DB SYSTEMS FOR ANALYTICS HOMEWORK-1

1.a. [10 points]
Compare what it is like to work with data in
(i) binary (1s and 0s) files
(ii) text files
(iii) text files containing key-value pairs such as, "Name: John Glow"
(iv) Excel Spreadsheets
(v) Relational Database

I. **Binary (1s and 0s) files**
- Binary files are not human readable, a special program or software is needed to read and write the files.
- There is no efficiency in binary files when it comes to text, but they are very useful for images.
- Binary files have better access speed and portability than text files.
- A binary file stores values in numeric format, which consumes less memory than a text file.

II. **Text files:**
- Unlike binary files, text files can be readable and modifiable because they contain content which is human readable.
- text files only contain textual data, so their limitations are greater than those of binary files. But in contrast, text files are less prone to corruption than binary files.
- A text file occupies more memory space than a binary file, plus it is not as portable as a binary file.

III. **Text files containing key-value pairs such as, "Name: John Glow":**
- It is easy to read and write the data because the format is basic.
- A key value pair contains information about a set of two data items. The unique identifier in a key-value pair identifies the location of the value.
- But few of their limitation are that the structures are difficult to maintain every time, the processing rate is slow, and the formatting is incorrect.
.
IV. **Excel Spreadsheets:**
- Spreadsheets provide us with the flexibility to store different types of data in the same field. A spreadsheet also allows for easier overwriting and deleting than a database. (Which can also be an integrity issue later).
- As spreadsheets require a lot of hard-drive space for data storage, they are unmanageable when dealing with large amounts of data. And also spreadsheets operate slower than database when dealing with large datasets.
- When data is in separate spreadsheets, it is difficult to analyze and compare the data. Data bases, on the other hand, allow for efficient data management and easy comparison (querying), analysis, and reporting.

V. Relational Database:

- Databases follow standard integrity properties(ACID) to provide accurate data and prevent future issues related to data cleanliness.
- Compared to spreadsheets, databases are more efficient and effective when dealing with huge amounts of data as they can handle volumes of data with their querying tools.
- By using stored procedures, triggers and many other features in database, queries and reports are automated, which is making databases faster than spreadsheets. Apart from this, databases also provide better accessibility and security than any other files.

1.b. [10 points]
Search for job descriptions which specifically ask for some of the terms listed on the syllabus sheet. Here's an example of one such search:

- https://www.linkedin.com/jobs/search/?currentJobId=3231898557&geoId=103644278&keywords=sql&location=United%20States&refresh=true

- https://www.linkedin.com/jobs/search/?currentJobId=3257009028&geoId=103644278&keywords=nosql&location=United%20States&refresh=true

- https://www.linkedin.com/jobs/search/?currentJobId=3245597804&geoId=103644278&keywords=etl&location=United%20States&refresh=true

- https://www.linkedin.com/jobs/search/?currentJobId=3261418359&geoId=103644278&keywords=data%20analytics&location=United%20States&refresh=true
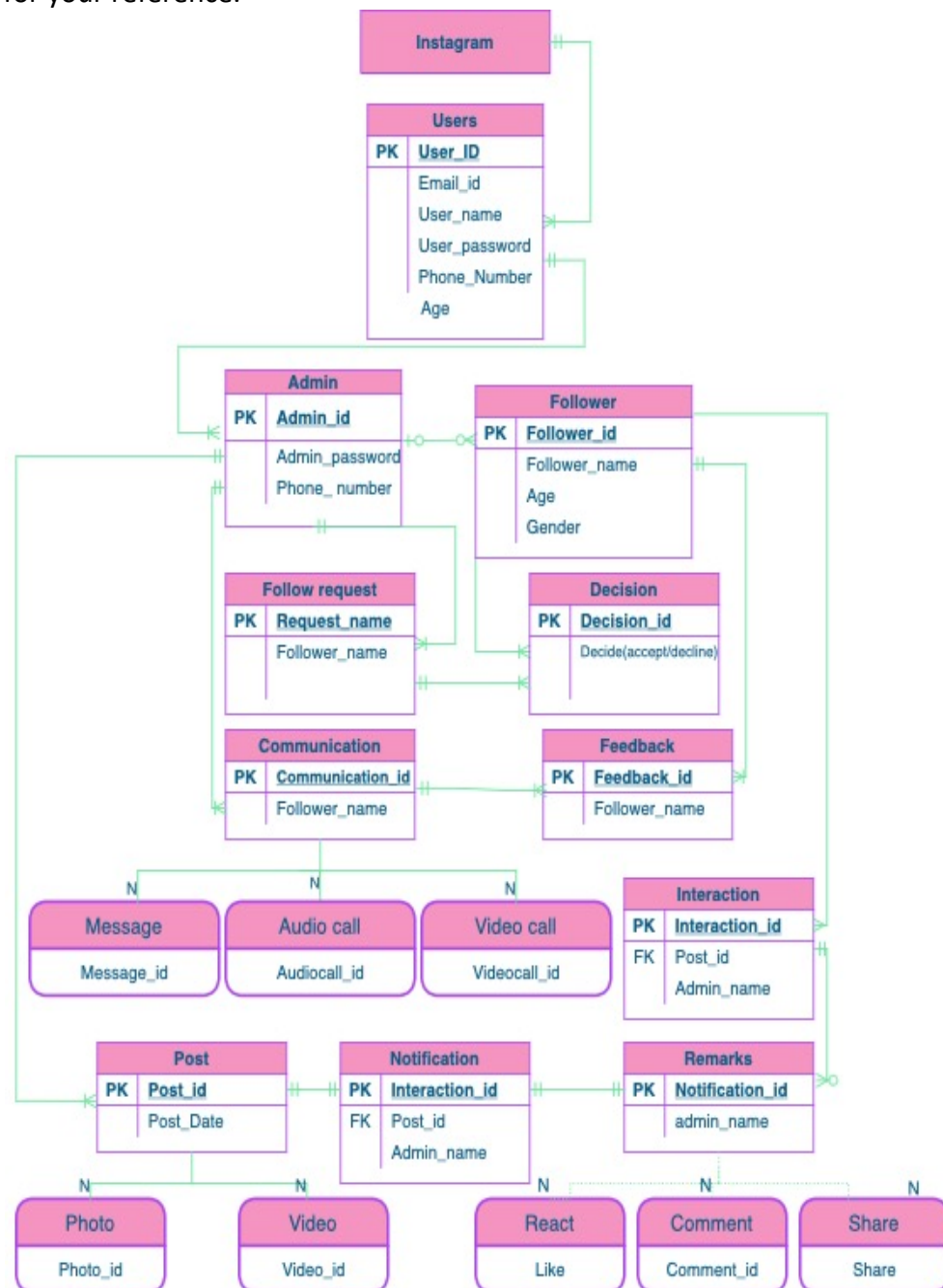
Describe your top 5 takeaways / observations based on the search results.

- SQL and PL SQL is one of the most demanding technologies that needs a thorough understanding of Oracle databases and also needs experience and knowledge of Oracle reports, forms, and performance tuning.
- The majority of companies require experience and knowledge of designing high-performance online analytical processing and online transaction processing databases.
- A Oracle database administrator is one of the most demanding roles in the industry. Specifically, they focus on Oracle Database Security features and Database Firewalls. So, we should develop skills on firewalls and security features.
- Data analysts and engineers are expected to develop data collection strategies, optimize data delivery, and identify and automate manual tasks.
- The most important skills for data analyst roles are Python, SQL, NOSQL, and tableau/PowerBI.
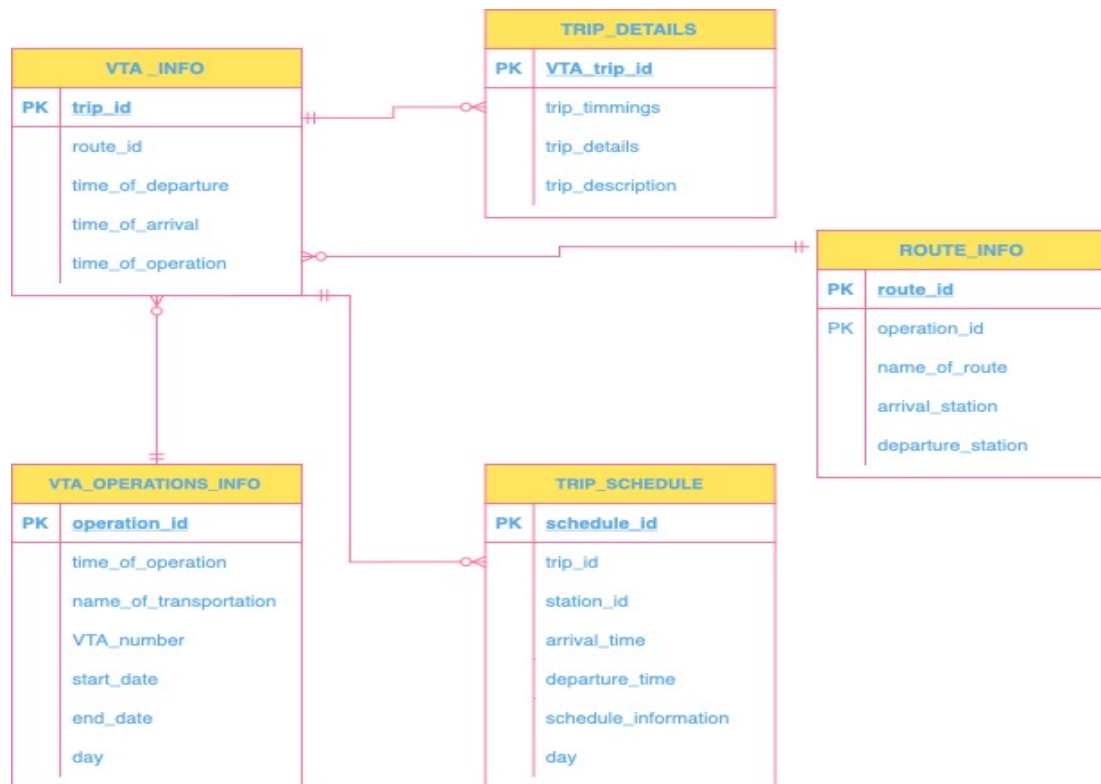
2. [20 points]
To the best of your knowledge, explain the data model and using tools such as draw.io or Visio, draw the ER diagram for the user data in anyone (a) social media company such as Facebook, Twitter, LinkedIn, Reddit, etc and any one (b) transportation company such as VTA, BART, Amtrak, etc

2.a The below picture explain the information of the Instagram Data Model. Please find below picture for your reference.
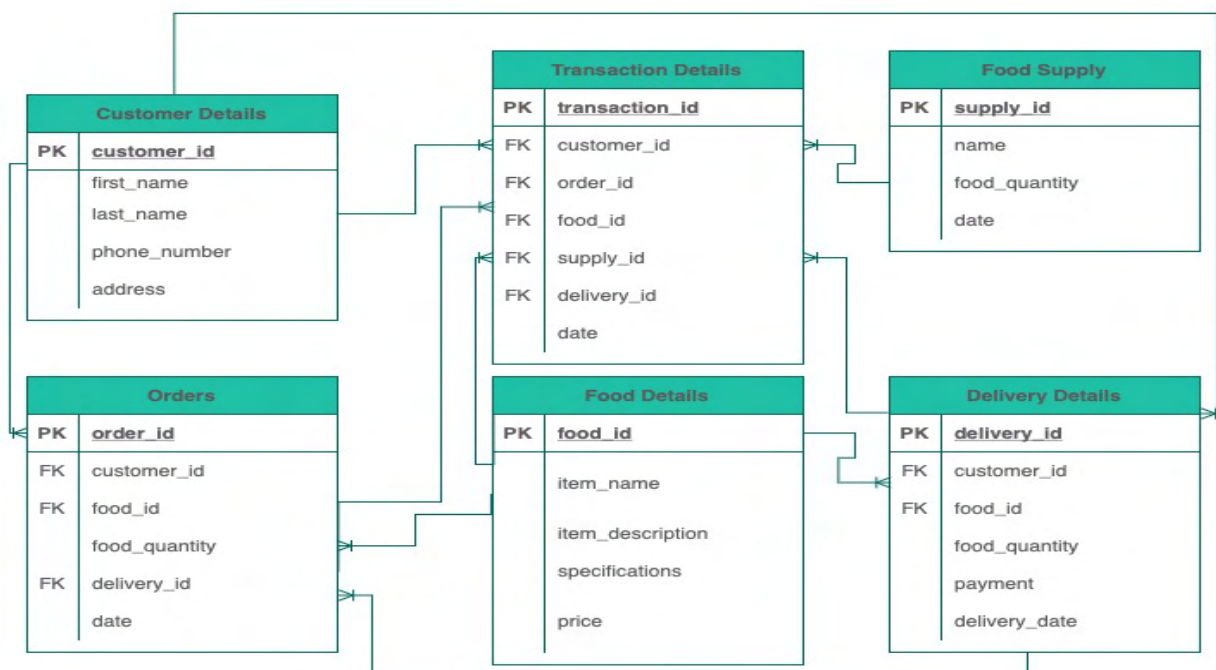

Basically Instagram have many number of users in it, which can be classified into two types of users, one is the admin user who have all the authority of his account where he can post, communicate anything, and the other type is follower user where he can decide the follower request of admin user and both can interact.

2.b The below picture explain the information of the VTA transportation Data Model.
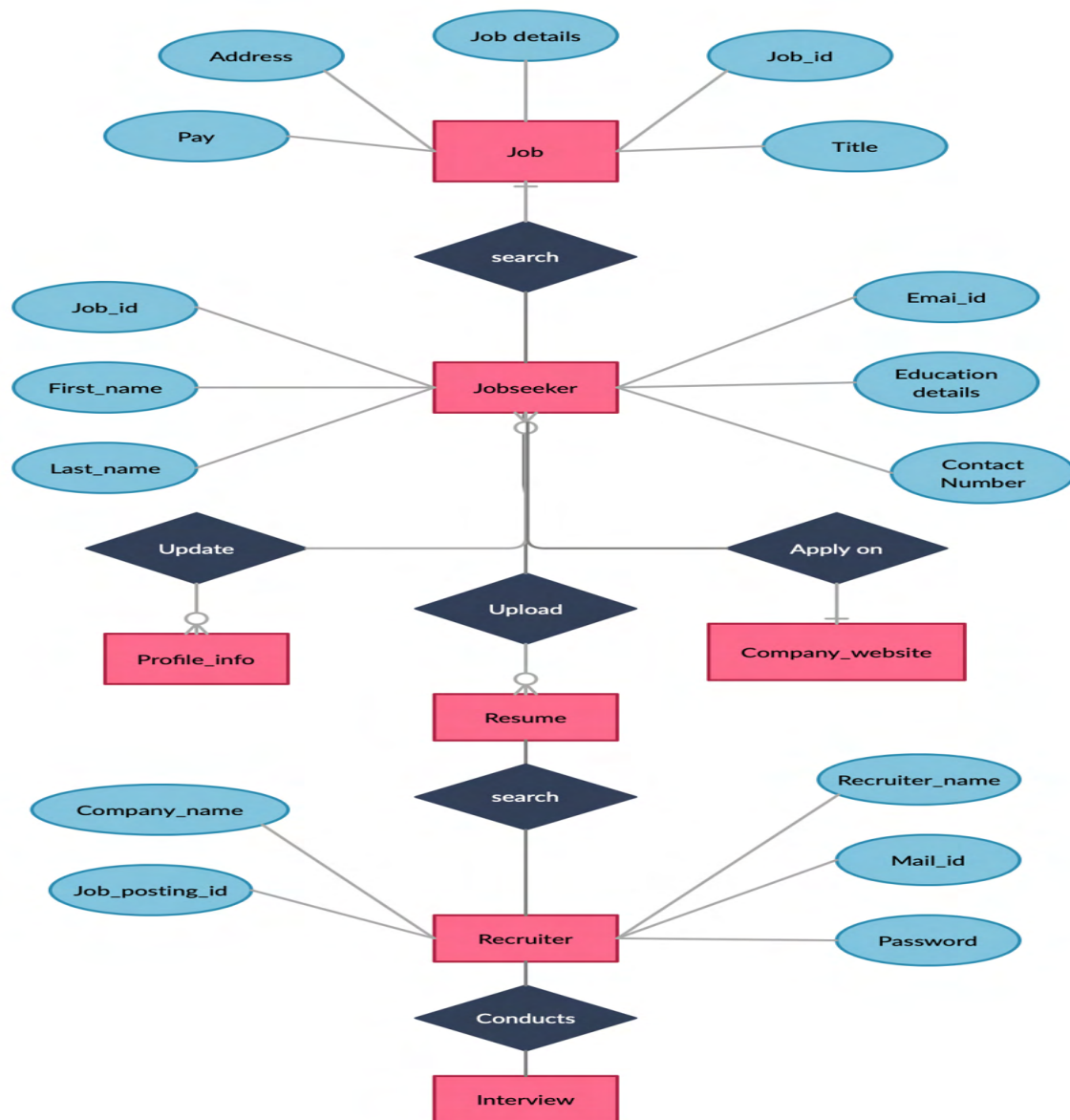

3. Identify the entities and the relationships between them in any 2 other very different scenarios in a common man's life where databases are or can be used. Using tools such as draw.io or Visio, model the entities, their attributes, and the relationships in an ER diagram, one for each of the scenarios (2 ER diagrams).

3.a. The below example depicts the information about the Online food ordering system.

In the above ER diagram of online food ordering system we have few important entities and their attributes as shown above. And all the relations are 1 : many in this case.

3.b : The below picture explains the information about online job search of a job seeker.

4. [20 points]

Follow the instructions at https://dev.mysql.com/doc/refman/8.0/en/installing.html (Links to an external site.) to install MySQL and then in README.md at https://github.com/datacharmer/test_db (Links to an external site.) to import the employees database into your MySQL instance. Examine the tables in the employee's database using the commands such as the following.

- **Installed MYSQL database and created employees database from the GitHub Successfully.**

```
nikhilmylarusetty@Nikhils-MacBook-Air ~ % mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 25
Server version: 8.0.30 MySQL Community Server - GPL

Copyright (c) 2000, 2022, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| employees          |
| information_schema |
| mysql              |
| performance_schema |
| sys                |
+--------------------+
5 rows in set (0.05 sec)


mysql> use employees;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A
```

```
Last login: Sun Sep 11 14:59:44 on ttys001
(base) nikhilmylarusetty@Nikhils-MacBook-Air ~ % mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 30
Server version: 8.0.30 MySQL Community Server - GPL

Copyright (c) 2000, 2022, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| employees          |
| information_schema |
| mysql              |
| performance_schema |
| sys                |
+--------------------+
5 rows in set (0.01 sec)

mysql> show tables;
ERROR 1046 (3D000): No database selected
mysql> use employees;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+----------------------+
| Tables_in_employees  |
+----------------------+
| current_dept_emp     |
| departments          |
| dept_emp             |
| dept_emp_latest_date |
| dept_manager         |
| employees            |
| salaries             |
| titles               |
+----------------------+
8 rows in set (0.01 sec)
```

```
Last login: Sun Sep 11 14:59:44 on ttys001
[(base) nikhilmylarusetty@Nikhils-MacBook-Air ~ % mysql -u root -p
[Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 30
Server version: 8.0.30 MySQL Community Server - GPL

Copyright (c) 2000, 2022, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

[mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| employees          |
| information_schema |
| mysql              |
| performance_schema |
| sys                |
+--------------------+
5 rows in set (0.01 sec)

[mysql> show tables;
ERROR 1046 (3D000): No database selected
[mysql> use employees;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
[mysql> show tables;
+----------------------+
| Tables_in_employees  |
+----------------------+
| current_dept_emp     |
| departments          |
| dept_emp             |
| dept_emp_latest_date |
| dept_manager         |
| employees            |
| salaries             |
| titles               |
+----------------------+
8 rows in set (0.01 sec)

[mysql> describe departments;
+-----------+-------------+------+-----+---------+-------+
| Field     | Type        | Null | Key | Default | Extra |
+-----------+-------------+------+-----+---------+-------+
| dept_no   | char(4)     | NO   | PRI | NULL    |       |
| dept_name | varchar(40) | NO   | UNI | NULL    |       |
+-----------+-------------+------+-----+---------+-------+
2 rows in set (0.01 sec)

mysql>
```
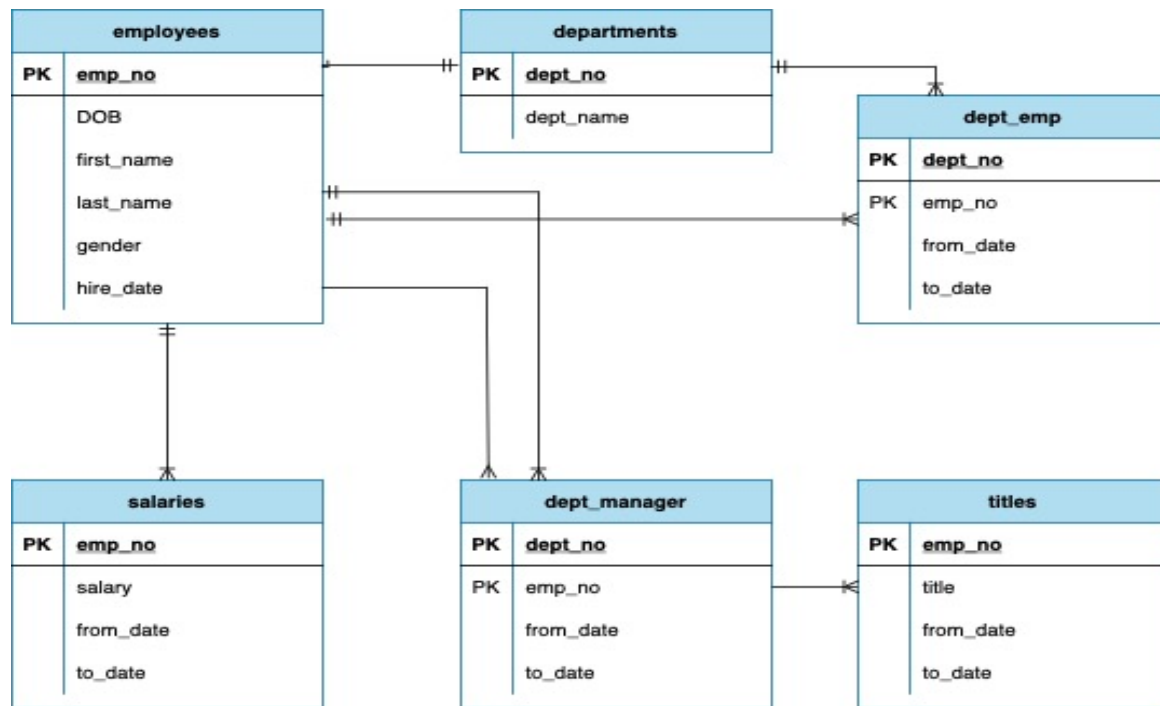
In the above 3 screenshots we have examined the tables in the employees database using commands :show databases; use employees; show table; describe departments;

The below picture depicts the information about the Employees database with all the relationships across the tables present in employee database. Please find the below screen shot for your reference.
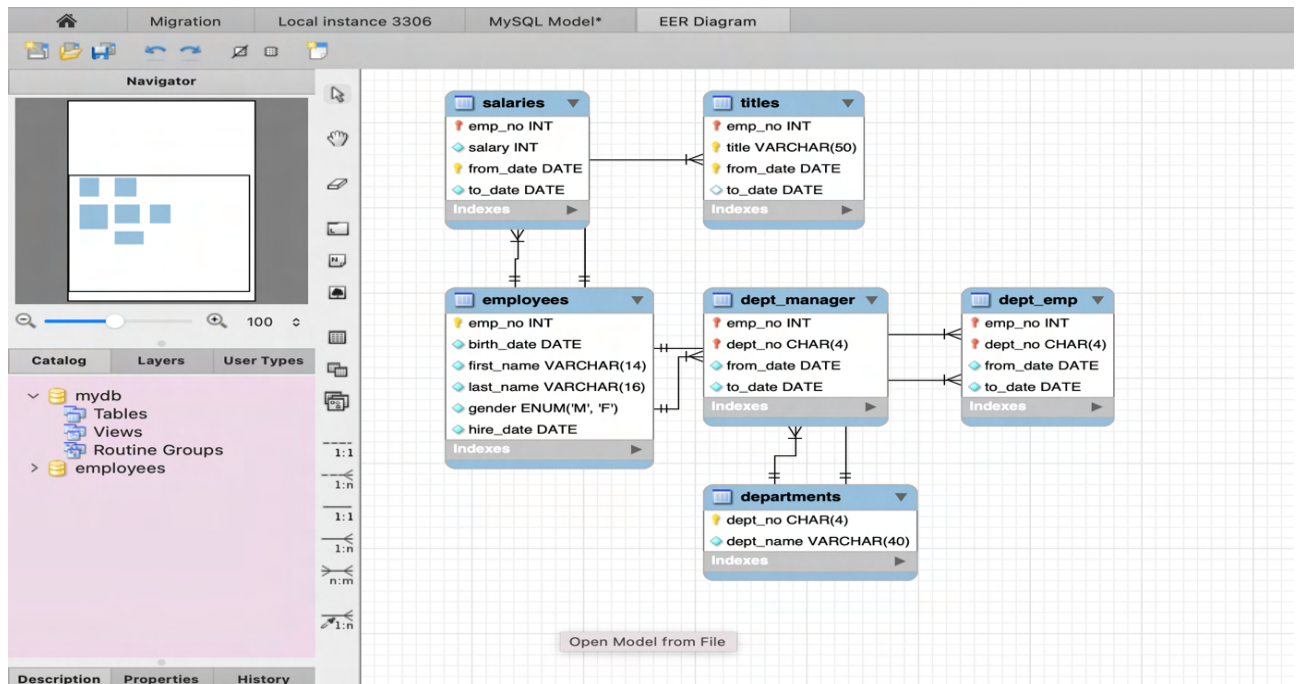

5.a. [10 points]

Install the appropriate version of MySQL workbench for your system from https://dev.mysql.com/downloads/ (Links to an external site.). For instance, if your MacBook is running Mojave (10.14), install 8.0.14 version of the workbench. Higher versions may not work on Mojave. Complete the initial setup to associate the employees database from (3) above with the workbench. Click on Database->Reverse Engineer and complete the steps to generate an ER diagram. Either export the ER diagram as a pdf or take a screenshot and submit the .png / .pdf. Do NOT revise the ER diagram that you manually generated in (3) above.

Installed the MYSQL workbench Successfully

ER Diagram: Up on performing the reverse engineering we were able to generate the Entity Relation Ship model for the employee's database by the system. Please find below Screenshot for your reference.

5.b. [10 points]
Explore the various options available in MySQL Workbench. Feel free to browse through the documentation in the help section of the tool to understand it better. Describe your top 5 takeaways from this exploration.

- By using the MySQL workbench, we can automate routine tasks, build tools, create schemas, manipulate them, and build customized workbenches.
- We can also utilize MySQL workbench to develop SQL queries, model data, manage tables and to migrate data and provide enterprise support.
- MySQL workbench makes it super-efficient when we are working with large SQL scripts, procedures, views, or functions as we can reformat the SQL statements in a readable format.
- MySQL workbench is also offering us with many other features like performance tuning, connection management, MySQL enterprise backup, audit and firewall.
- By using SQL workbench, we will be able to create, edit, reverse engineer, or forward engineer DB models graphically.
- Authentication can be checked before connecting to a database using MySQL's "Test Connection" button