**Lab Tutorial**

**NOTE**: Don’t forget to replace **syadav5** to your account name.

[$ ssh syadav5@129.150.79.29](mailto:$ ssh syadav5@129.150.79.29 )

$df –h

**Note**: If dev/shm directory is not present in the list, select any directory having size more than 9GB.

$ cd dev/shm

$ wget -O CarData.zip <http://usedcardataset.s3.amazonaws.com/archive.zip>

$ unzip CarData.zip

$ ls -al

$ hdfs dfs -mkdir GP2TermProject

$ hdfs dfs -ls

$ hdfs dfs -put /dev/shm/used\_cars\_data.csv  GP2TermProject

$ hdfs dfs –ls GP2TermProject

$ hdfs dfs -chmod -R o+w  .

$ hdfs dfs -ls   GP2TermProject/      // Write permissions

**Connect to Beeline CLI:**

> Beeline

!connect jdbc:hive2://summer2020-bdcsce-1:2181,summer2020-bdcsce-2:2181,summer2020-bdcsce-3:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2?tez.queue.name=interactive  bdcsce\_admin

**Note**: If database is not created in your name. Plese create one using following command. If already exists skip this command.

create database syadav5;

use syadav5;

show tables;

|  |
| --- |
| CREATE EXTERNAL TABLE IF NOT EXISTS cargurus\_usedcars(vin string, back\_legroom string, bed string, bed\_height  string, bed\_length string, body\_type string, cabin string, city string, city\_fuel\_economy float, combine\_fuel\_economy int, daysonmarket int, dealer\_zip string, description string, engine\_cylinders string, engine\_displacement float, engine\_type string, exterior\_color string, fleet string, frame\_damaged string, franchise\_dealer string, franchise\_make string, front\_legroom string, fuel\_tank\_volume string, fuel\_type string, has\_accidents string, height string, highway\_fuel\_economy float, horsepower int, interior\_color string, iscab string, is\_certified string, is\_cpo string, is\_new boolean, is\_oemcpo string, latitude double, length string, listed\_date date, listing\_color string, listing\_id bigint, longitude double, main\_picture\_url string, major\_options string, make\_name string, maximum\_seating string, mileage float, model\_name string, owner\_count int, power string, price float, salvage string, savings\_amount int, seller\_rating float, sp\_id bigint, sp\_name string, theft\_title string, torque string, transmission string, transmission\_display string, trimid string, trim\_name string, vehicle\_damage\_category string, wheel\_system  string, wheel\_system\_display string, wheelbase string, width string, year int) ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde' WITH SERDEPROPERTIES ("separatorChar" = "," , "quoteChar"  = "\"")  STORED AS TEXTFILE LOCATION  '/user/syadav5/GP2TermProject/' TBLPROPERTIES ('skip.header.line.count' = '1'); |

|  |
| --- |
| describe cargurus\_usedcars; |