INFO client.RMProxy: Connecting to ResourceManager at
/0.0.0.0:8032
25/03/25 23:56:27 INFO input.FileInputFormat: Total input paths to process : 1
25/03/25 23:56:27 INFO mapreduce.JobSubmitter: number of splits:1
25/03/25 23:56:27 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1742970151032_0004
25/03/25 23:56:28 INFO impl.YarnClientImpl: Submitted application
application_1742970151032_0004
25/03/25 23:56:28 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1742970151032_0004/
25/03/25 23:56:28 INFO mapreduce.Job: Running job: job_1742970151032_0004
25/03/25 23:56:37 INFO mapreduce.Job: Job job_1742970151032_0004 running in
uber mode : false
25/03/25 23:56:37 INFO mapreduce.Job:  map 0% reduce 0%
25/03/25 23:56:43 INFO mapreduce.Job:  map 100% reduce 0%
25/03/25 23:56:52 INFO mapreduce.Job:  map 100% reduce 100%
25/03/25 23:56:52 INFO mapreduce.Job: Job job_1742970151032_0004 completed
successfully
25/03/25 23:56:52 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=211
                FILE: Number of bytes written=221869
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=776
                HDFS: Number of bytes written=349
```

HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
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)=4641
Total time spent by all reduces in occupied slots (ms)=5486
Total time spent by all map tasks (ms)=4641
Total time spent by all reduce tasks (ms)=5486
Total vcore-seconds taken by all map tasks=4641
Total vcore-seconds taken by all reduce tasks=5486
Total megabyte-seconds taken by all map tasks=4752384
Total megabyte-seconds taken by all reduce tasks=5617664
Map-Reduce Framework
Map input records=16
Map output records=16
Map output bytes=173
Map output materialized bytes=211
Input split bytes=116
Combine input records=0
Combine output records=0
Reduce input groups=16
Reduce shuffle bytes=211
Reduce input records=16
Reduce output records=16
Spilled Records=32
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=118
CPU time spent (ms)=1150
Physical memory (bytes) snapshot=350507008
Virtual memory (bytes) snapshot=3006996480
Total committed heap usage (bytes)=226365440
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=660

File Output Format Counters
Bytes Written=349
[cloudera@quickstart Desktop]$ hadoop fs -ls MRDir6
Found 2 items
-rw-r--r-- 1 cloudera cloudera 0 2025-03-25 23:56 MRDir6/_SUCCESS
-rw-r--r-- 1 cloudera cloudera 349 2025-03-25 23:56 MRDir6/part-r-00000
[cloudera@quickstart Desktop]$ hadoop fs -cat MRDir6/part-r-00000
2000 Max: 62, Min: 23
2001 Max: 31, Min: 26
2002 Max: 36, Min: 31
2003 Max: 43, Min: 38
2004 Max: 41, Min: 0
2005 Max: 62, Min: 21
2006 Max: 39, Min: 26
2007 Max: 85, Min: 31
2008 Max: 42, Min: 20
2009 Max: 65, Min: 0
2010 Max: 55, Min: 0
2011 Max: 66, Min: 23
2012 Max: 99, Min: 26
2013 Max: 121, Min: 31
2014 Max: 53, Min: 38
2015 Max: 64, Min: 0
[cloudera@quickstart Desktop]$