Logic for Final Submission

<Explain the queries, list them and attach screenshots after successful execution of queries>

Hive Queries for Task 5 – 11.

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

Query:
select customer_id, count(driver_id) from yourowncabs.bookings_data group by customer_id ;

Explanation:
The above hive query gives the customer_id of each customer and total number of different (distinct) drivers for each customer.

Screenshot of the successful execution of the query:

```
hive> select customer_id, count(driver_id) from yourowncabs.bookings_data group by customer_id ;
Query ID = hadoop_20220328033455_6bab9b88-af2a-4f55-b69a-462bf53b1dac
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1648435655208_0009)

--------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 6.97 s
--------------------------------------------------------------------------------
OK
10022393        1
10555335        1
10592274        1
10678994        1
11264797        1
11418437        1
11438890        1
11518953        1
11580321        1
11764909        1
11860278        1
12312603        1
12334699        1
12367832        1
12885363        1
12914577        1
12966909        1
13015449        1
13229062        1
13262795        1
13356177        1
```

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

Query:
select customer_id, count(booking_id) from yourowncabs.bookings_data group by customer_id ;

Explanation:
The above hive query gives the customer_id of each customer and the total number of rides taken by the customer.

Screenshot of the successful execution of the query:

```
hive> select customer_id, count(booking_id) from yourowncabs.bookings_data group by customer_id ;
Query ID = hadoop_20220328034006_22083e7e-d176-4924-b061-5c278f250a4b
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1648435655208_0009)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 6.62 s
----------------------------------------------------------------------------------------------
OK
10022393        1
10555335        1
10592274        1
10678994        1
11264797        1
11418437        1
11438890        1
11518953        1
11580321        1
11764909        1
11860278        1
12312603        1
12334699        1
12367832        1
12885363        1
12914577        1
12966909        1
13015449        1
13229062        1
13262795        1
13356177        1
```

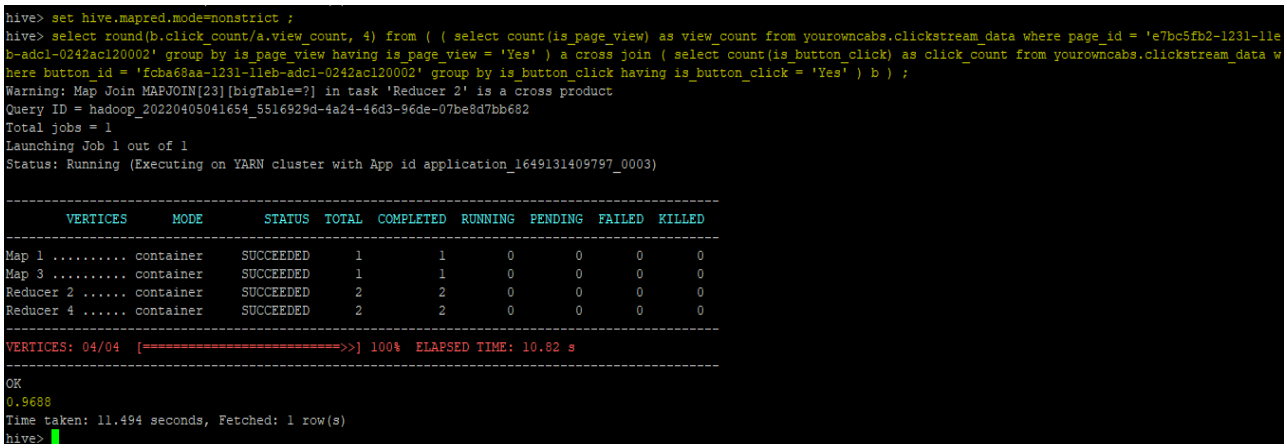
- **Task 7**: Find the total visits made by each customer on the booking page and the total ‘Book Now’ button presses. This can show the conversion ratio. The booking page id is **‘e7bc5fb2-1231-11eb-adc1-0242ac120002’**. The Book Now button id is **‘fcba68aa-1231-11eb-adc1-0242ac120002’**. You also need to calculate the conversion ratio as part of this task. Conversion ratio can be calculated as **Total 'Book Now' Button Press/Total Visits made by customer on the booking page**.

Note: Run the below command in Hive CLI before running the query.
Command: set hive.mapred.mode=nonstrict ;

Query:
select round(b.click_count/a.view_count, 4) from ((select count(is_page_view) as view_count from yourowncabs.clickstream_data where page_id = 'e7bc5fb2-1231-11eb-adc1-0242ac120002' group by is_page_view having is_page_view = 'Yes') a cross join (select count(is_button_click) as click_count from yourowncabs.clickstream_data where button_id = 'fcba68aa-1231-11eb-adc1-0242ac120002' group by is_button_click having is_button_click = 'Yes') b) ;

Explanation:
The above hive query calculates the total visits made by each customer on the booking page with page id ‘’e7bc5fb2-1231-11eb-adc1-0242ac120002’ and total count of ‘Book Now’ button presses by the customers with button id ‘fcba68aa-1231-11eb-adc1-0242ac120002’. It then calculates the Conversion ratio given by **Total 'Book Now' Button Press/Total Visits made by customer on the booking page and gives the conversion ratio value.**

Screenshot of the successful execution of the query:

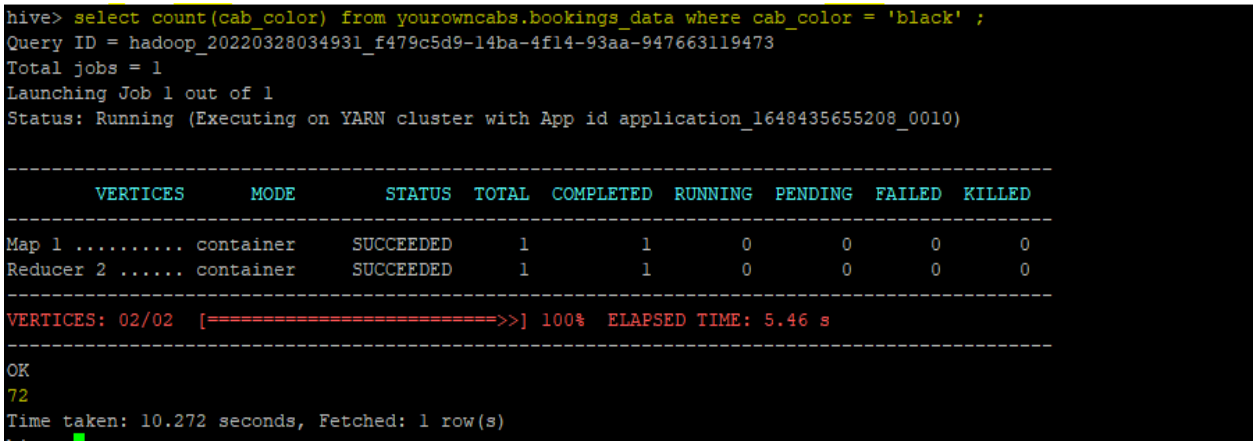


- **Task 8**: Calculate the count of all trips done on black cabs.

Query:
select count(cab_color) from yourowncabs.bookings_data where cab_color = 'black' ;

Explanation:
The above hive query calculates the total number of trips done where the cab colour was ‘black’.

Screenshot of the successful execution of the query:

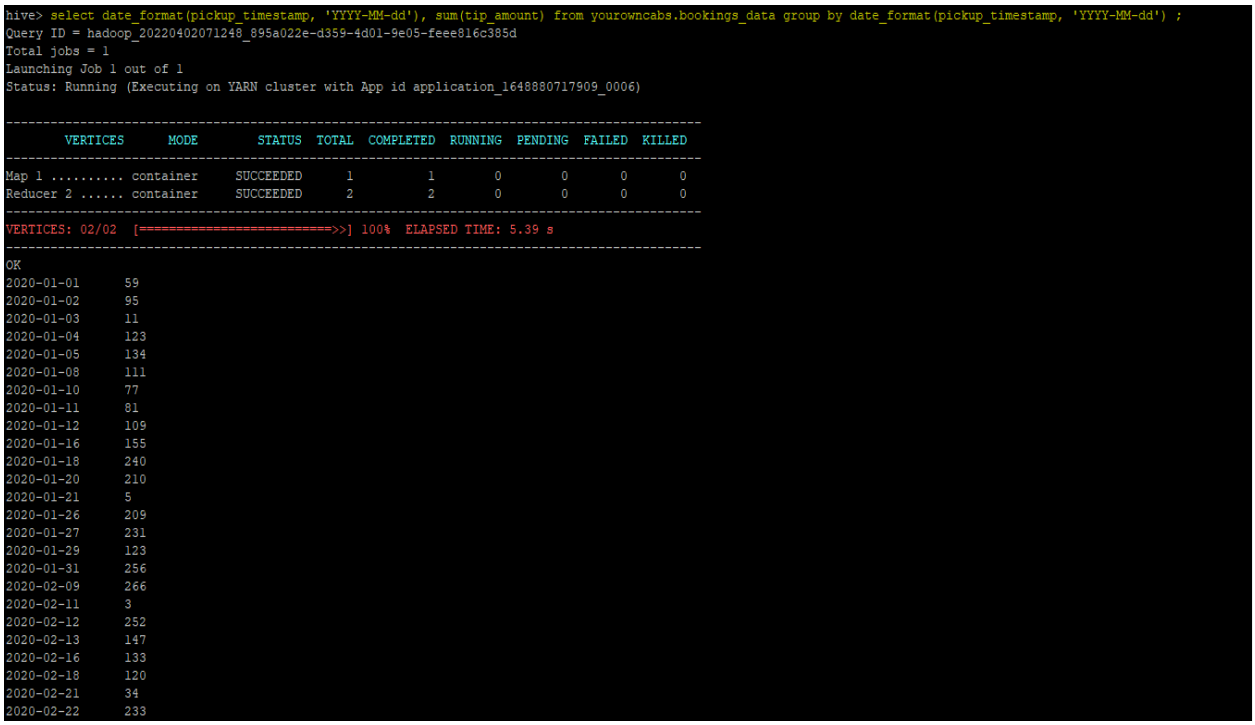


- **Task 9**: Calculate the total amount of tips given date wise to all drivers by customers.

Query:
select date_format(pickup_timestamp, 'YYYY-MM-dd'), sum(tip_amount) from yourowncabs.bookings_data group by date_format(pickup_timestamp, 'YYYY-MM-dd') ;

Explanation:
The above hive query calculates the date wise total amount of tips given by the customers to the drivers.

Screenshot of the successful execution of the query:

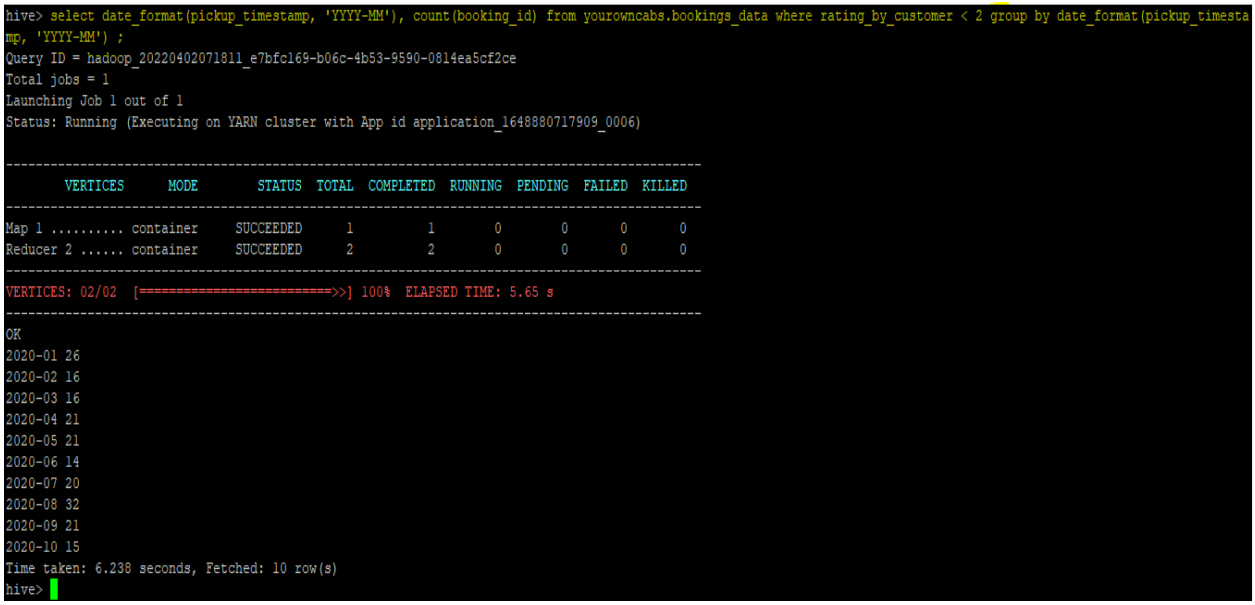


- **Task 10**: Calculate the total count of all the bookings with ratings lower than 2 as given by customers in a particular month.

Query:
select date_format(pickup_timestamp, 'YYYY-MM'), count(booking_id) from yourowncabs.bookings_data where rating_by_customer < 2 group by date_format(pickup_timestamp, 'YYYY-MM') ;

Explanation:
The above hive query calculates the month wise total number of bookings in which the customers have given the ratings lower than 2.

Screenshot of the successful execution of the query:



- **Task 11**: Calculate the count of total iOS users.

Query:
select count(os_version) from yourowncabs.clickstream_data where os_version = 'iOS' ;

Explanation:
The above hive query calculates the total number of customers whose phone OS version is ‘iOS’.

Screenshot of the successful execution of the query: