Queries

<Hive Query for Task 5>

Task 5: Calculate the total number of different drivers for each customer.

```
SELECT customer_id, count(DISTINCT driver_id)
FROM booking_data
GROUP BY customer_id
ORDER BY customer_id ASC;
```

```
Total MapReduce CPU Time Spent: 12 seconds 980 msec
OK
10022393        1
10058402        1
10339567        1
10435129        1
10555335        1
10592274        1
10614890        1
10678994        1
11264797        1
11353346        1
11418437        1
11438890        1
11454977        1
11479815        1
11518953        1
11580321        1
11596512        1
11608791        1
11655671        1
11757536        1
11764909        1
11860278        1
11981042        1
12106105        1
12142182        1
12312603        1
12334699        1
12367832        1
12856708        1
12885363        1
12913608        1
12914577        1
12966909        1
13015449        1
13229062        1
```

<Hive Query for Task 6>

Task 6: Calculate the total rides taken by each customer.

```
SELECT customer_id, count(DISTINCT booking_id)
FROM booking_data
GROUP BY customer_id
ORDER BY customer_id ASC;
```

```
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.18 sec   HDFS Read: 32766 HDFS Write: 220 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 180 msec
OK
10022393        1
10058402        1
10339567        1
10435129        1
10555335        1
10592274        1
10614890        1
10678994        1
11264797        1
11353346        1
11418437        1
11438890        1
11454977        1
11479815        1
11518953        1
11580321        1
11596512        1
11608791        1
11655671        1
```

<Hive Query for Task 7>

Task 7: Find the total visits made by each customer on the booking page and the total 'Book Now' button presses. This can show the conversion ratio.

```
SELECT (SUM(CASE WHEN button_id = "fcba68aa-1231-11eb-adc1-0242ac120002" AND
is_button_click = 'Yes' THEN 1 END) /
SUM(CASE WHEN page_id = "e7bc5fb2-1231-11eb-adc1- 0242ac120002" AND
is_page_view = 'Yes' THEN 1 END))
AS conversion_ratio
FROM clickstream_data;
```

```
Total MapReduce CPU Time Spent: 8 seconds 320 msec
OK
0.9688109161793372
```

<Hive Query for Task 8>

Task 8: Calculate the count of all trips done on black cabs.

```
SELECT COUNT (booking_id)
FROM booking_data
WHERE cab_color in ('black');
```

```
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.08 sec   HDFS Read: 25875 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 80 msec
OK
72
```

<Hive Query for Task 9>

Task 9: Calculate the total amount of tips given date wise to all drivers by customers.

```
SELECT date_format(pickup_timestamp,'yyyy-MM-dd'),
SUM(tip_amount)
FROM booking_data
GROUP BY date_format(pickup_timestamp,'yyyy-MM-dd');
```

```
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.7 sec   HDFS Read: 177106 HDFS Write: 4257 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 700 msec
OK
2020-01-01      59
2020-01-02      95
2020-01-03      11
2020-01-04      123
2020-01-05      134
2020-01-06      189
2020-01-07      148
2020-01-08      111
2020-01-09      48
2020-01-10      77
2020-01-11      81
2020-01-12      109
2020-01-14      142
2020-01-15      338
2020-01-16      155
2020-01-17      296
```

<Hive Query for Task 10>

Task 10: Calculate the total count of all the bookings with ratings lower than 2 as given by customers in a particular month.

```
SELECT date_format(pickup_timestamp,'yyyy-MM'),
COUNT(rating_by_customer)
FROM booking_data
WHERE rating_by_customer < 2
GROUP BY date_format(pickup_timestamp,'yyyy-MM') ;
```

```
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.31 sec   HDFS Read: 5375 HDFS Write: 110 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 40 msec
OK
2020-01 26
2020-02 16
2020-03 16
2020-04 21
2020-05 21
2020-06 14
2020-07 20
2020-08 32
2020-09 21
2020-10 15
```

<Hive Query for Task 11>

```
SELECT os_version, COUNT(DISTINCT customer_id)
FROM clickstream_data
WHERE os_version in ('iOS')
GROUP BY os_version;
```

```
Total MapReduce CPU Time Spent: 8 seconds 105 msec
OK
iOS   1503
Time taken: 32.875 seconds, Fetched: 1 row(s)
```