



Queries

<Hive Query for Task 5>

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

```
customer_id ,count( DISTINCT driver_id) from booking
Status: Running (Executing on YARN cluster with App id application_166533174127
Map 1: 0/1 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 0/1 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 0(+1)/1 Reducer 2: 0/2 Reducer 3: 0/1
Map 1: 1/1 Reducer 2: 0(+1)/2 Reducer 3:
Map 1: 1/1 Reducer 2: 1(+1)/2 Reducer 3:
Map 1: 1/1 Reducer 2: 2/2 Reducer 3: 0/1
Map 1: 1/1 Reducer 2: 2/2 Reducer 3: 0(+1)/1
Map 1: 1/1 Reducer 2: 2/2 Reducer 3: 1/1
Map 1: 1/1 Reducer 2: 2/2 Reducer 3: 1/1
                                                   Reducer 3: 0/1
Reducer 3: 0/1
10022393
10058402
 Map 1: 1/1
                             Reducer 2: 2/2
                                                         Reducer 3: 0(+1)/1
 Map 1: 1/1
                             Reducer 2: 2/2
                                                       Reducer 3: 1/1
 OK
 10022393
 10058402
 10339567
 10435129
 10555335
 10592274
 10614890
 10678994
 11264797
 11353346
 11418437
 11438890
 11454977
 11479815
 11518953
 11580321
 11596512
 11608791
 11655671
 11757536
 11764909
```





<Hive Query for Task 6>

select customer_id ,count(DISTINCT booking_id) from booking group by customer_id order by customer id asc;

<Screenshot after executing Query>

```
10022393
10058402
10339567
10435129
10555335
10592274
10614890
10678994
11264797
11353346
11418437
11438890
11454977
11479815
11518953
11580321
11596512
11608791
11655671
11757536
11764909
11860278
```

<Hive Query for Task 7>

select count(page_id), sum(case when is_button_click = 'Yes' then 1 else 0 end), sum(case when is_button_click = 'Yes' then 1 else 0 end)/count(page_id) from

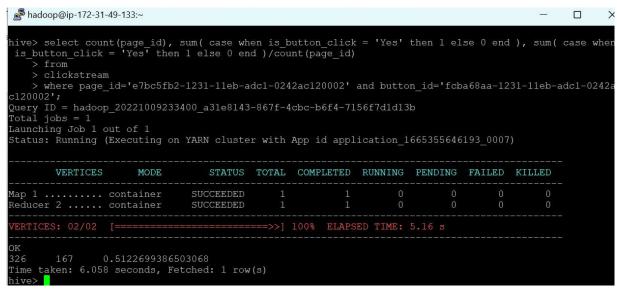




clickstream

where page_id='e7bc5fb2-1231-11eb-adc1-0242ac120002' and button_id='fcba68aa-1231-11eb-adc1-0242ac120002';

<Screenshot after executing Query>



<Hive Query for Task 8>

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

<Screenshot after executing Query>

<Hive Query for Task 9>





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

<Screenshot after executing Query>

```
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1665355646193_0007)
      VERTICES
                             STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
                   MODE
Map 1 ..... container
Reducer 2 ..... container
                           SUCCEEDED
2020-01-01
2020-01-02
              123
134
2020-01-04
2020-01-11
2020-01-12
2020-01-18
2020-01-20
              5
209
231
2020-01-21
2020-01-27
2020-01-29
```

<Hive Query for Task 10>

select date_format(pickup_timestamp,'yyyy-MM') ,count(rating_by_customer) from booking where rating_by_customer < 2 group by date_format(pickup_timestamp,'yyyy-MM');

<Screenshot after executing Query>





<Hive Query for Task 11>

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

<Screenshot after executing Query>