12-741: Data Management
Assignment#3

Instructor: Mario Bergés
TA: Rami Ariss
Due on: Sunday November 14th 11:59pm ET

November 8, 2021

Some notes before you begin:

When answering the following questions, please provide all of your calculations to arrive at the answer (in addition to the answer itself). Your calculations should be very clear and easy to understand. They should include your assumptions, and a step-by-step explanation of how you arrived at the solution. Also, make sure you type your name and AndrewID on the top of each page.

Some final recommendations:

- Before finding the answer to each question or looking at the next step in the solution, take some time to think about how you can come up with this on your own.

- Again, make sure you document everything you do, and not just write down the answer to the question. This will both help during grading as well as improving your learning process.

- Do not write down any solution or process that you do not understand. If you feel that you do not understand how to do something, seek some help. The preferred method for this is to post your questions on the discussion board for the course, in Canvas and/or Piazza.

1 Designing and Implementing your First Database

Suppose that you are required to develop a simple database for a construction design firm, using the following information as a starting point for your design:

- The database can be used by multiple projects, and each of these projects is composed of design items.

- The organization you are working under has several departments, namely: Architectural Department, Structural Department, Environmental Department, Land Development Department, etc.

- One or more departments may participate in any single project. Any department can participate in several projects at once.

- A department may employ many employees but each employee is employed by only one department

- Each department is managed by only one of its employees.

- The organization has a training program in which an employee may manage (or train) many other employees, and each employee is managed (or trained) by one employee.

- During the implementation of a project, project documents (reports or drawings) will be generated.

- One document is produced by one or many employees and an employee can produce several documents.

- A design item consists of multiple documents.

- The developed database can provide the user with the name of a project, its start date, its end date, contract price, design fee for each design item, the involved departments, basic information about the employees involved in the project, the number of employees in the organization, the number of Licensed Professional Engineers, etc.

1.1 Entity Relationship Diagram (40%)

Develop an Entity-Relationship diagram for the problem described above. Your E-R diagram should have several weak entity sets, one relationship with roles, many types of relationships (one-to-many, many-to-many, one-to-one) and one entity set that contains subclasses in an "isa" relationship. Please specify any additional assumptions you make while building this E-R diagram. You can use any software you wish to develop the diagram, including Microsoft (Powerpoint, Visio, Publisher), or resort to a neat hand-drawn figure that you later digitize and include in your submission.

1.2 From E-R to Relations (20%)

Based on the previous E-R diagram, and without resorting to SQL, generate a list of all the relations and their schema. Please use consistent notation here. For example, if one of your relations is named "Organization" and has attributes "Name" (the key), "Address" and "Phone Number", then you could write down:

```
Organization(Name, Address, Phone Number)
```
Make sure to generate relations and schemas for all entity sets and relationships in your E-R diagram.

1.3 Moving to SQL (20%)

1. **Set up MySQL environment:** Firstly, install MySQL following this link. Then (optionally), install a GUI such as MySQL Workbench. Finally, create a connection to a local database server, following this. Include a screenshot in your assignment to show your have set-up the environment successfully.

2. **Create a database:** Create a database in MySQL based on the relations you developed in the previous question. Include also additional SQL queries to `INSERT` sample data into each of the tables that you created. Save all the queries in a single *.sql script and submit it with your assignment. (The graders will run your *.sql script to validate your code.)

2 Revisiting Theory (20%)

2.1 Sets (10%)

Which of the following formulas are true for arbitrary sets A, B, C? (You can use Venn diagrams to support your answers):

1. $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

2. $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

3. $(A \cup B) \setminus C = (A \setminus C) \cup B$

4. $(A \cap B) \setminus C = (A \setminus C) \cap B$

5. $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$

6. $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$

2.2 Relational Algebra (10%)

Assume you were to write queries for the database you created in the first part of this assignment. Write down the relational algebra expression for the following queries:

1. Find the name and licensing status (i.e., whether he/she is a licensed professional engineer) of the epmployees working on projects since 2012

2. Find the project names for those projects that have fewer than 10 employees associated with them