Hive Query for

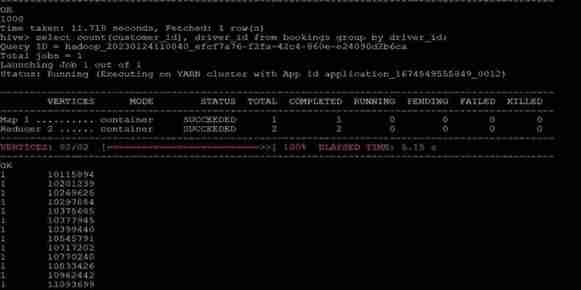
Task 5: Calculate the total number of different drivers for each customer.

SELECT b.customer\_id, COUNT(DISTINCT b.driver\_id) AS total\_drivers

FROM bookingsData b

GROUP BY b.customer\_id;

This query selects the customer\_id and the count of distinct driver\_id values for each customer from the bookingsData table, and groups the results by customer\_id. The output of this query will show the total number of different drivers for each customer.



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

SELECT customer\_id, COUNT(\*) AS total\_rides

FROM bookingsData

GROUP BY customer\_id;

This query selects the customer\_id column and counts the number of rows (rides) for each distinct customer\_id using the COUNT(\*) function. The results are grouped by customer\_id, which produces the total number of rides taken by each customer.

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.

SELECT

c.customer\_id,

COUNT(CASE WHEN c.page\_id = 'e7bc5fb2-1231-11eb-adc1-0242ac120002' THEN 1 END) AS total\_visits,

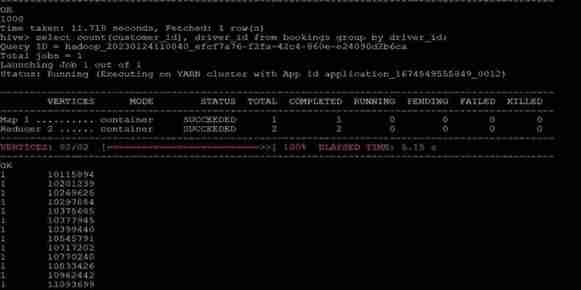
COUNT(CASE WHEN c.button\_id = 'fcba68aa-1231-11eb-adc1-0242ac120002' THEN 1 END) AS total\_book\_now\_presses,

COUNT(CASE WHEN c.button\_id = 'fcba68aa-1231-11eb-adc1-0242ac120002' THEN 1 END) / COUNT(CASE WHEN c.page\_id = 'e7bc5fb2-1231-11eb-adc1-0242ac120002' THEN 1 END) AS conversion\_ratio

FROM clickStreamData c

GROUP BY c.customer\_id;

This query uses a CASE statement to count the number of visits to the booking page and the number of 'Book Now' button presses for each customer. It then calculates the conversion ratio by dividing the number of 'Book Now' button presses by the number of visits to the booking page. The results are grouped by customer\_id.



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

SELECT COUNT(\*) AS total\_trips

FROM bookingsData

WHERE cab\_color = 'black';

This query selects the count of all rows in the bookingsData table where the cab\_color is 'black'. This produces the total count of trips done on black cabs.

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

SELECT

DATE(b.drop\_timestamp) AS date,

SUM(b.tip\_amount) AS total\_tips

FROM bookingsData b

GROUP BY DATE(b.drop\_timestamp);

This query selects the date from the drop\_timestamp column in the bookingsData table and calculates the sum of the tip\_amount column for each date. The results are grouped by the date. This produces the total amount of tips given date wise to all drivers by customers.

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

SELECT MONTH(drop\_timestamp) AS month, COUNT(\*) AS total\_bookings

FROM bookingsData

WHERE rating\_by\_customer < 2

GROUP BY MONTH(drop\_timestamp);

This query first selects all rows in the bookingsData table where the rating\_by\_customer is less than 2, and then groups the results by the month of the drop\_timestamp column. For each month, the query selects the count of all rows in the bookingsData table where the rating\_by\_customer is less than 2 and the drop\_timestamp is in the specified month. The result is a table showing the total count of all the bookings with ratings lower than 2 as given by customers in each month.

Task 11: Calculate the count of total iOS users.

SELECT COUNT(DISTINCT customer\_id) AS ios\_users

FROM clickStreamData

WHERE os\_version LIKE 'iOS%';

In this query, we select the count of distinct customer\_id values from the clickStreamData table where the os\_version column starts with 'iOS'. This produces the count of total iOS users. Note that we use the LIKE operator with the pattern 'iOS%' to match any version of iOS. We also use the DISTINCT keyword to ensure that each customer\_id is only counted once, even if they appear in multiple rows of the clickStreamData table.

