CSCI 4125/5125
Data Models and Database Systems
Fall 2021
Course Project
Phase1: E-R Modeling (8/23)

Due: Sunday, 9/5 @ 11:59pm
Reading: Silberschatz Chapters 1 & 6

Submission Guidelines:
1. This assignment is worth 100 points for all students.

2. All answers in the form of images or screenshots must be readable. Any additional files must be clearly referenced and labeled.

3. It is your responsibility to make sure all files are readable and submitted on time.

Submission:
- Part A requires you to answer a total of 5 short answer questions worth 34 points.

- Part B requires you to submit a single E-R diagram worth 66 points. Each group member submits their own diagram and group members should work on separate databases.

Part A. Short Answer (34 points)

1 (5 pts). List two advantages of using a database management system (DBMS) to store data over a flat file (e.g., csv or text). Briefly describe each advantage in your own words.

- **Data Entry Errors**: a flat file allows a person to write anything on it without regards to constraints. One could write a number where a word is required, a word where a single character is needed, a word where a number is needed, and so on. A DBMS allows the developer to introduce these types of constraints.

- **Poor File Sharing:** data inconsistency could occur if multiple people made changes to a flat file simultaneously and their current snapshots are - to them - the 'current' most recent version that needs to be stored. These data inconsistencies could add up overtime and eventually become disastrous.

2 (5 pts). What is the difference between a (logical) database schema and a database instance (aka snapshot)?

- A **database schema** is the overall template of what the database is formatted as with respect to things such as types of entities, attributes, and relationships. It does not include specific objects or data, but rather a sort of visualization of the types of objects or data that will be stored in the database.

- A **database instance or snapshot** is - as the name implies - is the set of objects and/or data that exists in the database at a specific moment in time.

3 (5 pts). What is physical data independence (aka program data independence)? Be sure to include how this concept is important to a DBMS.

- **Physical Data Independence** is the idea that the underlying hardware on which the data and such is stored is abstracted or 'hidden' from the Database Management System such that the DBMS application doesn't need to compensate in any way for specific hardware or hardware changes. This concept is important to a DBMS because instead of worrying about the low-level, underlying hardware details, the DBMS can focus on it's job - the actual data.

4 (5 pts). In later phases of the project, we will interact with our database using SQL. SQL consists of a few classes of commands; we will primarily be concerned with data-definition language (DDL) and data-manipulation language (DML). What is the difference between SQL DDL commands and SQL DML commands?

- **SQL DDL** is the set of SQL commands that allows a developer to create the underlying schema of a database which defines its data template and structure (as opposed to actual data). *Example: we define a 'student' schema with name, major, and GPA.*

- **SQL DML** is the set of SQL commands that allows a developer to interact with actual data on the database. This includes the basic CRUD commands. *Example: we insert a new student "Annie Edison" with a Healthcare Management major and a 4.0 GPA.*

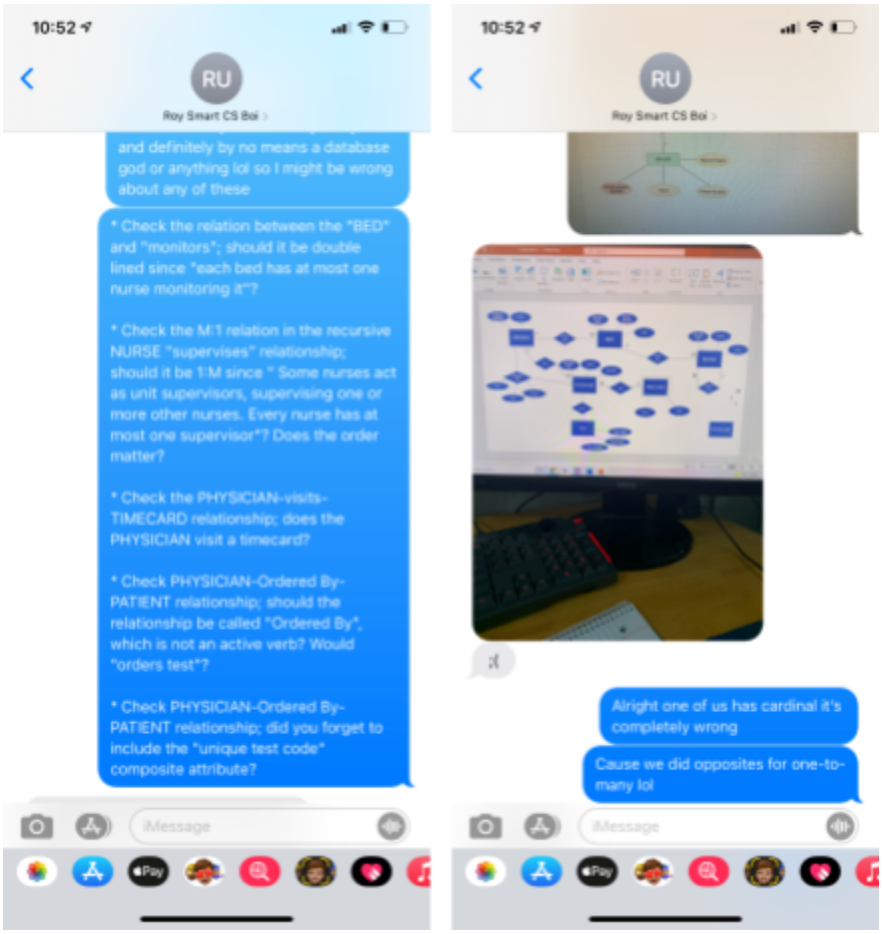
5 (14 pts). In this phase of the project, we are modeling the data given a set of written business requirements. In Phase 2 of the project, we will transform these models in a relational schema. In your own words, describe the following terms in an E-R model. You can use your own examples.

a. **Entity**: an object in the E-R model.
b. **Attribute:** adjectives or qualities that describe an object.
c. **Relationship:** connecting (adj.) relations between entities.
d. **Key:** an attribute or combination of attributes that could identify a unique entity.
e. **Weak Entity:** an entity that can't be uniquely defined purely by its own attributes but rather requires an 'owner' entity for its existence.
f. **Partial Key:** an attribute of a weak entity that could identify a unique weak entity.
g. **Unary (or recursive relationship):** a relationship where an entity could have a relation with another entity of the same type.

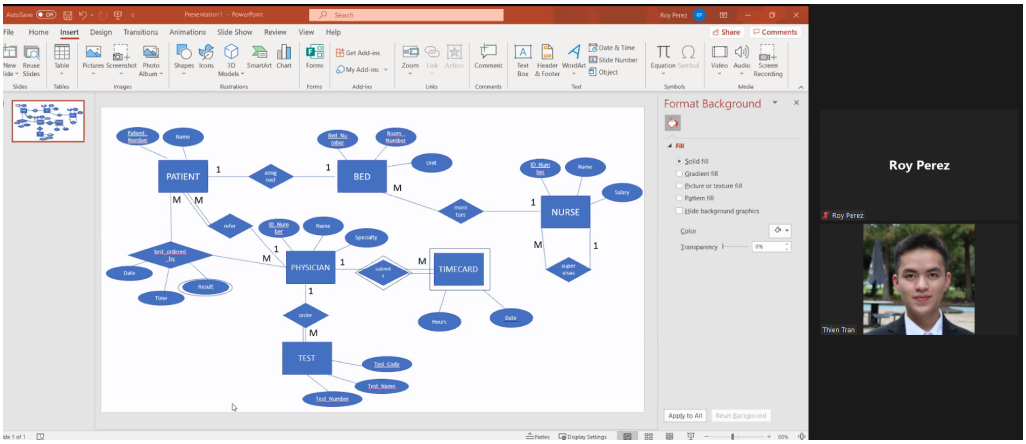
Extra Credit

I had the blessing of having my dearest friend since freshman semester, Roy Perez, be my partner for this database project. Roy Perez is one of the most ambitious and sharp-witted people that I know in the Computer Science department. Roy was very understanding of my situation where I was homeless in Texas for some time due to a car accident, and we both share a mutual desire to see each other succeed.

Anyways, I've enjoyed Roy's help during this project, and he's enjoyed mine. We both totaled about 40 texts related to Phase 1 of this project. Two highlights of our texts are shown in the picture below. On the left, I went through the project requirements and checked its compliance in Roy's E-R model. I listed possible suggestions and fixes to Roy for his consideration. On the right, I realized that we had completely different cardinalities for what should be the same kind of relationship logic, and we both double checked the lecture videos and such to see who had to apply a fix.



Texting eventually became too time consuming and non-intimate, so we decided to do a dual Zoom (for screen sharing) and phone call (for voice) session. We talked for 41 minutes about the class and the troubles that Roy was facing in his project. The topic of primary concern was that Roy had two different relationships where the PHYSICIAN-orders for-PATIENT, but then a PHYSICIAN-orders-TEST. We weren't too sure if TEST should just be a relationship attribute or if it should be a ternary relationship. I personally leaned towards a ternary relationship, but Roy is looking at it a bit more before deciding.



Overall, I think Roy has been a nice and wholesome partner, and I think he deserves the maximum amount of extra credit possible. I do truly hope that my mind remains unchanged as time passes and we go further into the unknown together!

Part B. E-R Modeling (66 points)

In this task, you will create an E-R Diagram for a database. Below is the verbal description of the data to be modeled. Your task is to draw an E-R diagram containing the entities discussed below, their attributes and keys (underlined), and the relationships (including cardinality and participation) among them. You can use software (e.g., PowerPoint or Visio) or draw a legible, free-hand diagram. Submit an image of your E-R diagram below. Each team member should only submit the database he/she is working on.

Database 1: Real Estate Company. Our database keeps track of **houses** and **sellers**. Each house has a unique home ID, a location (consisting of a street address and a state name), and a number of square feet. A seller has a Social Security number to identify him/her, a name, a phone number, and an optional spouse name. A seller may own one or more houses, but each house has just one seller that owns it.

Our agents list houses for the sellers – a seller may use the same agent to list many of their houses if they have several houses to sell, or they may list different houses with different agents. An agent can list many houses, of course, but each house is only listed by one agent at a time. Each agent has a unique AgentID, a name, one or more phone numbers, and a name for the office out of which he/she works. When a seller lists a house with an agent, an asking price and a commission percentage are determined.

We keep track of potential buyers too. A buyer has a Social Security number to identify him/her, a name, a phone number, and a price range (consisting of lower and upper limits). An agent can represent many buyers, but a buyer is represented by only one agent – in fact, we don't even keep track of the information on a potential buyer unless he/she is represented by one of our agents.

Database 2: Mountain Valley Community Hospital. For each patient, we record a unique patient number and his/her name. A given patient may or may not be assigned to a bed in the hospital (some are outpatients only). Each bed may or may not be assigned to a patient. Each bed has a unique bed number in our system, and we also keep track of the room number and which unit it is on.

For each of our nurses, we record a unique ID number, his/her Name, and a salary. Each nurse may monitor many beds or may not monitor any. Each bed has at most one nurse monitoring it (the beds that are not occupied don't have to be monitored). Some nurses act as unit supervisors, supervising one or more other nurses. Every nurse has at most one supervisor.

Each patient is referred to the hospital by his/her physician. A physician may refer any number of patients or may not refer any patients. We record a unique ID number, his/her name, and a specialty for each physician. To track each physician's activity in the hospital, each

day that a physician visits the hospital to see patients, he/she submits a daily timecard that records the date and the number of hours he/she spent at the hospital.

A physician may order any number of tests for any number of patients or may not order any tests. Each test has a unique test code that consists of a test name and test number. A patient may have tests ordered by any physician, not just the one who referred him/her. For each test ordered for a patient by a particular physician, the hospital records the date, the time, and a list of one or more results of the test.