CT3 Project – Big Data for Machine Learning

Car showroom Project:

Following are detailed step-by-step process for executing the **Car Showroom** project in Oracle VM using Hive, Sqoop, and MySQL.

**Step 1: Prerequisites**

**1.1 Install Required Software**

Ensure that the following are installed and configured in your VM:

1. **MySQL**: For data storage and management.
   * [Instructions for installing MySQL](https://chatgpt.com/c/67446052-bbb0-8011-9e54-e7d6508b50a0" \l "how-to-install-mysql-in-your-vm)
2. **Hadoop**: For the distributed file system (HDFS) required by Hive.
3. **Hive**: For querying datasets.
4. **Sqoop**: To transfer data between MySQL and Hive.

**1.2 Set Environment Variables**

Ensure your environment variables are set correctly in your VM:

1. Add Hadoop, Hive, and Sqoop to your PATH:
2. export HADOOP\_HOME=/path/to/hadoop
3. export HIVE\_HOME=/path/to/hive
4. export SQOOP\_HOME=/path/to/sqoop
5. export PATH=$PATH:$HADOOP\_HOME/bin:$HIVE\_HOME/bin:$SQOOP\_HOME/bin
6. Source your .bashrc file:
7. source ~/.bashrc

**Step 2: Prepare the Dataset**

**2.1 Create a CSV File**

1. Create a dataset file called car\_showroom.csv:
   * Use a text editor in your VM (e.g., nano) or transfer the file from your host machine.
   * Sample content for car\_showroom.csv:
   * CarID,Brand,Model,ShowroomID,Price
   * 1,Toyota,Corolla,101,20000
   * 2,Hyundai,Elantra,102,22000
   * 3,Toyota,Camry,101,30000
   * 4,Ford,Focus,103,18000
   * 5,Hyundai,Sonata,102,25000
2. Save the file in a directory, e.g., /home/cloudera/files/car\_showroom.csv.
3. **Verify the file content**:
4. cat /home/cloudera/files/car\_showroom.csv

**Step 3: Load Data into MySQL**

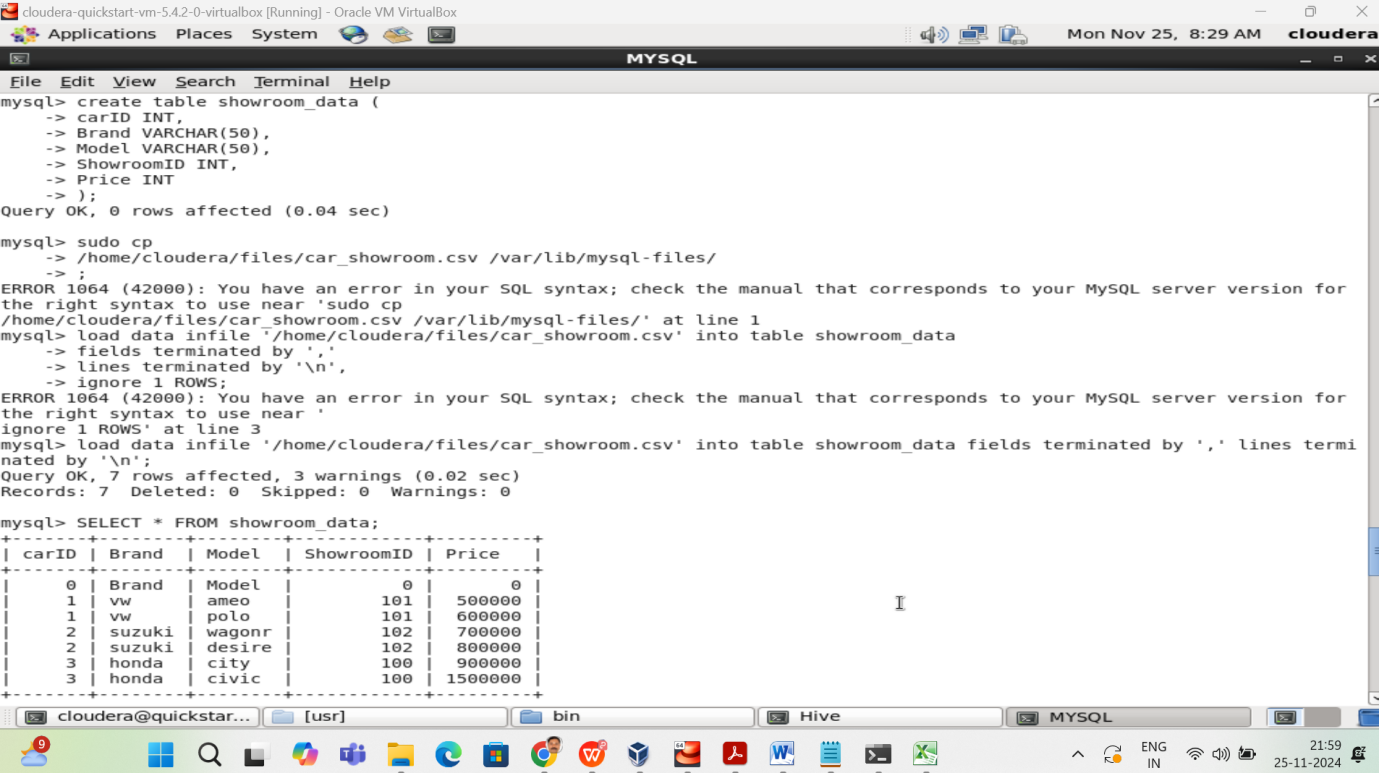
**3.1 Start MySQL**

1. Start the MySQL service:
2. sudo systemctl start mysql
3. Log in to MySQL as root:
4. mysql -u root -p

**3.2 Create a Database and Table**

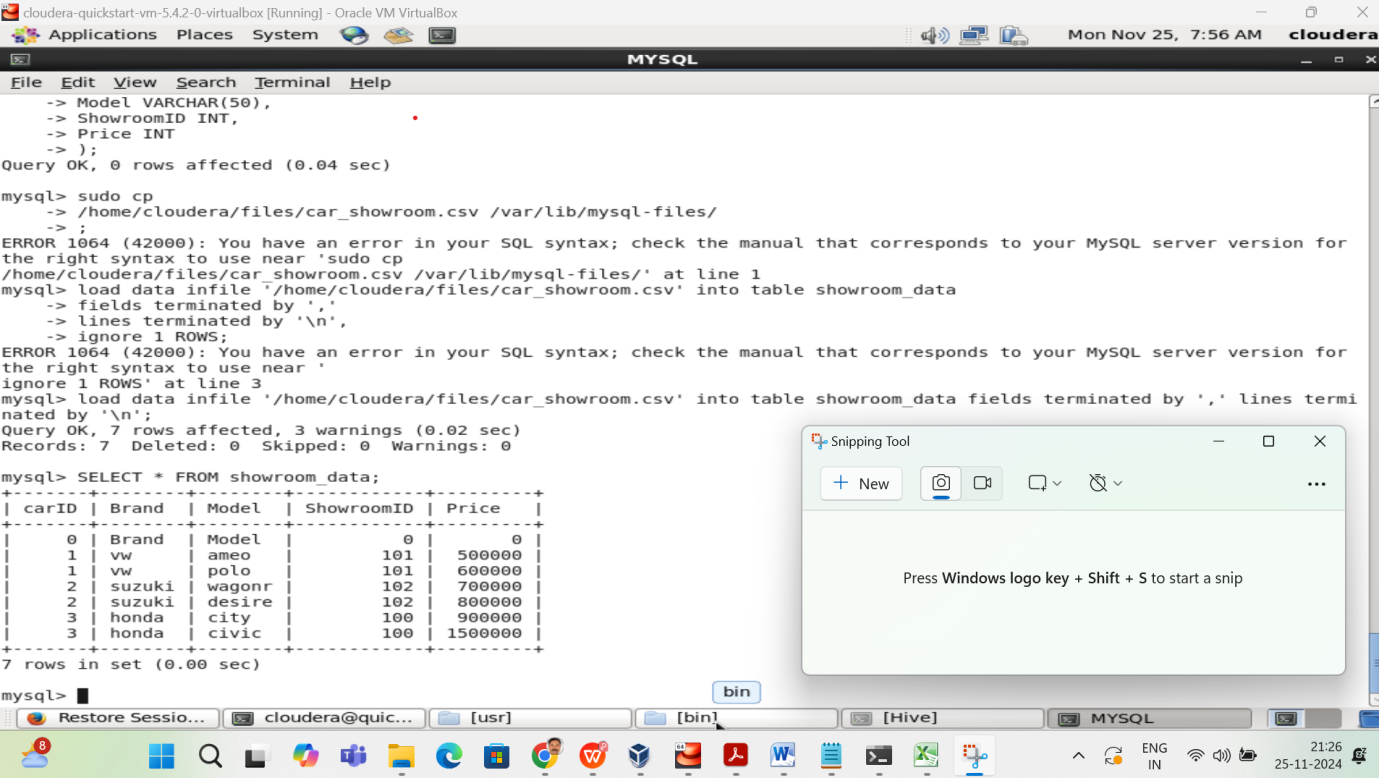
1. Create a database for your project:
2. CREATE DATABASE car\_showroom\_db;
3. Use the new database:
4. USE car\_showroom\_db;
5. Create a table for your data:
6. CREATE TABLE showroom\_data (
7. CarID INT,
8. Brand VARCHAR(50),
9. Model VARCHAR(50),
10. ShowroomID INT,
11. Price INT
12. );





**3.3 Load Data from CSV**

1. Move the CSV file to MySQL's accessible directory:
2. sudo cp /home/cloudera/files/car\_showroom.csv /var/lib/mysql-files/
3. Load the data into MySQL:
4. LOAD DATA INFILE '/var/lib/mysql-files/car\_showroom.csv'
5. INTO TABLE showroom\_data
6. FIELDS TERMINATED BY ','
7. LINES TERMINATED BY '\n'
8. IGNORE 1 ROWS;
9. Verify the data is loaded:
10. SELECT \* FROM showroom\_data;



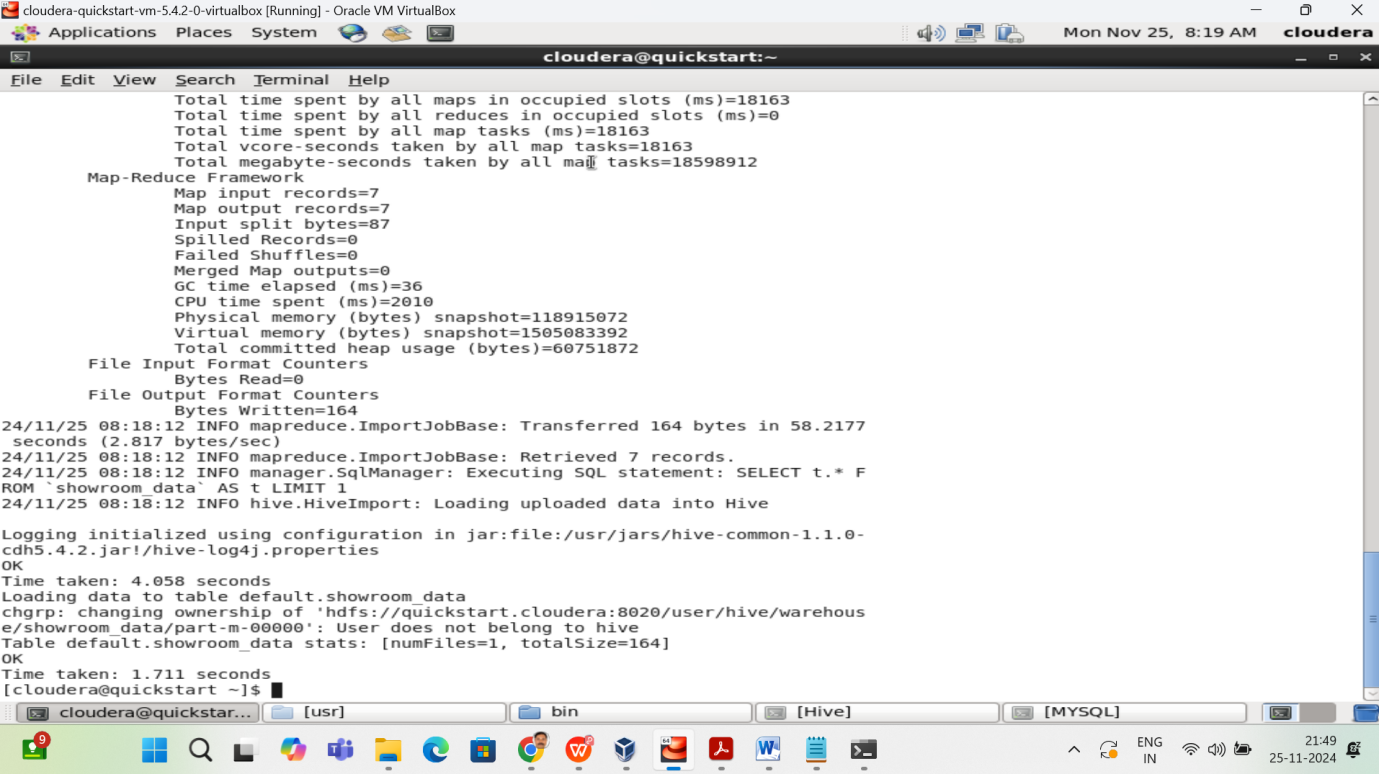
**Step 4: Import Data into Hive Using Sqoop**

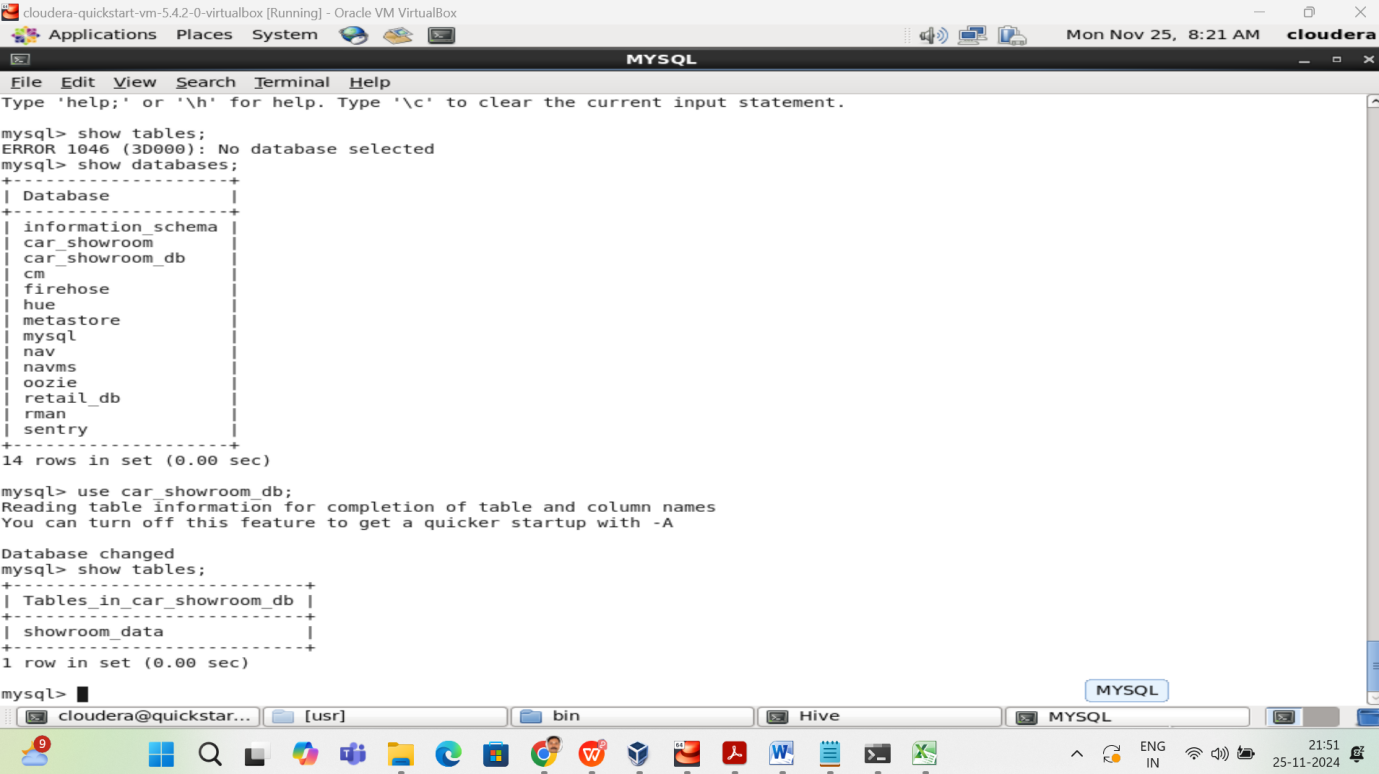
**4.1 Ensure MySQL JDBC Connector is in Sqoop's Library**

1. Verify the MySQL JDBC connector is present:
2. ls /usr/lib/sqoop/lib/mysql-connector-java-\*.jar
3. If not, download and place it in /usr/lib/sqoop/lib/.

**4.2 Run Sqoop Import**

1. Import the data into Hive:
2. sqoop import \
3. --connect jdbc:mysql://localhost:3306/car\_showroom\_db \
4. --username root \
5. --password your\_password \
6. --table showroom\_data \
7. --hive-import \
8. --create-hive-table \
9. --hive-table showroom\_data \
10. --m 1
11. Verify the data is in Hive:
12. hive
13. SHOW TABLES;
14. SELECT \* FROM showroom\_data;



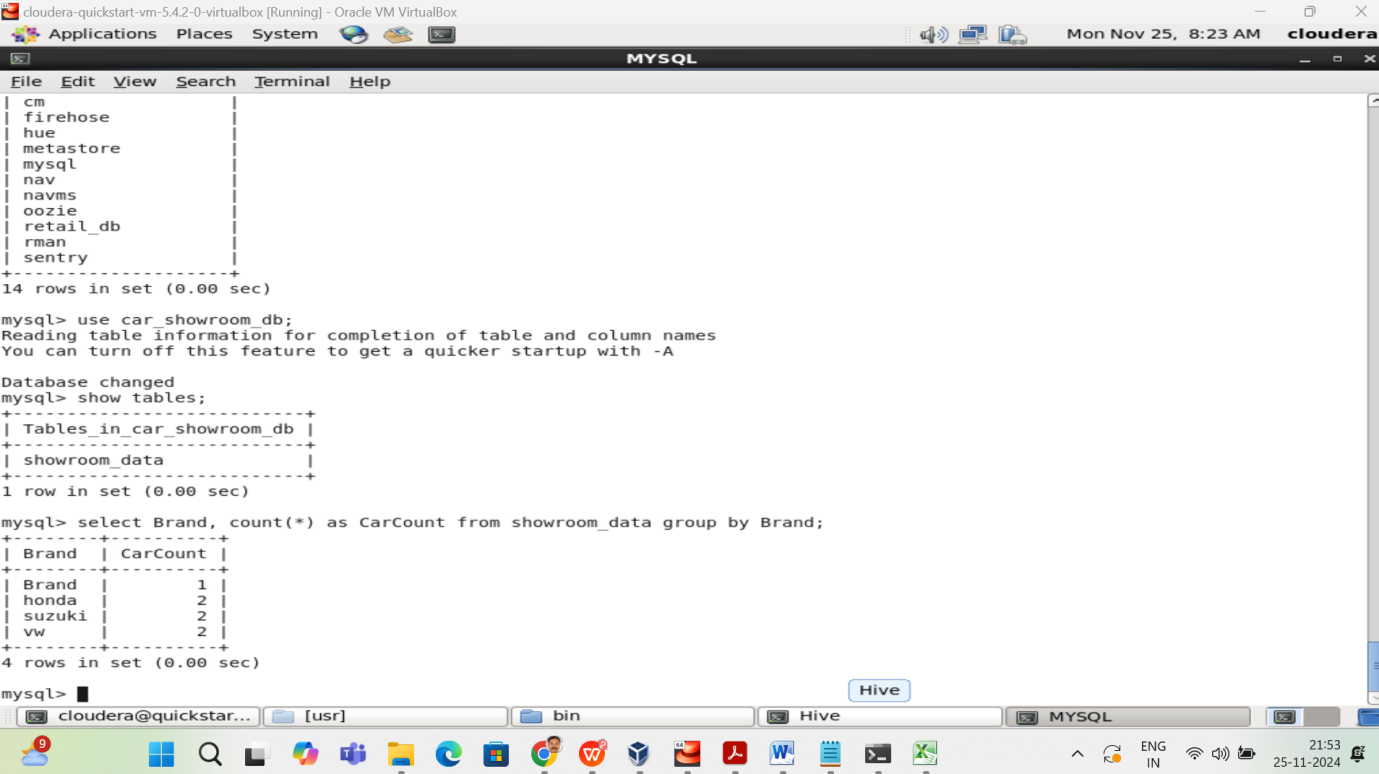


**Step 5: Perform Analysis Using Hive**

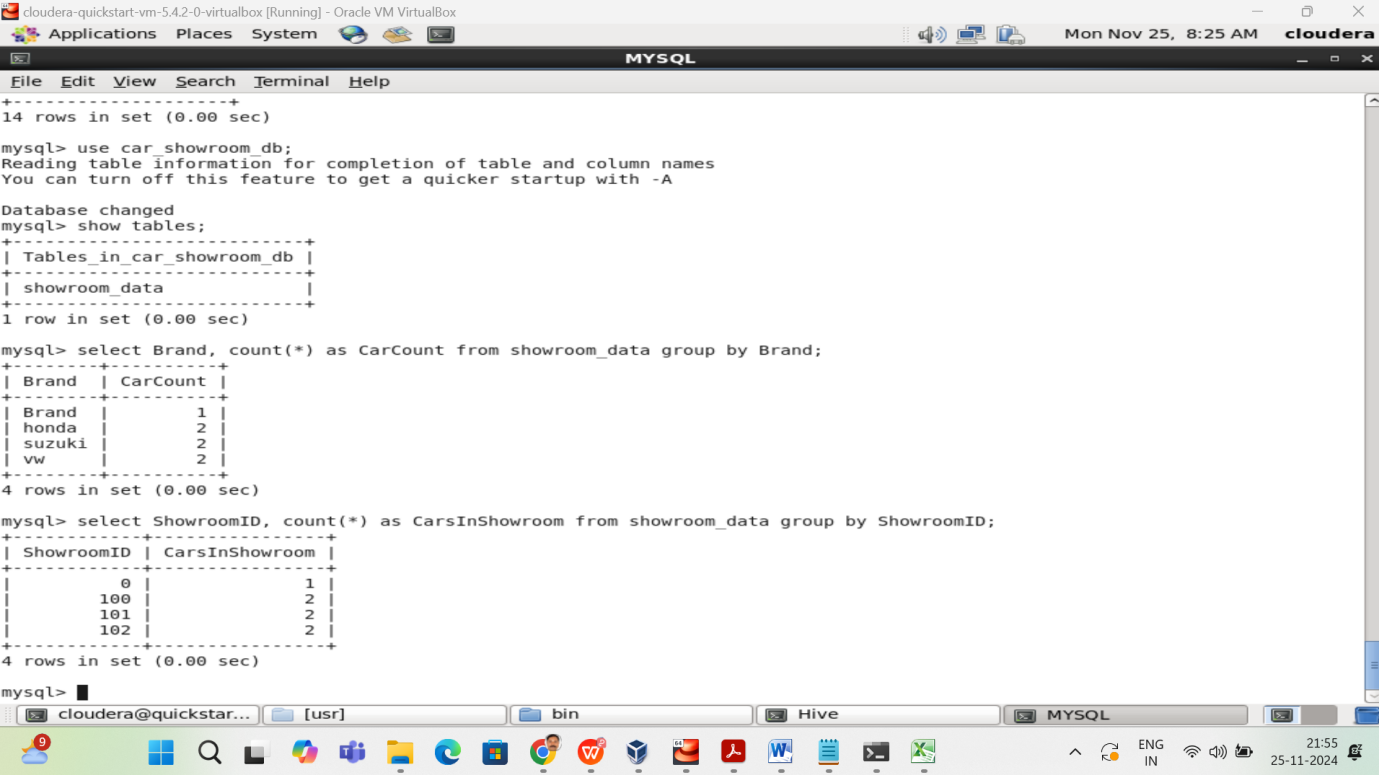
**5.1 Query Data**

Run the following HiveQL queries to analyze the data:

1. Count cars by brand:
2. SELECT Brand, COUNT(\*) AS CarCount FROM showroom\_data GROUP BY Brand;



1. Count cars in each showroom:
2. SELECT ShowroomID, COUNT(\*) AS CarsInShowroom FROM showroom\_data GROUP BY ShowroomID;



**5.2 Export Results to MySQL (Optional)**

If you want to export the aggregated data back to MySQL:

1. Save the Hive query output to a directory:
2. INSERT OVERWRITE DIRECTORY '/path/to/output'
3. SELECT Brand, COUNT(\*) AS CarCount FROM showroom\_data GROUP BY Brand;
4. Use Sqoop to export the data:
5. sqoop export \
6. --connect jdbc:mysql://localhost:3306/car\_showroom\_db \
7. --username root \
8. --password your\_password \
9. --table brand\_summary \
10. --export-dir /path/to/output \
11. --input-fields-terminated-by '\t'