Create Hive-Managed Tables

<Command to create the Hive tables>

    1. **Create a Hive-managed table for clickstream data.**

```
create table if not exists clickstream_data (
     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. **Create a Hive-managed table for bookings data.**

```
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 int,
     customer_rating_by_driver varchar(100),
```

```
     rating_by_customer int,
     passenger_count int)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
stored as textfile;
```

3. **Create a Hive-managed table for aggregated data date-wise total bookings.**

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

<Command to load the data into Hive tables>

1. **Clickstream data:**
```
load data local inpath '/hadoop/clickstream/clickstream_flattened.csv' into table
clickstream_data;
```

2. **Booking data:**
```
load data local inpath '/hadoop/booking_data/part-m-00000' into table booking_data;
```

3. **Aggregated data datewise:**
```
load data local inpath '/hadoop/agg_datewise' into table aggregate_datewise;
```