

Create Hive-Managed Tables

<Command to create the Hive tables>

Managed table for clickstream_data

1. create database if not exists cab_bookings ;
2. use cab_bookings ;
3. create table if not exists clickstream_data (customer_id int ,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)) row format delimited fields terminated by "," ;

Managed table for bookings data

4. create table if not exists booking_data (booking_id string ,customer_id int ,driver_id int , customer_app_version string, customer_phone_os_version string , pickup_lat double , pickup_lon double, drop_lat double, drop_lon double, 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 "," ;

Managed table for aggregated

5. create table if not exists datewise_data (date string , count int) row format delimited fields terminated by "," ;

<Command to load the data into Hive tables>

6. load data inpath '/user/root/user/root/clickstream_data_flatten/part-00000-8371bd57-dd00-4a55-8933-05ad5732a984-c000.csv' into table clickstream_data ;
7. load data inpath '/user/root/bookings_data_csv/part-00000-a38e960c-5cd8-4c9c-a134-62b4bbb3a6ac-c000.csv' into table booking_data ;
8. load data inpath '/user/root/datewise_aggregation/part-00000-90c0a9b6-347a-4288-bace-d15669abb937-c000.csv' into table datewise_data ;

<screenshot from HIVE>

```
hive> create database if not exists cab_bookings ;
OK
Time taken: 0.058 seconds
hive> show databases;
OK
cab_bookings
default
Time taken: 0.012 seconds, Fetched: 2 row(s)
hive>
```

```
hive> show tables;
OK
booking_data
clickstream_data
datewise_data
Time taken: 0.018 seconds, Fetched: 3 row(s)
hive>
```

```
hive> load data inpath '/user/root/user/root/clickstream_data_flatten/part-00000-8371bd57-dd00-4a55-8933-05ad5732a984-c000.csv' into table clickstream_data ;
Loading data to table cab_bookings.clickstream_data
Table cab_bookings.clickstream_data stats: [numFiles=1, totalSize=397733]
OK
Time taken: 0.469 seconds
```

```
hive> load data inpath '/user/root/bookings_data_csv/part-00000-a38e960c-5cd8-4c9c-a134-62b4bbb3a6ac-c000.csv' into table booking_data ;
Loading data to table cab_bookings.booking_data
Table cab_bookings.booking_data stats: [numFiles=1, totalSize=182961]
OK
Time taken: 0.262 seconds
hive> load data inpath '/user/root/datewise_aggregation/part-00000-90c0a9b6-347a-4288-bace-d15669abb937-c000.csv' into table datewise_data ;
Loading data to table cab_bookings.datewise_data
Table cab_bookings.datewise_data stats: [numFiles=1, totalSize=3769]
OK
Time taken: 0.28 seconds
```