

Problem Set #1 (due 2/15)

Note: In this homework, you do not need to create tables and execute queries using an actual DBMS. A written solution is sufficient.

Problem 1: Suppose you have a database modeling an online food delivery company such as Grubhub, where customers can order food that is then delivered by various restaurants. It is given by the following highly simplified schema:

CUSTOMER(cid, cname, ccity, cstate, cphone)

RESTAURANT(rid, rname, rcity, rstate)

MENUITEM(rid, itemname, price)

ORDER(oid, cid, rid, ordertime, deliverytime, totalprice, deliveryfee)

ORDERDETAIL(oid, rid, itemname, quantity)

In this schema, restaurants offer different menu items, and each item has a current price. Customers can place order that contains one or more menu items. The ORDERDETAIL table shows which and how many menu items are included in an order. In the ORDER table, we store the total price of an order and the delivery fee, and the time and date when the order was placed and delivered. For both customers and restaurants, we store the city and state where they are located. We assume that ordertime and deliverytime contain both time and date information, and that a single order can only contain menu items from a single restaurant.

(a) Identify suitable foreign keys for this schema

(b) In the MENUITEM table, why is (rid, itemname) the primary key? What happens if we remove rid from the key? What happens if we remove rid from the key for ORDERDETAIL?

(c) Write statements in SQL for the following queries.

- I. Output the cid and cname of any customer who has placed an order for more than \$50.
- II. Output the cid and cname for any customer living in "Queens" who has placed an order from a restaurant in "Brooklyn".
- III. Output the rid and rname of any restaurant that has delivered every order within 1 hour.
- IV. For each customer, output the cid, cname, and number of orders she has placed.
- V. Output the cid and cname of the customer(s) who has placed the most expensive order ever.
- VI. Output for each order the oid, the price that was paid, and the price that the order would now cost if we sum up the current prices of all the items in the order.
- VII. Output the cid and cname for any customer who has placed more than one order from a restaurant whose rname is "Little Sheep".
- VIII. For each state, output the cid and cname of the customer(s) living in this state who has placed the most orders.

(d) Write expressions in Relational Algebra for the above queries.

(e) Write either DRC or TRC queries for the above queries. Or explain the reason why you think a particular query cannot be done in DRC or TRC.

Problem 2: In this problem, you need to design a relational schema for a patient management system for a chain of hospitals specializing in infectious diseases. Each hospital in the chain has a name and an address. The system keeps track of patients that are admitted to the hospitals, their doctors, and their diseases.

For each patient, you need to store a unique id and their name, age, gender, height, weight, and the address where they currently live. For each medical staff member, you need to store their name, age, gender, address, job title (doctor, resident, intern, nurse, and so on), and in which hospital they work. Whenever a patient stays in a hospital, we need to store the patient ID, hospital name, and time and date when they were admitted and when they were released. Also, we need to store which staff members took part in the treatment of the patient during this particular stay. Also, for each time a patient stays in the hospital, one doctor is labeled the “attending physician”, and this doctor has primary responsibility for treating the patient.

We also need to store the diseases patients currently have, or had in the past. Each disease has a name and a short description, and is either contagious or not contagious. You also need to store when the disease was first diagnosed in the patient, and if it was cured already, when it was diagnosed as cured.

Obviously, this is a very simplified scenario, and we do not attempt to store information about the exact treatment steps taken, or about billing and insurances, or worry about updates such as address changes, name changes, or staff turnover.

(a) Design a relational database schema for this application that supports the above functionality. Specify all primary and foreign key constraints, and state any assumptions you are making. You can decide which exact attributes make sense for this schema.

(b) Write SQL statements for the following queries. If your schema does not support these, you need to modify it appropriately. (For this first homework, you may use informal expressions such as $\text{year}(\text{ts}) = '2016'$, where ts is a timestamp, to check if the year is 2016.)

- I. Output the names of all patients who have been admitted into “St. Rudolph’s Hospital” during 2016.
- II. For each doctor, output their name, and the number of distinct patients for whom they have served as attending physician during a stay.
- III. Output the names of doctors who have been in charge of more than 10 patients who either have or have had a contagious disease.
- IV. The BMI (Body Mass Index) is a measure of body fat based on height (h) and weight (w) that applies to adult men and women. Its formula is $\frac{w}{h^2}$. For each infectious disease, output the average BMI of patients that currently have this disease.