

Module 2: Assignment - 3

1. Find out the list of vehicles going above the speed of 65 km per hour
2. Find out the percentage of entries with vehicles going above 65 km per hour

Solution:

Putting dataset files in HDFS:

```
[training@sandbox-hdp ~]$ hadoop fs -put /home/training/Speed-data.txt /user/training/assignments/
```

Reading contents of speed.txt file:

```
training@sandbox-hdp ~]$ cat /home/training/speed/speed.txt
A-123,70,7.00
A-223,60,8.02
A-123,72,9.10
A-223,63,10.05
A-123,75,11.20
A-123,70,13.00
A-223,65,15.02
A-123,65,16.10
A-223,60,18.05
```

Viewing contents of JAR file:

```
training@sandbox-hdp ~]$ jar -tvf map_reduce_2.jar
 25 Tue Nov 02 02:22:08 UTC 2021 META-INF/MANIFEST.MF
2926 Tue Nov 02 02:21:36 UTC 2021 map_reduce_2/Speed_code$MapClass.class
2427 Tue Nov 02 02:21:36 UTC 2021 map_reduce_2/Speed_code$ReduceClass.class
2032 Tue Nov 02 02:21:36 UTC 2021 map_reduce_2/Speed_code.class
```

Running Map Reduce Job

```
[training@sandbox-hdp ~]$ hadoop jar map_reduce_2.jar map_reduce_2/Speed_code /user
/training/assignments/Speed-data.txt output_speed
21/11/01 20:58:59 INFO client.RMPProxy: Connecting to ResourceManager at sandbox-hdp
.hortonworks.com/172.18.0.2:8050
21/11/01 20:58:59 INFO client.AHSPProxy: Connecting to Application History server at
sandbox-hdp.hortonworks.com/172.18.0.2:10200
21/11/01 20:59:00 WARN mapreduce.JobResourceUploader: Hadoop command-line option pa
rsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
21/11/01 20:59:00 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for
path: /user/training/.staging/job_1635778639489_0009
21/11/01 20:59:00 INFO input.FileInputFormat: Total input files to process : 1
21/11/01 20:59:01 INFO mapreduce.JobSubmitter: number of splits:1
21/11/01 20:59:01 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16357
78639489_0009
21/11/01 20:59:01 INFO mapreduce.JobSubmitter: Executing with tokens: []
21/11/01 20:59:01 INFO conf.Configuration: found resource resource-types.xml at fil
e:/etc/hadoop/3.0.1.0-187/0/resource-types.xml
21/11/01 20:59:01 INFO impl.YarnClientImpl: Submitted application application_16357
78639489_0009
21/11/01 20:59:01 INFO mapreduce.Job: The url to track the job: http://sandbox-hdp.
hortonworks.com:8088/proxy/application_1635778639489_0009/
21/11/01 20:59:01 INFO mapreduce.Job: Running job: job_1635778639489_0009
21/11/01 20:59:11 INFO mapreduce.Job: Job job_1635778639489_0009 running in uber mo
de : false
21/11/01 20:59:11 INFO mapreduce.Job: map 0% reduce 0%
21/11/01 20:59:19 INFO mapreduce.Job: map 100% reduce 0%
21/11/01 20:59:28 INFO mapreduce.Job: map 100% reduce 100%
21/11/01 20:59:28 INFO mapreduce.Job: Job job_1635778639489_0009 completed successf
ully
21/11/01 20:59:28 INFO mapreduce.Job: Counters: 53
    File System Counters
        FILE: Number of bytes read=136
        FILE: Number of bytes written=470335
        FILE: Number of read operations=0
```

Viewing Output:

```
[training@sandbox-hdp ~]$ hadoop fs -ls /user/training/output_speed
Found 2 items
-rw-r--r--  1 training hdfs          0 2021-11-01 20:59 /user/training/output_speed/_SUCCESS
-rw-r--r--  1 training hdfs       70 2021-11-01 20:59 /user/training/output_speed/part-r-00000
```

```
[training@sandbox-hdp ~]$ hadoop fs -cat /user/training/output_speed/part-r-00000
GA-123,65
GA-123,70
GA-123,75
GA-123,72
GA-123,70
GA-223,75
GA-223,65
```