🔍 Step-by-Step Guide

📂 Step 1: Launch Hive

hive

📄 Step 2: Create Database (Optional)

CREATE DATABASE IF NOT EXISTS retail;
USE retail;

🖊️ Step 3: Create Table

Example: Create a table to store flight details (as in the flight assignment)

```
CREATE TABLE IF NOT EXISTS flight(
    flightdate STRING,
    origin STRING,
    destination STRING,
    airline STRING,
    flightnum INT,
    distance INT
)

ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;
```

```
Step 4: Crate csv using basic terminal:
$ cat <<EOF > flight_data.csv
> 2024-05-01,JFK,LAX,Delta,1001,2475
> 2024-05-01,LAX,ORD,United,2002,1744
> 2024-05-01,ORD,ATL,American,3003,606
> 2024-05-02,SFO,SEA,Alaska,4004,679
> 2024-05-02,DEN,MIA,Southwest,5005,1720
> EOF

$ hdfs dfs -mkdir -p /user/cloudera

$ hdfs dfs -put /home/cloudera/Desktop/flight.csv /user/cloudera/
```

LOAD DATA INPATH '/user/cloudera/flight.csv' INTO TABLE flight;

🔢 Step 5: Run Queries

A. View all records:
SELECT * FROM flight;

B. Count total rows:
SELECT COUNT(*) FROM flight;

C. Count flights by airline:
SELECT airline, COUNT(*) FROM flight GROUP BY airline;

D. Top 5 longest flights:
SELECT flightnum, distance FROM flight ORDER BY distance DESC LIMIT 5;