

CS 122A : Introduction to Data Management-Assignment 1

Fall 2019

While working on your CS122A course assignments you will gain experience in database modeling, design, loading, querying and updates. You are required to work on this assignment **in teams of three**.

Deadline : Friday, October 18 (23:45 pm)

Problem 1

Please answer the following multiple choice questions by **writing your choice in the empty box** given below the question.

1) Relationship constraints:

- A. It always indicates the exact number of entities that participate in a relationship.
- B. Can only be specified for binary relationships.
- C. Indicate whether a relationship between entities in the corresponding entity sets is mandatory or optional
- D. None of the above

2) A car dealership asked its database team to design a database that holds the following entities:

- Cars including VIN number, make, model, year and color.
- Salespersons including id, name, gender, commission rate and start date.
- Customers including id, name, gender, address, Date of Birth.

The dealership wants to store information of the two distinct types of cars which are fuel cell cars with its MPG and electric cars with its charging time. The dealership also wants to keep the information of all sales with selling price and date of sale. The policy of this dealership is that every car can be sold to one customer only and every customer will be helped by only one salesperson with his sales and inquiries (That means this salesperson is assigned to this customer and no other salesperson can help this customer but him). The dealership also wants to store information of potential buyers that are

interested in cars. Whenever a customer comes and asks for a car we store the car and customer's information along with the date of inquiry. A customer can be interested in different cars and vice versa.

2.1) You would represent the interested-in relationship as a :

- A. 1-1 relationship between Cars and Customers.
- B. 1-Many relationship between Cars and Customers.
- C. Many-1 relationship between Cars and Customers.
- D. Many-Many relationship between Cars and Customers.

2.2) You would represent the two types of cars as:

- A. Composite attribute of Cars.
- B. Multi-value attribute of Car.
- C. Two disjoint entities that inherit from Car.
- D. Two overlapping entities that inherit from Car.

2.3) You would represent the sale transaction as a:

- A. Three 1-to-1 relationships between Cars, Salespersons, Customers.
- B. 1-to-1 between Cars and Customers and 1-to-1 between Salespersons and Cars.
- C. 1-to-1-to-Many ternary relationship between Cars, Salespersons, Customers.
- D. Many-to-1 between Cars and Customers and Many-to-1 between Customers and Salespersons.

Problem 2

1. Goal

In this first step you will create a detailed description of a target application and create an entity-relationship model (E-R diagram) to describe the data for the application. To get you started, we are providing you with a sketch of the application that you will be working on, your job in this question is to add more specifics to the target application and to develop its E-R model. To make the assignment more interesting, we've intentionally left some aspects of the application description underspecified.

2. High Level Application Description

In this assignment, you are to design and create a database to support “smart waste management” applications built on sensor data. It will be clear momentarily what we mean by smart waste management applications.

Motivation: Over the past decade, sensing and data capture technologies have significantly advanced and such advances coupled with mechanisms for low-power wireless networking can be used to create deeply instrumented physical spaces. Such physical spaces may be shopping malls, office buildings, airports, critical infrastructures, electric grid, water distribution systems, hospitals, etc. Sensors and other data capture devices embedded into the physical spaces capture the state of the evolving physical systems and processes creating situational awareness of the activities in the instrumented space. Situational awareness, in a broad sense, refers to a continuum of knowledge that captures the current state of the physical environments being observed, to future projected states of these observed environments. Such situational awareness offers opportunities to realizing new functionalities and/or bringing transformational improvements in many application domains. One such application domain is **smart waste management**¹ where waste bins are augmented with an array of sensors such as load sensor, ultrasonic sensor, camera etc. This allows us to collect data about the type of trash thrown and fullness of a bin which is then pushed to a database system with the goal of assisting in waste management in a physical space. For this purpose, data collected from sensors will be used for scheduling pickup, learning diversion rates (percentage of waste diverted from landfills through recycling and source reduction activities), setting up incentivization mechanisms and so on. **Your goal in this assignment is to design and create a database that can store the following concepts related to smart waste management.**

To be able to build such smart waste management infrastructure, we first need to represent the key concepts concepts that arise in such an application.

The first such concept is that of users who use the waste bins to dispose trash.

Users are people who are/have been to the campus and are identified by a specific identifier. A user can either be a UCI-affiliate or a visitor. A UCI-affiliate can either be students, faculty or staff. Students, in turn,

¹ This is a real application scenario and you can read more about the project here if you are interested <https://zotbins.github.io/>

can either be a graduate or undergraduate but not both. Each UCI-affiliate is associated with exactly one department which, in turn, is part of a school (possibly consisting of several departments). (Notice that this is a simplification since sometimes staff are associated with a school and not to a specific department. We are making things simpler by assuming that all the UCI-affiliates are associated with some department).

At a given time, a user is either in the campus, or is outside the campus. If the user is in the campus, we assume that he/she is carrying a device (e.g., their cell phone) using which we can locate the user within the campus. We refer to such a device as a location sensor (will be discussed later). The location of a user may change as a function of time as they move through the campus.

Building the campus has several buildings (e.g., Bren Hall). Each building has a distinct building name and a location (which doesn't change, since buildings do not move :-). Buildings location are captured as a region or rectangle and will be clear when we discuss how we model location. Furthermore, some buildings may be associated with one or more departments. For instance, Bren Hall is associated with Computer science, statistics, and Informatics Departments. Not all buildings are required to be associated with a department though.

Department: a department is a part of a school. It has a name which is distinct within the school. Each department has a department chair and a department manager. The department chair must be a faculty member associated with the department, and the department manager is a staff belonging to the department.

School: each school has a distinct name, and a dean who is a faculty in one of the departments in the school.

Waste bin is a container with several sensors (described later) attached. Waste bins can be of one of recycle or landfill or compost type. A bin has a fixed location and may be associated with a building -- i.e., it might be located inside the building. Several bins of different types can be located within the same building. If a bin is not associated with a building it means that it is outside in the open area.

Sensors: There are two types of sensors -- waste bin sensors that are associated with each waste bin and location sensors, which correspond to devices carried by people. Waste bin sensors are of two types - load sensors that tell us the weight of the bin, and object recognition sensor that identifies the trash item and its type (i.e., compost, landfill, or recycle). A waste bin sensor produces a sensor reading whenever there is a someone throws trash in the waste bin (this will lead to observations by the corresponding sensors).

Each person, when in campus carries a device which we refer to as location sensor (it could be a cell phone with GPS, or any other localization device -- let us not worry about exactly what it is and simply refer to it as a location sensor). Each location sensor produces the location of a person every time a person changes his/her location. Notice that a device is associated uniquely with a person.

Sensor Observations: Sensors produce observations. Someone throwing trash into the waste bin will trigger the object recognition sensor to produce an observation which records what type of item was thrown into which bin at what time. Likewise load sensor will produce the observation of the new weight of the waste bin.

Similarly, whenever a person moves, that is, the location sensor/device changes location, it produces a new observation about the current location of the device which is interpreted to be the location of the individual.

Location: The key to modeling several of the concepts above is that of location. In general, location could be GPS coordinates, for instance. But we will keep things simple. We will consider that our campus is divided into a 10000 by 10000 Grid and each grid cell is identified by its X and Y values. Thus, grid cells range from $\langle 0,0 \rangle$ to $\langle 9999,9999 \rangle$. We will assume that people and waste bins are point objects - that is, they are located within a single cell and do NOT cross over through multiple cells. Buildings, on the other hand, as mentioned earlier are modeled as rectangles. For instance, Bren Hall may be modeled as a rectangle with a left-lower coordinate of $\langle 32,45 \rangle$ to a upper-right coordinate of $\langle 90,130 \rangle$. Given the above representation, we can tell by the location of a person (and the waste bin) if it is inside/outside a building. For instance, if we know that Peter is at location $\langle 34,50 \rangle$, we know he is at Bren Hall. If, instead, he was at a location $\langle 100,200 \rangle$ then he is not at Bren Hall.

Notice that, as mentioned above, a location sensor only produces a new location of a person when the person enters a cell. So for instance, if a person carrying his device D goes from cell $\langle 20,32 \rangle$ to a cell $\langle 20,33 \rangle$ at 2pm, the device (i.e., the location sensor) will produce one reading specifying that the device entered cell $\langle 20,33 \rangle$ at 2pm. Knowing which device is owned by which person, we can then determine the person who entered the specified region.

Your goal is to capture the above concepts in the database. As you can probably guess, we will use the database we create to write a lot of queries such as:

- a) Given a person at location $\langle x,y \rangle$ where is the nearest recycle bin which is less than $\frac{1}{2}$ full?
- b) How many trash events happened for each waste bin in the last month? How many of those events were correct in the sense that the trash disposed was of the same type as the bin.
- c) Which building recycles the most as a percentage of total trash produced by the building in the last week, where we consider the recycling/trash produced by building to correspond to the total such waste associated with any waste bin in the building.
- d) Can we identify the top 10 violators where the violators are people who throw recycling or compost waste into a landfill waste wrongly? Can we identify the buildings with the highest violations ?

Let us not worry too much about queries above for right now. We will worry about them when we get to relational algebra and SQL. We have provided the above just to give you a teaser on what kind of applications you will be writing later during the quarter.

3. Entity-Relationship Model Requirements

Create an ER diagram modeling the concepts in the above description. To help you in the task, we have provided a template (Assignment 1 Template.pdf). All of the entities will be in the template (If you need to, please add more entities). Also, none of the entities in the template are marked as weak; if you wish to change that, you may. You will need to specify two things: (a) for each entity set identified, you will need to specify all the attributes, including their keys (b) define all relationships, make sure to mark all constraints on entities and relationships (key constraints for entities, and cardinality and participation constraints for relationships.)

Be sure to download the template from the CS122a web site and use that as the basis for drawing the E-R schema that you turn in. (Your solution will NOT be accepted if you do not use our provided layout for your diagram). Please, do not change the spatial placement of the entities provided in the template. Again, you must use the concepts and their associated notation from class (see the lecture slides!)

Submission

1. Please submit an appended pdf file with your answers to Problem 1 and Problem 2 (template file). The name of the pdf file should be last names of each team member placed together. For example if Edgar Codd, Donald Chamberlin and Peter Chen were teammates, they would submit:
codd_chamberlin_chen_assignment1.pdf.
2. Upload pdf to **Gradescope**. Only one member of your team is required to submit the file. Be sure to identify all the team members in pdf file (**name and student ID**).