DATA 514 Homework 4: Entity Relationship Diagrams and Query Plans

Objectives: To be able to translate from entity relationship diagrams to a relational database and basics of query plans and RA.

Assignment tools:

- Pen and paper or any drawing tools you prefer (e.g., powerpoint, draw.io) for part 1

Assigned date: Feb. 12, 2019

Due date: Feb. 18, 2019.

What to turn in:

- `hw4-solutions.pdf`: A pdf containing some diagrams, some short decompositions, some free response questions

Resources:

Textbook chapter 3.1-3.4, 4.1-4.6

Assignment Details

Part 1: Theory

1. (20 points) Design an E/R diagram for geography that contains the following kinds of objects or entities together with the listed attributes.

   Model the relationships between the objects with edges. Note that edges between entities can be labeled with constraints. Make sure to label only the/those primary key(s) that is/are mentioned below.

   Entities

   - countries (with attributes): name, area, population, gdp ("gross domestic product")
     - a country's name uniquely identifies the country within all countries
   - cities: name, population, longitude, latitude
     - a city is uniquely identified by its (longitude, latitude) (not by name, since for instance there are 41 different cities and towns are named Springfield in the US!)
   - rivers: name, length
   - seas: name, max depth
     - rivers and seas are uniquely identified within all water entities by their name (e.g., "Ganges" would be a unique water entity)

   Relationships:

   - each city belongs to exactly one country
   - each river crosses one or several countries
   - each country can be crossed by zero or several rivers

- each river ends in either a river or in a sea

You can draw your diagrams on paper and scan them, take *quality* pictures of your drawn diagram, or use your favorite drawing tool such as Powerpoint, Keynote, or draw.io. (FYI: Google Slides lacks a few shapes that you might need such as rounded arrows.)
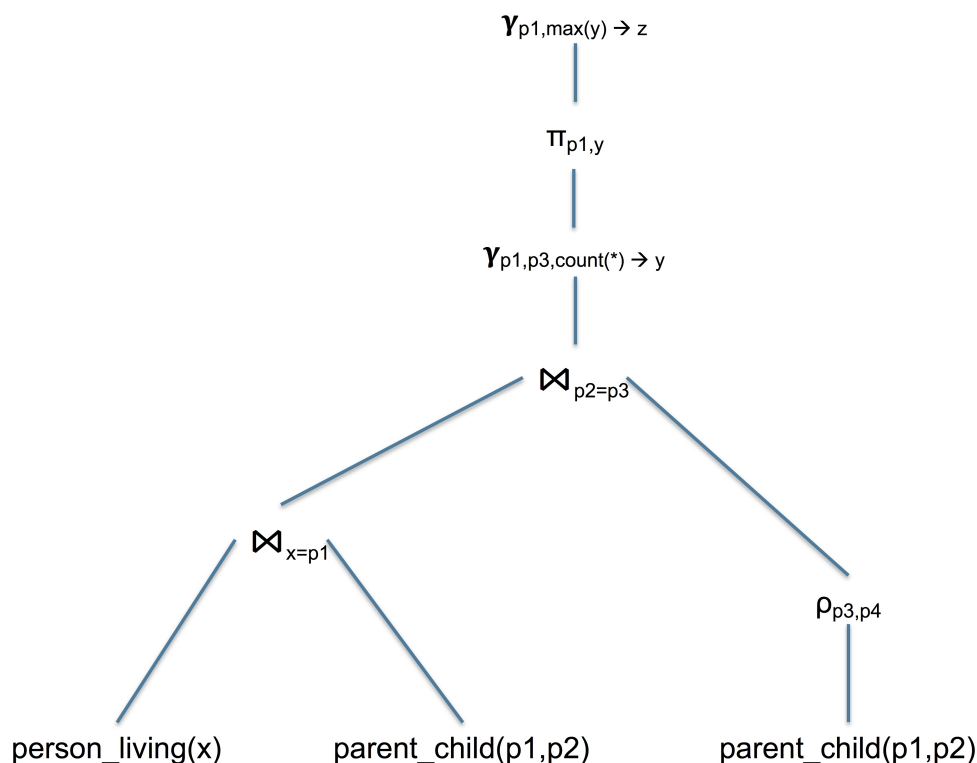
2. (40 points) Consider the following E/R diagram:
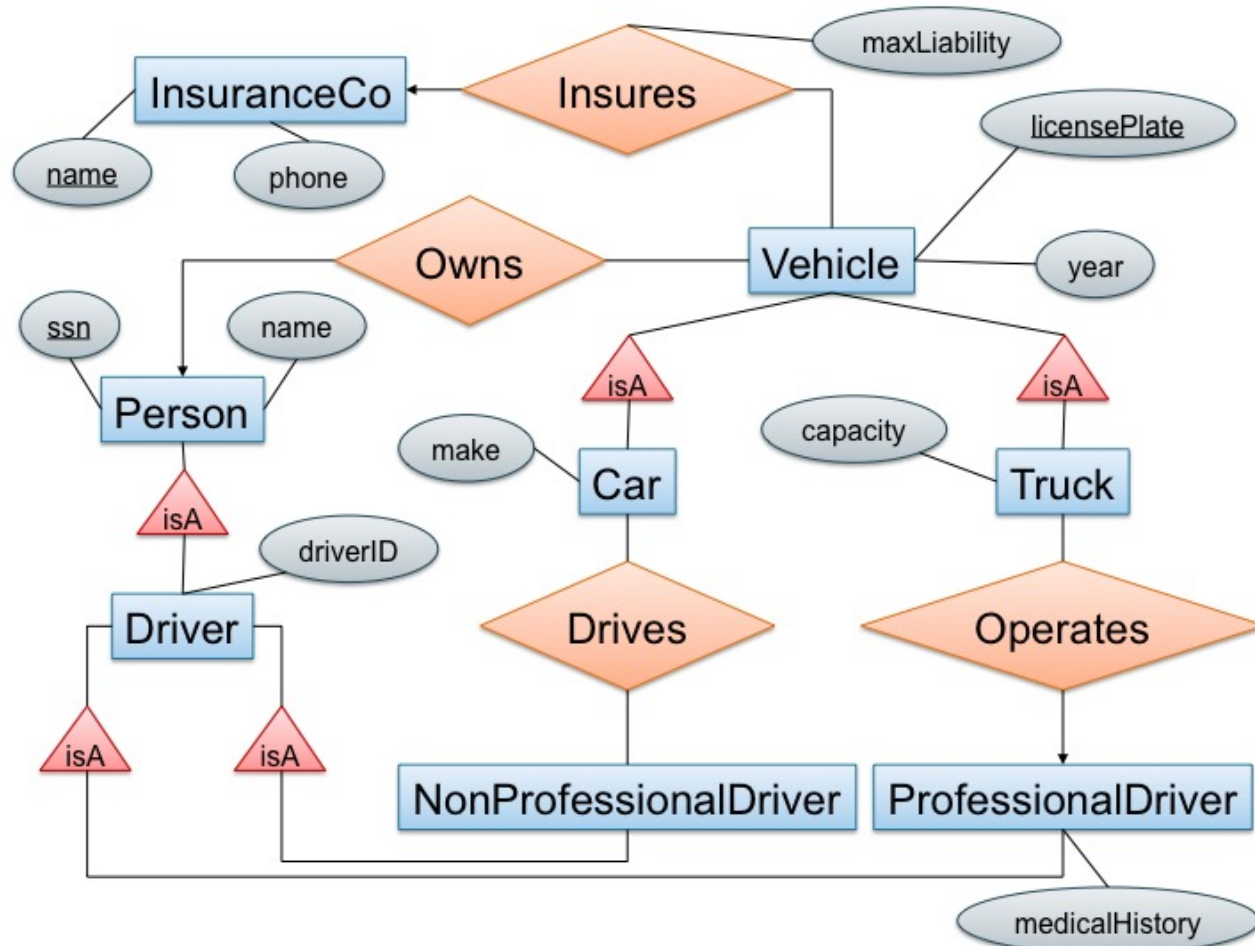
License plate can have both letters and numbers; *driverID* and *Social Security* contain only numbers; *maxLiability* is a real number; *year*, *phone* are integers; everything else are strings.

- Translate the diagram above by writing the SQL CREATE TABLE statements to represent this E/R diagram. Include all key constraints; you should specify both primary and foreign keys. Make sure that your statements are syntactically correct (you might want to check them using sqlite / Azure for instance). (20 points)
- Which relation in your relational schema represents the relationship "insures" in the E/R diagram and why is that your representation? (10 points)
- Compare the representation of the relationships "drives" and "operates" in your schema, and explain why they are different. (10 points)

3. (20 points) Write the equivalent SQL query to the following relational algebra query plan:

$$\gamma_{p1,max(y) \to z}$$

$$\pi_{p1,y}$$

$$\gamma_{p1,p3,count(*) \to y}$$

$$\bowtie_{p2=p3}$$

$$\bowtie_{x=p1}$$

$$\rho_{p3,p4}$$

person_living(x)   parent_child(p1,p2)   parent_child(p1,p2)

4. (20 points) Write a relational algebra plan for the following SQL query:

```
SELECT a.p
FROM   person_living AS a, male AS b
WHERE  a.p = b.name AND
       NOT EXISTS (SELECT *
                   FROM   parent_child AS c, female AS d
                   WHERE  c.p1=d.name AND c.p2=a.p)
```

You do not need to draw the query plan as a tree and can use the linear style instead. To make precedence clear, we ask you to break down your query plan by using *at most one* operator on each line. For example, given the query in question 1, you could write it as:

```
T1(x,p1,p2) = person_living(x) Join[x=p1] parent_child(p1,p2)
T2(p3,p4) = Rename[p3,p4] parent_child(p1,p2)
T3(x,p1,p2,p3,p4) = T1(x,p1,p2) Join[p2=p3] T2(p3,p4)
T4(p1,p2,y) = GroupBy[p1,p2,count(*)->y] T3(x,p1,p2,p3,p4)
T5(p1,y) = Project[p1,y](T4)
T6(p1,z) = GroupBy[p1,max(y)->z] T5(p1,y)
```

where T1, T2, etc are temporary relations. Note that each line has at most one relational operator. You do not need to use the Greek symbols if you prefer. You also don't need to distinguish among the different flavors of join (just make sure that you write out the full join predicate). We will be grading this part of the homework manually so don't worry too much about the exact syntax of the relation algebra function names etc, as long as we can tell what is going on.

Submission Instructions

Write your answers in a file `hw4-solutions.pdf`.