Qualifying Exam for Databases

Fall, 2014, Dec. 12, 2014

Open book. Your code: _____

Good luck! (2hrs, 100 points)

Q1. (25 pts) An environmental scientist needs to design a small database to store the data about (i) the trees that she is monitoring and (ii) the physical environment in which the trees grow. Please help this scientist design her database to keep the following information.

- The trees are planted in three different physical sites. Each site has a unique name and an address.
- Each physical site has many spots. The spots are labeled with globally unique spot ids. Spots may have different area sizes. According to the area size of the spot, each spot is categorized to Large, Medium, and Small.
- A tree is planted inside a spot. But a spot may not have any tree. (E.g., when it is a new spot).
- A tree has a unique tag, a tree type, and the year that this tree was planted.
- Once every month, this scientist needs to go to every site to measure the trees and soils. In particular, for each spot, the pH value and the humidity of the soil need to be measured and stored. For each tree, the radius and the height needs to be measured.

Questions:

1. (20 pts) Draw an E-R diagram to capture the above described requirements that can be represented by E-R diagram.
2. (5 pts) State two conditions that the above requirements describe, but cannot be captured by the E-R model.

Q2. (25 pts) Given a schema $R(A, B, C, D, E)$. Suppose that the following functional dependencies hold in R: F=$\{A \rightarrow CD$, $BC \rightarrow D$, $E \rightarrow AB\}$. Please answer the following questions.

1. (5 pts) Is $EC \rightarrow D \in F^+$? Please justify your answer succinctly.
2. (5 pts) Is R in BCNF? Please justify your answer succinctly.
3. (15 pts) Assume that R is decomposed to $R_1(A, B, E)$ and $R_2(B, C, D)$,
   (a) Is this decomposition a lossless-join decomposition? Please justify your answer succinctly.
   (b) Can the above decomposition preserve all the functional dependencies in F? Please justify your answer succinctly.
   (c) Are R_1 and R_2 in 3NF? Please justify your answer succinctly.

Answer Q3 and Q4 by using the following relational database schema, where the underlined attribute(s) form(s) the primary key of the corresponding schema.
Note: the data type `Time` is in the format of *hh:mm:ss* and it does not include the date information (in the format of *yyyy-mm-dd*).

- `Flight(flightno:char(10), boardTime:Time, landTime:Time, boardAirport:char(3), landAirport:char(3), distance:int, company:char(2))`
- `Ticket(id:int, passengerName:varchar(64)), passengerEmail:varchar(64), price:float)`
- `TripSegment(tid:int, rid:int, flightno:char(10), date:Date)`
  –Foreign key `tid` references `Ticket(id)`
  –Foreign key `flightno` references `Flight(flightno)`

Q3. (25pts) Express the following queries in relational algebra. Write ONE relational algebra expression (NOT a set of relational algebra expressions) to answer each of the following queries.

1. (12pts) Find all the airports that "John Smith" has visited. Please show the unique airports.
2. (13pts) "United Airline (UA)" wants to find their frequent passengers and send promotion information to them. The passengers who flew in more than ten UA flights in the past are treated as frequent passengers. You are asked to find all the frequent passengers. For each passenger, show his/her name and email address.

Q4. (25pts) Write an SQL statement (NOT a set of SQL statements) to answer each of the following queries.

1. (12pts) Assume that each trip corresponds to one and only one ticket. Find the long trips for a passenger named "John Smith". A trip is long if the total distance of the corresponding trip segments is more than 5,000 miles. Show the long trips' corresponding ticket ids and ticket prices.
2. (13pts) Assume that one flight can fly at most once in a day. For all the UA flights (i.e., flights belonging to the company "United Airline"), please find the average number of passengers in each flight.

END