Question Booklet

W4111 Introduction to Databases
Fall 2018
Midterm Exam
Instructor: Eugene Wu

Closed Book, 1 page notes: 8.5x11" letter paper, both sides
Duration: 75 minutes
501 Northwest Corner if the last digit of your UNI is 0,1,2,3,4,5
Pupin 329 if the last digit of your UNI is 6, 7, 8, or 9

Instructions

This is the question booklet, which contains questions for the exam.
There is a separate answer booklet for your answers.

1. You are supposed to write your answers on the answer sheets.

2. The staff will ignore text written on the question sheets.

3. You will submit the Quesion AND Answer booklets at the end of the exam. If we do not receive both booklets with your UNI, **you will receive a zero**.

Your Name: _____

Your UNI: _____

1 (16 points) Equivalences

For each of the following pairs of queries (in relational algebra or SQL), you will write the contents of two database instances. The databases have the following schemas:

```
A(a int, b int);
B(a int, b int)
```

(2 Points) The first database instance should be populated with one or more rows so that Q1 and Q2 output different results. If Q1 and Q2 are equivalent, then write "identical" next to the empty tables instead.

(2 Points) The second database instance should be populated with one or more rows so that Q1 and Q2 return the same results. If this is not possible, write "not possible" next to the empty tables instead.

1.1 (4 Points, 2 Per Database Instance)

Q1: $A \bowtie_a B$
Q2: $A \bowtie B$

1.2 (4 Points, 2 Per Database Instance)

Q1: $A \bowtie_a B$
Q2: $A \times \sigma_{B.a=' a'}(B)$

1.3 (4 Points, 2 Per Database Instance)

Q1: $\sigma_{\$1=\$3}(A \times B)$
Q2: `SELECT * FROM A, B WHERE A.a = B.a`

1.4 (4 Points, 2 Per Database Instance)

Q1: `SELECT * FROM A JOIN B ON A.a = B.a WHERE B.a = 1 or B.b = 2`
Q2: `SELECT * FROM A JOIN B ON A.a = B.a WHERE B.a = 1`
```
   UNION ALL
 SELECT * FROM A JOIN B ON A.a = B.a WHERE B.b = 2
```

2 (18 points) Entity-Relationship Models

2.1 Constraints (3 Points)

Your friend downloaded the following CSV file and shared it with you. What constraints, if any, can be inferred from this dataset? If there are none, simply write "none" and explain why in one sentence:
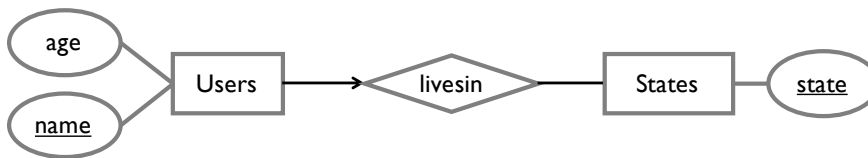
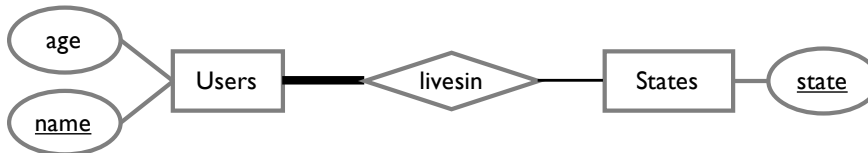| name	age	state
amy	18	TX
amy	18	CA
amy	18	FL
joe	20	MA
joe	20	NY

Table 1: Table for Problem 2

2.2 ER Constraints

For each of the following ER diagrams, select TRUE if Table 1 satisfies the constraints depicted in the diagram, and FALSE otherwise. If FALSE, write a short sentence about why.
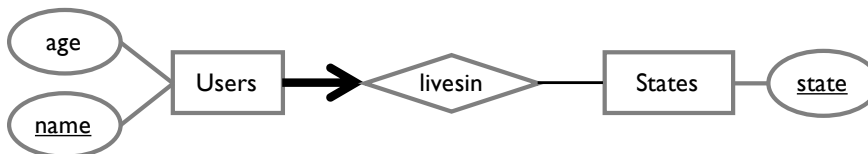
2.2.1 (3 Points)

2.2.2 (3 Points)

2.2.3 (3 Points)

2.3 ER to SQL

2.3.1 (3 Points)

Translate the ER diagram from Problem 2.2.1 into SQL.

2.3.2 (3 Points)

Translate the ER diagram from Problem 2.2.3 into SQL.

3 (8 points) Triggers

Consider the following database schema, and each table is empty:

```
CREATE TABLE A(a int);
CREATE TABLE B(b int);
```

In this question, we will present two implementations of triggers. Assuming the two tables are initially empty, and you run the following INSERT statements, you will write the final contents of A and B in the answer sheet.

```
INSERT INTO a VALUES(1);
INSERT INTO a VALUES(2);
INSERT INTO a VALUES(3);
```

3.1 Triggers

3.1.1 (4 Points, 2/table)

```
CREATE FUNCTION UDF1() RETURNS trigger      CREATE TRIGGER T1
AS $$                                       BEFORE INSERT ON a
BEGIN                                       FOR EACH ROW
  INSERT INTO b VALUES(NEW.*);                EXECUTE PROCEDURE UDF1();
  RETURN NEW;
END;
$$ language plpgsql;
```

3.1.2 (4 Points, 2/table)

```
CREATE FUNCTION UDF2() RETURNS trigger      CREATE TRIGGER T2
AS $$                                       AFTER INSERT ON a
BEGIN                                       FOR EACH ROW
  INSERT INTO b VALUES(NEW.*);                EXECUTE PROCEDURE UDF2();
  RETURN null;
END;
$$ language plpgsql;
```

4 (10 points) Misc. Questions

4.1 (2 Points)

In at most 2 short sentences, describe the significance of integrity constraints in database management systems as compared to writing code to check constraints within the application.

4.2 (2 Points)

List 2 important properties that the relational model provides that the Network/Hierarchical model does not provide. 4 words MAX for each property.

4.3 (2 Points)

In at most 2 short sentences, describe the difference between VIEW and WITH in SQL.

4.4 (2 Points)

In ONE sentence, explain multiset semantics.

4.5 (2 Points)

Write a creative example of joins In Real Life by filling in the sentence in the answer sheet. Most creative answer (subjectively judged by the staff) gets 2 extra credit points.

5 (14 points) Pass the SQL

The Warriors is the dominant basketball team in the National Basketball Association (NBA). Legend states that their dominance is not due to having a team of four (now five) NBA all stars, but instead due to their focus on passing and unselfish ball handling. Is this really the case? This problem will walk through an analysis to study how long players hold the ball before they pass to their teammates. When a player holds the ball, we say that the player *possesses* the ball.

Consider the following database schema, where `Players` contains information about each basketball player, `Teams` contains information about each team, and `Possessions` contains information about each time a player held the ball and how long the player possessed the ball.

```
Players(                Teams(                  Possessions(
  pid int primary key,    tid int primary key,    id int primary key,
  tid int not null        name text not null,     pid int not null
    references Teams,      westcoast bool not null     references Players,
  name text not null,    )                       -- when the possession started
  age int not null                               time timestamp not null,
                                                  -- number of seconds the player
)                                                 -- held the ball
                                                  held int not null
                                                )
```

5.1 Sports Never Ages (2 Points)

Write the SQL query to find the average age across all players on west coast teams (westcoast is true). The output should be the average age.

5.2 Hold It (4 Points)

Fill in the `CREATE TABLE TimeHeld(pid, name, held` statement by writing a query that computes the average possession time for each player. The average possession time is defined as the average number of seconds that a player possesses the ball before the ball is passed to a teammate. Return the `pid` and `name` of each player, and the player's average possession time.

5.3 Teams That Pass (4 Points)

Fill in the `CREATE TABLE TeamPasses(passer, passee)` statement so it contains the `pids` of the players that passed the ball (`passer`) and the teammate that received the pass (`passee`).

This is defined as `Possessions` where the passer that is in possession of the ball is on the same team as the passee that receives the ball, *and* if the time of the passee's possession is equal to the time of the passer's possession plus the amount of time that the passer held onto the ball. You may assume that `TimeHeld` has been correctly created and may use it in your answer if it helps.

5.4 The Control Tower (6 Points)

Fill in the `CREATE TABLE Control(tid, pid)` statement to identify the players on each team that have passed the ball to every teammate at least once. Return the team `tid` and player `pid`. You may assume that `TimeHeld` and `TeamPasses` have been correctly created and may use it in your answer if it helps.

5.5 The Passing-est (4 Points)

Let the team possession time be the average of the average possession time over all players on the team. Write the SQL query that returns the name of the team with the lowest team possession time. You may assume that `TimeHeld`, `TeamPasses`, and `Control` have been correctly created and may use it in your answer if it helps.