Homework Set 4 - Lou's List

Purpose: This week is about answering the question, "How should we structure our data?" In this assignment you will practice answering this question at four layers of abstraction (view, conceptual, logical, physical) as you build a relational database. In doing so you will gain proficiency with the skills necessary to go from a goal or problem statement to an implemented database. **N.B.** This assignment draws heavily from what you learned about driving MySQL databases in SYS 2022.

Instructions: Put your solutions for all problems in a google colaboratory notebook. Create a level-1 header for each problem and then present the solution in subsequent cells. The header must contain the problem number followed by "confidence x" where x is an integer from 1 to 10 with 10 representing high confidence. For full credit you must show your work. The submission is made via a link on UVACanvas. **N.B.** This set includes SQL code. You will run that locally as taught in DS 2022. Then take your code and include it in a code block in a text cell in your colab notebook.

Domain Extras: This week the data set is one you are already very familiar with, the course catalog at UVA. As a result you are already an expert on the domain.

Data

The data for this week is the course catalog of UVA. You already have access to this data in a variety of ways. One of the most popular is Lou's List `https://hooslist.virginia.edu/`. You are free to acquire the data in whatever way you prefer.

4.1 Data Preparation - Logical Level (10ψ)

Draw an ERD for your database. Your diagram must include:

- Entities with attributes for: Classes, Students, Professors, and Rooms
- Primary keys (clearly indicated as such)
- Foreign keys (clearly indicated as such)
- Relationship cardinality shown with crow's foot notation

You may use any diagramming tool (draw.io, Lucidchart, etc.). Hand-drawn diagrams are not accepted.

4.2 Data Preparation - Physical Level (20ψ)

4.2.1 Database Creation

Write MySQL code that implements your schema. Your code should:

- Create the database
- Create all tables with appropriate data types
- Define primary keys and foreign keys

4.2.2 Populate the Database

Populate your database with real course information. Your code should:

- Insert data into all tables
- Include at least 20 records in the classes table and at least 5 records in all other tables

4.3 Data Query (10ψ)

Write SQL queries to solve the following problems for different scenarios.

N.B. If a scenario cannot be accomplished with your current database schema, that is okay. State that the query cannot be handled and explain what is missing. The next section (Design) is where you will address these limitations.

4.3.1 Student: Finding Classes

A student is looking for classes to take next semester. Write a query that allows a student to search for available classes. Your query should:

- Filter by subject (e.g. DS, PHYS, CHEM, etc.)
- Show course name, meeting time, location, and instructor

4.3.2 Professor: Teaching Schedule

A professor needs to see the list of classes they are assigned to teach. Write a query that:

- Returns all classes assigned to a given professor
- Includes course name, meeting time, and location

4.3.3 Registrar: Room Conflict Detection

The registrar needs to ensure there are no scheduling conflicts. Write a query that:

- detects if two different classes are scheduled in the same room at the same time.
- Demonstrate it works by intentionally inserting a conflict and showing your query detects it.

4.3.4 Registrar: Professor Conflict Detection

The registrar needs to ensure there are no scheduling conflicts. Write a query that:

- detects if a professor is assigned to teach two different class sections at the same time.
- Demonstrate it works by intentionally inserting a conflict and showing your query detects it.

4.3.5 Dean: Cost to Deliver

A dean needs to calculate the cost to deliver a class. Write a query that:

- Calculates the total cost for a given course based on instructor salary, enrollment
- Computes cost per student for the class

4.4 Press Release (10ψ)

Chose one of the scenarios from the Analysis section (Student, Professor, Registrar, or Dean). In the spirit of Lou's List create a new product powered by your database. Write a press release to your chosen audience. Make sure it includes:

- Headline
- The problem it solves
- Statement of why a database is necessary
- Visualization in support of the statement
- Quote from satisfied customer

4.5 Apply in new context (10ψ)

Apply what you learned about relational database design to a different aspect of UVA student life. Choose one of the following domains:

- Dining (meal plans, dining halls, menus, etc.)
- Transportation (busses, scooters, parking, etc.)
- Housing (dorms, room assignments, RAs, etc.)
- Intramural Sports (teams, schedules, standings, etc.)
- Student Organizations (clubs, members, events, etc.)
- Library System (books, checkouts, study rooms, etc.)

For your chosen domain:

1. Draw an ERD with at least 4 entities and relationships
   - Identify the primary keys, foreign keys, and cardinality (using crow's foot) for each relationship
   - You may use any diagramming tool (draw.io, Lucidchart, etc.). Hand-drawn diagrams are not accepted.
2. Describe two different personas who would query this database and provide two example questions each of them would ask.

4.6 Review: Definitions (20ψ)

Provide definitions, or describe the distinction, for the following words:

1. Abstraction Level
2. Normal Forms
3. ERD
4. Schema
5. Primary Key vs. Foreign Key
6. SQL vs. MySQL
7. SQL vs. NoSQL
8. Entity vs. Attribute
9. Record
10. Query
11. Relational Algebra

12. Cardinality
13. The many to many problem
14. Junction/Bridge Table

4.7 Review: Multiple Choice (20ψ)

Instructions: Select the best answer for each question. Do not show work.

4.7.1 In the ANSI/SPARC three-level architecture, the level that describes how data is actually stored on disk is the:

(A) Conceptual level

(B) External level

(C) Logical level

(D) Physical level

4.7.2 The external (view) level in database abstraction is primarily responsible for:

(A) Defining the global schema for the entire database

(B) Describing the physical representation of data

(C) Providing relevant information to different users

(D) Managing index structures and file organization

4.7.3 The cardinality constraint "one-to-many" in an ER diagram means:

(A) Each entity instance participates in exactly one relationship instance

(B) One instance of entity A can be associated with many instances of entity B, and vice versa

(C) One instance of entity A can be associated with many instances of entity B, but each B is associated with at most one A

(D) The relationship is optional on both sides

4.7.4 Which of the following is a common problem when a relational schema is not properly normalized?

(A) Improved query performance

(B) Update anomalies: changing one fact may require updating many rows

(C) Fewer joins are needed

(D) Simpler constraint enforcement

4.7.5 Choosing a natural key (e.g., email) over a surrogate key (e.g., auto-increment ID) can be problematic because:

(A) Surrogate keys are always faster

(B) Natural keys may change or not be available at insert time, and updates can cascade

(C) Natural keys cannot be used in joins

(D) Surrogate keys take less storage

4.7.6 The primary purpose of creating an Entity-Relationship Diagram (ERD) before building a database is to:

(A) Make the database run faster

(B) Visualize and plan the structure, relationships, and constraints before implementation

(C) Satisfy a requirement for documentation

(D) Generate SQL code automatically

4.7.7 Database normalization is the process of:

(A) Making all data values fall within a normal distribution

(B) Organizing data to reduce redundancy and improve data integrity

(C) Converting all text to lowercase

(D) Ensuring all tables have the same number of columns

4.7.8 Many-to-many (M:N) relationships require a junction table because:

(A) SQL databases cannot store arrays in a single cell

(B) It improves query performance

(C) A foreign key in one table can only reference one row in another table

(D) It is required by the ERD standard

4.7.9 When designing a database that serves multiple user personas (e.g., students, professors, registrars), the best approach is to:

(A) Create a separate database for each persona

(B) Design the schema to store all necessary data and create different queries for each persona's needs

(C) Only design for the most important persona

(D) Let each persona design their own tables

4.7.10 Which of the following is NOT a valid reason to add a new table to an existing database schema?

(A) To eliminate a many-to-many relationship

(B) To store a new type of entity with its own attributes

(C) To make a query run in fewer lines of code

(D) To reduce data redundancy

4.7.11 This homework set is:

(A) Too short and Too easy

(B) Too short and Too hard

(C) Too long and Too easy

(D) Too long and Too hard