

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.

- **Controls Data Redundancy:** One of flat file's big problem is data redundancy, where there is duplicated data. For example, if a student is enrolled into different courses, they may be inputted multiple times within the flat file's data. As a result, storage space is wasted and access time is longer, making it insufficient. However, ensure there is no data duplication.
- **Data Consistency:** Given a scenario where an employee was to move and have a different address, their information is needed to be change in the system. However, if it is change in one record in the system, it has to be changed in all the records that store his address. If the data isn't updated in all the records, it will lead to data inconsistency. Database ensures that if a given data is updated in one area, it would be updated in all areas involving that data, keeping data consistent.

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

- **Database schema describes the database. Meaning, it gives a visual representation of the database, i.e., the structure, the data type, and any constraints the database may have. Database snapshot is the data itself that stored in the database at any given 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 what the program concerns are and what the data concerns are. For example, the program doesn't care where or how the data is stored. It is only concerned for what is in the database, how to get the data, and how the data is changed. Whereas data only concerns are where and how the data is stored and not what is in the database and vice versa. This concept is important to DBMS because whatever is changed in the lower level of the DBMS, won't affect the higher-level schemas.**

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?

- **DDL commands are commands that specify the database schema i.e. creating the table and constraints. DML expresses the database queries and updates i.e. the records and the insert/delete/update of the user's records.**

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:** any item which data would be captured and stored within the database. For example, within a market database, an employee and a manager would be considered an entity.
- b. **Attribute:** characteristics/properties of an entity. For example, given an employee entity, the attributes for the employee can be Name, Employee ID, Position, Salary. Attributes can be composite (where it consists of two or more attributes, such as a Price Range may also have a Min Range and a Max Range), derived (where it is based on two or more things, such as a Salary may be derived from highest pay to lowest pay value).
- c. **Relationship:** describes how two entities are associated to one another.
- d. **Key:** an attribute or group of attribute that are used to uniquely identify each data, such as a SSN or a Student ID
- e. **Weak Entity:** an entity that can't be uniquely identified by its attribute alone or may rely on another entity to exist, a Child entity may rely on a Parent entity to exist.
- f. **Partial Key:** attributes that are used to uniquely identify a weak entity.
- g. **Unary (or recursive relationship):** a relationship where both participants are the same entity. For example, if an Employee may assist another Employee, they would have a unary relationship.

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.