**QUESTION 1**

**1.a (1)**

A binary file includes some type of header that stipulates the type of file. Binary files can only be read by specialized software or hardware that has been programmed to read binary data.

Each type of data is encoded in binary according to a separate set of rules. Before we can write to a file, you must first open it. You can add data to the file's end using append mode. In write mode, the file contents are overwritten, or if none are present, a new file is created.

It is used to store data in the form of bytes (8 bits), which are often interpreted as something other than textual characters. The headers of these files typically include instructions that explain how to read the data they store. Any kind of data can be saved in a computer using them.

**(2)** A text file is a type of digital, non-executable file that contains letters, numbers, symbols and/or a combination. It enables the creation and storage of text without any special formatting. Text files are also known as flat files or ASCII files. A text file is a kind of digital file that cannot be executed and contains letters, numbers, symbols, and/or a mix of these. It makes it possible to create and save text without any additional formatting. ASCII files or flat files are other names for text files. Text files are used to store human readable, standardized, and structured text data. It is defined in a variety of formats, including the most widely used ANSI for Windows-based operating systems and the most common ASCII for cross-platform use. Text formatting features including text alignment, bold text, and font styles are absent.

**(3)** A key-value pair consists of two bits of information that are connected. A value is either the data being identified or a link to that data, while a key is an exclusive identifier that points to its associated value.

The straightforward data format makes read and write operations quick.

**(4)** A Microsoft Excel Open XML Spreadsheet (XLSX) file is one that has the.xlsx file extension and was produced using Microsoft Excel. Other spreadsheet programs including Apple Numbers, Google Docs, and OpenOffice can also open this format. They are kept in a compressed Zip file that also includes a number of other files required to open the document.

XLSX files are files used in Microsoft Excel, a spreadsheet application that uses tables to organize, analyze, and store data. Each cell can contain text or numerical data, including incorporating mathematical formulas.

**(5)** A database management system that is based on the relational paradigm is called an RDBMS (Relation Database Management System). Table, Record/Tuple/Row, Field, and Column/Attribute are its four main parts. Database with a hierarchy. These DBMS store the data via parent-child relationships. Network DBMS. Many-to-many relationships are supported by Network DBMS, creating intricate database architectures. Databases offer centralized security and storage.

1. **(b)**

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?keywords=sql&location=United%20States (Links to an external site.)](https://www.linkedin.com/jobs/search?keywords=sql&location=United%20States)

Describe your top 5 takeaways / observations based on the search results.

* Skilled at designing cluster topologies with MongoDB or NoSQL schema

<https://www.linkedin.com/jobs/search/?currentJobId=3243269605&keywords=NoSQL>

* Utilizing business analytics reporting technologies, analyze and understand large data sets pertaining to a customer's business and create reports for both internal and external audiences.

<https://www.linkedin.com/jobs/search/?currentJobId=3268488957&geoId=103644278&keywords=Data%20Analyst&location=United%20States&refresh=true>

* Using a variety of tools and technologies, such as Pentaho/SSIS/Informatica, PowerBI, SQL, and PostgreSQL/SQL Server, design, create, debug, and support complex data integration interfaces and processes (including stored procedures and complex queries).

<https://www.linkedin.com/jobs/search/?currentJobId=3253517183&geoId=103644278&keywords=Data%20Warehouse&location=United%20States&refresh=true>

* Create and manage the best cloud-based data pipeline systems that can securely, scalable, and effectively ingest both structured and unstructured data utilizing batch and streaming approaches.

<https://www.linkedin.com/jobs/search/?currentJobId=3260643882&geoId=103644278&keywords=etl%20developer&location=United%20States&refresh=true>

* Maintain conceptual logical and physical data models, together with the associated metadata, and validate business data objects for completeness and accuracy.

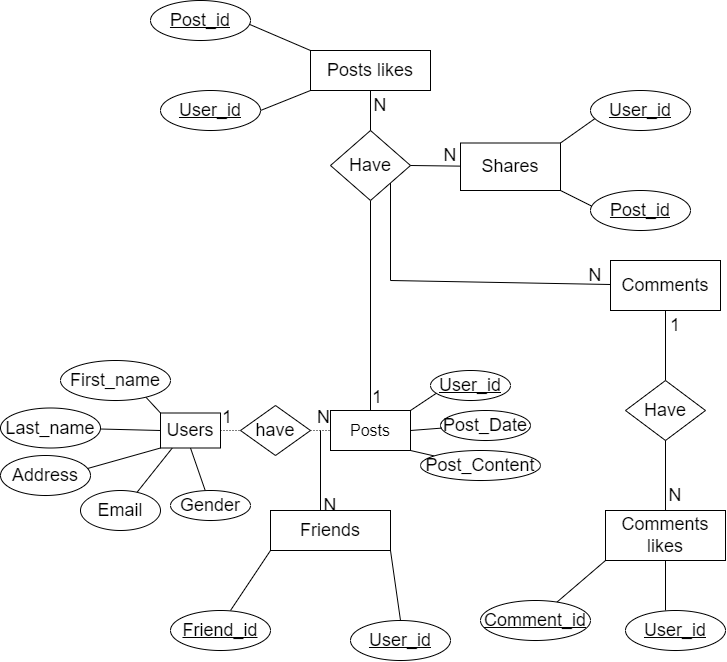
<https://www.linkedin.com/jobs/search/?currentJobId=3264584071&geoId=103644278&keywords=data%20modelling&location=United%20States&refresh=true>

**2.** 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 any one **(a)** social media company such as Facebook, Twitter, LinkedIn, Reddit, etc and any one **(b)** transportation company such as VTA, BART, Amtrak, etc

1. The most popular photo-focused social network now is Facebook. Users have the option of sharing data publicly or privately. Additionally, Facebook gives its users the ability to post content on a variety of social media websites, including Instagram, Twitter, and Tumblr. Facebook has various groups where one can join and connect with people in the group.

The Entities in ER Diagram are Friends, Users, Posts, Posts likes, Comments, Shares and Comment likes.

* One post can have any number of likes so it has 1:N relationship and attributes respectively.
* A user can post any number of posts and it has 1:N relationship between them
* A post can have any number of comments(1:N) and comments can have any number of likes(1:N)



1. transportation company such as VTA, BART, Amtrak, etc

* Entities for VTA transportation includes: VTA, Transport\_office, Registered\_customer, Route\_Details and Base Station.
* Attributes for respective Entities includes:

VTA (cost, capacity, vta\_id, number\_plate)

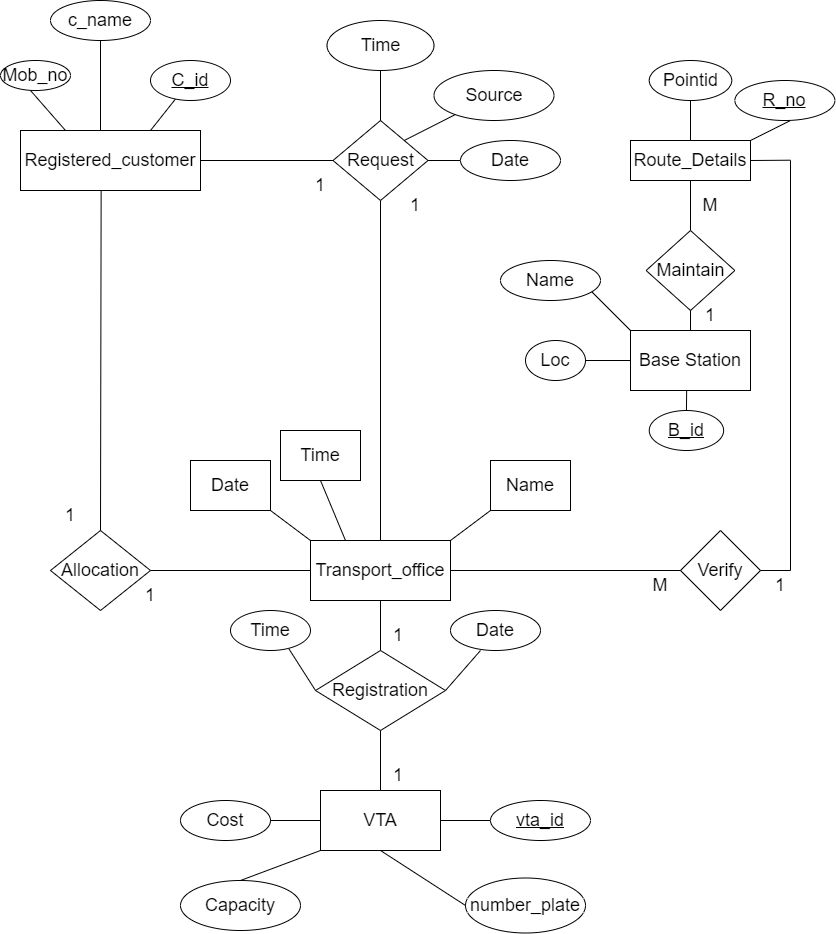
Transport\_office(Date, Time, Name)

Registered\_customer(Mob\_no, c\_name, c\_id)

Route\_Details(point\_id, R\_no)

Base\_station(Name, Loc, B-id)

* A VTA can be registered in only one office so it has 1:1 relationship
* A Registered customers will be allocated to one Transport office
* A base station has a many number of routes to it and it has 1:N relationship



**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

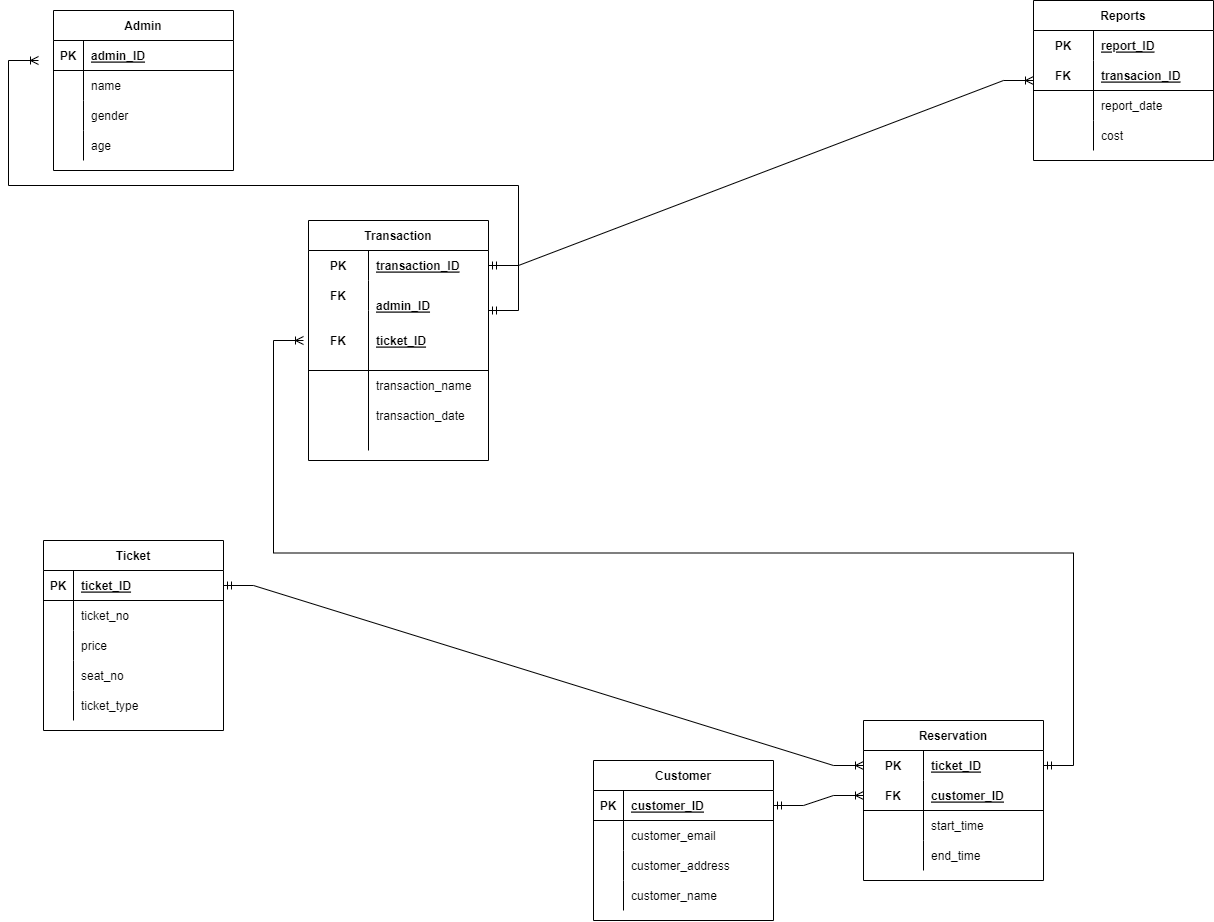
**MOVIE TICKET BOOKING**

Entities for Movie Ticket Booking includes:

Admin, Transaction, Reports, Ticket, Customer and Reservation

An admin has only one admin ID and also the Customer can have only one Customer ID

Transaction has transaction\_name and transaction\_date.



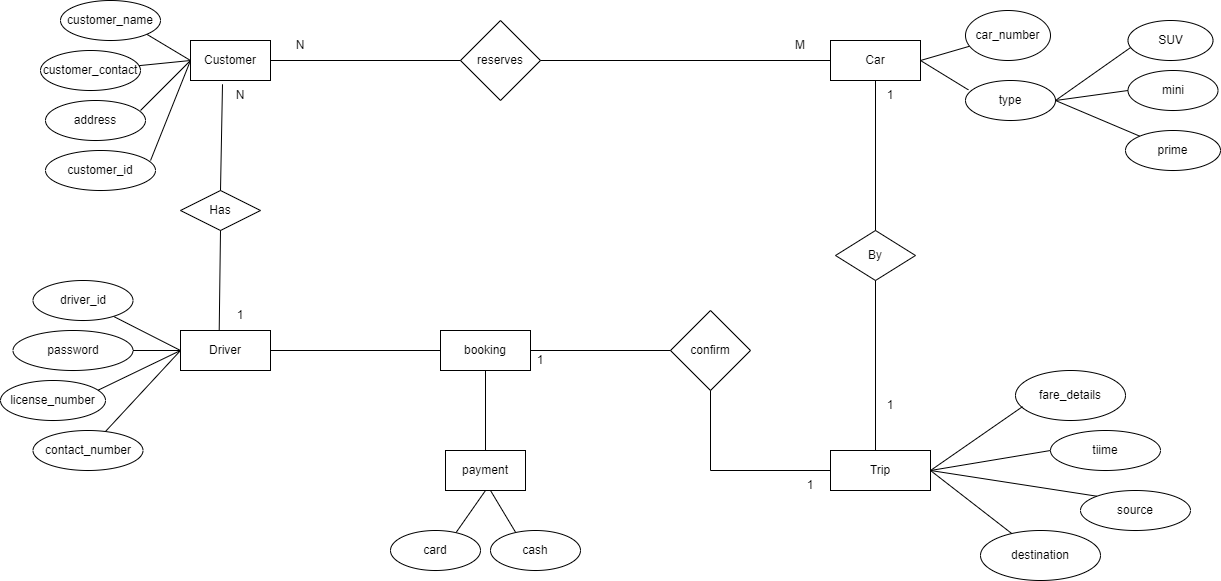
**OLA CAB BOOKING**

Entities for OLA Cab Booking includes: Customer, Driver, booking, payment, Trip and Car

A driver can have any number of Customers so it has 1:N relationship

Booking confirms the Trip so it is 1:1 relationship.

Any number of customers can reserve any number of car so it has N:M relationship.



**4**.

Follow the instructions at [https://dev.mysql.com/doc/refman/8.0/en/installing.html](https://dev.mysql.com/doc/refman/8.0/en/installing.html )to install MySQL and then in README.md at <https://github.com/datacharmer/test_db> to import the employees database into your MySQL instance. Examine the tables in the Employees database using the commands such as the following.

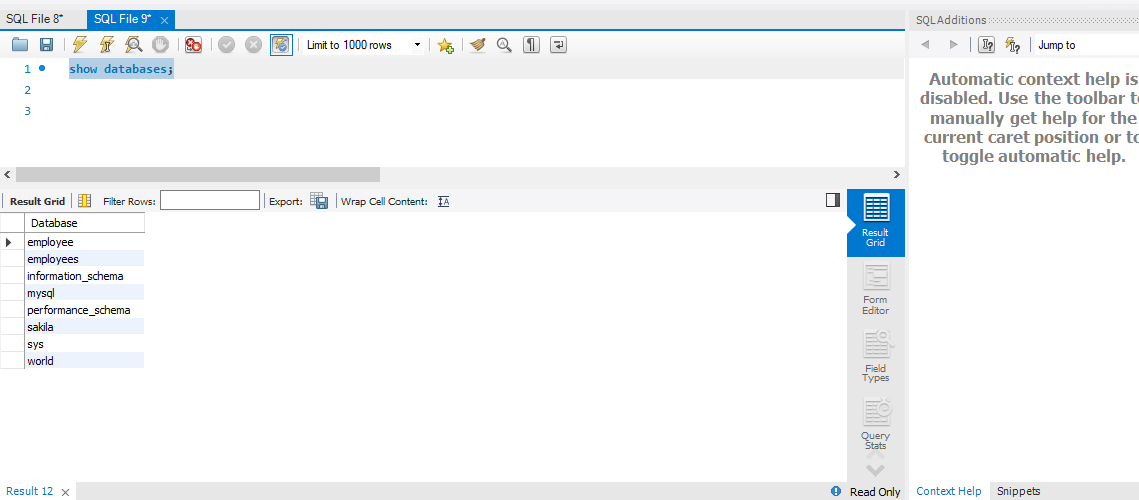
            show databases;

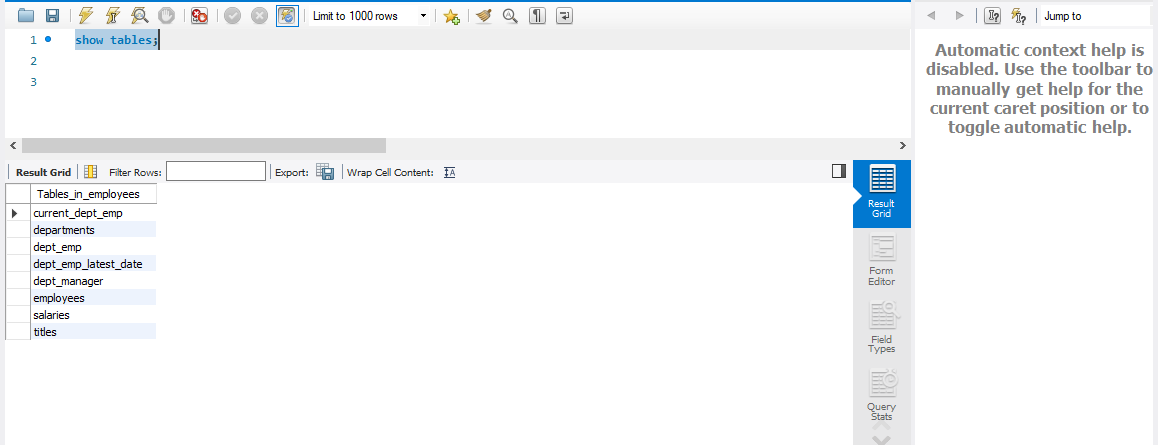
            use employees;

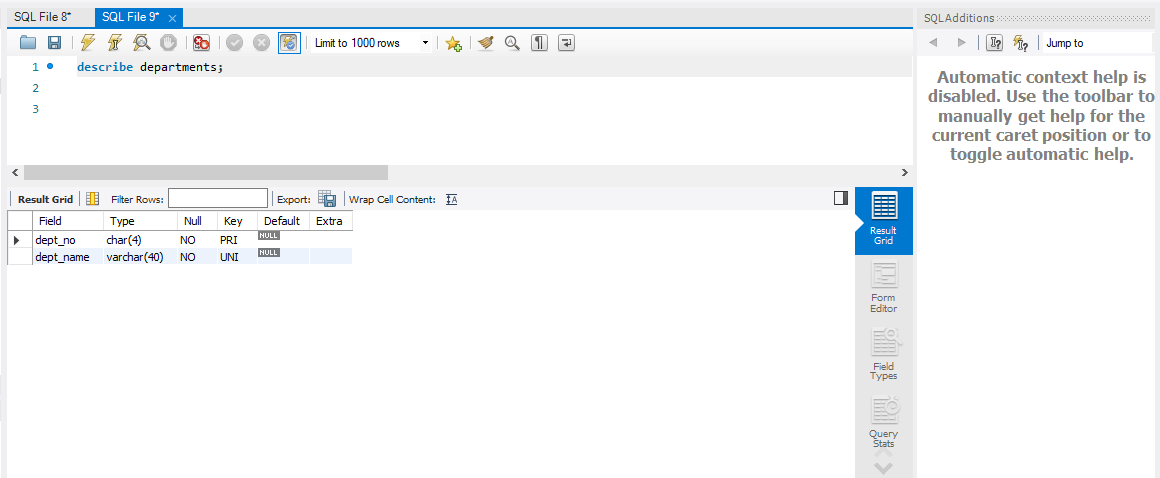
            show tables;

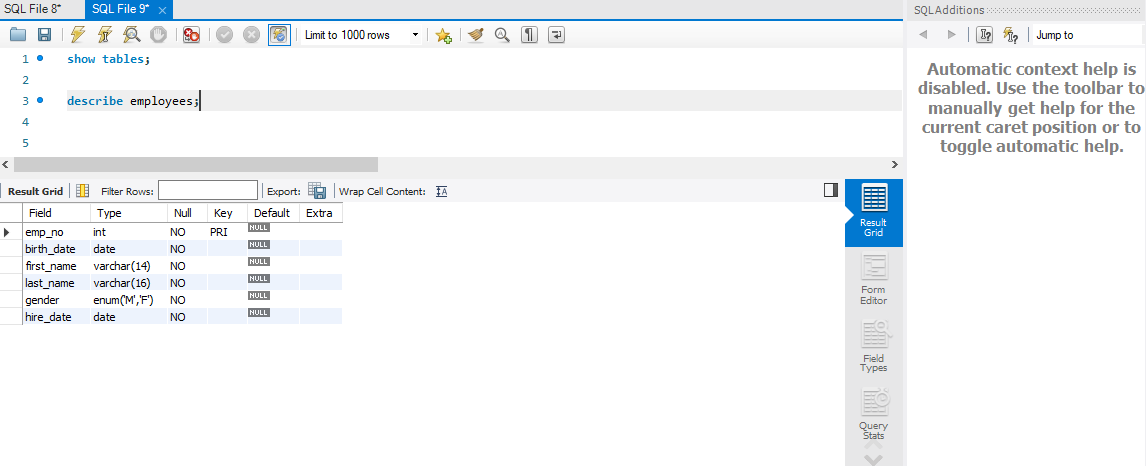
            describe departments;

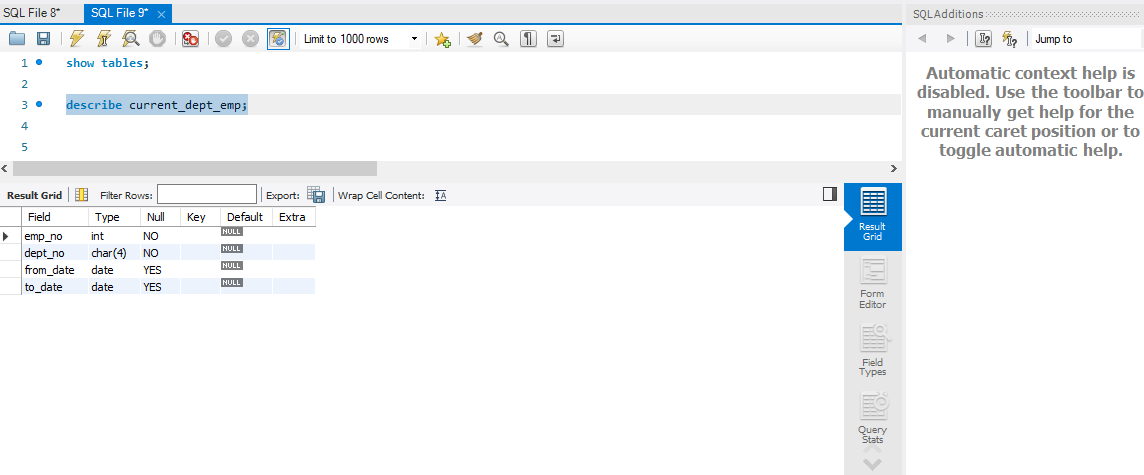
Replace “departments” in the last command with names of the other tables. Based on the output of the above commands, model the data in the Employees database by drawing an ER diagram (using draw.io, Visio or other software), to the best of your knowledge.

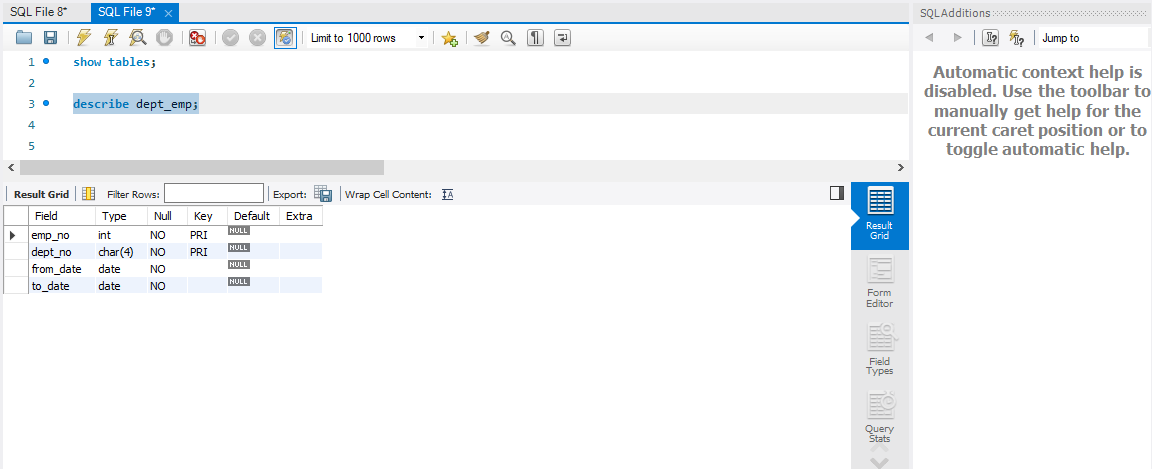


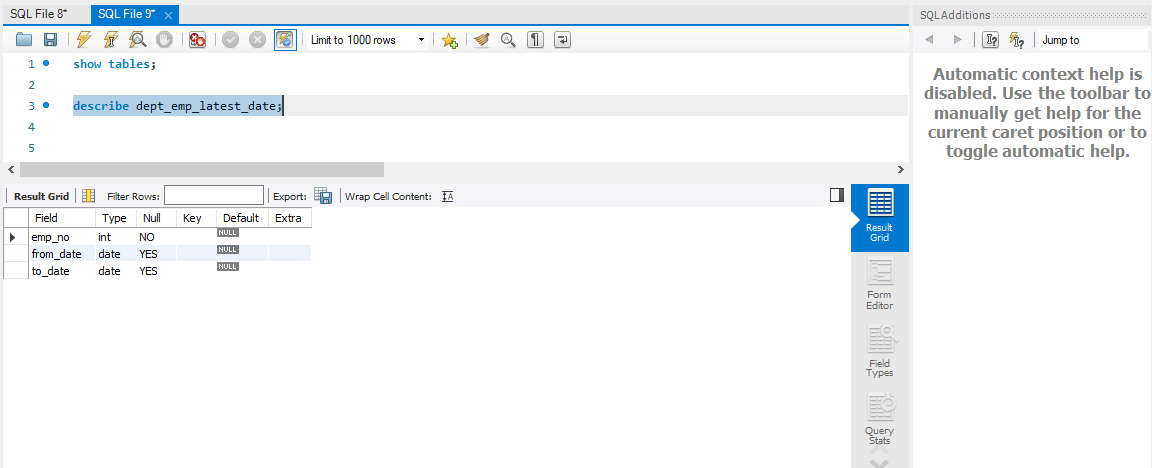


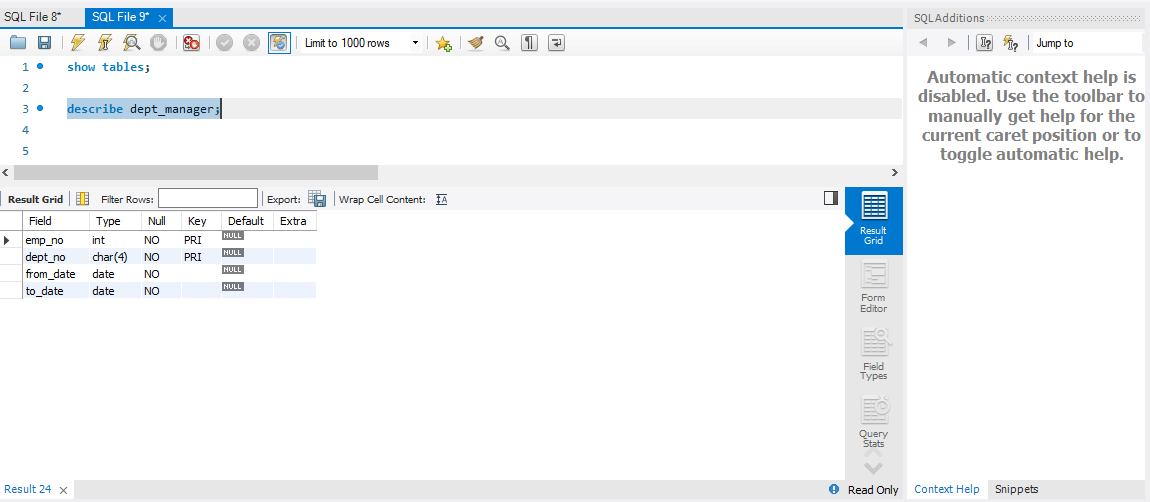


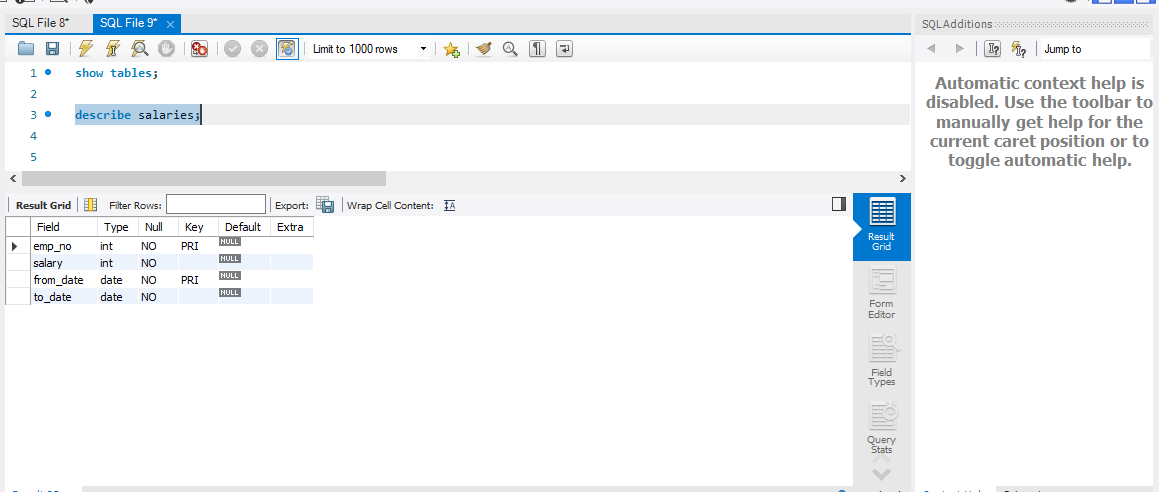


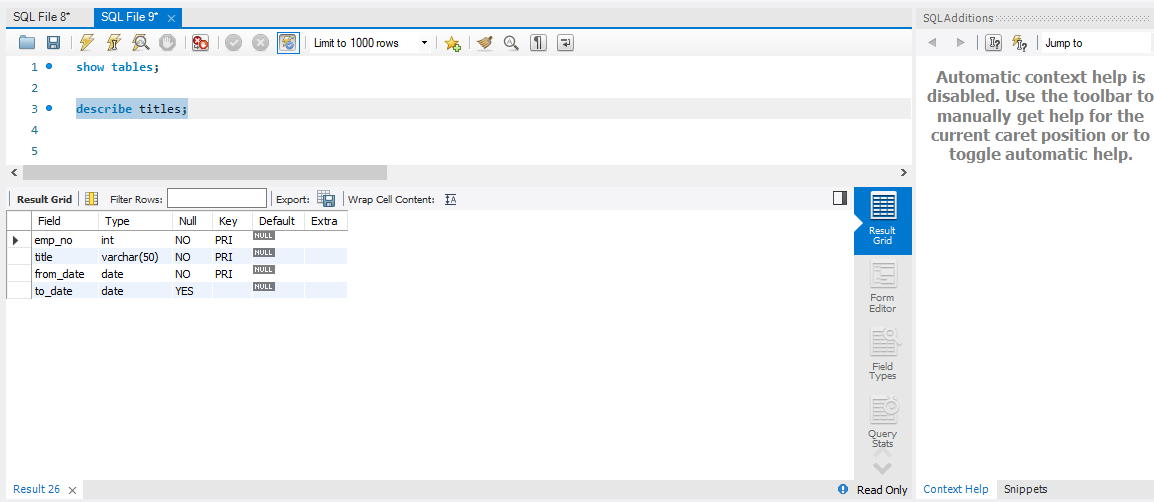


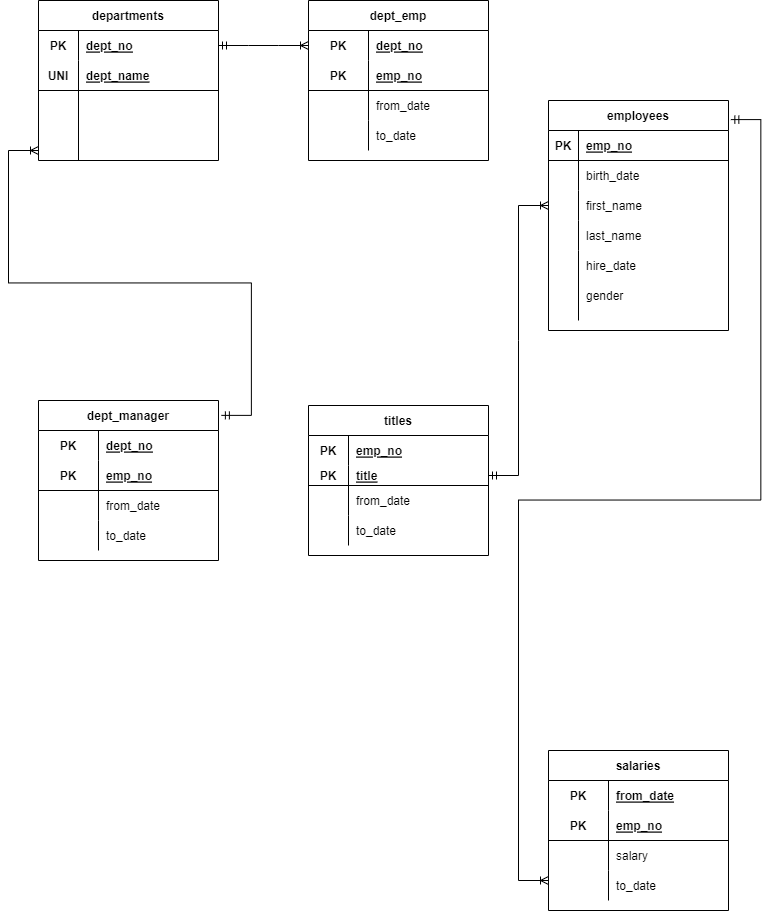




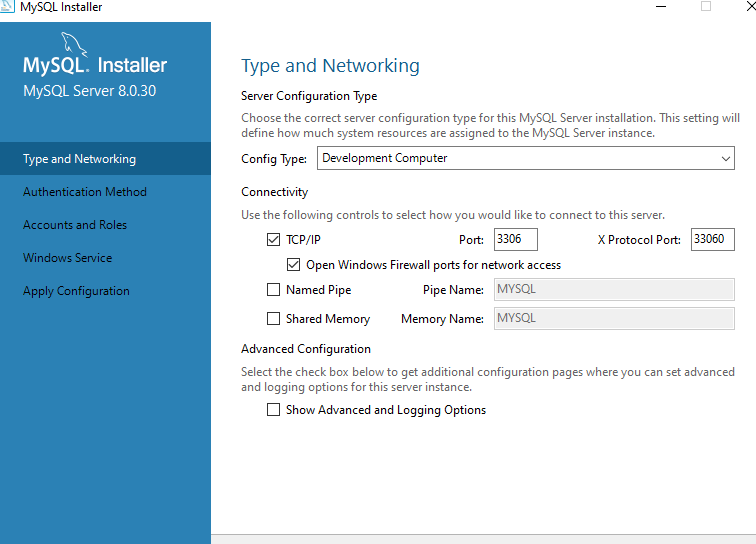


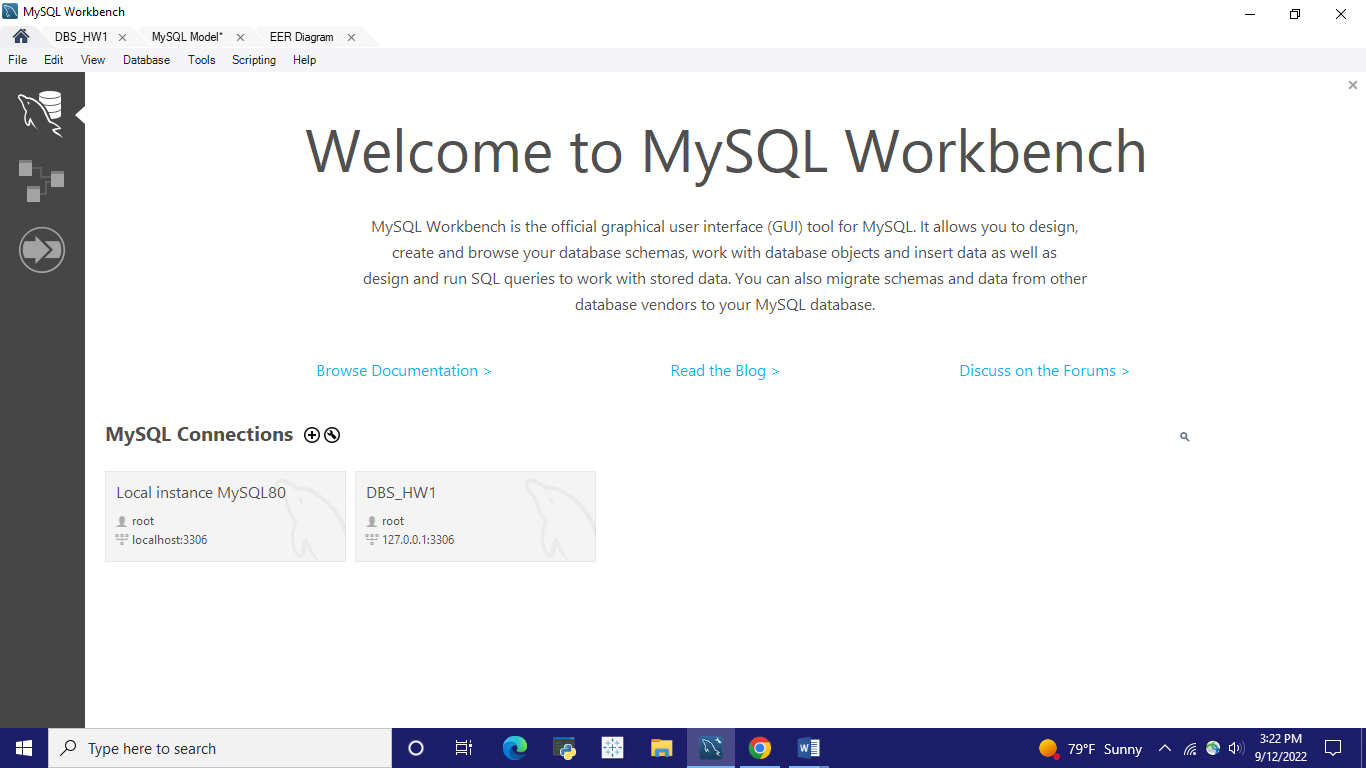


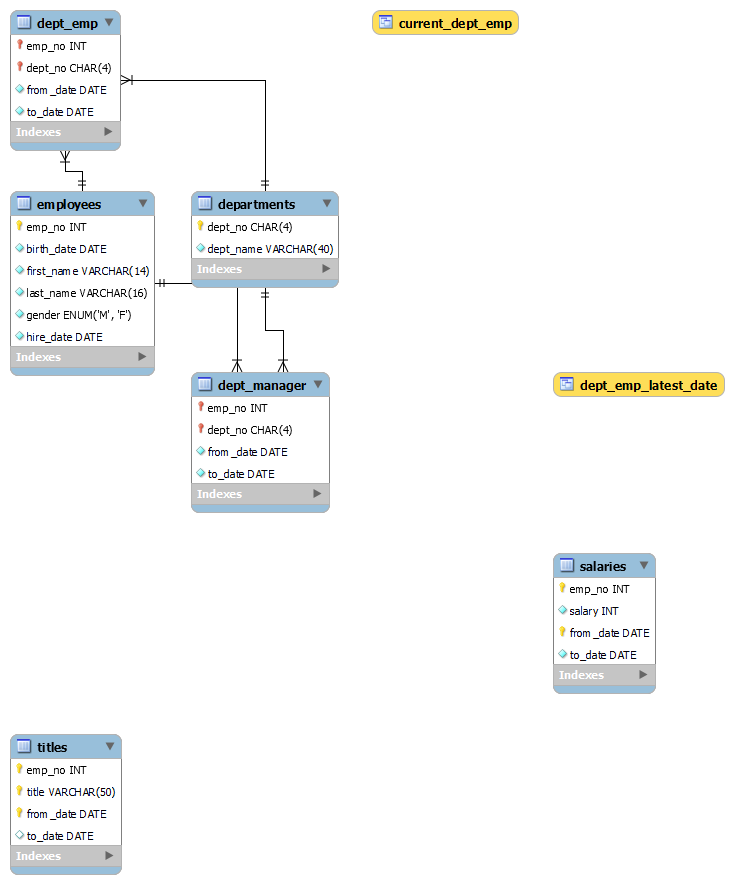




**5.a.** Install the appropriate version of MySQL workbench for your system from <https://dev.mysql.com/downloads/>. For instance, if your MacBook is running Mojave (10.14), install 8.0.14 version of the workbench. Higher versions may not work on Mojave. Complete the initial setup to associate the employees 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.







**5.b.** 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.

* Reverse engineering uses SQL text scripts to build models.
* The column width is automatically adjusted by pressing the "Refresh" button to fit the longest string in one of its cells. Additionally, one can manually change the column width.
* One can switch from an ODBC database to MySQL using My SQL Workbench.
* Only MyISAM tables are affected by the AUTO INCREMENT and delayed key update behaviour.
* The Item Info and Session subtabs in the Information area offer read-only details about a selected object and the active connection.