Queries

Tables present in the hive database 'cab_ride_analysis' are **aggBookings**, **bookings** and **clickstreamdata**.

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

Query: *select customer_id as Customer, count(driver_id) NoOfDrivers from bookings group by customer_id;*

```
hive> select customer_id as Customer, count(driver_id) NoOfDrivers from bookings group by customer_id;
Query ID = ec2-user_20211030194848_a205d5e3-0b3f-4534-9dd3-19d6b88cef8e
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1635601130967_0035, Tracking URL = http://ip-10-0-0-52.ec2.internal:8088/proxy/application_1635601130967_0035/
Kill Command = /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/hadoop/bin/hadoop job  -kill job_1635601130967_0035
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-30 19:48:12,005 Stage-1 map = 0%,  reduce = 0%
2021-10-30 19:48:17,186 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.27 sec
2021-10-30 19:48:23,433 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.26 sec
MapReduce Total cumulative CPU time: 5 seconds 260 msec
Ended Job = job_1635601130967_0035
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.26 sec   HDFS Read: 177289 HDFS Write: 11000 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 260 msec
OK
customer        noofdrivers
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
```

© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved

Task6: Calculate the total rides taken by each customer.

Query: *select customer_id as Customer, count(booking_id) as NoOfRides from bookings group by customer_id;*

```
hive> select customer_id as Customer, count(booking_id) as NoOfRides from bookings group by customer_id;
Query ID = ec2-user_20211030195050_a7cebbd3-2e24-4dd8-9910-135889947678
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1635601130967_0036, Tracking URL = http://ip-10-0-0-52.ec2.internal:8088/proxy/application_1635601130967_0036/
Kill Command = /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/hadoop/bin/hadoop job  -kill job_1635601130967_0036
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-30 19:51:03,119 Stage-1 map = 0%,  reduce = 0%
2021-10-30 19:51:09,310 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.42 sec
2021-10-30 19:51:16,531 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.63 sec
MapReduce Total cumulative CPU time: 5 seconds 630 msec
Ended Job = job_1635601130967_0036
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.63 sec   HDFS Read: 177266 HDFS Write: 11000 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 630 msec
OK
customer        noofrides
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
```

© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved

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.

Query 1: *select count(customer_id) as NoOfButtonClicks from clickstreamdata where button_id = 'fcba68aa-1231-11eb-adc1-0242ac120002' and is_button_click='Yes';*

```
hive> select count(customer_id) as NoOfButtonClicks from clickstreamdata where button_id = 'fcba68aa-1231-11eb-adc1-0242ac120002' and is_button_click='Yes';
Query ID = ec2-user_20211030204848_bfa77316-4289-44d0-a33e-9e5cfbb36374
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1635601130967_0050, Tracking URL = http://ip-10-0-0-52.ec2.internal:8088/proxy/application_1635601130967_0050/
Kill Command = /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/hadoop/bin/hadoop job  -kill job_1635601130967_0050
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-30 20:48:16,708 Stage-1 map = 0%,  reduce = 0%
2021-10-30 20:48:23,172 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.87 sec
2021-10-30 20:48:30,510 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.43 sec
MapReduce Total cumulative CPU time: 6 seconds 430 msec
Ended Job = job_1635601130967_0050
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.43 sec   HDFS Read: 409134 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 430 msec
OK
496
Time taken: 29.988 seconds, Fetched: 1 row(s)
```

© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved

Task 7. Cont.

Query 2: *select count(customer_id) as NoOfPageViews from clickstreamdata where page_id = 'e7bc5fb2-1231-11eb-adc1-0242ac120002' and is_page_view='Yes';*

```
hive> select count(customer_id) as NoOfPageViews from clickstreamdata where page_id = 'e7bc5fb2-1231-11eb-adc1-0242ac120002' and is_page_view='Yes';
Query ID = ec2-user_20211030204949_f3b85411-505c-40d2-8152-ecb61e49fc96
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1635601130967_0051, Tracking URL = http://ip-10-0-0-52.ec2.internal:8088/proxy/application_1635601130967_0051/
Kill Command = /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/hadoop/bin/hadoop job  -kill job_1635601130967_0051
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-30 20:50:00,885 Stage-1 map = 0%,  reduce = 0%
2021-10-30 20:50:07,333 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.46 sec
2021-10-30 20:50:13,613 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.22 sec
MapReduce Total cumulative CPU time: 6 seconds 220 msec
Ended Job = job_1635601130967_0051
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.22 sec   HDFS Read: 409194 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 220 msec
OK
515
Time taken: 24.43 seconds, Fetched: 1 row(s)
```

Conversion Ratio: Total 'Book Now' Button Press/Total Visits made by customer on the booking page.

496/515 = **0.9631**

The Conversion Ratio of people booking a ride is 96.31%.

© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved

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

Query: *select count(booking_id) BlackCarRides from bookings where cab_color = 'black';*

```
hive> select count(booking_id) BlackCarRides from bookings where cab_color='black';
Query ID = ec2-user_20211030195353_fdb5ebbe-46c6-4015-b2a3-5e5523803ef0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1635601130967_0037, Tracking URL = http://ip-10-0-0-52.ec2.internal:8088/proxy/application_1635601130967_0037/
Kill Command = /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/hadoop/bin/hadoop job  -kill job_1635601130967_0037
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-30 19:53:57,623 Stage-1 map = 0%,  reduce = 0%
2021-10-30 19:54:03,978 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.22 sec
2021-10-30 19:54:11,242 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.59 sec
MapReduce Total cumulative CPU time: 6 seconds 590 msec
Ended Job = job_1635601130967_0037
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.59 sec   HDFS Read: 177962 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 590 msec
OK
blackcarrides
72
Time taken: 27.151 seconds, Fetched: 1 row(s)
```

© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved

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

Query: *select to_date(pickup_timestamp) as Date, sum(tip_amount) TotalTip from bookings group by to_date(pickup_timestamp);*

```
hive> select to_date(pickup_timestamp) as Date, sum(tip_amount) TotalTip from bookings group by to_date(pickup_timestamp);
Query ID = ec2-user_20211030200202_40e2578f-e045-483e-9157-b04400a1dcd0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1635601130967_0040, Tracking URL = http://ip-10-0-0-52.ec2.internal:8088/proxy/application_1635601130967_0040/
Kill Command = /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/hadoop/bin/hadoop job  -kill job_1635601130967_0040
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-30 20:02:17,233 Stage-1 map = 0%,  reduce = 0%
2021-10-30 20:02:22,526 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.81 sec
2021-10-30 20:02:28,872 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.51 sec
MapReduce Total cumulative CPU time: 5 seconds 510 msec
Ended Job = job_1635601130967_0040
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.51 sec   HDFS Read: 177460 HDFS Write: 4257 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 510 msec
OK
date    totaltip
2020-01-01      59
2020-01-02      95
2020-01-03      11
2020-01-04      123
2020-01-05      134
2020-01-06      189
2020-01-07      148
2020-01-08      111
2020-01-09      48
2020-01-10      77
2020-01-11      81
2020-01-12      109
2020-01-14      142
2020-01-15      338
2020-01-16      155
2020-01-17      296
2020-01-18      240
2020-01-20      210
```

© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved

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') as Month, count(booking_id) as TotalTripsLT2 from bookings where rating_by_customer < 2 group by date_format(pickup_timestamp, 'yyyy-MM');*

```
hive> select date_format(pickup_timestamp, 'yyyy-MM') as Month, count(booking_id) as TotalTripsLT2 from bookings where rating_by_customer<2 group by date_format(pickup_timestamp, 'yyyy-MM')
;
Query ID = ec2-user_20211030200707_8f15765b-c934-45f6-9501-b8038bec3cc5
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1635601130967_0042, Tracking URL = http://ip-10-0-0-52.ec2.internal:8088/proxy/application_1635601130967_0042/
Kill Command = /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/hadoop/bin/hadoop job  -kill job_1635601130967_0042
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-30 20:07:34,864 Stage-1 map = 0%,  reduce = 0%
2021-10-30 20:07:42,298 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.52 sec
2021-10-30 20:07:47,503 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.26 sec
MapReduce Total cumulative CPU time: 6 seconds 260 msec
Ended Job = job_1635601130967_0042
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.26 sec   HDFS Read: 178397 HDFS Write: 110 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 260 msec
OK
month   totaltripslt2
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
```

© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved

Task 11: Calculate the count of total iOS users.

Query: *select count(distinct(customer_id)) as iOS_User_Base from clickstreamdata where os_version = 'iOS';*

```
hive> select count(distinct(customer_id)) as iOS_User_Base from clickstreamdata where os_version='iOS';
Query ID = ec2-user_20211030194444_620b7b4b-f5f1-40ad-bd8b-7e624bb55643
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1635601130967_0034, Tracking URL = http://ip-10-0-0-52.ec2.internal:8088/proxy/application_1635601130967_0034/
Kill Command = /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/lib/hadoop/bin/hadoop job  -kill job_1635601130967_0034
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-30 19:45:05,872 Stage-1 map = 0%,  reduce = 0%
2021-10-30 19:45:12,121 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.01 sec
2021-10-30 19:45:20,383 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.43 sec
MapReduce Total cumulative CPU time: 7 seconds 430 msec
Ended Job = job_1635601130967_0034
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.43 sec   HDFS Read: 409021 HDFS Write: 5 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 430 msec
OK
ios_user_base
1503
Time taken: 25.074 seconds, Fetched: 1 row(s)
```

© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved