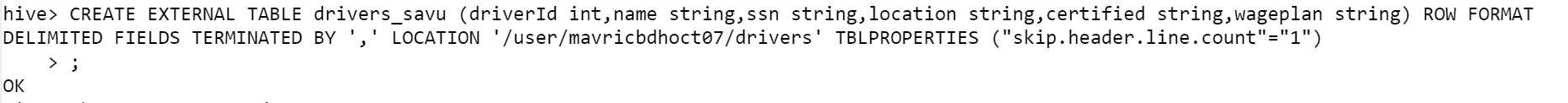
Create external tables for drivers and timesheet data and answer the following queries

hive> CREATE EXTERNAL TABLE drivers\_savu (driverId int,name string,ssn string,location string,certified string,wageplan string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',’ LOCATION '/user/mavricbdhoct07/drivers' TBLPROPERTIES ("skip.header.line.count"="1");

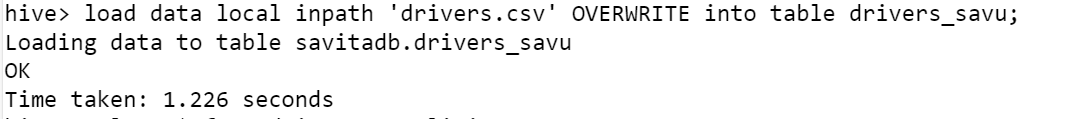
hive> desc drivers\_savu;



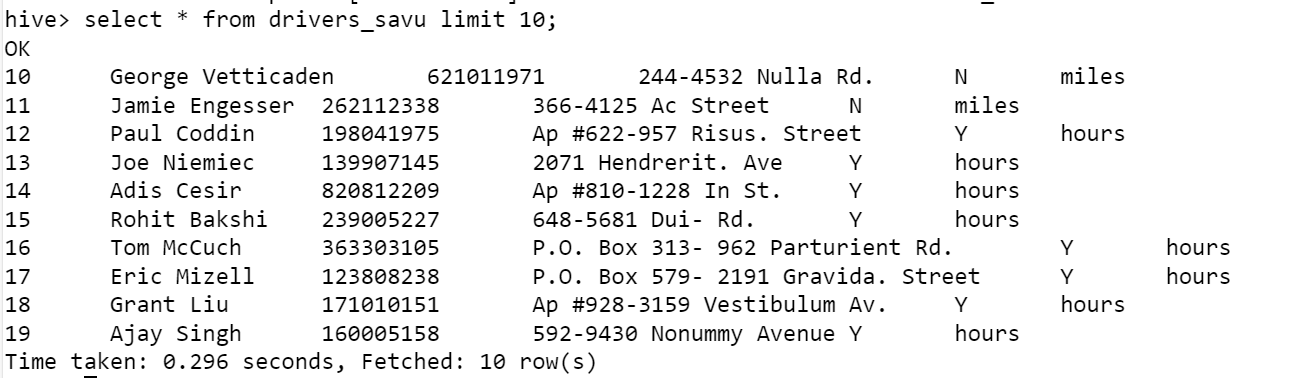
hive> load data local inpath 'drivers.csv' OVERWRITE into table drivers\_savu;

Loading data to table savitadb.drivers\_savu

OK



hive> select \* from drivers\_savu limit 10;

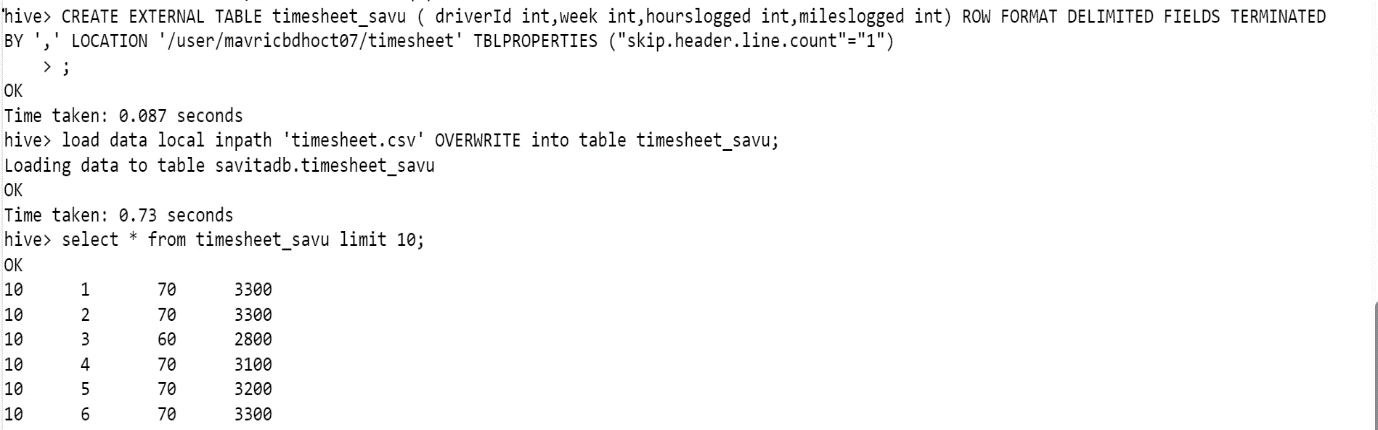


hive> CREATE EXTERNAL TABLE timesheet\_savu ( driverId int,week int,hourslogged int,mileslogged int) ROW FORMAT DELIMITED FIELDS TERMINATED

BY ',' LOCATION '/user/mavricbdhoct07/timesheet' TBLPROPERTIES ("skip.header.line.count"="1")

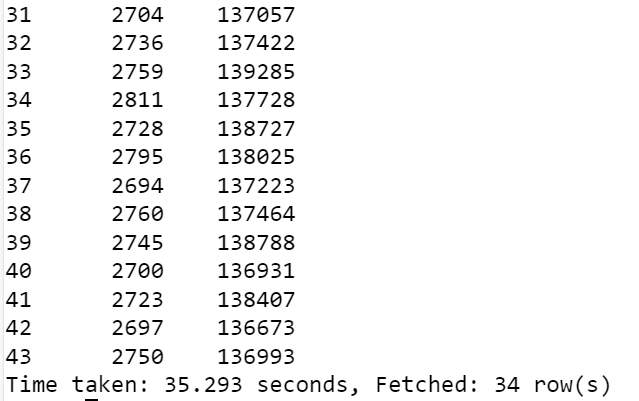
> ;

hive> load data local inpath 'timesheet.csv' OVERWRITE into table timesheet\_savu;



1. How many total hours and miles logged by each driver?

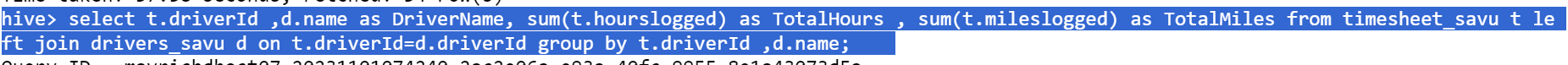
hive> select driverId, sum(hourslogged) as TotalHours , sum(mileslogged) as TotalMiles from timesheet\_savu group by driverId;

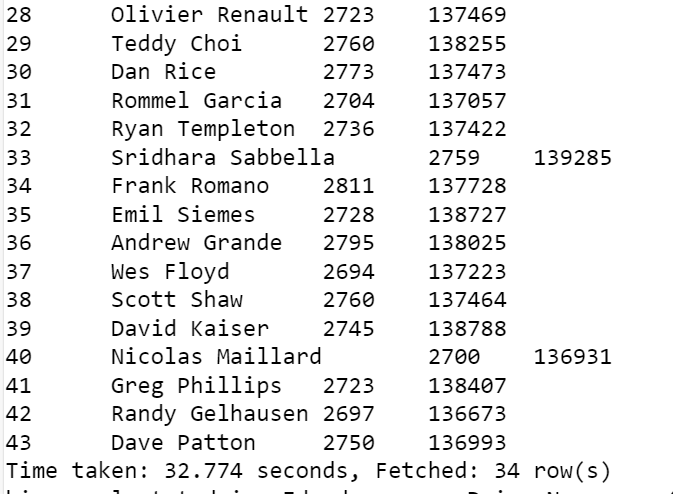


1. Print the name of the driver in the above query

hive> select t.driverId ,d.name as DriverName, sum(t.hourslogged) as TotalHours , sum(t.mileslogged) as TotalMiles from timesheet\_savu t

left join drivers\_savu d on t.driverId=d.driverId group by t.driverId ,d.name;





1. Write the above query output in to hdfs directory

**Created new table has driver\_a form using above query.**

1. hive> create table driver\_a as select t.driverId ,d.name as DriverName, sum(t.hourslogged) as TotalHours , sum(t.mileslogged) as TotalMiles

from timesheet\_savu t left join drivers\_savu d on t.driverId=d.driverId group by t.driverId ,d.name;

**and display output into hdfs directory**

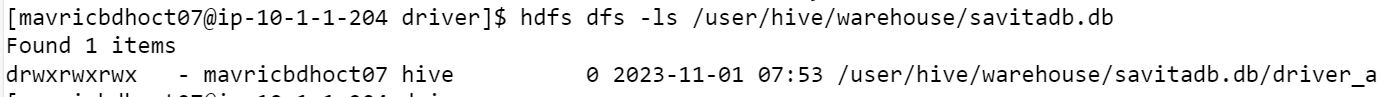
hive> exit;

[mavricbdhoct07@ip-10-1-1-204 driver]$ cd driver

[mavricbdhoct07@ip-10-1-1-204 driver]$ hdfs dfs -ls /user/hive/warehouse/savitadb.db

Found 1 items

drwxrwxrwx - mavricbdhoct07 hive 0 2023-11-01 07:53 /user/hive/warehouse/savitadb.db/driver\_a



1. come out of hive

in hdfs create one directory called output

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -mkdir output

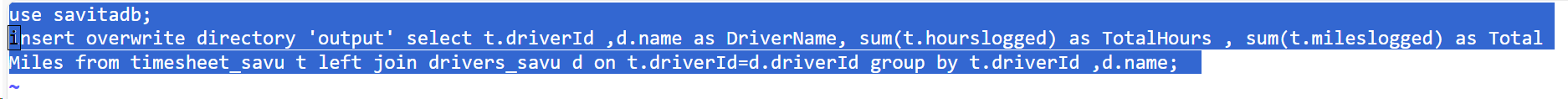
Using vi output.sql insert the above query

[mavricbdhoct07@ip-10-1-1-204 ~]$ vi output

use savitadb;

insert overwrite directory 'output' select t.driverId ,d.name as DriverName, sum(t.hourslogged) as TotalHours , sum(t.mileslogged) as Total

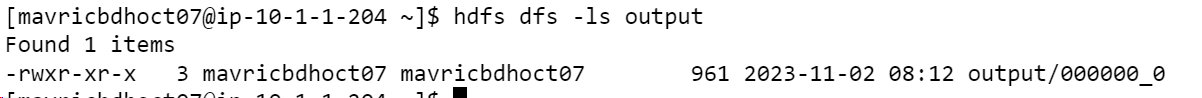
Miles from timesheet\_savu t left join drivers\_savu d on t.driverId=d.driverId group by t.driverId ,d.name;



[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -put output.sql

[mavricbdhoct07@ip-10-1-1-204 ~]$ hive -f output

[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -ls output



[mavricbdhoct07@ip-10-1-1-204 ~]$ hdfs dfs -cat output/000000\_0

A screenshot of a computer

Description automatically generated