\*\*Create table facebook :

create table if not exists facebook\_nish

(userid int, age int, dob\_day int, dob\_year int, dob\_month int, gender string, tenure int,

friend\_count int, friendships\_initiated int, likes int, likes\_received int, mob\_likes int,

mob\_likes\_received int, web\_likes int, web\_likes\_received int)

row format delimited

fields terminated by ','

lines terminated by '\n'

stored as textfile

TBLPROPERTIES ("skip.header.line.count"="1");

Loading data into facebook\_nish

load data local inpath '/nisha/hive/assignment/pseudo\_facebook.csv'

overwrite into table facebook\_nish;

Q1: Find out the total no of users in the dataset

select count(\*) from facebook\_nish;

Q2: Find out the no of users above the age of 25

select count(\*) from facebook\_nish where age >25;

Q3: Do male face book users tend to have more friends or female users

select gender, SUM(friend\_count) from facebook\_nish group by gender;

Q4: How many likes do young people receive on facebook as opposed to old people

select SUM(likes), 'likesTOyoung' from facebook\_nish where age <50

union

select SUM(likes), 'likesTOold' from facebook\_nish where age > 50;

Q5: Find out the count of facebook users for each birthday month

select dob\_month, count(\*) from facebook\_nish group by dob\_month;

Q6: Do young members use mobile phones or computers for facebook browsing

Assume Column web\_likes and mob\_likes as user using computer and mobile respectively.

select 'Young\_mobile\_users', count(\*) from facebook\_nish where mob\_likes > 0

UNION

select 'Young\_computer\_users', count(\*) from facebook\_nish where web\_likes > 0;

select 'Young\_mobile\_users', count(\*) from facebook\_nish where mob\_likes > 0 and age < 50

UNION

select 'Young\_computer\_users', count(\*) from facebook\_nish where web\_likes > 0 and age < 50;

Q7: Do adult members use mobile phones or computers for facebook browsing

select 'Old\_computer\_users', count(\*) from facebook\_nish where web\_likes > 0 and age > 50

UNION

select 'Old\_mobile\_users', count(\*) from facebook\_nish where mob\_likes > 0 and age > 50;

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

\*\*Creating driver table\*\*

hive> Create external table if not exists drivers\_ni (driverId int, name string,

ssn int, location string, certified string, wage\_plan string)

> row format delimited fields terminated by ',' lines terminated by '\n'

> stored as textfile

> tblproperties("skip.header.line.count"="1");

\*\*Creating timesheet table\*\*

hive> Create external table if not exists timesheet\_ni (driverId int, week int, hours\_logged int,miles\_logged int)

> row format delimited fields terminated by ',' lines terminated by '\n'

> stored as textfile

> tblproperties("skip.header.line.count"="1");

\*\*Creating truck\_event\_part table\*\*

hive> Create external table if not exists truck\_event\_part\_ni(driverId int, truckId int, eventTime string, eventType string, longitude float, lat

itude float, eventKey string, CorrelationId string, driverName string, routeId bigint, routeName string, eventDate string)

> row format delimited fields terminated by ',' lines terminated by '\n'

> stored as textfile

> tblproperties("skip.header.line.count"="1");

\*\*Loading files into hive tables\*\*

\*\*Drivers\_ni

hive> load data local inpath '/user/mavericbigdata13/nisha/drivers.csv'

> overwrite into table drivers\_ni;

\*\*Timesheet\_ni

hive> load data local inpath

'/user/mavericbigdata13/nisha/timesheet.csv'

> overwrite into table timesheet\_ni;

\*\*truck\_event\_part\_ni

load data local inpath '/user/mavericbigdata13/nisha/

truck\_event\_text\_partition.csv'

overwrite into table truck\_event\_part\_ni;

Query: How many total hours and miles logged by each driver?

hive> select driverid, sum(hours\_logged), sum(miles\_logged) from timesheet\_ni group by driverid;

Query: Print the name of the driver in the above query

select t.driverid, d.name, sum(t.hours\_logged), sum(t.miles\_logged)

from timesheet\_ni t inner join drivers\_ni d

on (t.driverid =d.driverid)

group by t.driverid, d.name;

Query: Write the above query output in to hdfs directory

Insert overwrite directory '/user/mavericbigdata13/nisha'

select t.driverid, d.name, sum(t.hours\_logged), sum(t.miles\_logged)

from timesheet\_ni t inner join drivers\_ni d

On (t.driverid =d.driverid)

Group by t.driverid, d.name;

hive> insert overwrite directory '/user/mavericbigdata13/nisha'

> select t.driverid, d.name, sum(t.hours\_logged),

>sum(t.miles\_logged) from timesheet\_ni t inner join >drivers\_ni d on (t.driverid =d.driverid) group by >t.driverid, d.name;