

Introduction

The purpose of the final project is to provide you with hands-on experience in the design, creation, development, and maintenance of a relational database. This project provides a realistic experience to apply your knowledge and skills you learn in class. Given these general goals, you are therefore given a wide variety of freedom in choices of data domain and interface technology. I want to encourage you to choose projects or data that are meaningful to you.

Requirements

Architecture

- Server/Client database system
 - Relational database using MySQL (or another DBMS with my permission)
 - Front-end interactive interface (e.g., PHP)
- Teams of Three
 - Exception: Group of two is allowed iff both are graduate students
 - Exception: Group of four is allowed iff one student is not a CS major

Data Sources

You may choose your own data sources. A significantly large dataset is required. Exceptions to this rule are made by permission for real-world problems you may wish to solve. When searching for data sources, consider how much work is needed to format and pre-process the data. Remember that the goal of this project is a relational database, therefore you should select your data appropriately. When considering a dataset, ask what types of queries the users of your system might be able to make.

Assignment

The goal of this assignment is to allow you to make progress on your class project in a series of stages. Please submit the different stages of the proposal into the appropriate D2L dropbox folder by 11:59 PM on the due date listed below:

- Project proposal due 9/28/19
- Project report 1 due 10/12/19
- Project report 2 due 10/26/19
- Project report 3 due 11/09/19
- Project report 4 due 11/25/19
- Project presentations on 11/25/19, 12/2/19 and 12/4/19

Although the stages of the project are due about two weeks apart, the amount of work required in later stages increases – please plan your time accordingly.

Your submissions should include the names of all teammates. Each team member will submit the same work. Each team member will also submit a text file with the relative contribution percentage to a given assignment made by all team members. This last requirement is to ensure that all the members of their team contribute to the project. Consistently low contribution will result in lower grade for the lagging team member.

Project proposal

Your proposal will consist of 1 – 2 pages describing your project. Use Chapter 1 as your guide.

Guidelines

1. You should specify:
 - The problem you plan to solve
 - Outline your approach to solving it
 - Explain your plan to acquire and pre-process (scrub) your data
 - Discuss the types of queries users could make to your system
 - Describe the implementation of your database and your interface
2. Although I want to give you a free hand in the problem and application you will choose, please be cognizant that data can be sensitive. Projects that propose or enable unethical use of data will not be allowed.

Project report 1 – ER diagram

Draw an ER, or EER, diagram for your database design. You may use any software tool of your choice, but handwritten ER diagrams will not be accepted. Use Chapters 3 and 4 as your guide.

Guidelines

1. Make sure to indicate:
 - keys,
 - all attributes,
 - composite attributes (if any),
 - meaningful derived attributes (if any),
 - relationship attributes (if any),
 - weak entities (if any),
 - cardinality constraints, and
 - participation constraints.

Hint: Consider twelve common ER diagram mistakes from <http://ceur-ws.org/Vol-848/ICTERI-2012-CEUR-WS-paper-15-p-222-227.pdf>.
2. If there are extra constraints that cannot be captured by the ER diagram, make sure you list them in supplementary text.
3. List any assumptions you make in the process. Your ER diagram should contain at least 6 (or more) entities; otherwise it is likely not of sufficient complexity for a CSCI 440 project.
4. Make sure your diagram captures a significant number of relationships, and all those required to accomplish your project goals.
5. For each entity set and relationship, write a short description in plain English of what it represents or models. One or two sentences per entity set and relationship is enough. These descriptions are primarily to help me understand what you are modeling and ensure that you are modeling it correctly.

Project report 2 – Relational schema and normalization

Convert your ER (or EER) diagram into a relational database schema diagram. Use Chapters 9 and 14 as your guide. Your diagram should be similar in style to Figure 9.2 (p.291) in the text. Do not overly complicate the task by using other diagram types (e.g., UML). Develop your diagram using Visio, or another drawing tool – handwritten diagrams will not be evaluated.

Guidelines

1. Transform the conceptual data into normalized relations:
 - Represent entities
 - Map multivalued attributes
 - Map weak entity types
 - Map all relationship types

- binary 1:1,
 - binary 1:N, and
 - binary M:N
 - n-ary relationships
 - Normalize the relations to 3NF (see Chapter 14)
 - Merge the relations
2. Ensure well-formed relations:
 - minimum amount of redundancy
 - permits users to insert, modify and delete the rows without errors or inconsistencies
 3. Identify all primary keys:
 - value of the key must uniquely identify every row in the relation
 - key should be nonredundant
 4. Ensure referential integrity is achieved (i.e., no dangling references)

Project report 3 – SQL and sample data

For this document, you will submit your data definition commands and examples of data in from your database. Use Chapters 6 and 7 as your guides.

Guidelines

1. Format data definition as SQL commands stored in a text document. Include a **CREATE SCHEMA** command and a **CREATE TABLE** command for each relation. If you use a framework to assist the interaction with SQL, do not submit any non-SQL code. You must generate (either programmatically or by hand) the underlying SQL schema commands and submit only those.
2. For each table created, you should include:
 - name (be descriptive) of each attribute
 - type of each attribute (consider the storage space implications of each choice)
 - attribute constraints (e.g., **NOT NULL**, uniqueness, default value)
 - primary key
 - foreign keys (all of them)
 - referential integrity constraints
 - **CHECK** clauses (if needed)
 - short comment linking the table creation command to the entity name of your ER diagram (use **#** character to comment)
3. Submit sample data for each and every one of your tables. For each table, write a query that returns any 5 records. Use the command **LIMIT 5** to control the result count. You should return all attributes and the default attribute names using **SELECT *** syntax. You may merely choose to return the first five records, or you use the **WHERE** clause to specify a more specific and interesting query. If accessing your database from the terminal, you should copy and paste the result of each query into a plain text file. If you are using PhpMyAdmin, you should export the results of your query into a text file (it is one of the options). For both approaches, you should collect the result for each query (one per table) into a single text file for submission.

Project report 4 – Transactions

For this document, you will submit five example transactions (use cases). These transactions should demonstrate interaction with your database as used in your project.

Guidelines

1. For each transaction, you should provide:
 - English language query
 - SQL query

- result set (`LIMIT 5`), including table headers

In other words, provide five example queries (both in English and SQL) and the result set limited to five rows.

2. These transactions should be representative of how your client might use your project. You should provide a diverse set of queries that highlights five different use case scenarios. If your queries are deemed too similar to each other, you may not receive full credit for this technical document.
3. Along with each transaction include a brief description of its purpose within the overall functionality of your project.
4. You do not need to give the high-level code (e.g., PHP, Python) used to assemble your SQL statements. If using prepared statements, please submit the only resulting SQL statement after your variables have been assigned.

Project Presentation

Each team will prepare a single powerpoint slide describing your project. Your slide should introduce and motivate the problem your project attempts to describe, discuss the current state of research and solutions in this area as related work, detail the approach of your solution. You will also demonstrate a working version of your system.

Do not leave the writing and design of the slide until the last minute. Slides should be well formatted, proofread, and spell-checked.

You will submit your final poster through a D2L dropbox.

Teams will present their project slide along with a short, working demo of the system in class over two days. You will have 5 minutes for your presentation. If any questions arise as to the correctness of the implementation, I may request individual demos during my office hours to complete the grading of the project.

Acknowledgements

Parts of this material were adopted from Mike Wittie's and Patrick Donnelly's CSCI 440 courses.