NYU Tandon School of Engineering
CS-GY 6083, Principles of Database Systems, Sections A and INET, Fall 2024
Prof Phyllis Frankl
HOMEWORK #1

Instructions: You may work alone or with one or two classmates. If you work with others, you should **not** divide them up between you. These exercises are intended to help you learn; if you are working with others, you should solve the problems individually, and then discuss them with your teammates. We will post something on Ed Discussion to help people find teammates.

Hand in your solutions via Gradescope as a single PDF file. Follow the Gradescope instructions to mark your solution to each problem or subproblem as indicated in the outline; otherwise, the graders will have trouble finding them and will apply a small penalty to your score. See the student workflow section in GradeScope help to learn how to do this. If you're working with a group, only one of you should hand in the assignment, using GradeScope's group submission feature to indicate all members of your group.
Note: You may find it useful to use draw.io or another drawing tool to draw ER diagrams. Alternatively, you may draw them neatly by hand. **You must use the notation used in class ... rectangles for entity sets, diamonds for relationship sets, double lines and arrows for participation/cardinality constraints, etc.**

Problem 1

a. Welcome to New York, home of Broadway Shows. The Tony awards are an annual award ceremony, like the Oscars, but for broadway shows, rather than movies. There are numerous award categories. Design a simple ER diagram modeling data representing which production won which award when. Your ER diagram does not need to model awards for individual actors. There should be entity sets representing Shows and Awards (each with a reasonable primary key and a couple of other attributes), and one or more relationship set, indicating which shows were nominated and which shows won which award in which year. The model doesn't have to indicate that the winner of an award was among the nominees, but it should indicate that there is exactly one winner of each award each year.

b. Show a little sample data for a few shows, a few award categories and which shows were nominated for and won those categories in the 77th annual Tony Awards. You may use any reasonable notation to indicate elements of the relationship set(s), e.g. (entity1, entity2) or lines connecting entities from the participating entity sets, etc.

Problem 2

A shoe store has hired you to design a database to keep track of their inventory and orders. The *Inventory* is a collection of (pairs of) shoes. Each type of shoe the store carries is identified by its *brand, styleID, size*, and *color*; in addition, the inventory keeps track of the quantity of each type of shoe that is currently *in stock.* Each *Customer* has a unique *email,* any number of *phone*

numbers, a number of *bonus points*, and an address composed of their *street address*, which in turn is composed of the *building number*, *street name,* and *apartment number,* and a *city, state,* and *zip code.*

a. Using the notation studied, design an ER diagram with entity sets representing *Customer* and *Inventory* and a relationship set representing *Current Orders*. The model should enable tracking of the date on which an order is made, the status of the order, the number of (pairs of) each type of shoe in an order, the price paid for each type of shoe, and the status of the order. For part (a), assume that customers cannot have separate current orders of the same shoe type (though they can order several identical pairs in a single order). Your ER diagram for part (a) should be as simple as possible. In particular, it should have only one relationship set and that relationship set should be binary.

b. The store manager realizes that they would also like to keep track of the history of all of the orders that have been made. Briefly explain why the ER diagram from part (a) does not allow representation of a customer who makes multiple orders of the same type of shoe on different dates. Then modify the ER diagram, adding any needed entity sets and replacing *Current Orders* by a relationship set *Orders* which can represent past and present orders of the same type of shoe by the same customer.

c. Now the manager realizes that in addition to storing the prices customers paid, they would like to store the *current price* of each type of shoe and a *description* of each type of shoe. Furthermore, the *current price* and the *description* are determined by the *brand* and *styleID.* In other words, shoes that have the same *brand* and *styleID* always have the same *current price* and *description*, even though they may have different *size* and/or *color.* In addition, the manager would like to allow each customer to designate one *Favorite* shoe type (which should not involve the size or the color). Modify the ER diagram accordingly. **HINT:** Use a weak entity set; think carefully about which weak or strong entity sets participate directly in each relationship set. [For part c) you may start with the assumptions either in part a) or part b)

d. The store is so successful that they've opened multiple locations. Modify the ER diagram to keep track of the inventories and orders at different locations. Customers can order from any location. Each Customer has a *home* location.

Problem 3

a) Design an ER diagram to model this situation: A tennis league has hired you to design a database for their upcoming tournament. In part (a), you will only consider *singles events, where one individual player plays against another individual player.*There are a large number of players, each with a unique *playerID* and a few other attributes (your choice.) Each player is enrolled in one or more *Events* (for example women's singles, men's singles, juniors, etc. – you don't need to know the specific events to design the model.) Each *Event* has a unique *eventID,* a *description*, and some other attributes

(again, your choice). Each player is enrolled in at least one event; some of the enrolled players are *seeded*, i.e. assigned a number indicating that they have a high ranking among all players in a given event. (Seeding is used in tournaments to arrange the schedule so the top players are unlikely to meet one another in early rounds.) The club has multiple *Courts*, each with a unique *courtNumber* and some other attributes. There are multiple *timeSlots* at which matches are played, each of which has a unique *date* and *startTime.*
A *SinglesMatch* takes place between two players enrolled in the same event in a particular time slot on a particular court; at the end the *score* is recorded along with a description of the *highlights.*

b) **Extra Credit:** Doubles anyone? The tournament also includes doubles events, in which two-player teams play one another (e.g. men's doubles, women's doubles, mixed doubles) . Augment your ER diagram to model doubles events and their matches. (You do not have to show all of the details for entity sets and relationship sets from part b) unless they're relevant here. Highlight the parts that are new.) There are several reasonable ways to approach this. In addition to designing the ER, comment on which aspects of doubles events and matches are or are not represented by your diagram.

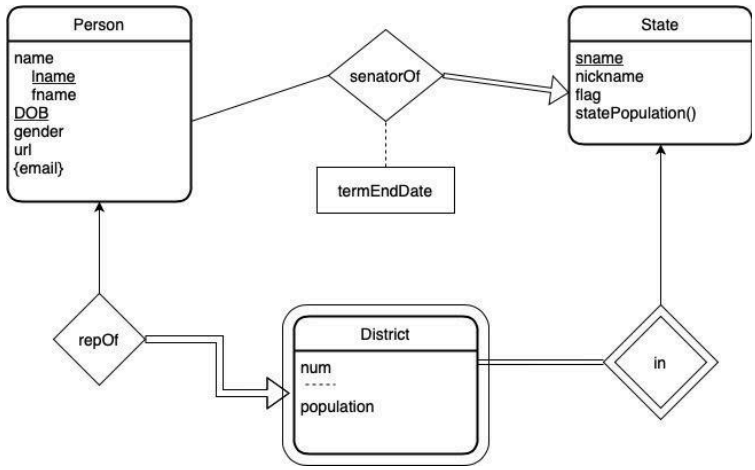
Problem 4

The United States Congress consists of two houses, the Senate and the House of Representatives:

- Each state elects two senators.
- Each state is divided into Congressional Districts and each Congressional District elects one representative (also called a Member of Congress).

The number of Districts in a state is based on the state's population (as of the most recent US Census). Each congressional district has a district number and a population. A congressional district is uniquely identified by its number ***along with*** its state. For example, New York 2, New York 39, Pennsylvania 2, etc.
(See link for more info about Congressional Districts.)

Consider this ER diagram, modeling the congress:



The *senatorOf* relationship set indicates who is a senator for which state
The *repOf* relationship set indicates who is a Member of Congress for which district.

1) Which of the following aspects of the data are represented in this model?

(a) Each district has exactly one representative

(b) Each state has exactly two senators

(c) Each state has at least one senator

(d) A person cannot be senator of two different states

(e) A person cannot be congressperson of two different districts

(f) A person can be the congressperson of two different districts, as long as they're in different states

(g) A person can be the congressperson of two different districts, as long as they have different numbers.

2) Derive a relational database schema from this ER diagram. You may either show it as a schema diagram or in a textual format. Clearly indicate all primary key and foreign key constraints.