# **Driver Data Analysis**

ls -l

csv file path: hdfs dfs -put /home/mavricbdhnov0433/drivers.csv

hdfs dfs -put /home/mavricbdhnov0433/timesheet.csv

hdfs dfs -put /home/mavricbdhnov0433/truck\_event\_text\_partition.csv

-rw-rw-r-- 1 mavricbdhnov0433 mavricbdhnov0433 2272077 Nov 29 09:26 truck\_event\_text\_partition (1).csv

[mavricbdhnov0433@ip-10-1-1-204 ~]$ hdfs dfs -put /home/mavricbdhnov0433/drivers.csv

[mavricbdhnov0433@ip-10-1-1-204 ~]$ hdfs dfs -put /home/mavricbdhnov0433/timesheet.csv

[mavricbdhnov0433@ip-10-1-1-204 ~]$ hdfs dfs -put /home/mavricbdhnov0433/truck\_event\_text\_partition (1).csv

-bash: syntax error near unexpected token `('

[mavricbdhnov0433@ip-10-1-1-204 ~]$ hdfs dfs -put /home/mavricbdhnov0433/truck\_event\_text\_partition.csv

Time taken: 1.687 seconds

hive> show databases like '\*khushboo\*';

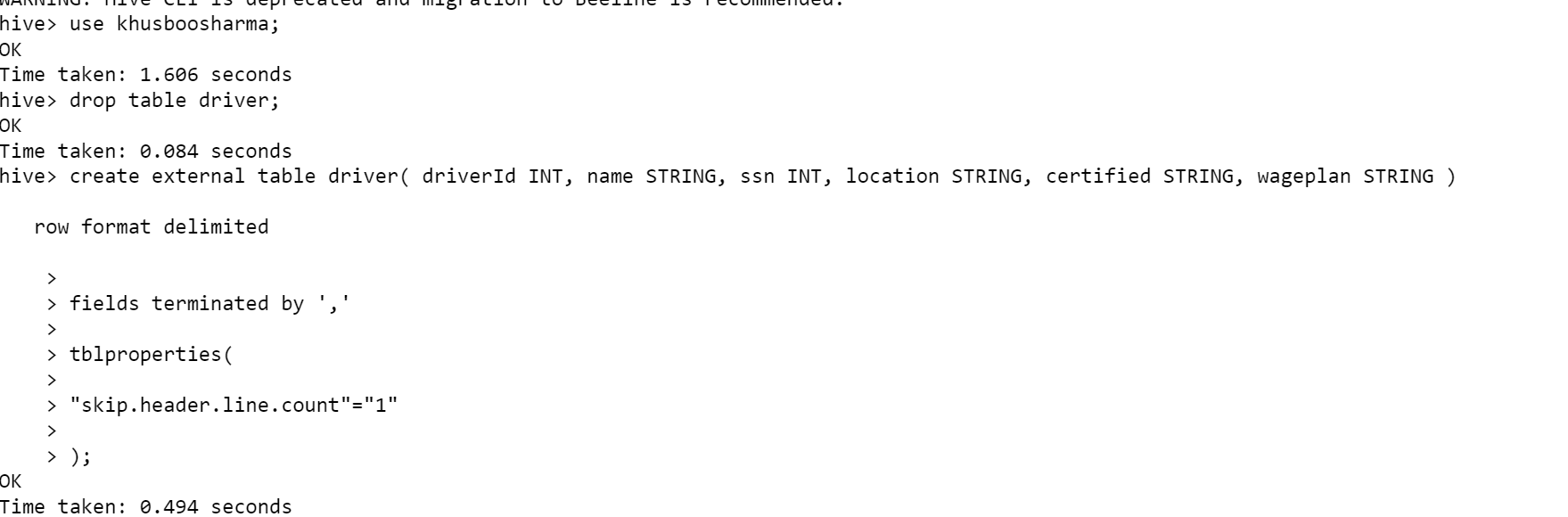
OK

Time taken: 0.035 seconds

hive> use khusboosharma;

OK

Time taken: 0.037 seconds



hive> create external table driver( driverId INT, name STRING, ssn INT, location STRING, certified STRING, wageplan STRING ) row format delimited

fields terminated by ',';

tblproperties(

"skip.header.line.count"="1"

);

OK

Time taken: 0.456 seconds

to load the data into hive :

LOAD DATA local INPATH 'drivers.csv' OVERWRITE INTO TABLE driver;

hive> LOAD DATA local INPATH 'drivers.csv' OVERWRITE INTO TABLE driver;

Loading data to table khusboosharma.driver

OK

Time taken: 1.469 seconds

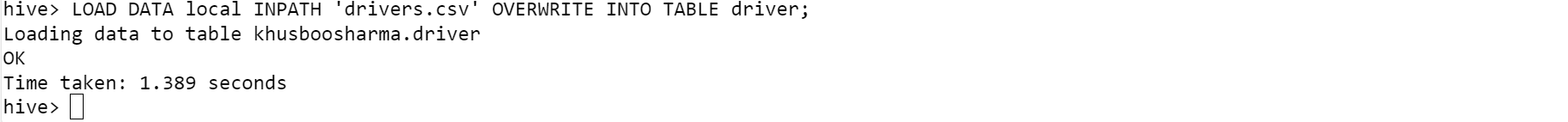
To see table structure:

describe driver;

to see the table data:

set hive.cli.print.header=true

select \* from driver limit 10;



A close-up of a document

Description automatically generated

create external table timesheet(driverId int, week int, hourslogged int, mileslogged int)

row format delimited

fields terminated by ','

tblproperties(

"skip.header.line.count"="1"

);

hive> describe timesheet;

OK

col\_name data\_type comment

driverid int

week int

hourslogged int

mileslogged int

Time taken: 0.086 seconds, Fetched: 4 row(s)

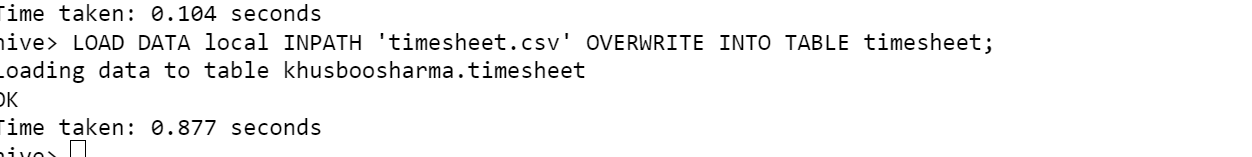
hive>

LOAD DATA local INPATH 'timesheet.csv' OVERWRITE INTO TABLE timesheet;

to see the table data:

set hive.cli.print.header=true;

select \* from timesheet limit 10;



**Query: 1**

**How many total hours and miles logged by each driver?**

set hive.cli.print.header=true;

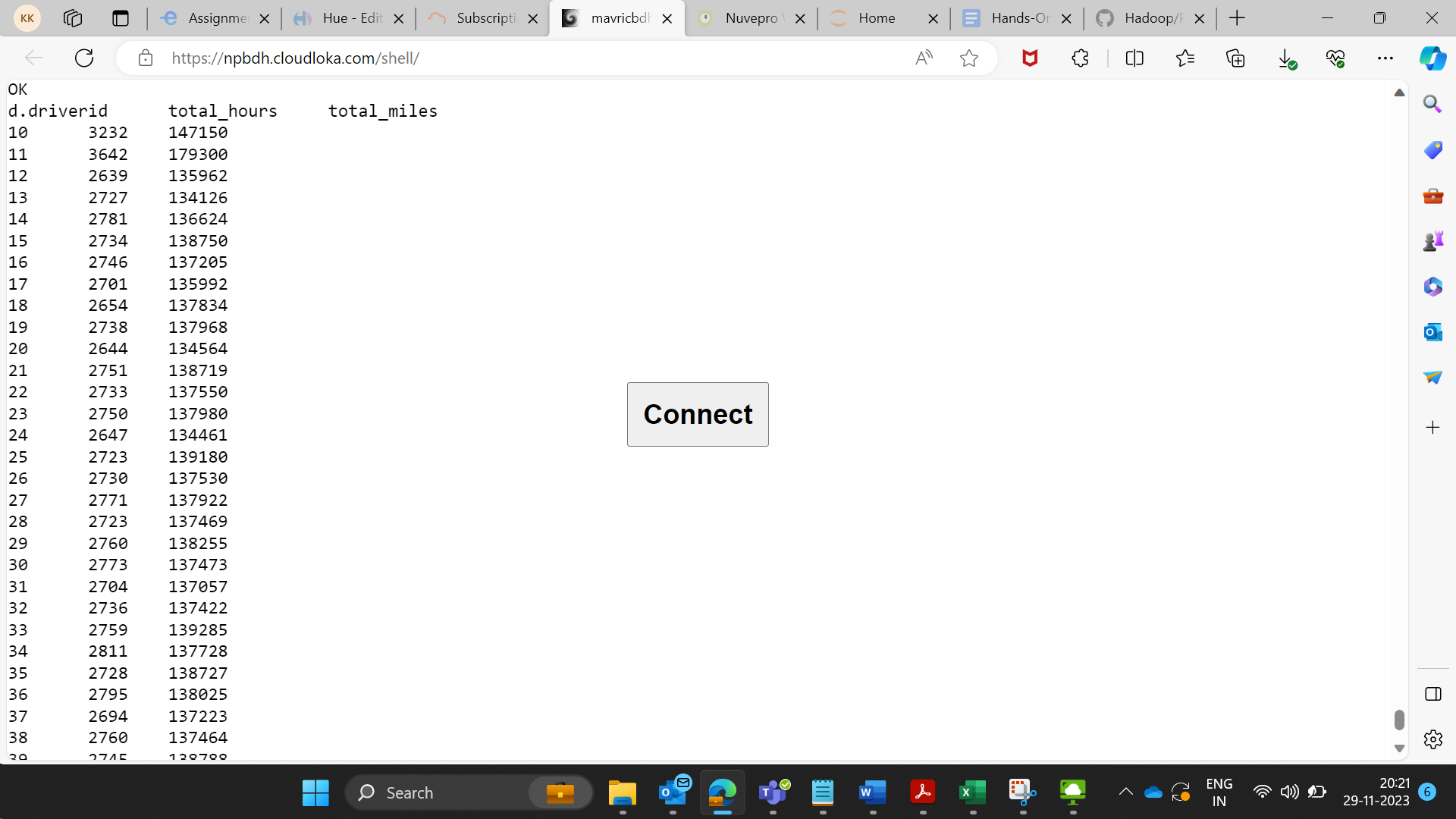
Select d.driverId,sum(t.hourslogged) as total\_hours,sum(t.mileslogged) as total\_miles

FROM driver d

JOIN timesheet t

ON d.driverId = t.driverId

GROUP BY d.driverId;



Query:2

Print the name of the driver in the above query

set hive.cli.print.header=true;

Select d.driverId,d.name,sum(t.hourslogged) as total\_hours,sum(t.mileslogged) as total\_miles

FROM driver d

JOIN timesheet t

ON d.driverId = t.driverId

GROUP BY d.driverId,d.name;





**Query: 3**

**Write the above query output in to hdfs directory**

**Hive> INSERT OVERWRITE DIRECTORY '/user/mavricbdhnov0433/driverquery'**

**Select d.driverId,d.name,sum(t.hourslogged) as total\_hours,sum(t.mileslogged) as total\_miles**

**FROM driver d**

**JOIN timesheet t**

**ON d.driverId = t.driverId**

**GROUP BY d.driverId,d.name;**

**HDFS ls -l**