Project, Phase 4
CS 4347
Fall 2023
Early Turn-In: November 17th, 11:59 pm
Due: November 30th, 11:59 pm

Instructions:
This assignment is a Cohort Assignment. Each cohort will be doing the assignment, and each cohort member will turn in their own copy.

Objective:
To continue through to the design aspect of construction a database for the Database Class. The goal of the project is to build the database, not the database application. To that end, certain phases must be completed. This phase consists of

- Final Design
- Data Generation
- Data Examples
- Query Examples

Bonus:
Turn in this phase of the project before the Fall Break, and 10 points will be added to the total score of part 4 of the project.

Final Design: (15 points)
By this stage all design diagrams and documentation are complete and have no further revisions.

1. the ER diagram.
2. the ER Dictionaries.
3. the Schema Diagram.
4. and the Schema Dictionary.

- This final version has no strike-through or other colors.
- This is your final design and description of your database with no further changes.
- This document should be beautiful.
- All four parts are in one PDF file.

Turn-in:
A ***ready-to read*** copy of your cohort's final document. Must be in the form <cohort>.FinalDesign.pdf

Data Generation (30 points)

In order for a database to function, it must be populated. And the population must adequately large explain the nature and complexity of the database structure. To that end, all the tables in the database must have values entered. For this section, the cohort has to do ***two particular tasks.***

Task 1: Size

The first task is the size of the database. Each database table must contain adequate entries. But how many entries should be in each database table? Depending on the design, the tables can be of different sizes.

If the data for the table is a large population.

- Items store in a warehouse
- Rooms in a skyscraper
- People attending a University.
- People going to work for a company.
- Vehicles being moved to and from a destination.

Then each table should be between 180 and 220 entries in length.

If the data for the table is a medium population

- Managers who oversee a dozen or more employees each.
- A vehicle delivering 50 packages.
- A teacher with a hundred students.

Then each table should have between 18 and 20 entries

If the data for the table is a small population

- Kings and rulers of countries
- CEO or Presidents of Corporations
- The singular actual office building that contains dozens of floors and hundreds of rooms

Then each table would be between 2 and 5 entries.

Task 2: Scripting

How is your cohort entering the data?

- An SQL script-A set of SQL commands would be used to load the information. Each cohort member could write a script for their data section, and use each script to load the database.
- A program-A program can be written, using random number generators to fill in the various attributes of each tuple. Some databases would then allow direct access for the insertion, or more likely, string concatenation can be used to create the INSERT SQL statements and save them to a file. Then that file can be loaded into the DBMS as an SQL script.
- A live stream-Some data, such as stock markets and weather information, is available on an internet socket. It could be that your cohort can use one of these systems to populate the database.
- Data from the internet, properly documented-Various institutions publish a "corpus" of data for data mining and database research. Such material may be used if the source is properly cited.
- Websites such as Mockaroo which can generate large fields of data.

Turn In:

You will be turning in a report called <cohort>.DataGeneration.pdf with the following sections:

1. A statement (about a paragraph/3 to 5 sentences) on why your data sizes are the size they are. Even if the reason is "Because Dr. Becker is a meanie."
2. A listing of all the relational tables in the database, and the ***size of the tables***. (SQL Count would be useful)
3. Turn in a copy of the script or program used to ***generate the population***. If a script or program was not used, describe how the data was entered, even if it was an all-nighter pizza party with four laptops.

Listing of Data (30 points)

Database software is getting better and better. By this time, you should be able to drop your entire database as a set of filled tables. Some tables will be large. Some tables will be small. Do a SELECT * for each of the tables in your database, copy the information to a file, and turn it in as a PDF.

Turn In:

You will be turning in a report called <cohort>.DataListing.pdf

Query Examples (5 examples, 5 points each)

For each of the following, show the SQL query that was written, and the result.
The examples are based on the hotel.
Show:

- The SQL Query
- and the results of the SQL Query
- You must have a different SQL statement for each problem.

1. Show a query that uses one table.
2. Show a query that uses two or more tables.
3. Show a query that uses an aggregate function. (COUNT, SUM, MIN,MAX, AVG, +,*,/,-) and a GROUP BY.
4. Show a query that uses either a LEFT OUTER JOIN, a RIGHT OUTER JOIN, or a FULL OUTER JOIN
5. Show a query that uses a set Operation, such as UNION or INTERSECTION

Turn In:

A document that, for each case, clearly shows the SQL query and the result of each query.

Naming Conventions for Turn In

Turn in each of the five sections with appropriate names. Each file should be a Portable Document File (.PDF) These files should be turned in as separate files (Which can be done during one session on Blackboard)

Note: The word *Cohort* in the table below should be replaced by the name of your cohort.

Section	*Filename*
Final Design	*Cohort*.FinalDesign.pdf
Data Generation	*Cohort*.DataGeneration.pdf
Data Examples	*Cohort*.DataListing.pdf
Query Examples	*Cohort*.QueryExamples.pdf

Additional:

The assignment may be turned in multiple times before the deadline.
Each member of each cohort must turn in their own copy.
The assignment should not be a zip file, the files should be independent for online grading.
Only the last submission will be counted. Upload all four files, please.