CS4.301: Data and Applications (Monsoon 2024)

Quiz 3

Maximum Marks: 16, Time: 45 minutes

- **Keep answers concise. State all assumptions.**
- **All questions are compulsory.**

Question 1

An online picture sharing company uses a database with the following schema:

```
create table Usr (
    uid int primary key,
    uname text not null,
    city text not null
);

create table Picture (
    pid int primary key,
    uid int not null references Usr(uid),
    size int not null,
    pdf text
);
```

Every user has a key (uid), a name (uname), and a city.

Every picture has a key (pid), an author (uid) that is a foreign key to Usr, a size, and the pdf content (which is plain text).

Consider the following:

1.1   For each query below, indicate if it is equivalent to the given query. You should answer 'yes' only if the query returns exactly the same answers (simply mention yes/no, no further elaboration is required here)   (0.5 * 4 = 2 marks)

```
SELECT x.uid, x.uname,
    (SELECT count(*)
     FROM Picture y
     WHERE x.uid = y.uid AND y.size > 1000000)
FROM Usr x
WHERE x.city = 'Denver';
```

a. Is this query equivalent?

```
SELECT x.uid, x.uname, count(*)
FROM Usr x, Picture y
WHERE x.uid = y.uid AND y.size > 1000000 AND x.city = 'Denver'
GROUP BY x.uid, x.uname;
```

b. Is this query equivalent?

```
SELECT x.uid, x.uname
FROM Usr x
WHERE x.city = 'Denver' AND exists
    (SELECT *
     FROM Picture y
     WHERE x.uid = y.uid AND y.size > 1000000);
```

c. Is this query equivalent?

```
SELECT x.uid, x.uname, count(y.pid)
FROM Usr x LEFT OUTER JOIN Picture y on x.uid = y.uid AND y.size
    > 1000000
GROUP BY x.uid, x.uname, x.city
having x.city = 'Denver';
```

d. Is this query equivalent?

```
SELECT x.uid, x.uname, count(*)
FROM Usr x LEFT OUTER JOIN Picture y on x.uid = y.uid AND y.size
        > 1000000
GROUP BY x.uid, x.uname, x.city
HAVING x.city = 'Denver';
```

1.2   For each query below indicate if it is equivalent to the given query. You should answer 'yes' only if the query returns exactly the same answers. (simply mention yes/no, no further elaboration is required here)   (0.5 * 4 = 2 marks)

```
SELECT distinct x.uid, x.uname
FROM Usr x, Picture u, Picture v, Picture w
WHERE x.uid = u.uid AND x.uid = v.uid AND x.uid = w.uid
AND u.size > 1000000 AND v.size < 3000000 AND w.size = u.size;
```

a. **Is this query equivalent?**
```
SELECT distinct x.uid, x.uname
FROM Usr x, Picture u, Picture v, Picture w
WHERE x.uid = u.uid AND x.uid = v.uid AND x.uid = w.uid
AND u.size > 1000000 AND v.size < 3000000 AND w.size > 1000000;
```

b. **Is this query equivalent?**
```
SELECT distinct x.uid, x.uname
FROM Usr x, Picture u, Picture v
WHERE x.uid = u.uid AND x.uid = v.uid
AND u.size > 1000000 AND v.size < 3000000;
```

c. **Is this query equivalent?**
```
SELECT distinct x.uid, x.uname
FROM Usr x, Picture u, Picture w
WHERE x.uid = u.uid AND x.uid = w.uid
AND u.size > 1000000 AND u.size < 3000000 AND u.size = w.size;
```

d. **Is this query equivalent?**
```
SELECT distinct x.uid, x.uname
FROM Usr x, Picture u, Picture v, Picture w
WHERE x.uid = u.uid AND x.uid = v.uid AND x.uid = w.uid
AND u.size > 1000000 AND v.size < 3000000 AND w.size = v.size;
```

Question 2 (2)

Consider a table People **with attributes** *SALARY* **and** *NAME*. **Given the relational algebra expression**
$\Pi_{\text{NAME}}(\sigma_{\text{SALARY}>10,000}(\Pi_{\text{NAME}\cup\text{SALARY}}(\text{People})))$, **Which of the following are valid rewrites?**
(Check all that apply.)

(simply circle the option(s) you think are correct, no further elaboration is required here)

a. $\Pi_{\text{NAME}}(\sigma_{\text{SALARY}>10,000}(\text{People}))$

b. $\Pi_{\text{NAME}}(\sigma_{\text{SALARY}>10,000}(\Pi_{\text{SALARY}}(\text{People})))$

c. $\sigma_{\text{SALARY}>10,000}(\Pi_{\text{NAME}\cup\text{SALARY}}(\text{People}))$

d. $\sigma_{\text{SALARY}>10,000}(\Pi_{\text{NAME}}(\Pi_{\text{NAME}\cup\text{SALARY}}(\text{People})))$

e. $\Pi_{\text{NAME}}(\sigma_{\text{SALARY}>10,000}(\Pi_{\text{NAME}}(\Pi_{\text{SALARY}}(\text{People}))))$

Question 3 (2)

Why does SQL not automatically eliminate duplicate tuples in the results of queries? Select all the correct statement(s):

(simply circle the option(s) you think are correct, no further elaboration is required here)

a. **Duplicate elimination is an expensive operation**
b. **When an aggregate function is applied to tuples, in most cases we do not want to eliminate duplicates**
c. **Duplicates are an implementation of redundancy in fault tolerance**
d. **The user may want to see duplicate tuples in the result of a query**

Question 4 (4)

You are given the following database tables:

Employees

EmpID	EmpName	Department	Salary
1	Alice	1	50000
2	Bob	2	70000
3	Charlie	1	60000
4	David	2	80000
5	Eva	3	55000

Departments

DeptID	DeptName
1	HR
2	IT
3	Sales

a. **For the SQL query below, provide the CODING Order (i.e., the logical way to write the query) and the EXECUTION Order (i.e., the order in which the query processor actually executes it):**

```
SELECT Department, COUNT(*) AS EmployeeCount
FROM Employees
WHERE Salary > 60000
GROUP BY Department
HAVING COUNT(*) > 1
ORDER BY EmployeeCount DESC;
```

b. **Examine the two SQL queries below. Identify which one(s) will fail and explain why. For each, write the expected output (if any). Keep the answer brief and concise (around 2-3 lines per query):**

Query A

```
SELECT EmpName, Salary
FROM Employees
WHERE Salary > (SELECT AVG(Salary) FROM Employees)
ORDER BY Salary DESC;
```

Query B

```
SELECT DeptName, COUNT(*) AS TotalEmployees
FROM Departments
JOIN Employees ON Departments.DeptName = Employees.Department
WHERE TotalEmployees > 1
GROUP BY DeptName;
```

Question 5 (4)

You are given two matrices `A`, `B` with schemas:

    A(i,j,v)

    B(j,k,v)

Write an SQL query that computes the product matrix *(A.B)*. **Your query should return a set of triples** `(i, k, v)` **representing the product matrix; you do not need to remove the entries with value 0. Assume 1 indexing of `i`, `j` and `k`. For example:**

$$\begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 4 \\ -2 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 10 \\ -7 & 2 \end{pmatrix}$$

A				B				C		
i	j	v		j	k	v		i	k	v
1	1	2		1	1	1		1	1	0
1	2	1		1	2	4		1	2	10
2	1	−1		2	1	−2		2	1	−7
2	2	3		2	2	2		2	2	2

(Please provide concise 2-3 lines of reasoning/justification for your final answer)