NYU School of Engineering September 13, 2017
Computer Science and Engineering
CS 6083, Fall 2017

Problem Set #1 (due 9/26)

Note: In this homework, you do not need to create tables and execute queries using an actual DBMS. This will however be needed for the second homework set. A written solution is sufficient for now. Also, while the problems talk about relational schemas for websites, you do not have to worry about how to build such a site, but only need to think about suitable underlying schemas.

Problem 1: Consider the following relational schema for a website for fans of live music that keeps track of artists (individual musicians or groups), concerts by artists, and the times and venues (places such as bars, concert halls, etc.) where concerts take place:

Artist (aid, aname, adesc)
ArtistGenre (aid, genre)
Concert (cid, vid, cdate, cstarttime, cendtime, cdesc)
ConcertArtist (cid, aid)
Venue (vid, vname, vaddress, vcity, vphone)
TicketType (cid, ticketclass, ticketcost)

Thus, for each artist we have an aid, a name, and a short description of their music. An artist can also play one or more genres, such as Jazz, Americana, Soul, etc. For each concert, we store where it takes place (the venue), the date, the start and end time, and a short description. A concert may involve several artists. For each venue we store its name, address, and contact phone number. Finally, we store information about the prices of different categories of tickets for each event. Note that the site is not keeping track of individual tickets and who purchased them; it only provides information about upcoming concerts and their ticket prices (probably with hyperlinks to the actual sites selling tickets).

(a) Identify suitable foreign keys for this schema.

(b) In the Concert table, if we did not have the attribute "cid" (meaning we only have vid, cdate, cstarttime, cendtime, and cdesc), how would you choose a suitable primary key?

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

I. Output the names of all artists that have the genre "Jazz"
II. Output all the cities in which artist "Bruno Mars" played in 2016.
III. For each venue, output the vid and name and the number of concerts it hosted in 2016.
IV. Output the cid, ticketclass, cost, and date for the most expensive tickets ever for a Jazz concert. (A Jazz concert is one where at least one artist has Jazz as a genre.)
V. Output the aids and names of all artists who have appeared together with "Bruno Mars" in at least two concerts during 2016.
VI. Output the aid and name of any artist who has never given a concert in New York City.
VII. Output the cid, data, and lowest cost for any Jazz concert in Chicago in October 2017. (Lowest cost means the cost of the cheapest class of tickets over all Jazz concerts.)

(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 babysitting agency website. This website is used to connect parents and babysitter, and to help parents book babysitters on-demand.

For each babysitter, you need to store detailed information, including a unique id, her/his name, age, gender, address, phone number, languages spoken, and a short text intro. For each customer (usually a parent needing a babysitter), you need to store a unique id, her/his name, address, and phone number. You also need to store information about each child, such as the parent/guardian, name, age, gender, and languages spoken.

Each babysitters has a schedule on the website, so that it is possible to search which babysitters are available at a particular time. For simplicity, you may assume that the schedule consists of slots lasting one hour. Parents may also upload babysitting requests, specifying the time and date, and which children need to be supervised.

When a babysitter and a customer come to an agreement on a babysitting job, (which will usually happen after some direct negotiation on the phone or in person outside this system), a booking is created and stored that specifies the customer, the babysitter, the children that will be supervised, the fee the babysitter charges for this job, a status (booked, cancelled, or completed), and a rating from one to five stars that the customer can assign after the job is completed. Note that each booking is for one babysitting session, so if the customer and babysitter agree on a job that takes place every Wednesday evening, this would result in one booking for each week. Also, the children that should be supervised need to be specified for each booking; in some cases, a customer with two kids may require babysitting only for one of the kids, or a neighbor’s child might join so the children might not all have the same legal guardian, and some babysitters might charge more for jobs involving more than one child.

(a) Design a relational database schema for this application scenario 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 Timestamp(ts) =’2017-09-08T11:00:00’, where ts is a timestamp, to check if the datetime is 2017-09-08 11:00:00, or use year(ts) or month(ts) to get the year or month of a timestamp, etc.)

I. Output the names of all babysitters who are available to work from 2017-09-20T15:00:00 to 2017-09-20T18:00:00. (They should be available all three hours, not just part of the time.)
II. For each babysitter, output their ids, and the number of distinct children they have babysat in 2016.
III. Output the babysitter(s) who earned the most money overall during 2016.
IV. Output the names of customers who have booked at least 5 completed jobs, but who have never given a rating of 4 stars or higher.