

Create Hive-Managed Tables

Command to create the Hive tables

Give the command “hive” to launch the Hive CLI.

- Command used to create the database.

Command: create database if not exists yourowncabs ;

```
[hadoop@ip-172-31-35-159 ~]$ hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: true
hive> set hive.cli.print.header=true ;
hive> create database if not exists yourowncabs ;
OK
Time taken: 0.605 seconds
hive> describe database yourowncabs ;
OK
db_name comment location owner_name owner_type parameters
yourowncabs hdfs://ip-172-31-35-159.ec2.internal:8020/user/hive/warehouse/yourowncabs.db hadoop USER
Time taken: 0.257 seconds, Fetched: 1 row(s)
hive> show databases ;
OK
database_name
default
yourowncabs
Time taken: 0.026 seconds, Fetched: 2 row(s)
hive>
```

- Command used to create Hive-managed table for clickstream data.

Command:

create table if not exists yourowncabs.clickstream_data (customer_id int, app_version string, os_version string, lat float, lon float, page_id string, button_id string, is_button_click string, is_page_view string, is_scroll_up string, is_scroll_down string, `timestamp` timestamp) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile ;

```
hive> create table if not exists yourowncabs.clickstream_data ( customer_id int, app_version string, os_version string, lat float, lon float, page_id string, button_id string, is_button_click string, is_page_view string, is_scroll_up string, is_scroll_down string, `timestamp` timestamp ) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile ;
OK
Time taken: 0.414 seconds
hive> describe yourowncabs.clickstream_data ;
OK
col_name data_type comment
customer_id int
app_version string
os_version string
lat float
lon float
page_id string
button_id string
is_button_click string
is_page_view string
is_scroll_up string
is_scroll_down string
timestamp timestamp
Time taken: 0.063 seconds, Fetched: 12 row(s)
hive>
```

- Command used to create hive managed table for bookings data.

Command:

create table if not exists yourowncabs.bookings_data (booking_id string, customer_id int, driver_id int, customer_app_version string, customer_phone_os_version string, pickup_lat float, pickup_lon float, drop_lat float, drop_lon float, pickup_timestamp timestamp, drop_timestamp timestamp, trip_fare int, tip_amount int, currency_code string, cab_color string, cab_registration_no string, customer_rating_by_driver int, rating_by_customer int, passenger_count int) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile ;

```
hive> create table if not exists yourowncabs.bookings_data ( booking_id string, customer_id int, driver_id int, customer_app_version string, customer_phone_os_version string, pickup_lat float, pickup_lon float, drop_lat float, drop_lon float, pickup_timestamp timestamp, drop_timestamp timestamp, trip_fare int, tip_amount int, currency_code string, cab_color string, cab_registration_no string, customer_rating_by_driver int, rating_by_customer int, passenger_count int ) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile ;
OK
Time taken: 0.094 seconds
hive> describe yourowncabs.bookings_data ;
OK
col_name      data_type      comment
booking_id    string
customer_id   int
driver_id     int
customer_app_version string
customer_phone_os_version string
pickup_lat    float
pickup_lon    float
drop_lat      float
drop_lon      float
pickup_timestamp timestamp
drop_timestamp timestamp
trip_fare     int
tip_amount    int
currency_code string
cab_color     string
cab_registration_no string
customer_rating_by_driver int
rating_by_customer int
passenger_count int
Time taken: 0.035 seconds, Fetched: 19 row(s)
hive>
```

- Command used to create hive managed table for date-wise total bookings.

Command:

create table if not exists yourowncabs.datewise_bookings_aggregates (trip_date date, total_bookings int) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile ;

```
hive> create table if not exists yourowncabs.datewise_bookings_aggregates ( trip_date date, total_bookings int ) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile ;
OK
Time taken: 0.069 seconds
hive> describe yourowncabs.datewise_bookings_aggregates ;
OK
col_name      data_type      comment
trip_date     date
total_bookings int
Time taken: 0.05 seconds, Fetched: 2 row(s)
hive>
```

- Below is the screenshot of the tables created.

```
hive> use yourowncabs ;
OK
Time taken: 0.027 seconds
hive> show tables ;
OK
tab_name
bookings_data
clickstream_data
datewise_bookings_aggregates
Time taken: 0.031 seconds, Fetched: 3 row(s)
hive>
```

Command to load the data into Hive tables

- Command used to load the data in **clickstream_data** table.

Command:

load data inpath '/user/hadoop/structuredclickStreamData/*.csv' into table
yourowncabs.clickstream_data ;

```
hive> load data inpath '/user/hadoop/structuredclickStreamData/*.csv' into table yourowncabs.clickstream_data ;
Loading data to table yourowncabs.clickstream_data
OK
Time taken: 0.631 seconds
hive> select * from yourowncabs.clickstream_data limit 5 ;
OK
clickstream_data.customer_id clickstream_data.app_version clickstream_data.os_version clickstream_data.lat clickstream_data.lon clickstream_data.page_id
clickstream_data.button_id clickstream_data.is_Button_click clickstream_data.is_page_view clickstream_data.is_scroll_up clickstream_data.is_scroll_down
clickstream_data.timestamp
26564820 3.2.35 Android 16.445486 99.90206 de545711-3914-4450-8c11-b17b8dabb5e1 fcba68aa-1231-11eb-adc1-0242ac120002 No Yes No Y
es 2020-09-14 08:59:07
31906387 2.4.7 iOS -64.81375 -133.52704 de545711-3914-4450-8c11-b17b8dabb5e1 a95dd57b-779f-49db-819d-b6960483e554 No No Yes Y
es 2020-05-16 16:30:21
25713677 3.4.12 Android 89.943436 127.313416 b328829e-17ae-11eb-adc1-0242ac120002 fcba68aa-1231-11eb-adc1-0242ac120002 No No Yes N
o 2020-02-09 00:52:13
83474293 3.1.8 Android -69.93907 -36.45167 e7bc5fb2-1231-11eb-adc1-0242ac120002 e1e99492-17ae-11eb-adc1-0242ac120002 Yes No Yes N
o 2020-06-17 10:42:50
63727807 2.2.9 iOS 64.08211 -81.822075 e7bc5fb2-1231-11eb-adc1-0242ac120002 fcba68aa-1231-11eb-adc1-0242ac120002 No Yes Yes Y
es 2020-07-06 02:51:53
Time taken: 1.725 seconds, Fetched: 5 row(s)
hive>
```

- Command used to load the data in **bookings_data** table.

Command:

load data inpath '/user/hadoop/bookings_data/part-m-00000' into table
yourowncabs.bookings_data ;

```
hive> load data inpath '/user/hadoop/bookings_data/part-m-00000' into table yourowncabs.bookings_data ;
Loading data to table yourowncabs.bookings_data
OK
Time taken: 0.809 seconds
hive> select * from yourowncabs.bookings_data limit 5 ;
OK
bookings_data.booking_id bookings_data.customer_id bookings_data.driver_id bookings_data.customer_app_version bookings_data.customer_phone os_versionb
bookings_data.pickup_lat bookings_data.pickup_lon bookings_data.drop_lat bookings_data.drop_lon bookings_data.pickup_timestamp bookings_data.drop_timestamp b
bookings_data.trip_fare bookings_data.tip_amount bookings_data.currency_code bookings_data.cab_color bookings_data.cab_registration_no bookings_data.cu
stomer_Rating_by_driver bookings_data.rating_by_customer bookings_data.passenger_count
BK8968087150 51811359 15055660 2.2.14 Android -49.431955 103.917854 -58.804386 146.47737 2020-06-23 19:33:10 2020-06-06 09:02
110 534 83 INR black 054-38-4479 4 3 3
BK629851904 31663218 60872180 3.4.1 iOS -83.54084 175.80086 86.20705 128.36723 2020-05-23 12:22:04 2020-08-09 19:02
156 126 67 INR lime 796-39-6801 3 2 4
BK1797410350 86869399 94276051 4.1.36 iOS -67.89307 55.234127 -51.1079 -31.07475 2020-05-19 14:14:32 2020-08-23 18:38
139 297 63 INR olive 748-73-1579 1 3 3
BK5788246325 58230837 45457227 2.4.27 Android 13.707887 113.49995 54.38129 -18.437752 2020-03-24 01:30:15 2020-05-19 11:16
145 932 32 INR white 558-80-6346 3 2 2
BK8342703255 84232510 86494681 4.1.34 Android -6.091461 -114.64979 22.84495 70.137825 2020-08-03 19:10:52 2020-03-24 08:25
140 260 7 INR blue 068-72-1637 3 3 3
Time taken: 0.187 seconds, Fetched: 5 row(s)
hive>
```

- Command used to load the data in **datewise_bookings_aggregates** table.

Command:

load data inpath '/user/hadoop/date-wiseBookingsAggregatesData/*.csv' into table
yourowncabs.datewise_bookings_aggregates ;

```
hive> load data inpath '/user/hadoop/date-wiseBookingsAggregatesData/*.csv' into table yourowncabs.datewise_bookings_aggregates ;
Loading data to table yourowncabs.datewise_bookings_aggregates
OK
Time taken: 0.459 seconds
hive> select * from yourowncabs.datewise_bookings_aggregates limit 5 ;
OK
datewise_bookings_aggregates.trip_date datewise_bookings_aggregates.total_bookings
2020-03-07 2
2020-08-22 6
2020-07-05 6
2020-04-19 4
2020-08-04 7
Time taken: 0.126 seconds, Fetched: 5 row(s)
hive>
```

