# CS486/586 Introduction to Databases

## Spring 2022 Quarter

---

## Database Implementation Project

## Due Dates:

1. Domain Description and Data Source

   Thursday, May 4th 2022, before class

2. ER Diagram and Relational Schema

   Tuesday, May 16th 2022, before class

3. Final Write-up with Queries and Results

   Friday, June 1st 2022 by 5pm (America/Los_Angeles)

4. Demo and live coding

   Scheduled during finals week

You may do this assignment individually or you may work with a group of no more than 5 people. If you work with a group, you should turn your assignments jointly, with all of your names on the paper. You should only talk to the instructor, the TA and your partner(s) about your project. You may also post comments and questions on the class Slack channel.

Please turn in the first two submissions as a PDF to Canvas. The third one should be submitted either as a PDF on Canvas or as a link to a web page.

## Project Overview

The goal of this project is to gain experience with database design and implementation. You need to choose a small domain (for example, food carts or campus sports teams) for which to implement a database, along with populating it with real data and producing example queries to execute over it.

## Submissions

There are three submissions for this project and a final meeting

**1. Domain Description and Data Source:** Describe in English the domain you intend to build a database for, along with 20 questions (in English) that someone might want to ask about the domain. (You will be permitted to revise these questions later.) Also describe what source you intend to use for data, and *how you intend to ingest the data into your database*. You should choose a domain where you can easily get several hundred rows of data. However, your source ***should not*** be an existing relational database. Excel spreadsheets from the Internet (or the like) is ultimately what you're looking for here.

**2. ER Diagram and Relational Schema:** Produce an ER diagram in UML for your domain, and its translation into a relational schema, including all keys and foreign keys. You should aim for a database with 6 – 10 tables. You should also submit evidence that you have created at least one table from your schema and populated it with at least one row.

**3. Final Write-up:** You are to implement your schema in a relational database and populate it with data. The preferred DBMS is Postgres. If you want to use a different DBMS, you must get advance approval. You then need to fully populate your database, translate your 20 questions into SQL, and execute them on the DB. Your write-up of these activities should include the following:

- Your schema diagram, showing any changes you made during the implementation process
- The CREATE TABLE statements for your database
- A brief description of how you populated the database and any code used for ETL

- For each of your 20 questions, the question in English, its translation to SQL and the (full) answer to the query. (If you needed to change any of your original questions, also list the originals and why you needed to change or replace them.)
- A listing of the contents of all of your tables

Please submit your write-up as a single PDF or as a link to a web page, which can in turn link to the different parts of the assignment.

**4. Demo and Live Coding:** You and your group will meet with me and Saba for 15-20 minutes during which we will all write and run some queries on your DB and talk about your experience on the project.

## Grading

All four parts of the project combined make up 30% of the final grade for the class.

Submission 1 is worth 10 points.

Submission 2 is worth 20 points. Part of the score will be on how realistically your design represents the corresponding real-world domain.

Submission 3 and the live demo together are worth 70 points.

The score will include the following factors, in addition to general correctness:

- Were you able to avoid large amounts of manual data entry with your approach to ETL?
- Does the data used do a good job of demonstrating the adequacy of your schema and the correctness of your queries?
- Does your set of queries make use of all the parts of your schema and a range of SQL features?

The points breakdown for submission 3 is
I. Readability and completeness: 15
II. Data entry and suitability–description and automation: 10
III. Sufficient tables and rows: 5

IV. Coverage of schema and SQL features: 10
V. Queries – correct translation of English: 30 (Note: You can lose up to two points on each missing or incorrect query.)
VI. Live demo – how comfortable is each member of your group navigating and querying the database you've built together? How well does each person understand the data and the results of the queries?