

## Create Hive-Managed Tables

### Command to create the Hive tables

#### 1. Clickstream table :

```
create table if not exists clickstream (  
    customer_id int,  
    app_version string,  
    os_version string,  
    lat double,  
    lon double,  
    page_id varchar(100),  
    button_id varchar(100),  
    is_button_click string,  
    is_page_view string,  
    is_scroll_up string,  
    is_scroll_down string,  
    `timestamp` timestamp )  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ','  
stored as textfile;
```

#### 2. Booking table:

```
create table if not exists booking (  
    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 int,  
    customer_rating_by_driver varchar(100),  
    rating_by_customer int,    passenger_count int)
```

```
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
stored as textfile;
```

### 3. Aggregated datewise table:

```
create table if not exists aggregate_datewise(
    pickup_date date,
    booking_id_count int
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
stored as textfile;
```

## Commands to load the data into Hive tables

### 1. Loading data into clickstream table

load data local inpath '/home/hadoop/clickstream/clickstream\_flattened.csv' into table clickstream;

### 2. Loading data into booking table

load data local inpath '/home/hadoop/cab\_rides/part-m-00000' into table booking;

### 3. Loading data into aggregate\_datewise table

load data local inpath '/home/hadoop/datewise\_bookings\_agg/datewise\_agg' into table aggregate\_datewise;

## Screenshots of created tables and data loading

```
hive> create table if not exists clickstream(
>     customer_id int,
>     app_version string,
>     os_version string,
>     lat double,
>     lon double,
>     page_id varchar(100),
>     button_id varchar(100),
>     is_button_click string,
>     is_page_view string,
>     is_scroll_up string,
>     is_scroll_down string,
>     `timestamp` timestamp)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> stored as textfile;
OK
Time taken: 0.547 seconds
```

```
Time taken: 0.147 seconds
hive> create table if not exists booking (
>     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 int,
>     customer_rating_by_driver varchar(100),
>     rating_by_customer int,
>     passenger_count int)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> stored as textfile;
OK
Time taken: 0.105 seconds
```

```
hive> load data local inpath '/home/hadoop/cab_rides/part-m-00000' into table booking;
Loading data to table capstone.booking
OK
Time taken: 1.566 seconds
hive> select * from capstone.booking;
OK
```

```
Time taken: 1.857 seconds, Fetched: 1000 row(s)
hive> load data local inpath '/home/hadoop/clickstream/clickstream_flattened.csv' into table clickstream;
Loading data to table capstone.clickstream
OK
Time taken: 0.442 seconds
```

```
FAILED: SemanticException Line 1:23 Invalid path '/home/hadoop/datewise_bookings_agg/datewise_agg'. No files mat
hive> load data local inpath '/home/hadoop/datewise_bookings_agg/datewise_agg.csv' into table aggregate_datewise;
Loading data to table capstone.aggregate_datewise
OK
Time taken: 0.426 seconds
```