**Assignment 2 Driver Analysis**

Create database indhudb;

Use indhudb;

CREATE EXTERNAL TABLE drivers(driverId int,name varchar(30),ssn int,location varchar(50),certified char(1),wage\_plan varchar(10))

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' location '/user/mavricbdhoct08/drivers';

LOAD DATA LOCAL INPATH 'drivers.csv' INTO TABLE drivers;

A screenshot of a computer

Description automatically generated

CREATE EXTERNAL TABLE timesheet(driverId int,week int,hours\_logged int,miles\_logged int)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' location '/user/mavricbdhoct08/timesheet';

LOAD DATA LOCAL INPATH 'timesheet.csv' INTO TABLE timesheet;

A screenshot of a computer

Description automatically generated

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

select driverId,sum(hours\_logged) as total\_hours, sum(miles\_logged) as total\_miles from timesheet group by driverId;

A screenshot of a computer

Description automatically generated

**Print the name of the driver in the above query**

select driverId,name,sum(hours\_logged) as total\_hours, sum(miles\_logged) as total\_miles from drivers D join timesheet T on D.driverId = T.driverId group by D.driverId,name;w

A screenshot of a computer

Description automatically generated

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

vi test.sql

use indhudb;

insert overwrite directory 'output' select driverId,name,sum(hours\_logged) as total\_hours, sum(miles\_logged) as total\_miles from drivers D join timesheet T on D.driverId = T.driverId group by D.driverId,name;

hive -f sql.test

hdfs dfs -ls

hdfs dfs -ls output

hdfs dfs -cat output/00000\_0

A screenshot of a computer

Description automatically generated