CSE 414 Midterm: Review Session

April, 2024

1 Writing SQL Queries

1. (50 points)

A doctor's office has several doctors, and several patients. They keep a database of appointments with the following schema:

```
Doctor(did,dname,speciality)
Patient(pid,pname,dob)
Visit(did,pid,date)
```

Doctors have an ID, a name, and a speciality (e.g. "orthopedics" or "dermatology" or "neurology" etc). The key is `Doctor.did`.

Patients have an ID, a name, and date of birth (dob). The key is `Patient`.*pid*.

`Visit.did` is a foreign key to `Doctor` and `Visid.pid` is a foreign key to `Patient`.

The attributes `did`, `pid` are integers. All other attributes are strings.

(This space is intentionally left blank))

In some of the questions below we will illustrate with the following test database instance. Please ensure that your answers are correct on *any* database instance, not just this one.

Doctor

did	DName	Speciality
10	Alice	surgeon
20	Bob	ophthalmologist
30	Carol	surgeon
40	David	surgeon
50	Eve	pediatrics

Patient

pid	PName	Dob
100	Agnes	1990-10-10
200	Brian	1998-12-01
300	Cathy	2001-01-15
400	Dylan	2003-11-12

Visit

did	pid	date
20	100	2018-05-10
30	100	2019-05-10
40	200	2020-05-10
10	300	2021-05-10
40	200	2022-01-02

```
Doctor(did,dname,speciality)
Patient(pid,pname,dob)
Visit(did,pid,date)
```

(a) (10 points) Write the sequence of SQL statements necessary to create the tables above. Include all keys or foreign keys declarations.

Write your answer here:

```
Doctor(did,dname,speciality)
Patient(pid,pname,dob)
Visit(did,pid,date)
```

For this and the remaining questions you may assume that the dates are stored in a normalize form as `yyyy-mm-dd`, for example `2019-01-05` means January, 5th, 2019. This applies to both `Patient.dob` and `Visit.date`.

(b) (10 points) Write a SQL query that computes, for each patient, the number of visits before 2021. Return the pid's, their names, and the number of times they visited before 2021.

On our test database instance your query should return the following result. Notice that the count for Cathy is 0, because her visit was during 2021.

Doctor

did	DName	Speciality
10	Alice	surgeon
20	Bob	ophthalmologist
30	Carol	surgeon
40	David	surgeon
50	Eve	pediatrics

Patient

pid	PName	Dob
100	Agnes	1990-10-10
200	Brian	1998-12-01
300	Cathy	2001-01-15
400	Dylan	2003-11-12

Visit

did	pid	date
20	100	2018-05-10
30	100	2019-05-10
40	200	2020-05-10
10	300	2021-05-10
40	200	2022-01-02

Query answer

pid	PName	count
100	Agnes	2
200	Brian	1
300	Cathy	0
400	Dylan	0

Write your answer here:

```
Doctor(did,dname,speciality)
Patient(pid,pname,dob)
Visit(did,pid,date)
```

(c) (10 points) Write a SQL query that returns all patients who have visited both a `surgeon` and an `ophthalmologist`. Return their PID, name, and date of birth.

On our test database instance your query should return:

Doctor

did	DName	Speciality
10	Alice	surgeon
20	Bob	ophthalmologist
30	Carol	surgeon
40	David	surgeon
50	Eve	pediatrics

Patient

pid	PName	Dob
100	Agnes	1990-10-10
200	Brian	1998-12-01
300	Cathy	2001-01-15
400	Dylan	2003-11-12

Visit

did	pid	date
20	100	2018-05-10
30	100	2019-05-10
40	200	2020-05-10
10	300	2021-05-10
40	200	2022-01-02

Query answer

pid	PName	dob
100	Agnes	1990-10-10

Write your answer here:

```
Doctor(did,dname,speciality)
Patient(pid,pname,dob)
Visit(did,pid,date)
```

(d) (20 points) Write a SQL query that retrieves, for each speciality, the name of the doctor with that speciality that had the first (earliest) appointment. Return the DID, name and speciality of the doctor, as well as his/her date of the earliest appointment. In case of ties (meaning: if there are two or more doctors in the same speciality that had the same earliest appointment), return all of them.

On our test database instance your query should look as follows. Notice that pediatrics is not included in the answer because there were no visits to pediatrics.

Doctor

did	DName	Speciality
10	Alice	surgeon
20	Bob	ophthalmologist
30	Carol	surgeon
40	David	surgeon
50	Eve	pediatrics

Patient

pid	PName	Dob
100	Agnes	1990-10-10
200	Brian	1998-12-01
300	Cathy	2001-01-15
400	Dylan	2003-11-12

Visit

did	pid	date
20	100	2018-05-10
30	100	2019-05-10
40	200	2020-05-10
10	300	2021-05-10
40	200	2022-01-02

Query answer

did	DName	speciality	date
20	Bob	ophthalmologist	2018-05-10
30	Carol	surgeon	2019-05-10

Write your answer here:

2 Analyzing SQL Queries

2. (20 points)

We will use the same database schema as in Part 1.

(a) Which of the following SQL queries are monontone?

i. (2 points) Is this query monotone?

```
select distinct x.did, x.dname
from doctor x, visit y
where not (x.did = y.did and y.date < ’2000’);
```

i. ______________

Yes or No?

ii. (2 points) Is this query monotone?

```
select x.did, x.dname
from doctor x
where exists
        (select *
         from visit y
         where x.did = y.did
           and y.date < ’2020’);
```

ii. ______________

Yes or No?

iii. (2 points) Is this query monotone?

```
select x.did, x.dname
from doctor x
where exists
        (select *
         from visit y
         where not (x.did = y.did
                    and y.date < ’2020’));
```

iii. ______________

Yes or No?

```
Doctor(did,dname,speciality)
Patient(pid,pname,dob)
Visit(did,pid,date)
```

iv. (2 points) Is this query monotone?

```
select x.did, x.dname
from doctor x
where not exists
        (select *
         from visit y
         where x.did = y.did
           and y.date < '2020');
```

iv. __________

Yes or No?

v. (2 points) Is this query monotone?

```
select x.did, x.dname
from doctor x
where not exists
        (select *
         from visit y
         where not (x.did = y.did
                    and y.date < '2020'));
```

v. __________

Yes or No?

```
Doctor(did,dname,speciality)
Patient(pid,pname,dob)
Visit(did,pid,date)
```

(b) We want to answer the following question:

> Retrieve all patients who did not visit the doctor's office before 2021, except for surgery appointments.

(This space is intentionally left blank)

In other words, we want the patients who did not visit before 2021, but the surgery appointments don't count. On our test database instance, the query should return Brian, Cathy and Dylan: Brian did visit before 2021 (on 2020-05-10) but that was for surgery, Cathy did not visit before 2021 (she only visited on 2021-05-10), while Dylan never visited. On the other hand, Agnes should not be returned because she visited an ophthalmologist on 2018-05-10.

as follows:

Doctor

did	DName	Speciality
10	Alice	surgeon
20	Bob	ophthalmologist
30	Carol	surgeon
40	David	surgeon
50	Eve	pediatrics

Patient

pid	PName	Dob
100	Agnes	1990-10-10
200	Brian	1998-12-01
300	Cathy	2001-01-15
400	Dylan	2003-11-12

Visit

did	pid	date
20	100	2018-05-10
30	100	2019-05-10
40	200	2020-05-10
10	300	2021-05-10
40	200	2022-01-02

Query answer

pid	PName
200	Brian
300	Cathy
400	Dylan

For each of the SQL queries below indicate if it returns the correct answer to this question. In each case, answer *Yes* or *No*.

i. (2 points) Does the query below return the correct answer?

```
select x.pid, x.pname
from patient x
where exists
        (select *
         from visit y, doctor z
         where x.pid = y.pid and y.did = z.did
          and z.speciality =  'surgeon'
          and y.date < '2021-01-01');
```

i. ______________

Yes or No?

ii. (2 points) Does the query below return the correct answer?

```
select x.pid, x.pname
from patient x
where exists
        (select *
         from visit y, doctor z
         where x.pid = y.pid and y.did = z.did
          and not(z.speciality =  'surgeon')
          and y.date < '2021-01-01');
```

ii. ______________

Yes or No?

iii. (2 points) Does the query below return the correct answer?

```
select x.pid, x.pname
from patient x
where not exists
        (select *
         from visit y, doctor z
         where x.pid = y.pid and y.did = z.did
          and z.speciality = 'surgeon'
          and y.date < '2021-01-01');
```

iii. ______________

Yes or No?

iv. (2 points) Does the query below return the correct answer?

```
select x.pid, x.pname
from patient x
where not exists
        (select *
         from visit y, doctor z
         where x.pid = y.pid and y.did = z.did
          and not(z.speciality = ’surgeon’)
          and y.date < ’2021-01-01’);
```

iv. ____________

Yes or No?

v. (2 points) Does the query below return the correct answer?

```
select x.pid, x.pname
from patient x, visit y
where x.pid = y.pid
  and y.date < ’2021-01-01’
  and not exists
        (select *
         from doctor z
         where y.did = z.did
          and not(z.speciality = ’surgeon’));
```

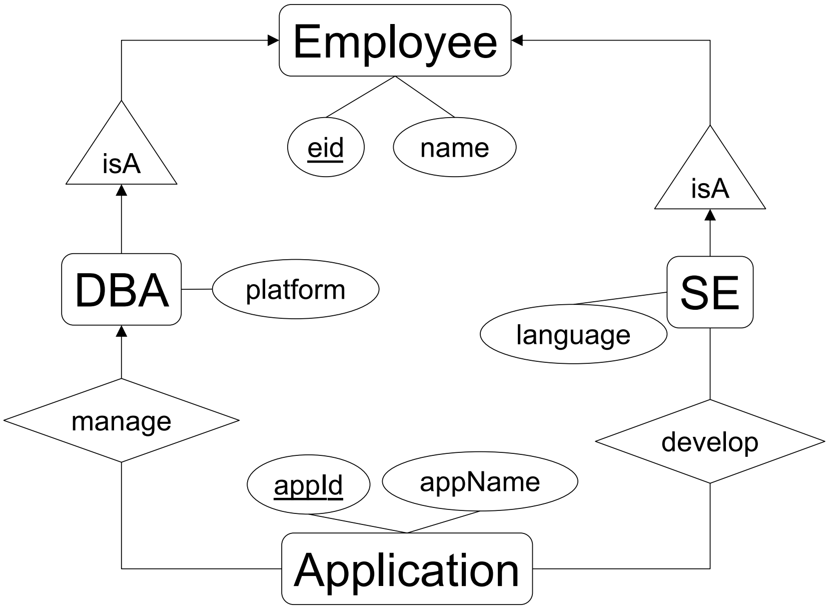
v. ____________

Yes or No?

3 Entity Relationship Diagrams

3. (20 points)

A company develops applications running on top of databases. Software Engineers (SE) develop the applications, and Database Administrators (DBA) manage the databases. This information is captured by the following Entity-Relationship diagram:



(a) For each of the statements below, indicate whether it is true or false:

i. (2 points) A DBA can manage at most one application.

i. ____________

Answer TRUE or FALSE:

ii. (2 points) An application can be managed by at most one DBA.

ii. ____________

Answer TRUE or FALSE:

iii. (2 points) Every application has a `platform` attribute.

iii. ____________

Answer TRUE or FALSE:

iv. (2 points) Every DBA has a `platform` attribute.

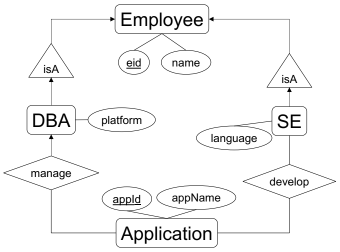
iv. ____________

Answer TRUE or FALSE:

v. (2 points) Every DBA has a `name` attribute.

v. ____________

Answer TRUE or FALSE:



(b) (10 points) Write the SQL statements to create the tables representing the E/R diagram above. The keys `eid, appId` are integers, `name, appName, platform, language` are strings. Include all key and foreign key statements.

Write your answer here:

4 Functional Dependencies

4. (10 points)

   Consider the following same relational schema, where all attributes are integers:

   ```
   R(A,B,C)
   S(A,B,D)
   ```

   (a) Let `Q1(A,B,C)` be the relation returned by the following query:

   ```
   select x.A, x.B, x.C
   from R x
   where x.B = 20;
   ```

   i. (1 point) Compute the closure $A+$ in the relation `Q1(A,B,C)`.

   i. ______________

   ii. (1 point) Compute the closure $B+$ in the relation `Q1(A,B,C)`.

   ii. ______________

   iii. (1 point) Compute the closure $C+$ in the relation `Q1(A,B,C)`.

   iii. ______________

   (b) Let `Q2(A,B,D)` be the relation returned by the following query:

   ```
   select y.A, y.B, y.D
   from S y
   where y.B = y.D;
   ```

   i. (1 point) Compute the closure $A+$ in the relation `Q2(A,B,D)`.

   i. ______________

   ii. (1 point) Compute the closure $B+$ in the relation `Q2(A,B,D)`.

   ii. ______________

   iii. (1 point) Compute the closure $D+$ in the relation `Q2(A,B,D)`.

   iii. ______________

(c) Let `Q3(A,B,C,D)` be the relation returned by the following query:

```
select x.A, x.B, x.C, y.D
from R x, S y
where x.A = y.A and x.B = y.B;
```

i. (1 point) Compute the closure $A+$ in the relation `Q3(A,B,C,D)`.

i. ______________

ii. (1 point) Compute the closure $B+$ in the relation `Q3(A,B,C,D)`.

ii. ______________

iii. (1 point) Compute the closure $C+$ in the relation `Q3(A,B,C,D)`.

iii. ______________

iv. (1 point) Compute the closure $D+$ in the relation `Q3(A,B,C,D)`.

iv. ______________