Virginia Tech. **CS 4604 – Introduction to DBMS**
Computer Science **Spring 2014, Prakash**

Homework 2: ER Models and More RA/SQL

(due February 18th, 2014, 3:30pm, in class—hard-copy please)

Reminders:

a. Out of 100 points. Contains 5 pages.
b. Rough time-estimates: 3~5 hours.
c. Please type your answers. Illegible handwriting may get no points, at the discretion of the grader. Only drawings may be hand-drawn, as long as they are neat and legible.
d. There could be more than one correct answer. We shall accept them all.
e. Whenever you are making an assumption, please state it clearly.
f. Unless otherwise mentioned, you may use any SQL/RA operator seen in class/in textbook.
g. Unless otherwise specified, assume set-semantics for RA and bag-semantics for SQL.
h. Feel free to use the linear notation for RA and create intermediate views for SQL.
i. **Important:**
    a. For E/R diagrams, use only the style and notation given in the lecture slides.
    b. A useful tool for creating E/R diagrams: http://logicnet.dk/DiagramDesigner/. You may have to manually draw-in some things though (like adding proper constraints etc.). There are other such programs too.
j. Lead TA for this HW: Qianzhou Du.

Q1. Querying a Book Store [20 points]

The following schema keeps track of information in a Book Store Chain scenario (e.g., say for Barnes and Noble):

Book (***ISBN: integer***, BTITLE: string, TYPE: string, YEAR: integer)

Customer (***CID: integer,*** CNAME: string, CADDRESS: string, CPHONE: integer)

Store (***SID***: integer, SNAME: string, SADDRESS: string)

Copy (***COPYID: integer,*** ISBN: integer, SID: integer)

Transaction (***CID: integer, COPYID: integer, TDATE: time,*** COST: numeric)

The Book table contains every released book's ISBN, title, type (e.g., a novel or a textbook), and year of publication. The Customer table records the information of customers' IDs, names, address, and phone numbers. The Store table includes stores' IDs, names, and address information. We store information about copies of a book in the table Copy: each copy has a unique COPYID and is located in a particular store, and there may be multiple copies of the same book in a single store. The Transaction table records each customer's ID, COPYID, the date when the book is purchased, and the total cost.

Q1.1. (10 points) Write the following query in **Relational Algebra**: List names of all customers who have purchased "Harry Potter 1" and "Harry Potter 2" but not "Harry Potter 3". (The store can then give coupons to such people)

Q1.2. (10 points) Write the following query in **SQL**: List the title of the most popular book (by total number of copies sold) for *each* store separately in 2013.

Q2. Where Art thou? [15 points]

Consider the situation in Exercise 2.8 in your textbook. We repeat it here for your convenience. Although you always wanted to be an artist, you ended up being an expert on databases because you love to cook data and you somehow confused "database" with "data baste". Your old love is still there, however, so you set up a database company, ArtBase that builds a product for art galleries. The core of this product is a database with a schema that captures all the information that galleries need to maintain. Galleries keep information about artists, their names (which are unique), birthplaces, age, and style of art. For each piece of artwork, the artist, the year it was made, its unique title, its type of art (e.g., painting, lithograph, sculpture, photograph), and its price must be stored. Pieces of artwork are also classified into groups of various kinds, for example, portraits, still-lifes, works by Picasso, or works of the 19th century; a given piece may belong to more than one group. Each group is identified by a name (like those just given) that describes the group. Finally, galleries keep information about customers. For each customer, galleries keep that person's unique name, address, total amount of dollars spent in gallery (very important!), and the artists and groups of art that the customer tends to like.

The ER model your DB engineer designed was lost, and he has resigned. So you need to step in:

Q2.1. (10 points) Draw an ER diagram for this database. Make sure to indicate primary keys, cardinality constraints, weak entities (if any), and participation constraints. List any assumptions you make in the process.

Q2.2. (5 points) Translate the ER diagram in Q2.1 into relational database tables (i.e. give the SQL DDL statements). Make sure that the translation captures key constraints (primary keys and foreign keys if applicable) and participation constraints in the ER diagram. Identify constraints, if any, that you are not able to capture.

Q3. Soccer DB [15 points]

You are hired as a consultant to design a database schema for a professional soccer league. The league consists of teams (clubs) identified by a name (e.g., Real Madrid) and located in a city (e.g., Madrid). Players have a unique name and a date of birth, and can be members of at most one team in any given season (e.g. 2013-2014 season), but can play in different teams in different seasons. Matches are played between two teams (e.g. a home team and an away team) on a particular date, and result in a certain score (e.g., 2:1). Each match also has a location (city and stadium name). We would also like to keep track of which players played in a given match, how long they played (e.g., a player played from the 33rd to the 75th minute), who scored the goals, and who received a red card.

Draw an ER diagram that models this database. Make sure to indicate primary keys, cardinality constraints, weak entities (if any), and participation constraints. List any assumptions you make in the process.

Q4. Airport Database [25 points]

Consider the situation in Exercise 2.6 in your textbook. We repeat it here for your convenience. Computer Sciences Department frequent fliers have been complaining to Dane County Airport officials about the poor organization at the airport. As a result, the officials decided that all information related to the airport should be organized using a DBMS, and you have been hired to design the database. Your first task is to organize the information about all the airplanes stationed and maintained at the airport. The relevant information is as follows:

- Every airplane has a registration number, and each airplane is of a specific model.
- The airport accommodates a number of airplane models, and each model is identified by a model number (e.g., DC-10) and has a capacity and a weight.
- A number of technicians work at the airport. You need to store the name, SSN, address, phone number, and salary of each technician.
- Each technician is an expert on one or more plane model(s), and his or her expertise may overlap with that of other technicians. This information about technicians must also be recorded.
- Traffic controllers must have an annual medical examination. For each traffic controller, you must store the date of the most recent exam.
- All airport employees (including technicians) belong to a union. You must store the union membership number of each employee. You can assume that each employee is uniquely identified by a social security number.

- The airport has a number of tests that are used periodically to ensure that airplanes are still airworthy. Each test has a Federal Aviation Administration (FAA) test number, a name, and a maximum possible score.
- The FAA requires the airport to keep track of each time a given airplane is tested by a given technician using a given test. For each testing event, the information needed is the date, the number of hours the technician spent doing the test, and the score the airplane received on the test.

Please answer the following questions:

Q4.1. (15 points) Draw an ER diagram for this database. Make sure to indicate primary keys, cardinality constraints, weak entities (if any), and participation constraints. List any assumptions you make in the process.
Hint: You may need an ISA hierarchy somewhere.

Q4.2. (10 points) Translate the ER diagram in Q4.1 into relational database tables (i.e. give the SQL DDL statements). Make sure that the translation captures key constraints (primary keys and foreign keys if applicable) and participation constraints in the ER diagram. Identify constraints, if any, that you are not able to capture.

Q5. Beverage (and Country/Color) for thought [25 points]

This is a tricky problem designed to help you think out of the box on the use of database programming for solving problems.

You are given a puzzle: There are 3 houses in one row in 3 different colors (Red/Blue/Yellow), and the owner of each house is of a different nationality (British/Danish/Norwegian). Every person drinks a (different) particular beverage (Milk/Tea/Water). We also know the following constraints:

1. The British person lives in the red house.
2. The person who lives in the yellow house drinks water.
3. The person who lives in the right-most house does not drink tea.
4. The Norwegian's house is not the middle one.
5. The Danish person does not drink milk.
6. The yellow house is located to the left of the blue house.
7. Those who drink tea and milk do not live adjacent to each other.

From this information, we want to figure out who-lives-in-which-house respectively i.e. list them based on their positions (Left/Middle/Right). Explain how you will solve this puzzle by creating database tables and writing a query.

Q5.1. (9 points) The schema of the tables you use.
Hint: Think about the solution space of the problem and how to get to that space using multiple small simple tables.

Q5.2. (10 points) Your SQL query.
Hint: Recall one can do cross-products in SQL.

Q5.3. (6 points) The solution you get for the puzzle when you use an SQL interpreter and RDBMS to solve this puzzle (e.g. you can use SQLite). Copy-paste the output you get.

Note:
The SQL query may be quite long so you may find it useful to create the query in a text file and use the source command (or equivalent) in your SQL interpreter to read in and execute the query.