

Logic For First Submission

<Properly explain the code, list the steps to run the code provided by you and attach screenshots of code execution>

Note: Be as descriptive as possible.

Task 1: Write a job to consume clickstream data from Kafka and ingest to Hadoop.

- Create cluster with Hadoop, Sqoop, Hive and Spark and with necessary hardware configurations like 1 master node and 2 core nodes.
- Login to hadoop user: **hadoop**

Steps to run the python file to load data from Kafka

- Created python file '**spark_kafka_to_local.py**' to load data from topic '**de-capstone3**' on kafka server.
- Load the data from kafka server to hadoop local path '**/home/hadoop/kafkadata**' folder in json format.
- **Command to execute the python file 'spark_kafka_to_local.py'**

```
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 spark_kafka_to_local.py
```

```
[hadoop@ip-172-31-8-86 ~]$ ls
kafkadata  spark_kafka_to_local.py  spark_local_flatten.py
[hadoop@ip-172-31-8-86 ~]$ ls kafkadata/
checkpoint  part-00000-965c96b0-1ea4-43f1-bb32-6cb6ddb1fc70-c000.json  _spark_metadata
[hadoop@ip-172-31-8-86 ~]$
```

```
hadoop@ip-172-31-8-68:~$ cat spark_local_flatten.py
{"value":{"customer_id":"71156145","app_version":"4.1.4","os_version":"Android","lat":"-24.5899605","lon":"67.62477","page_id":"de545711-3914-4450-8c11-b17b8dabb5e1","button_id":"fcba68aa-1231-11eb-adc1-0242ac120002","is_button_click":"No","is_page_view":"No","is_scroll_up":"No","is_scroll_down":"Yes","timestamp":"2020-01-09 14:54:48"},"value":{"customer_id":"22899996","app_version":"2.4.39","os_version":"iOS","lat":"-69.1155955","lon":"-143.163289","page_id":"de545711-3914-4450-8c11-b17b8dabb5e1","button_id":"e1e99492-17ae-11eb-adc1-0242ac120002","is_button_click":"Yes","is_page_view":"Yes","is_scroll_up":"Yes","is_scroll_down":"Yes","timestamp":"2020-06-26 09:08:00"},"value":{"customer_id":"85382952","app_version":"3.2.24","os_version":"Android","lat":"-47.734696","lon":"85.891429","page_id":"b328829e-17ae-11eb-adc1-0242ac120002","button_id":"fcba68aa-1231-11eb-adc1-0242ac120002","is_button_click":"Yes","is_page_view":"No","is_scroll_up":"Yes","is_scroll_down":"No","timestamp":"2020-04-11 10:41:05"},"value":{"customer_id":"42165322","app_version":"4.1.13","os_version":"iOS","lat":"80.8403425","lon":"-86.627180","page_id":"de545711-3914-4450-8c11-b17b8dabb5e1","button_id":"a95dd57b-779f-49db-819d-b6960483e554","is_button_click":"Yes","is_page_view":"No","is_scroll_up":"No","is_scroll_down":"Yes","timestamp":"2020-03-27 14:03:03"},"value":{"customer_id":"21222347","app_version":"1.2.27","os_version":"iOS","lat":"-3.6452735","lon":"173.446829","page_id":"b328829e-17ae-11eb-adc1-0242ac120002","button_id":"a95dd57b-779f-49db-819d-b6960483e554","is_button_click":"Yes","is_page_view":"No","is_scroll_up":"No","is_scroll_down":"Yes","timestamp":"2020-07-13 15:48:28"},"value":{"customer_id":"63872736","app_version":"1.1.2","os_version":"iOS","lat":"52.6143785","lon":"-4.783003","page_id":"e7bc5fb2-1231-11eb-adc1-0242ac120002","button_id":"fcba68aa-1231-11eb-adc1-0242ac120002","is_button_click":"Yes","is_page_view":"No","is_scroll_up":"No","is_scroll_down":"Yes","timestamp":"2020-08-16 13:27:47"},"value":{"customer_id":"25631527","app_version":"2.1.6","os_version":"Android","lat":"73.0767505","lon":"13.785813","page_id":"b328829e-17ae-11eb-adc1-0242ac120002","button_id":"e1e99492-17ae-11eb-adc1-0242ac120002","is_button_click":"No","is_page_view":"Yes","is_scroll_up":"Yes","is_scroll_down":"Yes","timestamp":"2020-01-23 08:41:36"},"value":{"customer_id":"92397542","app_version":"2.3.16","os_version":"iOS","lat":"-7.0993865","lon":"-40.763662","page_id":"de545711-3914-4450-8c11-b17b8dabb5e1","button_id":"fcba68aa-1231-11eb-adc1-0242ac120002","is_button_click":"Yes","is_page_view":"No","is_scroll_up":"No","is_scroll_down":"No","timestamp":"2020-08-02 07:44:14"},"value":{"customer_id":"15962691","app_version":"4.1.24","os_version":"iOS","lat":"44.1034375","lon":"-158.429047","page_id":"b328829e-17ae-11eb-adc1-0242ac120002","button_id":"a95dd57b-779f-49db-819d-b6960483e554","is_button_click":"Yes","is_page_view":"Yes","is_scroll_up":"Yes","is_scroll_down":"Yes","timestamp":"2020-08-24 17:58:13"},"value":{"customer_id":"85593174","app_version":"4.1.6","os_version":"Android","lat":"41.599276","lon":"-133.706095","page_id":"de545711-3914-4450-8c11-b17b8dabb5e1","button_id":"fcba68aa-1231-11eb-adc1-0242ac120002","is_button_click":"Yes","is_page_view":"No","is_scroll_up":"No","is_scroll_down":"No","timestamp":"2020-09-13 13:21:43"},"value":{"customer_id":"75574770","app_version":"4.1.39","os_version":"iOS","lat":"-82.045519","lon":"-60.819861","page_id":"b328829e-17ae-11eb-adc1-0242ac120002","button_id":"a95dd57b-779f-49db-819d-b6960483e554","is_button_click":"Yes","is_page_view":"Yes","is_scroll_up":"Yes","is_scroll_down":"Yes","timestamp":"2020-11-08 12:15:57"},"value":{"customer_id":"21922061","app_version":"3.1.35","os_version":"iOS","lat":"-72.0056295","lon":"119.587231","page_id":"de545711-3914-4450-8c11-b17b8dabb5e1","button_id":"e1e99492-17ae-11eb-adc1-0242ac120002","is_button_click":"No","is_page_view":"Yes","is_scroll_up":"No","is_scroll_down":"No","timestamp":"2020-07-25 00:16:07"},"value":{"customer_id":"27178502","app_version":"2.1.39","os_version":"iOS","lat":"25.9298745","lon":"125.573580","page_id":"de545711-3914-4450-8c11-b17b8dabb5e1","button_id":"a95dd57b-779f-49db-819d-b6960483e554","is_button_click":"Yes","is_page_view":"No","is_scroll_up":"Yes","is_scroll_down":"No","timestamp":"2020-11-07 11:45:41"},"value":{"customer_id":"51186077","app_version":"3.2.20","os_version":"Android","lat":"71.250386","lon":"141.830952","page_id":"e7bc5fb2-1231-11eb-adc1-0242ac120002","button_id":"a95dd57b-779f-49db-819d-b6960483e554","is_button_click":"No","is_page_view":"Yes","is_scroll_up":"No","is_scroll_down":"No","timestamp":"2020-10-24 23:26:26"},"value":{"customer_id":"26893884","app_version":"2.4.40","os_version":"Android","lat":"62.963317","lon":"-115.642762","page_id":"b328829e-17ae-11eb-adc1-0242ac120002","button_id":"fcba68aa-1231-11eb-adc1-0242ac120002","is_button_click":"No","is_page_view":"No"}}
```

Steps to load the data into Hadoop

- Created python file 'spark_local_flatten.py' to clean the loaded Kafka data to a more structured format like csv and store the file in hdfs
'/user/capstone/viji/clickstream' from hadoop local path
'/home/hadoop/kafkadata' folder.
- Command to execute the python file 'spark_local_flatten.py'

spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 spark_local_flatten.py

hadoop@ip-172-31-8-86:~										
customer_id	app_version	OS_version	lat	lon	page_id	button_id	is_button_click	is_page_view	is_scroll_up	is_scroll_down
265648201	3.2.35	Android	16.4454865	99.902065	de545711-3914-445...fcba68aa-1231-11e...		No	Yes	No	Yes
31906387	2.4.7	iOS	-64.813749	-133.527040	de545711-3914-445...a95dd57b-779f-49d...		No	No	Yes	Yes
25713677	3.4.12	Android	89.943435	127.313415	b328829e-17ae-11e...fcba68aa-1231-11e...		No	No	Yes	No
83474293	3.1.8	Android	-69.939070	-36.451670	e7bc5fb2-1231-11e...ele99492-17ae-11e...		Yes	No	Yes	No
63727807	2.2.9	iOS	64.082108	-81.822078	e7bc5fb2-1231-11e...fcba68aa-1231-11e...		No	Yes	Yes	Yes
73737907	4.3.19	Android	-18.850508	-116.358375	b328829e-17ae-11e...ele99492-17ae-11e...		No	Yes	No	Yes
36927433	3.2.26	iOS	-84.6857245	-146.507678	de545711-3914-445...a95dd57b-779f-49d...		Yes	Yes	No	Yes
12691783	3.3.11	Android	54.3852925	-37.411814	de545711-3914-445...ele99492-17ae-11e...		Yes	Yes	No	No
22635021	4.4.36	iOS	-31.805500	150.655650	e7bc5fb2-1231-11e...a95dd57b-779f-49d...		No	No	No	No
23593546	1.2.16	Android	8.8918475	-83.929878	de545711-3914-445...ele99492-17ae-11e...		Yes	No	Yes	No
16929816	1.2.25	Android	-34.8576385	-136.639835	e7bc5fb2-1231-11e...fcba68aa-1231-11e...		No	No	No	Yes
41929865	1.1.14	iOS	-34.3613155	-137.467833	b328829e-17ae-11e...a95dd57b-779f-49d...		No	Yes	No	No
68940320	1.1.17	Android	81.738239	-104.207763	de545711-3914-445...fcba68aa-1231-11e...		Yes	No	Yes	No

- Check the files created on hdfs

`hdfs dfs -ls /user/capstone/clickstream/`

- Command to print the records

`hdfs dfs -cat /user/capstone/clickstream/part-00000-9950f612-ebbd-4ccf-b88e-bda25394a87a-c000.csv`

- Screenshot of the records

```

520836611,1.3.6,Android,-,20.051056,-,115.098725,b328829e-17ae-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,No,No,Yes,No,"2020-01-03 22:44:45"
121488172,2.2.14,iOS,25.7465345,-,115.096705,e7bc5fb2-1231-11eb-adcl-0242ac120002,e199492-17ae-11eb-adcl-0242ac120002,Yes,No,Yes,No,"2020-06-09 09:32:52"
52887956,2.3.23,Android,-,14.8909615,-,161.408474,b328829e-17ae-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,No,Yes,No,No,"2020-03-24 01:12:52"
52887956,4.1.11,iOS,3.769129,-,128.422422,b328829e-17ae-11eb-adcl-0242ac120002,e199492-17ae-11eb-adcl-0242ac120002,No,Yes,No,No,"2020-06-20 21:33:07"
17040232,3.1.1,iOS,6.79099,-,26.258484,b328829e-17ae-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,No,No,No,No,"2020-04-13 21:38:54"
5243436,1.1.1,Android,86.2803345,-,4.664262,e7bc5fb2-1231-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,No,No,No,No,"2020-04-05 03:05:45"
83053436,1.1.25,iOS,69.6748285,-,93.018506,e7bc5fb2-1231-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,No,Yes,Yes,No,"2020-02-26 16:34:29"
63425257,3.4.9,iOS,88.945103,171.605415,b328829e-17ae-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,No,No,No,No,"2020-04-01 17:49:14"
89571802,1.1.39,iOS,6.4794635,4.974110,de457111-3914-4450-bc11-b7b8dbab5e1,e199492-17ae-11eb-adcl-0242ac120002,Yes,No,No,Yes,"2020-03-01 13:46:05"
87051710,4.2.3,iOS,-,5.041729,29.969392,e7bc5fb2-1231-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,No,No,No,No,"2020-09-12 17:21:42"
8407369,1.3.6,Android,83.702958,11.236421,e7bc5fb2-1231-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,No,Yes,No,No,"2020-10-23 07:43:55"
30901771,1.3.32,iOS,50.861921,-,18.663633,de457111-3914-4450-bc11-b7b8dbab5e1,fcba68aa-1231-11eb-adcl-0242ac120002,No,No,Yes,Yes,"2020-06-02 12:01:25"
7047855,4.3.3,iOS,50.861761,21.621528,de457111-3914-4450-bc11-b7b8dbab5e1,fcba68aa-1231-11eb-adcl-0242ac120002,Yes,No,Yes,Yes,"2020-03-03 08:39:54"
50945957,7.2.21,Android,-,19.018395,-,20.792102,de457111-3914-4450-bc11-b7b8dbab5e1,fcba68aa-1231-11eb-adcl-0242ac120002,No,No,No,No,"2020-07-18 15:55:53"
7036669,4.1.9,Android,-,68.149568,37.836674,b328829e-17ae-11eb-adcl-0242ac120002,e199492-17ae-11eb-adcl-0242ac120002,No,Yes,No,No,"2020-11-03 05:29:18"
71315595,4.1.18,iOS,-,52.566665,-,177.147545,de457111-3914-4450-bc11-b7b8dbab5e1,a95dd57b-779f-49db-819d-b696043e554,Yes,Yes,No,No,"2020-01-02 22:57:15"
70426642,2.3.13,iOS,76.746763,93.548616,b328829e-17ae-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,Yes,No,Yes,Yes,"2020-03-11 17:45:41"
10895363,1.1.1,Android,-,10.027164,46.705580,de457111-3914-4450-bc11-b7b8dbab5e1,e199492-17ae-11eb-adcl-0242ac120002,Yes,No,No,No,"2020-01-26 19:23:55"
99247595,2.3.15,Android,-,90.32165,49.306042,e7bc5fb2-1231-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,No,Yes,No,No,"2020-01-07 08:34:08"
63762620,3.2.1,Android,60.507770,56.985170,de457111-3914-4450-bc11-b7b8dbab5e1,e199492-17ae-11eb-adcl-0242ac120002,No,No,Yes,No,"2020-09-23 00:48:06"
84418220,4.1.40,iOS,64.6043925,-,53.285161,b328829e-17ae-11eb-adcl-0242ac120002,fcba68aa-1231-11eb-adcl-0242ac120002,No,No,No,No,"2020-06-29 19:29:04"
27581545,1.1.21,Android,-,63.363132,-,71.333768,b328829e-17ae-11eb-adcl-0242ac120002,e199492-17ae-11eb-adcl-0242ac120002,Yes,No,Yes,Yes,"2020-06-22 15:38:59"
68562728,1.1.35,Android,-,28.0262385,118.632567,e7bc5fb2-1231-11eb-adcl-0242ac120002,fcba68aa-1231-11eb-adcl-0242ac120002,Yes,Yes,Yes,Yes,"2020-03-22 06:03:07"
39596628,1.1.10,iOS,38.562533,120.214812,e7bc5fb2-1231-11eb-adcl-0242ac120002,a95dd57b-779f-49db-819d-b696043e554,No,No,No,No,"2020-08-02 18:20:59"
99647462,2.3.13,Android,-,12.2980225,-,78.433987,b328829e-17ae-11eb-adcl-0242ac120002,e199492-17ae-11eb-adcl-0242ac120002,No,No,Yes,No,"2020-10-26 05:22:30"
43144896,4.1.2,iOS,72.646367,-,168.699501,e7bc5fb2-1231-11eb-adcl-0242ac120002,e199492-17ae-11eb-adcl-0242ac120002,No,Yes,No,No,"2020-06-30 15:58:09"
67203137,3.1.8,Android,23.0276285,-,73.294984,e7bc5fb2-1231-11eb-adcl-0242ac120002,fcba68aa-1231-11eb-adcl-0242ac120002,No,Yes,No,Yes,"2020-08-27 21:53:38"
96621736,2.3.18,iOS,19.6529285,4.539062,de457111-3914-4450-bc11-b7b8dbab5e1,fcba68aa-1231-11eb-adcl-0242ac120002,Yes,No,Yes,No,"2020-05-18 23:20:28"
64088393,1.19,Android,2.1040565,79.576637,b328829e-17ae-11eb-adcl-0242ac120002,e199492-17ae-11eb-adcl-0242ac120002,No,Yes,No,No,"2020-09-22 14:00:46"
52887956,2.3.35,iOS,79.0659475,146.564208,b328829e-17ae-11eb-adcl-0242ac120002,e199492-17ae-11eb-adcl-0242ac120002,No,Yes,No,No,"2020-06-20 20:59:34"
6619292,4.1.1,iOS,46.693072,145.619751,de457111-3914-4450-bc11-b7b8dbab5e1,e199492-17ae-11eb-adcl-0242ac120002,No,Yes,No,No,"2020-01-01 10:11:11"
49644489,4.2.23,Android,-,73.7454135,-,36.176752,de457111-3914-4450-bc11-b7b8dbab5e1,e199492-17ae-11eb-adcl-0242ac120002,No,No,No,No,"2020-01-05 05:29:03"
44643264,2.3.18,Android,-,29.0877635,-,127.722778,b328829e-17ae-11eb-adcl-0242ac120002,e199492-17ae-11eb-adcl-0242ac120002,Yes,Yes,Yes,No,"2020-01-19 13:32:23"

283
```

Task 2: Write a script to ingest the relevant bookings data from AWS RDS to Hadoop.

- Download and install mysql-connector-java-5.1.46 jar files using the below commands
 - `wget https://dev.mysql.com/get/Downloads/Connector-J/mysql-connector-java-5.1.46.tar.gz`

```
[hadoop@ip-10-0-0-16 ~]$ wget https://dev.mysql.com/get/Downloads/Connector-J/mysql-connector-java-5.1.46.tar.gz
--2023-02-15 17:10:02-- https://dev.mysql.com/get/Downloads/Connector-J/mysql-connector-java-5.1.46.tar.gz
Resolving dev.mysql.com (dev.mysql.com)... 23.0.26.132, 2600:1408:c400:1889::2e31, 2600:1408:c400:188c::2e31
Connecting to dev.mysql.com (dev.mysql.com)|23.0.26.132|:443... connected.
HTTP request sent, awaiting response... 302 Moved Temporarily
Location: https://cdn.mysql.com//archives/mysql-connector-java-5.1/mysql-connector-java-5.1.46.tar.gz [following]
--2023-02-15 17:10:03-- https://cdn.mysql.com//archives/mysql-connector-java-5.1/mysql-connector-java-5.1.46.tar.gz
Resolving cdn.mysql.com (cdn.mysql.com)... 23.62.172.238
Connecting to cdn.mysql.com (cdn.mysql.com)|23.62.172.238|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 4434926 (4.2M) [application/x-tar-gz]
Saving to: 'mysql-connector-java-5.1.46.tar.gz'

100%[=====>] 4,434,926 --.-K/s in 0.03s

2023-02-15 17:10:03 (152 MB/s) - 'mysql-connector-java-5.1.46.tar.gz' saved [4434926/4434926]
```

- o tar zxvf mysql-connector-java-5.1.46.tar.gz

```
mysql-connector-java-5.1.46/src/testsuite/simple/SimpleTransformer.java
mysql-connector-java-5.1.46/src/testsuite/simple/SplitDBdotNameTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/StatementsTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/StringUtilsTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/TestBug57662Logger.java
mysql-connector-java-5.1.46/src/testsuite/simple/TestLifecycleInterceptor.java
mysql-connector-java-5.1.46/src/testsuite/simple/TransactionTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/TraversalTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/UpdatabilityTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/UtilsTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/XATest.java
mysql-connector-java-5.1.46/src/testsuite/simple/jdbc4/StatementsTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/jdbc42/ConnectionTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/jdbc42/ResultSetTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/jdbc42/StatementsTest.java
mysql-connector-java-5.1.46/src/testsuite/simple/tb2-data.txt.gz
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/ca-cert.pem
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/ca-key.pem
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/ca-truststore
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/certs_howto.txt
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/client-cert.pem
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/client-key.pem
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/client-keystore
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/ssl-test-certs/mykey.pem
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/ssl-test-certs/mykey.pub
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/server-cert.pem
mysql-connector-java-5.1.46/src/testsuite/ssl-test-certs/server-key.pem
[hadoop@ip-10-0-0-16 ~]$
```

- o cd mysql-connector-java-5.1.46


```
[hadoop@ip-10-0-0-16 ~]$ cd mysql-connector-java-5.1.46
[hadoop@ip-10-0-0-16 mysql-connector-java-5.1.46]$
```

- o `sudo cp mysql-connector-java-5.1.46-bin.jar /usr/share/java/mysql-connector-java.jar`

```
[hadoop@ip-10-0-0-16 mysql-connector-java-5.1.46]$ sudo cp mysql-connector-java-5.1.46-bin.jar /usr/share/java/mysql-connector-java.jar
[hadoop@ip-10-0-0-16 mysql-connector-java-5.1.46]$
```

- import bookings data from RDS to HDFS folder '/user/capstone/carride1'

```
sqoop import --connect jdbc:mysql://upgraddetest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase --username student --password STUDENT123 --table bookings -m 1 --target-dir /user/capstone/ carride1
```

```
File Output Format Counters
      Bytes Written=165678
23/02/15 17:19:23 INFO mapreduce.ImportJobBase: Transferred 161.7949 KB in 14.69
17 seconds (11.0127 KB/sec)
23/02/15 17:19:23 INFO mapreduce.ImportJobBase: Retrieved 1000 records.
[hadoop@ip-10-0-0-16 ~]$
```

- Checked the files that got created after RDS data ingestion.

```
hdfs dfs -ls /user/capstone/ carride1
```

- Checked the bookings data that got imported from RDS

```
hadoop fs -cat /user/capstone/carride1/part-m-00000
```

```
hadoop@ip-172-31-2-189:~$ hdfs dfs -ls /user/capstone/viji
Found 2 items
-rw-r--r-- 1 hadoop hadoop 0 2023-02-17 11:19 /user/capstone/viji/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 165678 2023-02-17 11:19 /user/capstone/viji/part-m-00000
hadoop@ip-172-31-2-189:~$ hadoop fs -cat /user/capstone/viji/part-m-00000
BK9969087150,51811359,15055660,2.2.14,Android,-49.4319655,103.917851,-58.8043875,146.477367,2020-06-23 19:33:10.0,2020-06-06 09:02:10.0,534,83,INR,black,054-38-4479,4,3,3
BK629851904,31663218,60872180,3.4.1,iOS,-83.5408405,175.80085,86.20705,128.367238,2020-05-23 12:22:04.0,2020-08-09 19:02:56.0,126,67,INR,lime,796-39-6801,3,2,4
BK1797410350,86869399,94276051,4.1.36,iOS,-67.8930645,55.234128,-51.1079,-31.07475,2020-05-19 14:14:32.0,2020-08-23 18:38:39.0,297,63,INR,olive,748-73-1579,1,3,3
BK5788246325,58230837,45457227,2.4.27,Android,13.707887,113.499943,54.3812915,-18.437751,2020-03-24 01:30:15.0,2020-05-19 11:16:45.0,932,32,INR,white,558-80-6346,3,2,2
BK8342703255,84232510,86494681,4.1.34,Android,-6.091461,-114.649789,22.8449505,70.137827,2020-08-03 19:10:52.0,2020-03-24 08:25:40.0,260,7,INR,blue,068-72-1637,3,3,3
BK6015582453,11981042,35862658,2.4.39,iOS,-18.910034,-70.193103,-10.182921,173.877213,2020-07-17 05:33:48.0,2020-04-30 04:54:27.0,907,53,INR,purple,102-10-5639,3,2,3
BK4529355854,60071878,78022360,2.1.9,iOS,1.215274,-56.014903,35.152876,104.324905,2020-01-02 01:48:40.0,2020-02-16 04:28:55.0,547,17,INR,teal,866-83-4349,2,3,4
BK9720088219,14327312,94427067,3.1.2,Android,-55.4822225,173.362256,65.0121265,51.390751,2020-04-10 15:11:07.0,2020-01-20 21:17:42.0,259,33,INR,maroon,572-73-6526,3,3,2
BK7157532607,46407210,43160003,1.3.4,Android,46.005843,-16.826146,7.6126015,-156.428577,2020-06-09 05:56:31.0,2020-03-19 01:53:16.0,787,21,INR,olive,667-23-5880,2,2,3
BK5014871433,65861573,64708618,1.3.28,iOS,-29.565326,64.843709,84.068109,-49.820835,2020-08-14 20:43:42.0,2020-06-03 09:39:59.0,586,5,INR,fuchsia,255-52-5654,5,5,1
BK9051488736,37721758,27297770,2.3.13,Android,61.9364605,83.249705,0.0281895,115.469099,2020-04-07 04:27:59.0,2020-09-29 10:51:41.0,912,80,INR,aqua,739-09-9569,2,1,2
BK243762319,62552969,45877457,3.3.9,iOS,-62.6515155,-139.154028,28.0299995,-62.8556,2020-07-01 00:36:05.0,2020-09-30 17:40:23.0,821,23,INR,black,590-44-6613,2,3,4
BK46833595168,56801961,53401707,4.2.34,iOS,-5.860265,-100.004839,25.016591,70.471358,2020-05-03 10:17:56.0,2020-06-08 09:11:27.0,71,10,INR,fuchsia,454-04-0608,5,2,3
BK9783284253,66909721,40509554,2.2.22,Android,36.1913155,5.686264,88.988393,36.580599,2020-03-05 16:02:01.0,2020-05-29 13:36:15.0,26,81,INR,black,600-17-7043,3,1,3
BK2880021380,50163555,34005420,3.4.23,Android,-83.06599,106.268689,8.8300855,74.872352,2020-01-15 02:00:07.0,2020-05-12 21:53:04.0,571,99,INR,navy,506-09-4981,1,5,3
BK4537426043,91111754,59250769,1.2.19,iOS,-43.1188435,-99.935719,3.7026225,-46.828716,2020-04-28 05:18:34.0,2020-02-12 11:31:40.0,650,81,INR,white,362-35-8054,5,5,2
BK998130731,67875357,14562526,3.1.15,Android,-10.861959,-111.989853,57.233121,95.469986,2020-01-25 01:37:22.0,2020-04-28 09:42:00.0,590,3,INR,teal,359-51-9362,1,1,4
BK5645323730,18442993,84939946,3.1.29,Android,-81.472235,-88.404916,12.690818,-140.99768,2020-09-24 08:18:31.0,2020-07-16 05:12:24.0,515,1,INR,blue,024-35-8771,1,3,4
BK6163608413,36591778,11946210,4.2.38,Android,60.2036385,120.988501,32.103263,-50.551889,2020-07-26 06:12:56.0,2020-04-23 06:57:20.0,810,58,INR,silver,833-16-1378,3,5,1
BK6803373649,28382306,97222676,2.1.18,iOS,6.540056,161.083998,-12.943502,-148.232621,2020-09-20 15:52:49.0,2020-09-17 03:13:26.0,927,74,INR,maroon,747-70-555
```

Task 3: Create aggregates for finding date-wise total bookings using the Spark script.

- Created python file named '[datewise_bookings_aggregates_spark.py](#)' to aggregate the booking data datewise and storing the aggregate data on csv format in '[/home/hadoop/datewise_booking/](#)' local path.
- Command to move the csv file to HDFS

[hadoop fs -put /home/hadoop/datewise_booking /user/capstone/aggregatedata](#)

Screenshot of the aggregated data records

```
hadoop@ip-172-31-2-189:~$ hadoop fs -put /home/hadoop/datewise_booking /user/capstone/viji/aggregatedata
[hadoop@ip-172-31-2-189 ~]$ hdfs dfs -ls /user/capstone/viji/aggregatedata
Found 2 items
-rw-r--r-- 1 hadoop hadoop          0 2023-02-17 11:56 /user/capstone/viji/aggregatedata/_SUCCESS
-rw-r--r-- 1 hadoop hadoop    3777 2023-02-17 11:56 /user/capstone/viji/aggregatedata/part-00000-27506ddc-7568-4052-934c-adbb5316cb29-c000.csv
[hadoop@ip-172-31-2-189 ~]$ hdfs dfs -cat /user/capstone/viji/aggregatedata/part-00000-27506ddc-7568-4052-934c-adbb5316cb29-c000.csv^C
[hadoop@ip-172-31-2-189 ~]$ hdfs dfs -cat /user/capstone/viji/aggregatedata/part-00000-27506ddc-7568-4052-934c-adbb5316cb29-c000.csv
booking_date,count
2020-01-01,1
2020-01-02,3
2020-01-03,2
2020-01-04,2
2020-01-05,2
2020-01-06,3
2020-01-07,2
2020-01-08,4
2020-01-09,2
2020-01-10,2
2020-01-11,3
2020-01-12,3
2020-01-14,2
2020-01-15,5
2020-01-16,3
2020-01-17,4
2020-01-18,4
2020-01-20,4
2020-01-21,1
2020-01-23,4
2020-01-24,0
2020-01-25,4
2020-01-26,4
2020-01-27,3
2020-01-28,9
2020-01-29,2
2020-01-30,4
2020-01-31,5
2020-02-01,7
2020-02-02,6
2020-02-03,3
2020-02-04,4
2020-02-05,3
2020-02-06,3
2020-02-07,3
2020-02-08,6
```

```
hadoop@ip-172-31-2-189:~$ hdfs dfs -cat /user/capstone/viji/aggregatedata/part-00000-27506ddc-7568-4052-934c-adbb5316cb29-c000.csv
2020-09-13,5
2020-09-14,1
2020-09-15,5
2020-09-16,2
2020-09-17,4
2020-09-18,2
2020-09-19,4
2020-09-20,8
2020-09-21,4
2020-09-22,2
2020-09-23,3
2020-09-24,6
2020-09-25,1
2020-09-26,3
2020-09-27,1
2020-09-28,4
2020-09-29,2
2020-09-30,3
2020-10-01,1
2020-10-02,2
2020-10-03,4
2020-10-04,5
2020-10-05,4
2020-10-06,1
2020-10-07,1
2020-10-08,4
2020-10-09,5
2020-10-10,3
2020-10-11,3
2020-10-12,2
2020-10-13,2
2020-10-14,5
2020-10-15,3
2020-10-16,4
2020-10-17,4
2020-10-18,8
2020-10-20,3
2020-10-21,4
2020-10-22,3
2020-10-23,1
2020-10-24,3
2020-10-25,5
2020-10-26,6
[hadoop@ip-172-31-2-189 ~]$
```

Task 4:

- Command to move to hive from hadoop user

hive


```
[hadoop@ip-172-31-2-189 ~]$ hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> █
```

- **Create new database**

create database if not exists car;

- **Create a Hive-managed table for clickstream data.**

- **Command to load the data into Hive tables**

```
create table car.clickstream(customer_id string, app_version string, os_version
string, lat decimal(10,2), lon decimal(10,2), page_id string, button_id string,
is_button_click string, is_page_view string, is_scroll_up string, is_scroll_down
string, date_timestamp string) ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' LINES TERMINATED BY '\n';
```

```
hive> create table capstoneviji.clickstream(customer_id string, app_version string, os_version string, lat decimal(10,2), lon decimal(10,2), page_id string,
button_id string, is_button_click string, is_page_view string, is_scroll_up string, is_scroll_down string, date_timestamp string) ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' LINES TERMINATED BY '\n';
OK
Time taken: 0.065 seconds
hive> █
```

- **Command to show the created Hive tables**

show tables ;

```
hive> show tables;
OK
bookings
clickstream
datewisebookings
Time taken: 0.013 seconds, Fetched: 3 row(s)
hive> █
```

- **Command to load the data into Hive tables**

```
LOAD DATA INPATH '/user/capstone/clickstream/part-00000-9950f612-ebbd-
4ccf-b88e-bda25394a87a-c000.csv' INTO TABLE car.clickstream ;
```

79671395	4.3.3	iOS	-51.97	-4.68	b328829e-17ae-11eb-adc1-0242ac120002	e1e99492-17ae-11eb-adc1-0242ac120002	No	Yes	Yes	Yes
2020-01-21 20:11:49"										
91806237	2.4.38	Android	88.80	150.14	de545711-3914-4450-8c11-b17b8dabb5e1	fcba68aa-1231-11eb-adc1-0242ac120002	No	Yes	No	No
2020-02-21 08:59:10"										
35688284	2.1.40	iOS	-85.00	-44.52	b328829e-17ae-11eb-adc1-0242ac120002	fcba68aa-1231-11eb-adc1-0242ac120002	Yes	No	Yes	No
2020-05-18 12:09:33"										
63099569	2.2.1	iOS	17.38	-160.56	b328829e-17ae-11eb-adc1-0242ac120002	e1e99492-17ae-11eb-adc1-0242ac120002	Yes	No	No	Yes
2020-03-03 09:25:53"										
73730503	1.1.7	Android	78.15	167.20	de545711-3914-4450-8c11-b17b8dabb5e1	fcba68aa-1231-11eb-adc1-0242ac120002	No	Yes	No	No
2020-11-08 19:30:31"										
99590722	4.3.16	iOS	64.03	-179.54	de545711-3914-4450-8c11-b17b8dabb5e1	fcba68aa-1231-11eb-adc1-0242ac120002	No	Yes	No	Yes
2020-05-12 19:46:29"										
19325527	3.4.25	Android	-0.23	-131.18	e7bc5fb2-1231-11eb-adc1-0242ac120002	a95dd57b-779f-49db-819d-b6960483e554	No	Yes	Yes	No
2020-02-18 22:23:35"										
55553360	1.4.19	iOS	-42.85	41.31	e7bc5fb2-1231-11eb-adc1-0242ac120002	e1e99492-17ae-11eb-adc1-0242ac120002	Yes	Yes	No	No
2020-05-30 11:53:00"										
47059814	1.1.8	Android	49.66	-62.36	de545711-3914-4450-8c11-b17b8dabb5e1	fcba68aa-1231-11eb-adc1-0242ac120002	Yes	No	Yes	Yes
2020-04-21 02:49:54"										
98218956	1.4.16	Android	-58.40	83.38	de545711-3914-4450-8c11-b17b8dabb5e1	fcba68aa-1231-11eb-adc1-0242ac120002	No	No	No	No
2020-07-15 14:37:58"										
89287805	3.2.17	iOS	24.99	-62.30	e7bc5fb2-1231-11eb-adc1-0242ac120002	a95dd57b-779f-49db-819d-b6960483e554	No	No	No	No
2020-10-16 04:32:15"										
54265902	1.1.31	Android	55.20	-134.38	b328829e-17ae-11eb-adc1-0242ac120002	e1e99492-17ae-11eb-adc1-0242ac120002	No	No	Yes	Yes
2020-10-09 22:03:40"										
76699549	2.4.24	iOS	6.98	96.08	b328829e-17ae-11eb-adc1-0242ac120002	e1e99492-17ae-11eb-adc1-0242ac120002	Yes	No	No	Yes
2020-08-10 11:39:23"										
77933414	2.3.5	iOS	37.42	141.44	de545711-3914-4450-8c11-b17b8dabb5e1	a95dd57b-779f-49db-819d-b6960483e554	No	No	No	Yes
2020-04-26 18:11:28"										
13066424	4.1.22	iOS	-6.72	-99.92	e7bc5fb2-1231-11eb-adc1-0242ac120002	a95dd57b-779f-49db-819d-b6960483e554	Yes	Yes	Yes	No
2020-01-21 19:54:28"										
29685263	2.2.5	iOS	-16.83	-32.39	de545711-3914-4450-8c11-b17b8dabb5e1	fcba68aa-1231-11eb-adc1-0242ac120002	Yes	Yes	Yes	No
2020-09-14 23:58:18"										
26135567	1.3.5	Android	42.78	22.47	e7bc5fb2-1231-11eb-adc1-0242ac120002	e1e99492-17ae-11eb-adc1-0242ac120002	Yes	Yes	No	Yes
2020-03-16 01:56:04"										
68742799	2.3.33	iOS	11.32	170.02	b328829e-17ae-11eb-adc1-0242ac120002	fcba68aa-1231-11eb-adc1-0242ac120002	No	No	Yes	Yes
2020-08-31 00:43:05"										
69179497	2.1.25	iOS	89.63	-114.09	de545711-3914-4450-8c11-b17b8dabb5e1	fcba68aa-1231-11eb-adc1-0242ac120002	Yes	Yes	No	No
2020-01-30 06:54:38"										
86174555	2.1.37	iOS	40.41	-27.31	e7bc5fb2-1231-11eb-adc1-0242ac120002	fcba68aa-1231-11eb-adc1-0242ac120002	Yes	Yes	No	Yes
2020-01-24 23:51:38"										
31096758	4.4.9	iOS	-3.36	111.98	b328829e-17ae-11eb-adc1-0242ac120002	a95dd57b-779f-49db-819d-b6960483e554	No	No	No	Yes
2020-03-17 05:36:23"										
33113235	3.3.4	Android	52.71	-11.01	e7bc5fb2-1231-11eb-adc1-0242ac120002	a95dd57b-779f-49db-819d-b6960483e554	No	No	No	No
2020-01-20 14:18:29"										

89445384	2.3.29	Android	-74.42	-171.23	de545711-3914-4450-8c11-b17b8dabb5e1	e1e99492-17ae-11eb-adc1-0242ac120002	No	Yes	No	No
2020-03-09 06:06:38"										
23030502	2.1.33	Android	-72.53	-77.18	de545711-3914-4450-8c11-b17b8dabb5e1	e1e99492-17ae-11eb-adc1-0242ac120002	Yes	No	Yes	Yes
2020-09-17 21:43:14"										
81061658	3.3.1	iOS	32.23	154.13	de545711-3914-4450-8c11-b17b8dabb5e1	e1e99492-17ae-11eb-adc1-0242ac120002	Yes	No	Yes	No
2020-03-30 04:32:53"										
19413541	2.1.17	iOS	33.93	106.13	e7bc5fb2-1231-11eb-adc1-0242ac120002	fcba68aa-1231-11eb-adc1-0242ac120002	No	No	Yes	Yes
2020-01-25 16:44:03"										
65047706	2.4.27	iOS	-77.11	140.42	b328829e-17ae-11eb-adc1-0242ac120002	a95dd57b-779f-49db-819d-b6960483e554	No	Yes	Yes	No
2020-04-25 16:12:37"										
92497528	3.2.31	iOS	-69.01	46.26	b328829e-17ae-11eb-adc1-0242ac120002	a95dd57b-779f-49db-819d-b6960483e554	No	Yes	Yes	No
2020-05-24 16:05:43"										
53978698	1.3.34	Android	0.22	-8.53	de545711-3914-4450-8c11-b17b8dabb5e1	e1e99492-17ae-11eb-adc1-0242ac120002	Yes	Yes	Yes	Yes
2020-09-26 12:26:34"										
22360184	3.1.11	iOS	-50.99	-15.75	b328829e-17ae-11eb-adc1-0242ac120002	a95dd57b-779f-49db-819d-b6960483e554	Yes	Yes	No	Yes
2020-05-20 21:32:54"										
44585909	2.4.37	iOS	-23.10	-156.44	b328829e-17ae-11eb-adc1-0242ac120002	a95dd57b-779f-49db-819d-b6960483e554	No	No	Yes	No
2020-02-07 00:57:01"										
54823847	4.3.2	iOS	-48.66	-63.72	e7bc5fb2-1231-11eb-adc1-0242ac120002	fcba68aa-1231-11eb-adc1-0242ac120002	Yes	No	Yes	Yes
2020-03-08 14:03:15"										
21621758	2.2.22	iOS	-56.26	33.94	e7bc5fb2-1231-11eb-adc1-0242ac120002	fcba68aa-1231-11eb-adc1-0242ac120002	No	Yes	No	Yes
2020-08-01 02:46:47"										
33090503	2.1.4	Android	-53.88	150.94	e7bc5fb2-1231-11eb-adc1-0242ac120002	fcba68aa-1231-11eb-adc1-0242ac120002	Yes	No	No	Yes
2020-02-10 07:12:55"										
83436917	1.3.19	iOS	29.61	91.53	de545711-3914-4450-8c11-b17b8dabb5e1	e1e99492-17ae-11eb-adc1-0242ac120002	No	No	No	Yes
2020-07-30 05:04:13"										
68796997	2.3.29	iOS	43.30	-42.65	b328829e-17ae-11eb-adc1-0242ac120002	e1e99492-17ae-11eb-adc1-0242ac120002	Yes	Yes	Yes	Yes
2020-10-17 20:00:19"										
84392327	1.3.12	iOS	40.75	-105.60	de545711-3914-4450-8c11-b17b8dabb5e1	fcba68aa-1231-11eb-adc1-0242ac120002	Yes	No	No	No
2020-01-07 12:48:35"										
39542434	2.1.11	iOS	-8.89	31.78	e7bc5fb2-1231-11eb-adc1-0242ac120002	fcba68aa-1231-11eb-adc1-0242ac120002	Yes	Yes	Yes	No
2020-04-05 18:02:34"										
16813062	3.2.18	Android	-88.34	-176.29	b328829e-17ae-11eb-adc1-0242ac120002	e1e99492-17ae-11eb-adc1-0242ac120002	No	Yes	No	No
2020-05-05 23:08:24"										
13307480	4.3.36	iOS	26.66	109.84	b328829e-17ae-11eb-adc1-0242ac120002	e1e99492-17ae-11eb-adc1-0242ac120002	No	Yes	Yes	Yes
2020-06-12 19:23:21"										
66537285	3.4.37	Android	-43.48	108.72	de545711-3914-4450-8c11-b17b8dabb5e1	e1e99492-17ae-11eb-adc1-0242ac120002	Yes	Yes	No	Yes
2020-09-07 17:10:58"										
48434687	4.1.28	Android	-7.89	-54.73	b328829e-17ae-11eb-adc1-0242ac120002	a95dd57b-779f-49db-819d-b6960483e554	Yes	Yes	No	No
2020-02-18 19:48:48"										
99030364	2.3.35	iOS	59.45	-66.16	e7bc5fb2-1231-11eb-adc1-0242ac120002	e1e99492-17ae-11eb-adc1-0242ac120002	Yes	Yes	No	Yes
2020-04-15 17:33:44"										
Time taken: 0.122 seconds, Fetched: 3001 row(s)										
hive>										

- Create a Hive-managed table for bookings data.

- **Command to create the Hive tables**

create table car.bookings(booking_id string, customer_id int, driver_id int, customer_app_version string, customer_phone_os_version string, pickup_lat decimal(5,2), pickup_lon decimal(5,2), drop_lat decimal(5,2), drop_lon decimal(5,2),

```
pickup_timestamp timestamp, drop_timestamp timestamp, trip_fare decimal(8,2),
tip_amount decimal(8,2), currency_code string, cab_color string,
cab_registration_no string, customer_rating_by_driver int, rating_by_customer int,
passenger_count int) ROW FORMAT SERDE
'org.apache.hadoop.hive.contrib.serde2.MultiDelimitSerDe' WITH
SERDEPROPERTIES ("field.delim"=",","line.delim" = '\n') STORED AS TEXTFILE;
```

```
hive> create table capstoneviji.bookings(booking_id string, customer_id int, driver_id int, customer_app_version string, customer_phone_os_version string, pi
ckup_lat decimal(5,2), pickup_lon decimal(5,2), drop_lat decimal(5,2), drop_lon decimal(5,2), pickup_timestamp timestamp, drop_timestamp timestamp, trip_fare
decimal(8,2), tip_amount decimal(8,2), currency_code string, cab_color string, cab_registration_no string, customer_rating_by_driver int, rating_by_customer
int, passenger_count int) ROW FORMAT SERDE 'org.apache.hadoop.hive.contrib.serde2.MultiDelimitSerDe' WITH SERDEPROPERTIES ("field.delim"=",","line.delim" =
'\n') STORED AS TEXTFILE;
OK
Time taken: 0.485 seconds
hive> █
```

- **Command to show the created Hive tables**

show tables ;

```
hive> show tables ;
OK
bookings
Time taken: 0.036 seconds, Fetched: 1 row(s)
hive> █
```

- **Command to load the data into Hive tables**

LOAD DATA INPATH '/user/capstone/bookings/part-m-00000' INTO TABLE
car.bookings ;

- **Screenshot of the records**

```
hadoop@ip-172-31-2-189:~$ hive> select * from bookings;
```

BK0968087150	51811359	15055660	2.2.14	Android	-49.43	103.92	-58.80	146.48	2020-06-23 19:33:10	2020-06-06 09:02:10	534.00	83.00
INR	black	054-38-4479	4	3								
BK629851904	31663218	60872180	3.4.1	iOS	-83.54	175.80	86.21	128.37	2020-05-23 12:22:04	2020-08-09 19:02:56	126.00	67.00
INR	lime	796-39-6801	3	4								
BK1797410350	86869399	94276051	4.1.36	iOS	-67.89	55.23	-51.11	-31.07	2020-05-19 14:14:32	2020-08-23 18:38:39	297.00	63.00
INR	olive	748-73-1579	1	3								
BK5788246325	58230837	45457227	2.4.27	Android	13.71	113.50	54.38	-18.44	2020-03-24 01:30:15	2020-05-19 11:16:45	932.00	32.00
INR	white	558-80-6346	3	2								
BK8342703255	84232510	86494681	4.1.34	Android	-6.09	-114.65	22.84	70.14	2020-08-03 19:10:52	2020-03-24 08:25:40	260.00	7.00
NR	blue	068-72-1637	3	3								
BK6015582453	11981042	35862658	2.4.39	iOS	-18.91	-70.19	-10.18	173.88	2020-07-17 05:33:48	2020-04-30 04:54:27	907.00	53.00
INR	purple	102-10-5639	3	2								
BK4529355854	60071878	78022360	2.1.9	iOS	1.22	-56.01	35.15	104.32	2020-01-02 01:48:40	2020-02-16 04:28:55	547.00	17.00
INR	teal	866-83-4349	2	3								
BK9720088219	14327312	94427067	3.1.2	Android	-55.48	173.36	65.01	51.39	2020-04-10 15:11:07	2020-01-20 21:17:42	259.00	33.00
INR	maroon	572-73-6526	3	2								
BK7157532607	46407210	43160003	1.3.4	Android	46.01	-16.83	7.61	-156.43	2020-06-09 05:56:31	2020-03-19 01:53:16	787.00	21.00
INR	olive	667-23-5880	2	2								
BK5014871433	65861573	64708618	1.3.28	iOS	-29.57	64.84	84.07	-49.82	2020-08-14 20:43:42	2020-06-03 09:39:59	586.00	5.00
NR	fuchsia	255-52-5654	5	1								
BK9051488736	37721758	27297770	2.3.13	Android	61.94	83.25	0.03	115.47	2020-04-07 04:27:59	2020-09-29 10:51:41	912.00	80.00
INR	aqua	739-05-9569	2	1								
BK243762319	62552969	45877457	3.3.9	iOS	-62.65	-139.15	28.03	-62.86	2020-07-01 00:36:05	2020-09-30 17:40:23	821.00	23.00
INR	black	590-44-6613	2	3								
BK4683595168	56801961	53401707	4.2.34	iOS	-5.86	-100.00	25.02	70.47	2020-05-03 10:17:56	2020-06-08 09:11:27	71.00	10.00
INR	fuchsia	454-04-0608	5	2								
BK9783284253	66909721	40509554	2.2.22	Android	36.19	5.69	88.99	36.58	2020-03-05 16:02:01	2020-05-29 13:36:15	26.00	81.00
INR	black	600-17-7043	3	1								
BK2880021380	50163555	34005420	3.4.23	Android	-83.07	106.27	8.83	74.87	2020-01-15 02:00:07	2020-05-12 21:53:04	571.00	99.00
INR	navy	506-09-4981	1	5								
BK4537426043	91111754	59250769	1.2.19	iOS	-43.12	-99.94	3.70	-46.83	2020-04-28 05:18:34	2020-02-12 11:31:40	650.00	81.00
INR	white	362-35-8054	5	2								
BK998130731	67875357	14562526	3.1.15	Android	-10.86	-111.99	57.23	95.47	2020-01-25 01:37:22	2020-04-28 09:42:00	590.00	3.00
NR	teal	359-51-9362	1	1								
BK5645323730	18442993	84939946	3.1.29	Android	-81.47	-88.40	12.69	-141.00	2020-09-24 08:18:31	2020-07-16 05:12:24	515.00	1.00
NR	blue	024-35-8771	1	3								
BK6163608413	36591778	11946210	4.2.38	Android	60.20	120.99	32.10	-50.55	2020-07-26 06:12:56	2020-04-23 06:57:20	810.00	58.00
INR	silver	833-16-1378	3	5								
BK6803373649	28382306	97222676	2.1.18	iOS	6.54	161.08	-12.94	-148.23	2020-09-20 15:52:49	2020-09-17 03:13:26	927.00	74.00
INR	maroon	747-70-5557	2	2								
BK9764570097	61225539	15265942	4.4.14	iOS	80.72	179.70	-33.35	134.01	2020-01-26 02:20:39	2020-06-12 15:05:49	246.00	72.00
INR	blue	332-71-7565	5	1								

```
hadoop@ip-172-31-2-189:~$ hive> select * from bookings;
```

BK7038242108	16698708	27674043	4.4.4	Android	-11.97	-167.37	70.40	173.93	2020-05-06 18:31:33	2020-07-14 02:15:47	467.00	94.00
INR	green	236-56-9503	2	1								
BK1805019782	72602005	80532337	3.2.26	iOS	-80.12	151.35	37.10	-143.29	2020-08-01 21:33:09	2020-04-08 18:07:14	693.00	62.00
INR	olive	709-92-6170	1	3								
BK1718641612	67559017	61308810	4.4.30	iOS	35.81	118.79	36.32	-69.32	2020-06-21 01:20:34	2020-03-25 12:11:42	932.00	63.00
INR	fuchsia	723-15-9493	2	5								
BK7168466183	15278180	42292257	3.1.40	iOS	50.21	142.63	19.39	145.43	2020-01-28 16:32:38	2020-05-28 10:25:29	263.00	21.00
INR	fuchsia	568-30-4169	2	4								
BK4797431513	28119681	42859819	1.1.8	Android	36.39	-69.68	45.99	-141.02	2020-04-17 10:46:25	2020-10-17 00:56:14	481.00	1.00
NR	yellow	410-04-8620	2	5								
BK817041033	43759137	43883071	3.1.5	Android	82.20	53.59	-43.36	-115.99	2020-06-02 22:13:15	2020-10-18 23:20:14	260.00	81.00
INR	maroon	151-30-6951	2	3								
BK3091366811	78172665	13232332	1.1.19	iOS	-54.24	-77.78	-22.02	71.88	2020-09-17 21:17:22	2020-07-25 03:14:58	751.00	89.00
INR	blue	062-69-8682	5	3								
BK2211742187	98294088	38321934	3.4.35	Android	-60.74	-21.29	-25.73	-12.31	2020-02-08 11:34:53	2020-04-17 04:14:10	295.00	22.00
INR	maroon	379-54-1940	4	3								
BK5718782371	91413624	50796177	1.2.17	iOS	12.21	-49.50	35.84	21.18	2020-08-02 15:24:44	2020-06-02 16:24:03	649.00	1.00
NR	olive	687-83-0160	4	1								
BK6925041861	16754182	65101252	2.3.4	iOS	-50.39	-156.10	81.07	-135.87	2020-09-11 10:49:04	2020-03-05 22:44:59	168.00	69.00
INR	purple	711-68-5952	4	2								
BK1273485977	77394438	78242097	3.4.20	iOS	-81.81	-2.09	-46.86	-69.44	2020-01-24 03:41:51	2020-02-11 06:38:16	442.00	99.00
INR	green	210-68-5025	3	4								
BK1840869068	70959823	34539083	1.1.28	Android	-62.96	6.00	-11.00	81.95	2020-08-23 05:50:43	2020-07-31 14:24:55	875.00	40.00
INR	olive	401-09-0475	5	3								
BK997882598	58504148	46158041	4.3.35	iOS	29.94	-124.22	-58.44	-89.12	2020-01-12 08:48:05	2020-04-20 20:47:16	230.00	3.00
NR	green	266-60-7424	3	1								
BK5662087214	77440134	39470327	3.2.17	Android	13.97	16.86	-42.64	172.76	2020-02-02 12:03:37	2020-09-26 06:29:56	941.00	88.00
INR	gray	279-59-5349	5	4								
BK8865887274	71215906	71671654	2.3.21	iOS	-85.65	-6.88	9.91	-11.56	2020-08-12 05:42:41	2020-06-15 18:24:22	894.00	72.00
INR	green	228-40-4173	2	2								
BK9708496297	29394176	20731780	2.4.7	Android	-45.19	-33.41	-67.90	78.63	2020-05-31 12:33:42	2020-04-12 21:21:13	938.00	15.00
INR	aqua	616-72-4971	2	5								
BK605227090	94946719	98071218	3.4.8	iOS	-4.81	-76.31	76.07	-136.37	2020-07-25 14:13:40	2020-07-04 14:08:20	716.00	19.00
INR	teal	112-71-5489	2	5								
BK8602928713	33696426	68228240	4.2.33	iOS	29.87	84.00	-43.82	88.59	2020-07-10 22:27:16	2020-06-07 20:32:40	763.00	65.00
INR	fuchsia	069-53-2145	5	1								
BK6243816121	74681879	25278350	2.3.24	iOS	-53.20	-50.22	85.72	-25.82	2020-05-11 20:53:51	2020-08-26 11:52:47	196.00	78.00
INR	purple	874-93-2459	4	1								
BK9843664360	67564464	71913052	2.1.30	iOS	-10.03	-49.10	73.86	157.58	2020-09-20 22:51:48	2020-01-27 22:11:07	289.00	49.00
INR	maroon	876-76-7598	3	3								
BK6282269780	86988153	87739332	4.4.35	iOS	64.29	95.84	-84.54	-96.12	2020-06-23 05:38:14	2020-07-11 16:40:57	196.00	89.00
INR	purple	740-96-7931	4	5								

Time taken: 1.231 seconds, Fetched: 1000 row(s)

- Create a Hive-managed table for aggregated data in Task 3.

○ Command to create the Hive tables

create table datewisebookings(booking_date date, count int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n';

```
hive> create table datewisebookings(booking_date date, count int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n';
OK
Time taken: 0.045 seconds
hive> █
```

- **Command to show the created Hive tables**

show tables ;

```
hive> show tables;
OK
bookings
datewisebookings
Time taken: 0.016 seconds, Fetched: 2 row(s)
hive> █
```

- **Command to load the data into Hive tables**

LOAD DATA INPATH '/user/capstone/aggregatedata/part-00000-27506ddc-7568-4052-934c-adbb5316cb29-c000.csv' INTO TABLE car.datewisebookings ;

- **Screenshot of the records**

Select count(*) from car.datewisebookings;

```
Query ID = hadoop_20230217131211_f856174a-b2b3-4475-892b-37b3eb071c61
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1676632604307_0006)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED   1         1           0         0         0         0
Reducer 2 ..... container  SUCCEEDED   1         1           0         0         0         0
-----
VERTICES: 02/02  [=====>>>] 100%  ELAPSED TIME: 4.27 s
-----
OK
290
Time taken: 4.696 seconds, Fetched: 1 row(s)
hive> █
```

Select * from car.datewisebookings;

```
hive> select * from capstoneviji.datewisebookings;
OK
NULL      NULL
2020-01-01      1
2020-01-02      3
2020-01-03      2
2020-01-04      2
2020-01-05      2
2020-01-06      3
2020-01-07      2
2020-01-08      4
2020-01-09      2
2020-01-10      2
2020-01-11      3
2020-01-12      3
2020-01-14      2
2020-01-15      5
2020-01-16      3
2020-01-17      4
2020-01-18      4
2020-01-20      4
2020-01-21      1
2020-01-23      4
2020-01-24      8
2020-01-25      4
2020-01-26      4
2020-01-27      3
2020-01-28      9
2020-01-29      2
2020-01-30      4
2020-01-31      5
2020-02-01      7
2020-02-02      6
2020-02-03      3
2020-02-04      4
2020-02-05      3
2020-02-06      3
2020-02-07      3
2020-02-08      6
2020-02-09      6
2020-02-10      3
2020-02-11      1
2020-02-12      3
2020-02-13      2
```

```
hadoop@ip-172-31-2-189~
2020-09-14      1
2020-09-15      5
2020-09-16      2
2020-09-17      4
2020-09-18      2
2020-09-19      4
2020-09-20      8
2020-09-21      4
2020-09-22      2
2020-09-23      3
2020-09-24      6
2020-09-25      1
2020-09-26      3
2020-09-27      1
2020-09-28      4
2020-09-29      2
2020-09-30      3
2020-10-01      1
2020-10-02      2
2020-10-03      4
2020-10-04      5
2020-10-05      4
2020-10-06      1
2020-10-07      1
2020-10-08      4
2020-10-09      5
2020-10-10      3
2020-10-11      3
2020-10-12      2
2020-10-13      2
2020-10-14      5
2020-10-15      3
2020-10-16      4
2020-10-17      4
2020-10-18      8
2020-10-20      3
2020-10-21      4
2020-10-22      3
2020-10-23      1
2020-10-24      3
2020-10-25      5
2020-10-26      6
Time taken: 0.11 seconds, Fetched: 290 row(s)
hive>
```