Final Project for Graduate Students of CSCI E59

Proposal Due: May 1 11:59 PM ET (Note: NO late days are permitted)
Project Due: May 11 4:59 PM ET (Note: NO late days are permitted)

Note: The Final Project is only for students enrolled for graduate credit. If you are enrolled as non-credit or undergraduate, you will not be graded for this Final Project.

Goals

The goal of the final project is to design and implement a database using a dataset not previously used in this course. The key objectives are to:

- identify the purpose of your database:
    - what information is your database responsible for storing?
    - what will be the most frequently used functions of your database? For example: handling account transactions for specific types of records, or answering specific questions about a dataset
- select the right database for your use case (Relational or NoSQL)
- design an efficient schema
- show meaningful queries from your database
- document and present your work

This is a good opportunity to work on something that is relevant to your job, or that you'd like to showcase in a future job interview. It's also a good opportunity to have a lot of fun working with something new or interesting to you. This assignment is fairly free format, as long as you cover the key points identified above. You may choose to use any database you wish, including ones not previously used in this course.

Data Sources

It should not be difficult to find free data collections online to work with. Do a simple Google search for 'free data sources' and you should find several links such as this one:
https://www.searchenginejournal.com/free-data-sources/302601/
You may use data from whichever source you choose, including data you make up by hand if you choose to do so. We do expect that your dataset is not too sparse however. If using an RDBMS, for example, we'd like to see at least 50 records per table. It is possible for you to populate your database manually, it just means you'll be performing a lot of inserts.

Project Proposal - 25 Points

The first part of the final project is to **write a one page proposal defining your project and its objectives**. You will submit your proposal online as you would any other assignment. You should wait to get approval from the teaching staff before moving forward on your project. The proposal is essentially a pass/fail assignment - if you pass, you may proceed to work on your project. If you fail, you must re-do it until you are given the go-ahead to work on your project. You can submit multiple times and receive full credit up until the deadline. After the deadline you can still submit a proposal for no points.

Your project proposal should do the following:

1) Define the purpose of the database. What data are you going to manage? What are you storing this data for? Some examples of large scale applications you could try to implement:
    - Healthcare customer database
    - Inventory database, similar to our bike store
    - Animals of the world
    - Database that serves as a digital twin of an organization's network/server topology
2) Provide a description of the data. What are your entities, their attributes, what are the significant relationships?
3) Identify which database you are choosing to use for this data, and (very importantly) why are you choosing this particular database? What makes you believe it is most suitable for the data you are going to be working with?

Final Project - 125 Points

The required submission artifacts for your final project are three things:

1) **Final Project Document. (65 Points)**

   Your document is a write-up similar to past homework assignments where you should clearly explain what you did and how you did it. Code snippets should be shared, commands should be listed, screen shots should be shown. Write your document as if you were writing an instruction manual for someone trying to replicate your work. Be very detailed. Your document should also include:

   - overview of what it is your project does (your proposal is probably a great reference for this).
   - Answers to the following questions:
     - How will this database scale as it grows?
     - As more data is inserted over time, how will your system handle the growth?
   - For a relational database: Diagram the schema using a common method. We have used Crow's Foot notation with Creately in-class, but you can use another standard format if you choose.
   - For a NoSQL database it can be a bit trickier since we do not have a standard ER model to follow, and the schema is flexible. However, diagram key entities, attributes and relationships to the best of your knowledge.
   - At least five queries against your data. We are looking for meaningful queries that demonstrate the purpose of your database. Think in terms of what kinds of questions you would like to ask against the information you have stored.

   Your document must also contain links to the following video:

2) **3-5 minute YouTube Presentation. (60 Points)**

A short summary of your project, and results. Summary videos will be shared with the class on the final day of class, and students should make an effort to be present for the live lecture that day as there will be Q&A after each video. Important: ***You will lose points if you are under three minutes, or over five minutes.***

3) (Where applicable) Include Source Code, Configuration Files, Input Data
   All source code, config files, scripts, and input data should be submitted in a separate zip file.

Video creation software can be found for free on the internet, take a look here as a starting point:
https://www.iskysoft.com/video-editing/video-creator-free.html