
Databasesystemer, forår 2005
IT Universitetet i København

Forelæsning 9: Mere om SQL

7. april 2005

Forelæser: Rasmus Pagh

— Today's lecture —

- Subqueries in SQL.
- Set operators in SQL.
- Security and authorization in SQL.
- Object orientation in databases and database modeling.

— Subqueries —

Until now, you have seen SQL queries of the form

```
SELECT <list of attributes>
FROM <list of relations>
WHERE <condition>
```

What we haven't used is that:

- In any place where a relation is allowed, we may put an SQL query (a **subquery**) computing a relation.
- In any place where an atomic value is allowed, we may put an SQL subquery computing a relation with one attribute and one tuple.

2

— Subqueries in FROM clauses —

Instead of just relations, we may use SQL queries in the FROM clause of SELECT-FROM-WHERE.

Such a subquery must be an independent query that does not refer to other relations or attributes in the SELECT-FROM-WHERE.

If we need a name for referring to the relation computed by the subquery, a tuple variable is used.

Subqueries should always be surrounded by parentheses.

3

— Subqueries producing scalar values —

When a query produces a relation with one attribute and one tuple, it can be used in any place where we can put an atomic (or **scalar**) value.

Semantics:

In places where an atomic value is expected, SQL regards a relation instance containing one atomic value x to be the same as the value x .

If such a subquery does not result in exactly one tuple, it is a **run-time error**, and the SQL query cannot be completed.

4

— Subqueries in conditions —

One common use of subqueries is in the WHERE part of SELECT-FROM-WHERE. There are several operators in SQL that apply to a relation R and produce a boolean result:

- EXISTS R is true if and only if R is not empty.
- s IN R is true if and only if s is a tuple in R .

If R is **unary** (has just one attribute):

- s > ALL R is true if and only if s is greater than *all* values in R .
- s > ANY R is true if and only if s is greater than *some* value in R .

... and similarly for other comparison operators (<, >=, <=, <>).

5

Correlated subqueries

Sometimes a subquery in a WHERE part depends on (is **correlated** with) tuple variables/relations of the surrounding SELECT-FROM-WHERE.

Semantics:

The query is evaluated once for each **binding** of tuple variables in the surrounding SELECT-FROM-WHEREs.

Scoping rule:

In case several tuple variables/relations have the same name x , an occurrence of x refers to the *closest* such tuple variable/relation.

6

Semantics of SELECT-FROM-WHERE

```
SELECT *  
FROM E1 V1, E2 V2, ..., Ek Vk  
WHERE <condition>
```

1. Determine the values of $E1, \dots, Ek$ (which are subqueries or relations).
2. Form all possible combinations of tuples $V1, \dots, Vk$, where $V1$ comes from $E1$, etc:
 - (a) For each combination determine if $\langle \text{condition} \rangle$ is true.
This may involve computing subqueries. If (attributes from) one of $V1, \dots, Vk$ is referred to in a subquery and is *not* a tuple variable in the subquery, we substitute in the appropriate value.
 - (b) If the condition is true, we output the combination of $V1, \dots, Vk$.

7

— Problem session (5 minutes) —

What does the below SQL query compute?

```
SELECT title, year
FROM Movie
WHERE EXISTS (SELECT *
              FROM Movie M2
              WHERE Movie.year = year + 1
              AND EXISTS (SELECT *
                        FROM Movie M3
                        WHERE M2.year = year + 1));
```

Tip: Read from inside out.

8

— Subqueries and efficiency —

Generally speaking, queries involving subqueries are more difficult to deal with for a DBMS.

A good DBMS will execute most such queries efficiently, but there may be performance problems, especially for correlated subqueries.

Some, but not all, queries can be rewritten as a SELECT-FROM-WHERE with no subquery. This will sometimes improve efficiency.

Example: “Return all students that do not yet have any registered grade.”

In some cases, such as the above, a correlated subquery can be avoided by using SQL’s set operators. (Coming next.)

9

Next: Set operators in SQL

— Set operators in SQL —

SQL offers a number of *set operators* that can be used to combine results of several queries, if the queries return relations with *compatible schemas*.

- $R_1 \text{ UNION } R_2$ gives the relation with all tuples of R_1 and R_2 .
- $R_1 \text{ INTERSECT } R_2$ gives the relation with all tuples can be found in both R_1 and R_2 .
- $R_1 \text{ EXCEPT } R_2$ gives the relation with all tuples can be found in both R_1 but not in R_2 .

For all operators, the result relation will have **no** duplicate tuples.

— Relations in SQL are bags —

Relations in SQL are **bags** (or **multisets**), i.e., tuples may appear more than once.

The reason why bags are used is that *duplicate elimination* is relatively costly (requires time and memory), so it is generally an advantage to use it only when necessary.

We can force duplicate elimination in a SELECT-FROM-WHERE by adding the keyword DISTINCT.

The UNION operator involves duplicate elimination. To avoid duplicate elimination use $R_1 \text{ UNION ALL } R_2$, which corresponds to “putting one relation on top of the other”.

Next: Security and authorization in SQL.

— Database security issues —

- Data loss and unavailability – DBMSs provide good protection.
- Theft, fraud, sabotage – We won't address this issue.
- Data integrity – database constraints and triggers may aid in this direction (cf. also lecture on transactions).
- Loss of privacy/confidentiality – SQL has an *authorization* mechanism.

14

— Authorization in SQL —

Because databases often have many users, not all of which are allowed to do any database operation, SQL has an **authorization** system.

Every user (called a **module** in case the user is a program) has certain rights, or **privileges**, to access and modify database elements. The basic privileges are:

SELECT, INSERT, DELETE, UPDATE

It is possible to have privileges for certain attributes in a relation, e.g., a secretary might have the UPDATE(address, city) privilege for relation with customer information.

15

— Granting privileges —

Basics of managing privileges:

- If a user defines a new schema, she has all possible privileges for the tables (and other database elements) of that schema.
- Users may **grant** (“copy”) privileges of their own to other users.
- Being able to grant a privilege is a special privilege in itself that can be passed on.

Syntax for granting privileges:

```
GRANT <privilege list> ON <database element>  
TO <user list> [WITH GRANT OPTION]
```

16

— Revoking privileges —

Granted privileges can be withdrawn (or **revoked**) by a user at any time.

Basics of revoking privileges:

- Privileges given without the GRANT OPTION can simply be removed.
- Otherwise we would like to revoke any privilege in the database that is *only possible because of the privilege that was revoked*.
- What happens when revoking is *independent* on the order in which privileges were given.

Syntax for revoking privileges:

```
REVOKE <privilege list> ON <database element>  
FROM <user list> CASCADE
```

17

Next: Object orientation in databases and database modeling.

— Object orientation —

Object orientation is a way of thinking about *data* and *behavior*.

Object oriented programming languages (e.g. Java) support object oriented concepts such as:

- Classes, objects
- Inheritance
- Methods
- Polymorphism

UML is an increasingly popular notation that can be used for object-oriented design and analysis, also for database systems.

— Enhanced E-R model vs UML —

Similar features:

- Entity type \approx class.
- Entity instance \approx object (instance).
- Relationship \approx association.
- Associative entity \approx association class.
- Super/subtypes \approx inheritance

Some differences/additional features of UML:

- Implicit primary key in objects.
- Class-scope attributes.
- Notation for associating *actions/behavior* to objects.
- Notation for showing example object instances.

20

— Object-oriented and object-relational databases

Around 15 years ago many people believed that object-oriented databases and query languages would take over the world.

Several data models and query languages have been proposed (see MDM ch. 15), but have found limited use.

Instead, backwards-compatible *object-relational* databases dominate the market, integrating essential parts of object-oriented languages into SQL:

- Ability to store code with data.
- Ability to define new, complex data types.

It seems likely that object-relational hybrids will continue to dominate.

21

— Stored procedures in SQL —

The major DBMSs, and the SQL-99 standard, give a way of including functions and procedures in the database schema.

This means that tuples can be thought of as objects with methods.

It is not clear to what extent this will replace embedded and dynamic SQL in systems communicating with DBMSs.

22

— Most important points in this lecture —

As a minimum, you should after this week:

- Be able to understand and form SQL expressions using several levels of subqueries.
- Tell the difference between correlated and uncorrelated subqueries.
- Understand the mechanism for granting and revoking privileges in SQL.
- Have knowledge of the basic tools for implementing OO concepts in object-relational DBMSs.

23