#BIG DATA PROJECT#

Step 1----------

#In Local system's command prompt

/*Go to the location where file is stored*/
cd downloads

/*Check for the file*/
dir geolocation.csv **#or**
dir geolocation*

/*Copy the geolocation file from your home system to the vmware image*/
scp -o HostKeyAlgorithms=+ssh-rsa -o PubkeyAcceptedAlgorithms=+ssh-rsa geolocation.csv training@192.168.136.128:/home/training

#In vmware image

/*Check for the file*/
ls -l #in this path of local /home/training

/*Create the table in hive*/
```
CREATE TABLE geolocation (
truckid STRING,
driverid STRING,
event STRING,
latitude DOUBLE,
longitude DOUBLE,
city STRING,
state STRING,
velocity BIGINT,
event_ind BIGINT,
idling_ind BIGINT )
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
TBLPROPERTIES ("skip.header.line.count"="1");
```

/*Create a separate folder in hdfs for our project and the copy the file there from local path*/
hdfs dfs -mkdir /project
hdfs dfs -put /home/training/geolocation.csv /project

/*Load the data from your file to the table created*/
LOAD DATA INPATH '/project/geolocation.csv'
OVERWRITE INTO TABLE geolocation;

/*Check the data loaded*/
SELECT * FROM geolocation;

/*Then connect impala to PowerBI to load the data.*/

How to video:
https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Futdallas.box.com%2Fs%2Fgltlok00d9dvky85cyggxfa0vrghdto5&data=05%7C02%7CDebasmita.Ray%40UTDallas.edu%7C9e5c1dfad87e4fa4e2a008dd6837d513%7C8d281d1d9c4d4bf7b16e032d15de9f6c%7C0%7C0%7C638781311198946876%7CUnknown%7CTWFpbGZsb3d8eyJFbXB0eU1hcGkiOnRydWUsIlYiOiIwLjAuMDAwMCIsIlAiOiJXaW4zMiIsIkFOIjoiTWFpbCIsIldUIjoyfQ%3D%3D%7C0%7C%7C%7C&sdata=ARsFlQYUYkEvd5gQmtsBi7biJNJePcgNVzUdz6cM7HU%3D&reserved=0