

Queries

<Hive Query for Task 5>

```
select customer_id ,count( DISTINCT driver_id) from booking_data group by customer_id order by customer_id asc;
```

<Screenshot after executing Query>

```
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.92 sec HDFS Read: 193168 HDFS Write: 23250 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.61 sec HDFS Read: 28232 HDFS Write: 11005 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 530 msec
OK
NULL 0
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
13262795 1
13356177 1
13387493 1
13389366 1
13442644 1
13500355 1
```

<Hive Query for Task 6>

```
select customer_id ,count( DISTINCT booking_id) from booking_data group by customer_id order by customer_id asc;
```

<Screenshot after executing Query>

```
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.93 sec HDFS Read: 193232 HDFS Write: 23250 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.61 sec HDFS Read: 28234 HDFS Write: 11005 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 540 msec
OK
NULL 1
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
```

<Hive Query for Task 7>

```
select count(b.button_id)/count(a.booking_id) from booking_data a full outer join
clickstream_data b on a.customer_id = b.customer_id;
```

<Screenshot after executing Query>

```
MapReduce Total cumulative CPU time: 4 seconds 990 msec
Ended Job = job_1646119500330_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 8.73 sec HDFS Read: 599352 HDFS Write: 119 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.99 sec HDFS Read: 5843 HDFS Write: 18 SUCCESS
Total MapReduce CPU Time Spent: 13 seconds 720 msec
OK
2.998001998001998
Time taken: 60.537 seconds, Fetched: 1 row(s)
```

<Hive Query for Task 8>

```
select cab_color ,count(distinct driver_id ) from booking_data where cab_color in ('black') group
by cab_color ;
```

<Screenshot after executing Query>

```
MapReduce Total cumulative CPU time: 6 seconds 30 msec
Ended Job = job_1646119500330_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.03 sec HDFS Read: 193878 HDFS Write: 9 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 30 msec
OK
black 72
Time taken: 28.219 seconds, Fetched: 1 row(s)
```

<Hive Query for Task 9>

select date_format(pickup_timestamp,'yyyy-MM-dd'),sum(tip_amount) from booking_data group by date_format(pickup_timestamp,'yyyy-MM-dd');

<Screenshot after executing Query>

```
Total MapReduce CPU Time Spent: 7 seconds 960 msec
OK
2020-01-01 295
2020-01-02 475
2020-01-03 55
2020-01-04 615
2020-01-05 670
2020-01-06 945
2020-01-07 740
2020-01-08 555
2020-01-09 240
2020-01-10 385
2020-01-11 405
2020-01-12 545
2020-01-14 710
2020-01-15 1690
2020-01-16 775
```

<Hive Query for Task 10>

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') ;

<Screenshot after executing Query>

```
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 9.24 sec HDFS Read: 926531 HDFS Write: 116 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 240 msec
OK
2020-01 130
2020-02 80
2020-03 80
2020-04 105
2020-05 105
2020-06 70
2020-07 100
2020-08 160
2020-09 105
2020-10 75
```

<Hive Query for Task 11>

select os_version ,count(distinct customer_id) from clickstream_data where os_version in ('iOS') group by os_version;

<Screenshot after executing Query>

```
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.03 sec HDFS Read: 407324 HDFS Write: 9 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 30 msec
OK
ios 1515
Time taken: 29.12 seconds, Fetched: 1 row(s)
```