Create Hive-Managed Tables

<Command to create the Hive tables>

1. First create a database

create database if not exists cab_booking_data ;
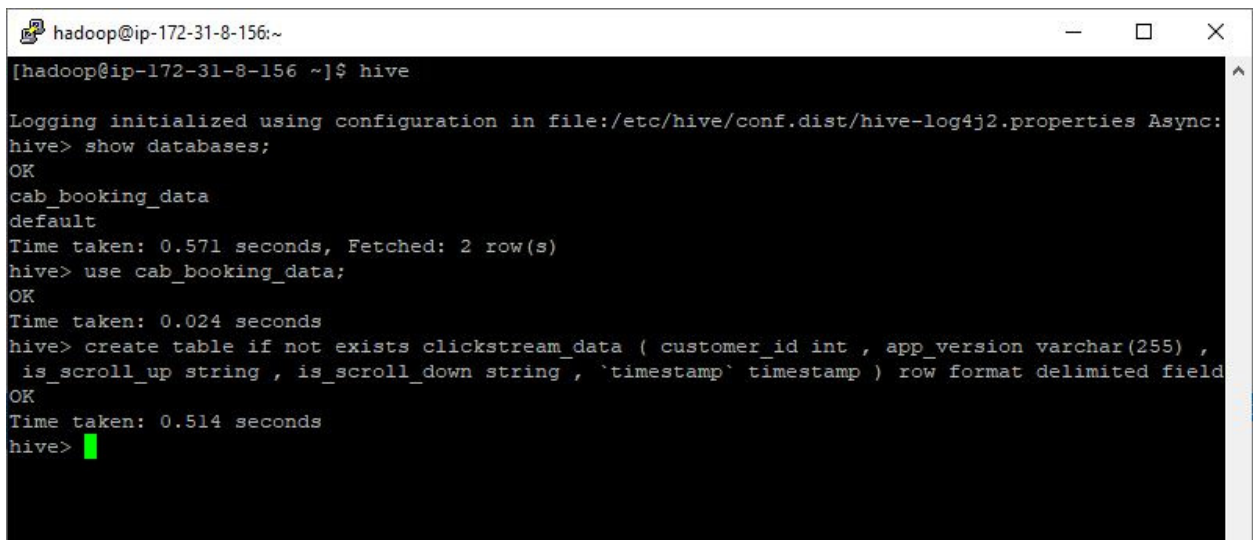use cab_booking_data ;

2. Creating a Hive-managed table for clickstream data

```
create table if not exists clickstream_data (
        customer_id int ,
        app_version varchar(255),
        os_version string,
        lat varchar(255),
        lon varchar(255),
        page_id varchar(255),
        button_id varchar(255),
        is_button_click string,
        is_page_view string,
        is_scroll_up string,
        is_scroll_down string,
        `timestamp` timestamp
)
row format delimited fields terminated by "," ;
```

3. Creating a Hive-managed table for bookings data

```
create table if not exists booking_data (
        booking_id varchar(255),
        customer_id int,
        driver_id int,
        customer_app_version varchar(255),
        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 varchar(255),
        customer_rating_by_driver int,
        rating_by_customer int,
        passenger_count int,
)
row format delimited fields terminated by "," ;
```

```
hive> create table if not exists booking_data ( booking_id varchar(255), customer_id int, driver_
id int, customer_app_version varchar(255), customer_phone_os_version string, pickup_lat double, p
ickup_lon double, drop_lat double, drop_lon double, pickup_timestamp timestamp, drop_timestamp ti
mestamp, trip_fare int, tip_amount int, currency_code string, cab_color string, cab_registration_
no varchar(255), customer_rating_by_driver int, rating_by_customer int, passenger_count int )
row format delimited fields terminated by "," ;
OK
Time taken: 0.106 seconds
hive>
```

4. Creating a Hive-managed table for aggregated data in Task 3

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

```
hive> create table if not exists datewise_aggregated_data ( `date` string, count int ) row format
 delimited fields terminated by "," ;
OK
Time taken: 0.096 seconds
hive>
```

<Command to load the data into Hive tables>

1. load data inpath 'clickstream_data_flatten/ part-00000-ec5ef800-8491-400f-9149-dc9a9158c46a-c000.csv ' into table clickstream_data ;
2. load data inpath 'booking_data_csv/ part-00000-ee83ab78-57ec-4982-9d50-e2ea2e04d9bf-c000.csv' into table booking_data ;
3. load data inpath 'date_aggregated_data/ part-00000-a085f9bc-ecd9-4ee7-b21b-3329d4fabfd3-c000.csv' into table datewise_aggregated_data ;

```
hive> load data inpath 'clickstream_data_flatten/part-00000-ec5ef800-8491-400f-9149-dc9a9158c46a-
c000.csv' into table clickstream_data ;
Loading data to table cab_booking_data.clickstream_data
OK
Time taken: 0.995 seconds
hive>
```

```
hive> load data inpath 'booking_data_csv/part-00000-ee83ab78-57ec-4982-9d50-e2ea2e04d9bf-c000.csv
' into table booking_data ;
Loading data to table cab_booking_data.booking_data
OK
Time taken: 0.445 seconds
hive>
```

```
hive> load data inpath 'datewise_aggregated_data/part-00000-a085f9bc-ecd9-4ee7-b21b-3329d4fabfd3-
c000.csv' into table datewise_aggregated_data ;
Loading data to table cab_booking_data.datewise_aggregated_data
OK
Time taken: 0.962 seconds
hive>
```

```
hive> select * from clickstream_data limit 5;
OK
customer_id    app_version    OS_version       lat    lon      page_id button_id      is_button_click is_page_view    is_scroll_up   is_scroll_down timestamp
26564820       3.2.35 Android 16.4454865       99.902065      de545711-3914-4450-8c11-b17b8dabb5e1    fcba68aa-1231-11eb-adc1-0242ac120002    No      Yes     No     Yes        "2020-09-14 09:59:07"
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 16:30:21"
25713677       3.4.12 Android 89.943435        127.313415     b328829e-17ae-11eb-adc1-0242ac120002    fcba68aa-1231-11eb-adc1-0242ac120002    No      No      Yes    No         "2020-02-09 00:52:13"
83474293       3.1.8  Android -69.939070       -36.451670     e7bc5fb2-1231-11eb-adc1-0242ac120002    e1e99492-17ae-11eb-adc1-0242ac120002    Yes     No      Yes    No         "2020-06-17 10:42:50"
Time taken: 0.207 seconds, Fetched: 5 row(s)
hive>
```

```
hive> select * from booking_data limit 5;
OK
booking_id    customer_id    driver_id      customer_app_version    customer_phone_os_version       pickup_lat     pickup_lon     drop_lat       drop_lon       pickup_timestamp        drop_timestamp trip_fare       tip_amount c
urrency_code   cab_color      cab_registration_no     customer_rating_by_driver      rating_by_customer     passenger_count pickup_date
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       534     83     INR    black   054-38-4479  4
3      3       2020-06-23
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       126     67     INR    lime    796-39-6801  3
2      4       2020-05-23
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       297     63     INR    olive   748-73-1579  1
3      3       2020-05-19
BK578246325    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       932     32     INR    white   558-80-6346  3
2      2       2020-03-24
Time taken: 0.114 seconds, Fetched: 5 row(s)
hive>
```

```
hive> select * from datewise_aggregated_data limit 5;
OK
2020-03-07        2
2020-08-22        6
2020-07-05        6
2020-04-19        4
2020-08-04        7
Time taken: 0.112 seconds, Fetched: 5 row(s)
hive>
```

```
[hadoop@ip-172-31-8-156 ~]$ hadoop fs -ls /user/hive/warehouse/
Found 1 items
drwxrwxrwt   - hadoop hadoop          0 2022-07-22 14:38 /user/hive/warehouse/cab_booking_data.db
[hadoop@ip-172-31-8-156 ~]$ hadoop fs -ls /user/hive/warehouse/cab_booking_data.db
Found 3 items
drwxrwxrwt   - hadoop hadoop          0 2022-07-22 14:40 /user/hive/warehouse/cab_booking_data.db/booking_data
drwxrwxrwt   - hadoop hadoop          0 2022-07-22 14:39 /user/hive/warehouse/cab_booking_data.db/clickstream_data
drwxrwxrwt   - hadoop hadoop          0 2022-07-22 14:59 /user/hive/warehouse/cab_booking_data.db/datewise_aggregated_data
[hadoop@ip-172-31-8-156 ~]$
```