Roll No.: T190058699

Code -

```
package Package;

import java.io.*;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.util.GenericOptionsParser;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class LogFileProcessing {

	public static void main(String [] args) throws IOException, ClassNotFoundException,
InterruptedException{
		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 = new Job(c,"logfileprocessing");

		j.setJarByClass(LogFileProcessing.class);
		j.setMapperClass(MapForLogFileProcessing.class);

		j.setReducerClass(ReducerForLogFileProcessing.class);
		j.setOutputKeyClass(Text.class);
		j.setOutputValueClass(IntWritable.class);

		FileInputFormat.addInputPath(j, input);
		FileOutputFormat.setOutputPath(j, output);

		System.exit(j.waitForCompletion(true)?0:1);

	}

	public static class MapForLogFileProcessing extends Mapper<LongWritable ,Text,Text,IntWritable>{

		public void map(LongWritable key, Text values,Context con) throws IOException,
InterruptedException{
			String [] data = values.toString().split(",");


				Text ip = new Text(data[1]);
				System.out.println(data[1]);

				String [] dt1 = data[5].split(" ");
				String [] dt2 = data[7].split(" ");
```

```
                                String[] intime = dt1[1].split(":");
                                String[] outtime = dt2[1].split(":");

                                int intime_sec = Integer.parseInt(intime[0])*3600 + Integer.parseInt(intime[1])*60
+ Integer.parseInt(intime[2]);
                                int outtime_sec = Integer.parseInt(outtime[0])*3600 +
Integer.parseInt(outtime[1])*60 + Integer.parseInt(outtime[2]);
                                int period = outtime_sec - intime_sec;

                                IntWritable time = new IntWritable(period);

                                con.write(ip,time);

                        }
                }


                public static class ReducerForLogFileProcessing extends Reducer<Text ,IntWritable,Text,IntWritable>{

                        Text maxIP = new Text();
                        int maxTime = 0;

                        Text minIP = new Text();
                        int minTime = Integer.MAX_VALUE;
                        public void reduce(Text ip, Iterable<IntWritable> time,Context con) throws IOException,
InterruptedException{

                                int total_period = 0;

                                for(IntWritable period : time){


                                        total_period +=  period.get();
                                }

                                if(maxTime < total_period)
                                {
                                        maxTime = total_period;
                                        maxIP.set(ip);
                                }

                                if(minTime > total_period)
                                {
                                        minTime = total_period;
                                        minIP.set(ip);
                                }

                                con.write(ip, new IntWritable(total_period));
                        }


                        protected void cleanup(Context con) throws IOException, InterruptedException{
                                con.write(new Text("User IP who logged max time in seconds: "+maxIP.toString()), new
IntWritable(maxTime));
                                con.write(new Text("User IP who logged min time in seconds: "+minIP.toString()), new
```

```
IntWritable(minTime));
                }
        }


}
```

Output -

```
[cloudera@quickstart T190058699]$ hadoop fs -put logTime22.csv hlogTime22.csv

[cloudera@quickstart T190058699]$ hadoop jar LogFileProcessing.jar Package.LogFileProcessing
hlogTime22.csv out_dir1
24/05/10 02:57:25 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
24/05/10 02:57:26 INFO input.FileInputFormat: Total input paths to process : 1
24/05/10 02:57:26 INFO mapreduce.JobSubmitter: number of splits:1
24/05/10 02:57:26 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1715334875657_0001
24/05/10 02:57:27 INFO impl.YarnClientImpl: Submitted application application_1715334875657_0001
24/05/10 02:57:27 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1715334875657_0001/
24/05/10 02:57:27 INFO mapreduce.Job: Running job: job_1715334875657_0001
24/05/10 02:57:36 INFO mapreduce.Job: Job job_1715334875657_0001 running in uber mode : false
24/05/10 02:57:36 INFO mapreduce.Job:  map 0% reduce 0%
24/05/10 02:57:43 INFO mapreduce.Job:  map 100% reduce 0%
24/05/10 02:57:51 INFO mapreduce.Job:  map 100% reduce 100%
24/05/10 02:57:52 INFO mapreduce.Job: Job job_1715334875657_0001 completed successfully
24/05/10 02:57:52 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=1864
                FILE: Number of bytes written=224487
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=9683
                HDFS: Number of bytes written=295
                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)=5862
                Total time spent by all reduces in occupied slots (ms)=5146
                Total time spent by all map tasks (ms)=5862
                Total time spent by all reduce tasks (ms)=5146
                Total vcore-seconds taken by all map tasks=5862
                Total vcore-seconds taken by all reduce tasks=5146
                Total megabyte-seconds taken by all map tasks=6002688
                Total megabyte-seconds taken by all reduce tasks=5269504
        Map-Reduce Framework
                Map input records=100
                Map output records=100
                Map output bytes=1658
                Map output materialized bytes=1864
                Input split bytes=125
```

```
                Combine input records=0
                Combine output records=0
                Reduce input groups=10
                Reduce shuffle bytes=1864
                Reduce input records=100
                Reduce output records=12
                Spilled Records=200
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=81
                CPU time spent (ms)=630
                Physical memory (bytes) snapshot=329940992
                Virtual memory (bytes) snapshot=3007397888
                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=9558
        File Output Format Counters
                Bytes Written=295
[cloudera@quickstart T190058699]$ hadoop fs -ls out_dir1
Found 2 items
-rw-r--r--   1 cloudera cloudera          0 2024-05-10 02:57 out_dir1/_SUCCESS
-rw-r--r--   1 cloudera cloudera        295 2024-05-10 02:57 out_dir1/part-r-00000
[cloudera@quickstart T190058699]$ hadoop fs -cat out_dir1/part-r-00000
10.10.10.103    120288
10.10.10.14     105108
10.10.10.145    62996
10.10.10.220    69972
10.10.10.221    204954
10.10.10.79     70072
10.10.13.167    13368
10.10.13.83     8912
10.10.15.9      76
10.10.15.98     114
User IP who logged max time in seconds: 10.10.10.221   204954
User IP who logged min time in seconds: 10.10.15.9      76
[cloudera@quickstart T190058699]$
```