**Problem Statement:**

You work for Spring Analytics Pvt. Ltd., a company that specializes in Big Data Analytics. Your organization’s latest client is a well-known social media enterprise. Based on the immense amount of data generated in the company, it is planning a targeted advertising campaign for its benefactors. Your firm is provided with this dataset, and it has put forth a set of queries based on that dataset. Use your data warehousing expertise in Apache Hive to help your client with its use case.

**Dataset:**

https://drive.google.com/file/d/1BzPFG0\_vdLtF\_bpApo9cssMKliqAyqQe/view?pli=1

**Tasks and Solutions:**

#Launching the hive metastore

$ hive –service metastore

# Starting hive in another tab

$hive

# Loading the dataset

create database hive\_assignment2;

CREATE EXTERNAL TABLE pseudo\_facebook (userid bigint, 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, mobile\_likes int, mobile\_likes\_received int, www\_likes int, www\_likes\_received int) row format delimited fields terminated by ',' TBLPROPERTIES ("skip.header.line.count"="1");

LOAD DATA LOCAL INPATH '/home/hadoop-intellipaat/Downloads/HiveAssignment2/pseudo\_facebook.csv' INTO TABLE pseudo\_facebook;

**1. What is the average age of a social media account user?**

select avg(age) as avg\_age from pseudo\_facebook;



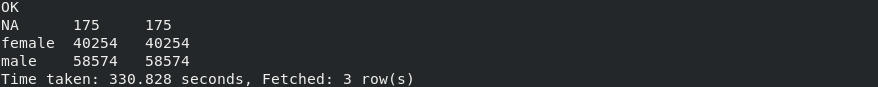
Ans: 37 years

**2. Does the social media platform have a higher number of male users, female users, or gender undisclosed users?**

select gender, count(\*) as row\_count, count(distinct userid) as user\_count

from pseudo\_facebook

group by gender;



Ans: Higher number of male users.

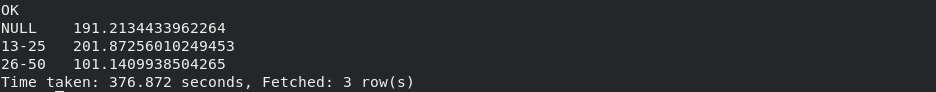
**3. In male users, on average, does the age demographic of 13–25 have more or less friends than the demographic of 26–50? Assess this with appropriate statistical reasoning.**

select case when age >= 13 and age <= 25 then '13-25' when age >= 26 and age <= 50 then '26-50' end as age\_demographic, avg(friend\_count) as avg\_friend\_count

from pseudo\_facebook

where gender = 'male'

group by case when age >= 13 and age <= 25 then '13-25' when age >= 26 and age <= 50 then '26-50' end;



Ans: 13-25 age demographic has more friends.

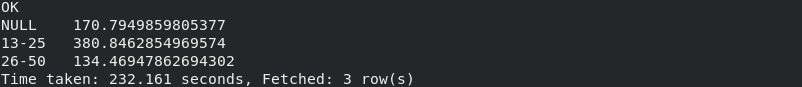
**4. In female users, on average, does the age demographic of 13–25 have more or less friends than the demographic of 26–50? Assess this with appropriate statistical reasoning.**

select case when age >= 13 and age <= 25 then '13-25' when age >= 26 and age <= 50 then '26-50' end as age\_demographic, avg(friend\_count) as avg\_friend\_count

from pseudo\_facebook

where gender = 'female'

group by case when age >= 13 and age <= 25 then '13-25' when age >= 26 and age <= 50 then '26-50' end;



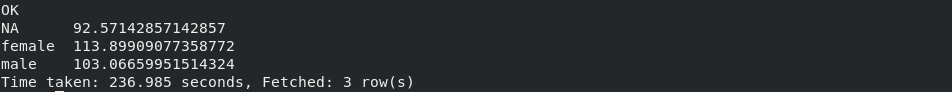
Ans: 13-25 age demographic has more friends.

**5. Which gender is more likely to send out a higher number of friend requests on average?**

select gender, avg(friendships\_initiated) as avg\_friendships\_initiated

from pseudo\_facebook

group by gender;



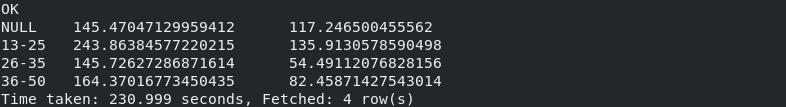
Ans: Females are more likely to send out higher friend requests.

**6. With the age demographics of 13–25, 26–35, and 36–50 as focal points, evaluate the comparison between mobile application usage and web browser usage when accessing the social media website. Use this to determine if mobile phones have indeed taken over the digital marketspace.**

select case when age >= 13 and age <= 25 then '13-25' when age >= 26 and age <= 35 then '26-35' when age >= 36 and age <= 50 then '36-50' end as age\_demographic, avg(mobile\_likes + mobile\_likes\_received) as avg\_mobile\_engagement, avg(www\_likes + www\_likes\_received) as www\_likes\_engagement

from pseudo\_facebook

group by case when age >= 13 and age <= 25 then '13-25' when age >= 26 and age <= 35 then '26-35' when age >= 36 and age <= 50 then '36-50' end;



Ans: Yes, mobile phones have taken over the digital marketplace as they have greater average engagement across all age demographics.