Module 2: Assignment – 2

Find out the list of customers who have the total call time in long-distance calls of more than
60 minutes.

Solution:

Viewing contents of JAR and phone.txt

```
[training@sandbox-hdp ~]$ jar -tvf map_reduce_1.jar
    25 Tue Nov 02 01:22:02 UTC 2021 META-INF/MANIFEST.MF
  3336 Tue Nov 02 01:22:02 UTC 2021 map_reduce_1/STDCalls$MapClass.class
  2410 Tue Nov 02 01:22:02 UTC 2021 map_reduce_1/STDCalls$ReduceClass.class
  2022 Tue Nov 02 01:22:02 UTC 2021 map_reduce_1/STDCalls.class
   588 Tue Nov 02 01:18:46 UTC 2021 .classpath
   388 Tue Nov 02 01:07:02 UTC 2021 .project
[training@sandbox-hdp ~]$ cat phone.txt
9665128505,8983006310,2015-03-01 09:08:10,2015-03-01 10:12:15,1
9665128505,8983006310,2015-03-01 07:08:10,2015-03-01 08:12:15,0
9665128505,8983006310,2015-03-01 09:08:10,2015-03-01 09:12:15,1
9665128505,8983006310,2015-03-01 09:08:10,2015-03-01 10:12:15,0
9665128506,8983006310,2015-03-01 09:08:10,2015-03-01 10:12:15,1
9665128507,8983006310,2015-03-01 09:08:10,2015-03-01 09:12:15,1
```

Running MapReduce Job

```
[training@sandbox-hdp ~]$ hadoop jar map_reduce_1.jar map_reduce_1/STDCalls /user/t
21/11/01 20:26:39 INFO client.RMProxy: Connecting to ResourceManager at sandbox-hdp
21/11/01 20:26:39 INFO client.AHSProxy: Connecting to Application History server at
21/11/01 20:26:40 WARN mapreduce.JobResourceUploader: Hadoop command-line option pa
olRunner to remedy this.
21/11/01 20:26:40 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for
21/11/01 20:26:40 INFO input.FileInputFormat: Total input files to process : 1
21/11/01 20:26:41 INFO mapreduce.JobSubmitter: number of splits:1
21/11/01 20:26:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16357
21/11/01 20:26:41 INFO mapreduce.JobSubmitter: Executing with tokens: []
21/11/01 20:26:42 INFO conf.Configuration: found resource resource-types.xml at fil
21/11/01 20:26:42 INFO impl.YarnClientImpl: Submitted application application_16357
21/11/01 20:26:42 INFO mapreduce.Job: The url to track the job: http://sandbox-hdp.
21/11/01 20:26:42 INFO mapreduce.Job: Running job: job_1635778639489_0005
21/11/01 20:26:51 INFO mapreduce.Job: Job job_1635778639489_0005 running in uber mo
21/11/01 20:26:51 INFO mapreduce.Job:  map 0% reduce 0%
21/11/01 20:27:42 INFO mapreduce.Job:  map 100% reduce 0%
21/11/01 20:27:51 INFO mapreduce.Job:  map 100% reduce 100%
21/11/01 20:27:52 INFO mapreduce.Job: Job job_1635778639489_0005 completed successf
21/11/01 20:27:52 INFO mapreduce.Job: Counters: 53
        File System Counters
                FILE: Number of bytes read=74
                FILE: Number of bytes written=470187
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=524
                HDFS: Number of bytes written=28
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
```

Viewing Output:
List of customers who have total call time in long-distance calls of more than 60 minutes

```
[training@sandbox-hdp ~]$ hadoop fs -ls /user/training/outputSTD
Found 2 items
-rw-r--r--   1 training hdfs          0 2021-11-01 20:27 /user/training/outputSTD/_
SUCCESS
-rw-r--r--   1 training hdfs         28 2021-11-01 20:27 /user/training/outputSTD/p
art-r-00000
[training@sandbox-hdp ~]$ hadoop fs -cat /user/training/outputSTD/part-r-00000
9665128505,68
9665128506,64
[training@sandbox-hdp ~]$
```