Queries

Hive Query for Task 5

Query:

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

Screenshot after executing Query

Hive Query for Task 6

Query:

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

Screenshot after executing 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
```

Hive Query for Task 7

Run the below command before executing query for task 7.

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) ;

Screenshot after executing Query

```
hive> set hive.mapred.mode=nonstrict ;
hive> 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 = 'e7bc5fb2-1231-11e
b-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 w
here button_id = 'fcba68aa-1231-11eb-adc1-0242ac120002' group by is_button_click having is_button_click = 'Yes' ) b ) ;
Warning: Map Join MAPJOIN[23][bigTable=?] in task 'Reducer 2' is a cross product
Query ID = hadoop_20220405041654_5516929d-4a24-46d3-96de-07be8d7bb682
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1649131409797_0003)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container    SUCCEEDED      1          1        0        0       0       0
Map 3 .......... container    SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container    SUCCEEDED      2          2        0        0       0       0
Reducer 4 ...... container    SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 04/04  [==========================>>] 100%  ELAPSED TIME: 10.82 s
----------------------------------------------------------------------------------------------
OK
0.9688
Time taken: 11.494 seconds, Fetched: 1 row(s)
hive>
```

Hive Query for Task 8

Query:

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

Screenshot after executing Query

```
hive> select count(cab_color) from yourowncabs.bookings_data where cab_color = 'black' ;
Query ID = hadoop_20220328034931_f479c5d9-14ba-4f14-93aa-947663119473
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1648435655208_0010)

----------------------------------------------------------------------------------------------
        VERTICES       MODE        STATUS   TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container    SUCCEEDED       1          1        0        0       0       0
Reducer 2 ...... container    SUCCEEDED       1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 5.46 s
----------------------------------------------------------------------------------------------
OK
72
Time taken: 10.272 seconds, Fetched: 1 row(s)
```

Hive Query for Task 9

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') ;

Screenshot after executing Query

```
hive> select date_format(pickup_timestamp, 'YYYY-MM-dd'), sum(tip_amount) from yourowncabs.bookings_data group by date_format(pickup_timestamp, 'YYYY-MM-dd') ;
Query ID = hadoop_20220402071248_895a022e-d359-4d01-9e05-feee816c385d
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1648880717909_0006)

----------------------------------------------------------------------------------------------
        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: 5.39 s
----------------------------------------------------------------------------------------------
OK
2020-01-01      59
2020-01-02      95
2020-01-03      11
2020-01-04      123
2020-01-05      134
2020-01-08      111
2020-01-10      77
2020-01-11      81
2020-01-12      109
2020-01-16      155
2020-01-18      240
2020-01-20      210
2020-01-21      5
2020-01-26      209
2020-01-27      231
2020-01-29      123
2020-01-31      256
2020-02-09      266
2020-02-11      3
2020-02-12      252
2020-02-13      147
2020-02-16      133
2020-02-18      120
2020-02-21      34
2020-02-22      233
```

Hive Query for Task 10

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') ;

Screenshot after executing Query

```
hive> select date_format(pickup_timestamp, 'YYYY-MM'), count(booking_id) from yourowncabs.bookings_data where rating_by_customer < 2 group by date_format(pickup_timesta
mp, 'YYYY-MM') ;
Query ID = hadoop_20220402071811_e7bfc169-b06c-4b53-9590-0814ea5cf2ce
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1648880717909_0006)

----------------------------------------------------------------------------------------------
        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: 5.65 s
----------------------------------------------------------------------------------------------
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
Time taken: 6.238 seconds, Fetched: 10 row(s)
hive>
```

Hive Query for Task 11

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

Screenshot after executing Query

```
hive> select count(os_version) from yourowncabs.clickstream_data where os_version = 'iOS' ;
Query ID = hadoop_20220328035638_0abc82f4-e9fa-4080-8cal-c936ld103596
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_0011)

--------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED     1         1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED     1         1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 6.11 s
--------------------------------------------------------------------------------
OK
1515
Time taken: 14.08 seconds, Fetched: 1 row(s)
hive>
```