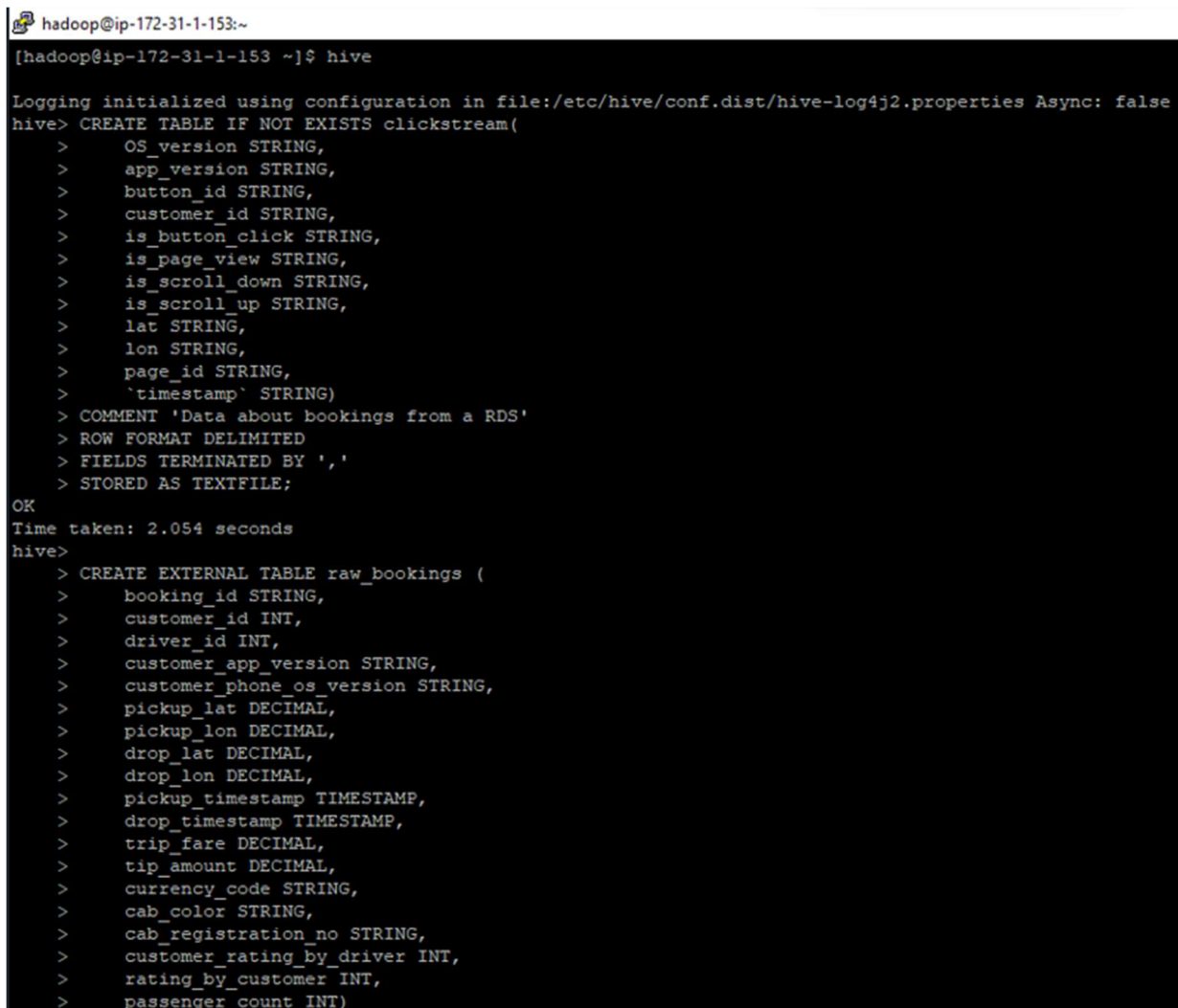


Create Hive database and Hive tables

```
>> hive
>> show tables;
>> select * from bookings limit 20;
>> select * from clickstream limit 20;
>> select * from bookings_per_day limit 20;
```

Screenshots of command process



```
hadoop@ip-172-31-1-153:~
[hadoop@ip-172-31-1-153 ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> CREATE TABLE IF NOT EXISTS clickstream(
  >   OS_version STRING,
  >   app_version STRING,
  >   button_id STRING,
  >   customer_id STRING,
  >   is_button_click STRING,
  >   is_page_view STRING,
  >   is_scroll_down STRING,
  >   is_scroll_up STRING,
  >   lat STRING,
  >   lon STRING,
  >   page_id STRING,
  >   `timestamp` STRING)
  > COMMENT 'Data about bookings from a RDS'
  > ROW FORMAT DELIMITED
  > FIELDS TERMINATED BY ','
  > STORED AS TEXTFILE;

OK
Time taken: 2.054 seconds
hive>
  > CREATE EXTERNAL TABLE raw_bookings (
  >   booking_id STRING,
  >   customer_id INT,
  >   driver_id INT,
  >   customer_app_version STRING,
  >   customer_phone_os_version STRING,
  >   pickup_lat DECIMAL,
  >   pickup_lon DECIMAL,
  >   drop_lat DECIMAL,
  >   drop_lon DECIMAL,
  >   pickup_timestamp TIMESTAMP,
  >   drop_timestamp TIMESTAMP,
  >   trip_fare DECIMAL,
  >   tip_amount DECIMAL,
  >   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 ','
> STORED AS TEXTFILE
> location '/home/hadoop/bookings-data';
```

OK

Time taken: 0.104 seconds

hive>

```
> CREATE TABLE IF NOT EXISTS bookings(
>   booking_id STRING,
>   customer_id INT,
>   driver_id INT,
>   customer_app_version STRING,
>   customer_phone_os_version STRING,
>   pickup_lat DECIMAL,
>   pickup_lon DECIMAL,
>   drop_lat DECIMAL,
>   drop_lon DECIMAL,
>   pickup_timestamp TIMESTAMP,
>   drop_timestamp TIMESTAMP,
>   trip_fare DECIMAL,
>   tip_amount DECIMAL,
>   currency_code STRING,
>   cab_color STRING,
>   cab_registration_no STRING,
>   customer_rating_by_driver INT,
>   rating_by_customer INT,
>   passenger_count INT)
> COMMENT 'Data about bookings from a RDS'
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> STORED AS TEXTFILE;
```

OK

Time taken: 0.124 seconds

hive>

```
> CREATE TABLE IF NOT EXISTS bookings_per_day(
>   pickup_timestamp DATE,
>   bookings INT)
> COMMENT 'Aggregated Data about bookings from a Bookings table'
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> STORED AS TEXTFILE;
```

OK

Time taken: 0.104 seconds

```

hive>
>
> INSERT OVERWRITE TABLE bookings SELECT * FROM raw_bookings;
Query ID = hadoop_20220615140906_aca35735-159d-4641-8c57-37b10f968b61
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1655298923838_0002)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01  [=====>>] 100%  ELAPSED TIME: 5.74 s
-----
Loading data to table default.bookings
OK
Time taken: 11.734 seconds
hive>
> LOAD DATA INPATH './bookings' INTO TABLE bookings_per_day;
Loading data to table default.bookings_per_day
OK
Time taken: 0.995 seconds
hive>
> LOAD DATA INPATH './clickstream' INTO TABLE clickstream;
Loading data to table default.clickstream
OK
Time taken: 0.404 seconds
hive> █

```

All the tables are created and data is loaded successfully!!!!