

## Create Hive-Managed Tables

### Command to create the Hive tables;

```
create database if not exists cab_booking ;  
use cab_booking ;
```

#### 1) Managed table for clickstream\_data

```
create table if not exists clickstream_data  
(customer_id string ,app_version string, os_version string,lat string ,lon string ,page_id  
string,button_id string , is_button_click varchar(3) ,is_page_view varchar(3) ,is_scroll_up  
varchar(3) ,is_scroll_down varchar(3), click_timestamp string )  
row format delimited fields terminated by ","  
location '/home/hadoop/clickstream_data_flatten/';
```

#### 2) Managed table for bookings data

```
create table if not exists booking_data  
(booking_id string ,customer_id string ,driver_id string , customer_app_version string,  
customer_phone_os_version string , pickup_lat double , pickup_lon double, drop_lat double,  
drop_lon double, pickup_timestamp string , drop_timestamp string ,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 ","  
location '/home/hadoop/booking_data_csv/';
```

#### 3) Managed table for aggregated data

```
create table if not exists datewise_data  
(booking_date string , count int)  
row format delimited fields terminated by ","  
location '/home/hadoop/datewise_aggreation/';
```

### Command to load the data into Hive tables

The data is loaded in hive tables using the '**location**' command used during table creation.

## Screenshot of hive table creation and loaded data:

```
hadoop@ip-172-31-33-247:~
[hadoop@ip-172-31-33-247 ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> create database if not exists cab_booking ;
OK
Time taken: 0.91 seconds
hive> use cab_booking ;
OK
Time taken: 0.073 seconds
hive> create table if not exists clickstream_data
> (customer_id string ,app_version string, os_version string,lat string ,lon string ,page_id
> string,button_id string , is_button_click varchar(3) ,is_page_view varchar(3) ,is_scroll_up
> varchar(3) ,is_scroll_down varchar(3), click_timestamp string )
> row format delimited fields terminated by ","
> location '/home/hadoop/clickstream_data_flatten/';
OK
Time taken: 0.422 seconds
hive> create table if not exists booking_data
> (booking_id string ,customer_id string ,driver_id string , customer_app_version string,
> customer_phone_os version string , pickup_lat double , pickup_lon double, drop_lat double,
> drop_lon double, pickup_timestamp string , drop_timestamp string ,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 ","
> location '/home/hadoop/booking_data_csv/';
OK
Time taken: 0.07 seconds
hive> create table if not exists datewise_data
> (booking_date string , count int)
> row format delimited fields terminated by ","
> location '/home/hadoop/datewise_aggregation/';
OK
Time taken: 0.05 seconds
hive>
```

```
hive> use cab_booking;
OK
Time taken: 0.078 seconds
hive> select * from clickstream_data limit 5;
OK
26564820      3.2.35  Android 16.4454865      99.902065      de545711-3914-4450-8c11-b17b8dabb5e1      fcb68aa-1231-11eb-adc1-0242ac120002      No      Yes      No      Yes      "2020-09-14 0
9:59:07"
31906387      2.4.7   iOS      -64.813749      -133.527040      de545711-3914-4450-8c11-b17b8dabb5e1      a95dd57b-779f-49db-819d-b6960483e554      No      No      Yes      Yes      "2020-05-16 1
6:30:21"
25713677      3.4.12  Android 89.943435      127.313415      b328829e-17ae-11eb-adc1-0242ac120002      fcb68aa-1231-11eb-adc1-0242ac120002      No      No      Yes      No      "2020-02-09 0
0:52:13"
83474293      3.1.8   Android -69.939070      -36.451670      e7bc5fb2-1231-11eb-adc1-0242ac120002      ele99492-17ae-11eb-adc1-0242ac120002      Yes      No      Yes      No      "2020-06-17 1
0:42:50"
63727807      2.2.9   iOS      64.082108      -81.822078      e7bc5fb2-1231-11eb-adc1-0242ac120002      fcb68aa-1231-11eb-adc1-0242ac120002      No      Yes      Yes      Yes      "2020-07-06 0
2:51:53"
Time taken: 2.124 seconds, Fetched: 5 row(s)
hive> select * from booking_data limit 5;
OK
BK8968087150      51811359      15055660      2.2.14  Android -49.4319655      103.917851      -58.8043875      146.477367      2020-06-23T19:33:10.000Z      2020-06-06T09:02:10.000Z      5
34      83      INR      black      054-38-4479      4      3      3
BK629851904      31663218      60872180      3.4.1   iOS      -83.5408405      175.80085      86.20705      128.367238      2020-05-23T12:22:04.000Z      2020-08-09T19:02:56.000Z      1
26      67      INR      lime      796-39-6801      3      2      4
BK1797410350      86869399      94276051      4.1.36  iOS      -67.8930645      55.234128      -51.1079      -31.07475      2020-05-19T14:14:32.000Z      2020-08-23T18:38:39.000Z      2
97      63      INR      olive      748-73-1579      1      3      3
BK5788246325      58230837      45457227      2.4.27  Android 13.707887      113.499943      54.3812915      -18.437751      2020-03-24T01:30:15.000Z      2020-05-19T11:16:45.000Z      9
32      32      INR      white      558-80-6346      3      2      2
BK8342703255      84232510      86494681      4.1.34  Android -6.091461      -114.649789      22.8449505      70.137827      2020-08-03T19:10:52.000Z      2020-03-24T08:25:40.000Z      2
60      7      INR      blue      068-72-1637      3      3      3
Time taken: 0.16 seconds, Fetched: 5 row(s)
hive> select * from datewise_data limit 5;
OK
2020-06-20      1
2020-04-20      3
2020-05-14      3
2020-04-22      2
2020-03-16      2
Time taken: 0.138 seconds, Fetched: 5 row(s)
hive>
```