

Create Hive-Managed Tables

<Command to create the Hive tables>

Create custom DB

```
hive> create database cabrides_db;
OK
Time taken: 0.398 seconds
hive> show databases;
OK
cabrides_db
default
Time taken: 0.036 seconds, Fetched: 2 row(s)
```

<Command to load the data into Hive tables>

Create 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;
```

```
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: 0.711 seconds
```

Load clickstream data to this table

```
hive> load data inpath "/user/root/clickstream_flattened/part-00000-aa5dcf04-898
a-4189-b284-5f57b3789b9b-c000.csv" into table Clickstream;
Loading data to table cabrides_db.clickstream
OK
Time taken: 1.232 seconds
hive>
```

Create 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;
```

```
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: 0.094 seconds
```

Load booking data to this table

```
hive> load data inpath "/user/root/cab_rides/part-m-00000" into table booking;
Loading data to table cabrides_db.booking
OK
Time taken: 0.581 seconds
```

Create Aggregate_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;
```

```
hive> create table if not exists aggregate_datewise(
>     pickup_date date,
>     booking_id_count int
> )
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> stored as textfile;
OK
Time taken: 0.088 seconds
```

Load aggregate_datewise data to this table

```
hive> load data inpath "/user/root/datewise_bookings_agg" into table aggregate_datewise;
Loading data to table cabrides_db.aggregate_datewise
OK
Time taken: 0.543 seconds
```