Structured Query Language SQL

Lecture 6

SELECT FROM WHERE
Structured Query Language (SQL)

- SQL is a high-level special-purpose language for manipulating relations
- SQL is mostly a declarative language:
  you declare what you want without specifying how you want to get answer
- SQL provides a limited set of operations:
  mostly implementations of Relational Algebra operators
- SQL programmer needs to focus on readability and on getting the right results – do not need to worry about efficiency:
  because the DMBS optimizes every query and chooses the most efficient implementation for each operation
Sub-sets of SQL

- Data Manipulation Language (DML): INSERT, UPDATE, DELETE, SELECT, Transaction control: COMMIT, ROLLBACK
- Data Definition Language (DDL): CREATE, ALTER, DROP, RENAME
- Data Control Language (DCL): GRANT, REVOKE
Language elements

- Clauses
- Expressions - produce either scalar values, or tables
- Predicates - specify conditions that can be evaluated according to SQL three-valued logic (3VL) to true/false/unknown
- Queries
- Statements
SELECT clause corresponds to projection $\pi$ in RA
Query: list student names with GPA >3

<table>
<thead>
<tr>
<th>Student</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>GPA</td>
</tr>
<tr>
<td>Bob</td>
<td>3</td>
</tr>
<tr>
<td>John</td>
<td>3</td>
</tr>
<tr>
<td>Tom</td>
<td>3.5</td>
</tr>
<tr>
<td>Maria</td>
<td>4</td>
</tr>
</tbody>
</table>

$S = \pi_{name} \sigma_{gpa>3}(Student)$

<table>
<thead>
<tr>
<th>$S$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>GPA</td>
</tr>
<tr>
<td>Tom</td>
<td>3.5</td>
</tr>
<tr>
<td>Maria</td>
<td>4</td>
</tr>
</tbody>
</table>
How the query is evaluated

- Each tuple of *Student* is inspected
- Each attribute of WHERE clause is substituted with the actual tuple value
- The condition is then evaluated, and if true – this tuple is added to the output relation

<table>
<thead>
<tr>
<th>Name</th>
<th>GPA</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>3</td>
<td>Canada</td>
</tr>
<tr>
<td>John</td>
<td>3</td>
<td>Britain</td>
</tr>
<tr>
<td>Tom</td>
<td>3.5</td>
<td>Canada</td>
</tr>
<tr>
<td>Maria</td>
<td>4</td>
<td>Mexico</td>
</tr>
</tbody>
</table>

WHERE 3 > 3
FALSE

WHERE 4 > 3
TRUE
How to parse SQL query

SELECT a,b
FROM X,Y,Z
WHERE X.c=Y.c AND Z.d > 12

1. What relations are involved: FROM clause
2. Selection condition on rows: WHERE clause
3. Projection on columns: SELECT clause
FROM clause
FROM clause

FROM is always followed by name(s) of input relation(s):

SELECT * FROM Student
FROM clause: sub-queries

- You can construct a new relation using a sub-query, give it a name (optional in most DBMSs), and use it in FROM clause.

- Thus, the result of one query (sub-query) becomes an input to another.

<table>
<thead>
<tr>
<th>Name</th>
<th>GPA</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>3</td>
<td>Canada</td>
</tr>
<tr>
<td>John</td>
<td>3</td>
<td>Britain</td>
</tr>
<tr>
<td>Tom</td>
<td>3.5</td>
<td>Canada</td>
</tr>
<tr>
<td>Maria</td>
<td>4</td>
<td>Mexico</td>
</tr>
</tbody>
</table>

SELECT name FROM (SELECT * FROM Student WHERE gpa > 3) AS goodStudent
FROM clause: table alias I

- We can **rename** input relations and their attributes to use in SELECT and WHERE clauses

- In that way we can perform queries on self-relationships

```
SELECT e.name [AS] employee, s.name [AS] supervisor
FROM Faculty AS e, Faculty AS s
WHERE e.SupID = s.ID
```

<table>
<thead>
<tr>
<th>Faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>
FROM clause: table alias II

- We can **rename** input relations and their attributes to use in SELECT and WHERE clauses

- Or perform join of table with itself

```sql
SELECT S1.name, S2.name
FROM Student S1, Student S2
WHERE S1.address = S2.address
AND S1.name < S2.name;
```

<table>
<thead>
<tr>
<th>Student</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Address</td>
</tr>
<tr>
<td>Bob</td>
<td>Canada 1</td>
</tr>
<tr>
<td>John</td>
<td>Britain 2</td>
</tr>
<tr>
<td>Tom</td>
<td>Canada 1</td>
</tr>
<tr>
<td>Maria</td>
<td>Britain 2</td>
</tr>
</tbody>
</table>
Producing a new table from multiple tables
FROM clause: list of tables

• List of tables without any condition in the WHERE clause produces ...

```
<table>
<thead>
<tr>
<th>Student</th>
<th>Professor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name</td>
</tr>
<tr>
<td>Bob</td>
<td>Dr. Monk</td>
</tr>
<tr>
<td>John</td>
<td>Dr. Pooh</td>
</tr>
<tr>
<td>Tom</td>
<td>Dr. Patel</td>
</tr>
<tr>
<td>Maria</td>
<td></td>
</tr>
</tbody>
</table>
```

SELECT * FROM Student, Professor
Unexpected result?

<table>
<thead>
<tr>
<th>Student</th>
<th>Professor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name</td>
</tr>
<tr>
<td>Bob</td>
<td>Dr. Monk</td>
</tr>
<tr>
<td>John</td>
<td>Dr. Pooh</td>
</tr>
<tr>
<td>Tom</td>
<td>Dr. Patel</td>
</tr>
<tr>
<td>Maria</td>
<td></td>
</tr>
</tbody>
</table>

```
SELECT * FROM Student, Professor
T=Student x Professor
```
FROM clause: list of tables - warning

• List of tables without any condition in the WHERE clause produces **Cartesian product**

The implicit writing of Cartesian product - a dangerous illusion that you are asking the list of Professors to be appended to the end of the list of students, while in fact you are asking to pair each tuple in Student with each tuple in Professor
Combination of 2 tables: Cartesian product in SQL

• Results from multi-table query that does not have a WHERE clause

• The product results in a huge output which normally is not very useful

• To avoid a Cartesian product, we use one or more valid join conditions
Joins: NATURAL JOIN

<table>
<thead>
<tr>
<th>Student</th>
<th>RegisteredFor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name</td>
</tr>
<tr>
<td>Canada</td>
<td>Topic</td>
</tr>
<tr>
<td>GPA</td>
<td>Topic</td>
</tr>
<tr>
<td>Bob</td>
<td>Algorithms</td>
</tr>
<tr>
<td>John</td>
<td>Algorithms</td>
</tr>
<tr>
<td>Tom</td>
<td>Algorithms</td>
</tr>
<tr>
<td>Maria</td>
<td>Python</td>
</tr>
<tr>
<td></td>
<td>Python</td>
</tr>
<tr>
<td></td>
<td>Databases</td>
</tr>
<tr>
<td></td>
<td>Databases</td>
</tr>
<tr>
<td></td>
<td>Databases</td>
</tr>
<tr>
<td></td>
<td>GUI</td>
</tr>
<tr>
<td></td>
<td>GUI</td>
</tr>
</tbody>
</table>

SELECT *  
FROM Student **NATURAL JOIN** RegisteredFor;

More explicit:

SELECT *  
FROM Student **JOIN** RegisteredFor **USING** (name);
If you want to join only on a single common attribute – specify it with USING:

SELECT name, Teacher.score, Student.score
FROM Teacher JOIN Student USING (name);
SELECT name
FROM Teacher JOIN Student
ON Teacher.score > Student.score
AND Teacher.name = Student.name
Multiple joins are required to collect information from multiple tables

<table>
<thead>
<tr>
<th>Student</th>
<th></th>
<th>Teaches</th>
<th></th>
<th>RegisteredFor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Country</td>
<td>GPA</td>
<td>Name</td>
<td>Topic</td>
</tr>
<tr>
<td>Bob</td>
<td>Canada</td>
<td>3</td>
<td>Dr. Monk</td>
<td>Algorithms</td>
</tr>
<tr>
<td>John</td>
<td>Britain</td>
<td>3</td>
<td>Dr. Pooh</td>
<td>Python</td>
</tr>
<tr>
<td>Tom</td>
<td>Canada</td>
<td>3.5</td>
<td>Dr. Patel</td>
<td>Databases</td>
</tr>
<tr>
<td>Maria</td>
<td>Mexico</td>
<td>4</td>
<td>Dr. Patel</td>
<td>GUI</td>
</tr>
</tbody>
</table>

It is preferably to write joining attributes explicitly, using WHERE clause - to avoid mistakes:

```
SELECT s.name AS student, r.topic AS course, t.name AS professor
FROM Student s, RegisteredFor r, Teaches t
WHERE s.name = r.name
AND r.topic = t.topic
```
NULL values in joined columns

We use NULL to indicate:

• Value unknown
• Value inapplicable
• Value withheld
NULL is a special value

• When joining on condition involving attributes A and B:
  • If both A and B are NULL:
    • A=B returns false
    • A<>B returns false
  • If one of A or B is NULL
    • A=B returns false
    • A<>B returns false
• The reason is that DBMS uses a 3-valued logic – discussion on slides 36-37
• The NULLs do not generally appear in the results of joins
OUTER JOIN

• Preserves dangling tuples (that did not match any tuple in another table) by padding them with NULL

• Has 3 types:
  • **Full**: Pad dangling tuples in both tables.
    • L FULL OUTER JOIN R
  
  • **Left outerjoin**: Only pad dangling tuples of L.
    • L LEFT OUTER JOIN R

  • **Right outerjoin**: Only pad dangling tuples of R.
    • L RIGHT OUTER JOIN R
Keywords INNER and OUTER

• There are keywords INNER and OUTER, but you never need to use them.

• Your intentions are clear anyway:
  • You get an OUTER join iff you use the keywords LEFT, RIGHT, or FULL.
  • If you don’t use these keywords you get an inner join – normal join.
OUTER JOIN example: LEFT JOIN

```sql
SELECT t.name, country 
FROM Teacher t 
  LEFT JOIN Student s 
  ON t.name = s.name
```
OUTER JOIN example: FULL JOIN

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Student</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name</td>
<td>Name</td>
</tr>
<tr>
<td>Bob</td>
<td>Bob</td>
<td>Bob</td>
</tr>
<tr>
<td>John</td>
<td>John</td>
<td>John</td>
</tr>
<tr>
<td>Tom</td>
<td>Tom</td>
<td>Tom</td>
</tr>
<tr>
<td>Kim</td>
<td>Maria</td>
<td>Kim</td>
</tr>
<tr>
<td>Score</td>
<td>Score</td>
<td>Score</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3.5</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>NULL</td>
<td>3</td>
</tr>
</tbody>
</table>

SELECT *
FROM Teacher t FULL JOIN Student s
ON t.name = s.name
Subquery or Join?

• We can achieve the same result by using both subqueries and joins

• Which one is better?

• The one which is more readable – both queries will be parsed and optimized into the same code by DBMS
Example 2

• What does this do?
  SELECT studentID, courseID, grade
  FROM Took,
  (SELECT *
   FROM Offering
   WHERE instructor='David') Doffering
  WHERE Took.courseID = Doffering.courseID;

• Can you suggest another version?
WHERE clause
WHERE clause

The predicates (conditions) can be written using:

• Column names
• Logical and comparison operators
• Mathematical expressions
• Constants
• Built-in DBMS functions
• Sub-queries
Building Boolean expressions

• We can build Boolean expressions with operators that produce Boolean results.
  • comparison operators:  =, <>, <, >, <=, >=
  • and many other operators: see section 6.1.2 of the text and chapter 9 of the PostgreSQL documentation.

• Compound conditions are constructed using logical operators:  AND, OR, NOT.
Checking for NULLs

- Can’t meaningfully use = or <>
- Should use:
  - IS NULL
  - IS NOT NULL

```
SELECT *
FROM Students
WHERE age IS NOT NULL;
```
Operations involving NULL

• A tuple is in a query result iff the WHERE clause evaluates to TRUE.

• When we compare using any comparison operators: (for example $a < b$), and $a$ or $b$ or both are NULL, the result is UNKNOWN – the third truth value, SQL special

• But a query only produces a tuple in the answer if its truth value for the WHERE clause is TRUE (not FALSE or UNKNOWN).
3-valued truth of databases

Rule to remember:
TRUE = 1, FALSE = 0, UNKNOWN (NULL) = ½
AND: min, OR: max, NOT: 1-x

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x AND y (min)</th>
<th>x OR y (max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FALSE (0)</td>
<td>FALSE (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FALSE (0)</td>
<td>NULL(½ )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FALSE (0)</td>
<td>TRUE (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NULL(½ )</td>
<td>NULL(½ )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NULL(½ )</td>
<td>TRUE (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRUE (1)</td>
<td>TRUE (1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

x | NOT x (1-x) |
---|-------------|
FALSE (0) |             |
NULL(½ )  |              |
TRUE (1)  |              |
3-valued truth of databases

Rule to remember:
TRUE = 1, FALSE = 0, UNKNOWN (NULL) = ½
AND: min, OR: max, NOT: 1-x

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x AND y (min)</th>
<th>x OR y (max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FALSE (0)</td>
<td>FALSE (0)</td>
<td>FALSE (0)</td>
<td>FALSE (0)</td>
</tr>
<tr>
<td>FALSE (0)</td>
<td>NULL(½ )</td>
<td>FALSE (0)</td>
<td>NULL(½ )</td>
</tr>
<tr>
<td>FALSE (0)</td>
<td>TRUE (1)</td>
<td>FALSE (0)</td>
<td>TRUE (1)</td>
</tr>
<tr>
<td>NULL(½ )</td>
<td>NULL(½ )</td>
<td>NULL(½ )</td>
<td>NULL(½ )</td>
</tr>
<tr>
<td>NULL(½ )</td>
<td>TRUE (1)</td>
<td>NULL(½ )</td>
<td>TRUE (1)</td>
</tr>
<tr>
<td>TRUE (1)</td>
<td>TRUE (1)</td>
<td>TRUE (1)</td>
<td>TRUE (1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x</th>
<th>NOT x (1-x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FALSE (0)</td>
<td>TRUE (1)</td>
</tr>
<tr>
<td>NULL(½ )</td>
<td>NULL(½ )</td>
</tr>
<tr>
<td>TRUE (1)</td>
<td>FALSE (0)</td>
</tr>
</tbody>
</table>
Example

SELECT *
FROM course
WHERE year <=3 OR year >3

Meaning:

SELECT *
FROM course
WHERE year is NOT NULL

<table>
<thead>
<tr>
<th>Topic</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Databases</td>
<td>3</td>
</tr>
<tr>
<td>HTML</td>
<td></td>
</tr>
<tr>
<td>GUI</td>
<td>2</td>
</tr>
</tbody>
</table>
Comparison of strings

• Strings can be compared (lexicographically) using the same operators:

\[=\]
\[<>\]
\[<\]
\[>\]
\[\leq\]
\[\geq\]

\text{BETWEEN } A \text{ and } B \quad \text{is equivalent to } \geq A \text{ and } \leq B

<table>
<thead>
<tr>
<th>Student</th>
<th>Country</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>Canada</td>
<td>3</td>
</tr>
<tr>
<td>John</td>
<td>Britain</td>
<td>3</td>
</tr>
<tr>
<td>Joan</td>
<td>Canada</td>
<td>3.5</td>
</tr>
<tr>
<td>Maria</td>
<td>Mexico</td>
<td>4</td>
</tr>
</tbody>
</table>

\begin{verbatim}
SELECT * 
FROM student 
WHERE name <= 'John'
\end{verbatim}

\begin{verbatim}
SELECT * 
FROM student 
WHERE name > 'Job'
\end{verbatim}
Comparison of dates

- Default date data type format in PostgreSQL is ‘YYYY-MM-DD’: for example ‘1990-04-12’

- Dates can be compared against string literal using function `to_date`

```
SELECT name
FROM student
WHERE birthdate < to_date('28-03-1989','DD-MM-YYYY')
```

<table>
<thead>
<tr>
<th>Student</th>
<th>Birthdate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>‘1990-12-04’</td>
</tr>
<tr>
<td>John</td>
<td>‘1987-11-30’</td>
</tr>
<tr>
<td>Joan</td>
<td>‘1993-12-09’</td>
</tr>
<tr>
<td>Maria</td>
<td>‘1989-02-28’</td>
</tr>
</tbody>
</table>
Patterns

• General form:
  \(<Attribute> \text{ LIKE } \text{<pattern>}\>
  or
  \(<Attribute> \text{ NOT LIKE } \text{<pattern>}\>

• \(<pattern>\) is a quoted string which may contain
  
  \(\%\) = meaning “any string”
  
  \(\_\) = meaning “any character.”

```
SELECT *
FROM student
WHERE name LIKE 'Jo%';
```
Patterns: apostrophe

- Two consecutive apostrophes represent one apostrophe and not the end of the string.

<table>
<thead>
<tr>
<th>Student</th>
<th>Name</th>
<th>Birthdate</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>Bob</td>
<td>‘1990-12-04’</td>
<td>Mike’s brother</td>
</tr>
<tr>
<td>John</td>
<td>John</td>
<td>‘1987-11-30’</td>
<td></td>
</tr>
<tr>
<td>Joan</td>
<td>Joan</td>
<td>‘1993-12-09’</td>
<td>John’s sister</td>
</tr>
<tr>
<td>Maria</td>
<td>Maria</td>
<td>‘1989-02-28’</td>
<td></td>
</tr>
</tbody>
</table>

SELECT name
FROM student
WHERE comment LIKE '%''s%';
Patterns: % and _

- What if the pattern contains the characters % or _?
  We should “escape” their special meaning preceding them by some escape character. SQL allows us to use a custom escape character.

- Syntax: s LIKE 'x%%x%%' ESCAPE 'x'
  x will be the escape character.
  Example of matched string: '%aaaa%bb'

<table>
<thead>
<tr>
<th>Student</th>
<th>Name</th>
<th>Birthdate</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>‘1990-12-04’</td>
<td></td>
<td>Mike’s brother</td>
</tr>
<tr>
<td>John</td>
<td>‘1987-11-30’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joan</td>
<td>‘1993-12-09’</td>
<td></td>
<td>John’s sister</td>
</tr>
<tr>
<td>Maria</td>
<td>‘1989-02-28’</td>
<td></td>
<td>m_1</td>
</tr>
</tbody>
</table>

SELECT name
FROM student
WHERE comment LIKE 'my_%'
ESCAPE ‘y';
Pattern example with dates

- Born in 1980s:

```sql
SELECT name
FROM student
WHERE Birthdate > '1979-12-31'
AND Birthdate < '1990-01-01'
```

- We can use LIKE:

```sql
SELECT name
FROM student
WHERE Birthdate LIKE '__8%'
```

<table>
<thead>
<tr>
<th>Name</th>
<th>Birthdate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>'1990-12-04'</td>
</tr>
<tr>
<td>John</td>
<td>'1987-11-30'</td>
</tr>
<tr>
<td>Joan</td>
<td>'1993-12-09'</td>
</tr>
<tr>
<td>Maria</td>
<td>'1989-02-28'</td>
</tr>
</tbody>
</table>
PostgreSQL-specific escaping

• PostgreSQL also accepts "escape" string constants - extension to the SQL standard.

• *An escape string constant* is specified by writing letter E before the opening single quote: e.g. *E'foo'*.  

• Within an escape string, a backslash character (\) begins a C-like backslash escape sequence:
  
  \n for newline
  \t for a tab etc.

• Any other character following a backslash is taken literally.
  include a backslash character - write two backslashes (\\).
  Include a single quote –write \', in addition to the standard way of "
Conditions involving lists

SELECT name
FROM student
WHERE country = ‘Canada’
  OR country = ‘Britain’
  OR country = ‘Australia’

• Instead:

SELECT name
FROM Student
WHERE country IN (‘Canada’, ‘Australia’, ‘Britain’)

<table>
<thead>
<tr>
<th>Name</th>
<th>Country</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>Canada</td>
<td>3</td>
</tr>
<tr>
<td>John</td>
<td>Britain</td>
<td>3</td>
</tr>
<tr>
<td>Tom</td>
<td>Canada</td>
<td>3.5</td>
</tr>
<tr>
<td>Maria</td>
<td>Mexico</td>
<td>4</td>
</tr>
</tbody>
</table>
WHERE clause

The conditions can be written using
• Column names
• Logical and comparison operators
• Mathematical expressions
• Constants
• Built-in DBMS functions
• Sub-queries
Sub-queries in WHERE clause

• We can compare the value in the column in the current tuple to a value in another column (of the same tuple)
• We can also compare it to the result of a subquery

• Syntax:
  • The subquery must be parenthesized.
  • Must name the result (in PostgreSQL), so you can refer to it in the outer query.
Subquery as a value in a WHERE clause

- If a subquery is guaranteed to produce exactly one tuple, then the subquery can be used as a value.
- Simplest situation: that one tuple has only one component.
Example

• Find all students with a gpa greater than that of John.

```
SELECT name
FROM Student
WHERE gpa >
(SELECT gpa
 FROM Student
 WHERE name = 'John');
```

• This is analogous to something we can’t do in RA:

```
\pi_{\text{name}} \sigma_{\text{gpa} > (\text{subquery})} \text{Student}
```
What if the subquery returns more than one value?

• We can make comparisons using a special quantifier.

```sql
SELECT name
FROM Student
WHERE gpa >
  (SELECT gpa
   FROM Student
   WHERE name = 'John');
```

• We can require that

```sql
  gpa >= all of them, or
gpa > at least one of them.
```
SQL operators on subquery that returns multiple tuples - to produce a Boolean result

- ANY
- ALL
- IN
- EXISTS

- These operators can be negated by putting NOT in front of the entire expression.
ANY

• Suppose subquery returns relation R. If R is a unary relation (on a single column) then

• Condition \( s > \text{ANY} \ R \) is true if \( s \) is greater than at least one value in unary relation \( R \).

  • Similarly we can use any other comparison operators in place of \( > \). For instance, \( s = \text{ANY} \ R \) is the same as \( s \text{ IN} \ R \).

• If R is not unary we could match the entire tuple, but this feature is not supported by most DBMSs.
• Suppose subquery returns relation R.

• $s > \text{ALL } R$ is true if $s$ is greater than every value in the unary (one column) relation $R$.

• Similarly, the $>$ operator could be replaced by any other comparison operator with the analogous meaning. For instance, $s <> \text{ALL } R$ is the same as $s \text{ NOT IN } R$. 
Example with ANY

```
SELECT name
FROM student
WHERE GPA > ANY
    (SELECT GPA
     FROM Student)
```

What is the result?

<table>
<thead>
<tr>
<th>Name</th>
<th>Country</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>Canada</td>
<td>3</td>
</tr>
<tr>
<td>John</td>
<td>Britain</td>
<td>3</td>
</tr>
<tr>
<td>Tom</td>
<td>Canada</td>
<td>3.5</td>
</tr>
<tr>
<td>Maria</td>
<td>Mexico</td>
<td>4</td>
</tr>
</tbody>
</table>
Example with ANY

SELECT name
FROM student
WHERE GPA > ANY
(SELECT GPA
FROM Student)

• "Any" sounds a lot like "every" in this query. But it means "any one or more".

• Remember that ANY is existentially quantified.

• This query sounds much more like what it actually is when we express it instead with the keyword SOME, which is a synonym for ANY in SQL.
Example with ALL

```
SELECT name
FROM student
WHERE GPA > ALL
  (SELECT GPA
   FROM student)
```

What is the result?

<table>
<thead>
<tr>
<th>Name</th>
<th>Country</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>Canada</td>
<td>3</td>
</tr>
<tr>
<td>John</td>
<td>Britain</td>
<td>3</td>
</tr>
<tr>
<td>Tom</td>
<td>Canada</td>
<td>3.5</td>
</tr>
<tr>
<td>Maria</td>
<td>Mexico</td>
<td>4</td>
</tr>
</tbody>
</table>
Example with ALL

SELECT name
FROM student
WHERE GPA >= ALL (SELECT GPA
FROM Student)

What is the result?

<table>
<thead>
<tr>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Bob</td>
</tr>
<tr>
<td>John</td>
</tr>
<tr>
<td>Tom</td>
</tr>
<tr>
<td>Maria</td>
</tr>
</tbody>
</table>
• For subquery $R$:

  • $s \text{ IN } R$ is true if $s$ is equal to one of the tuples in $R$. Likewise, $s \text{ NOT IN } R$ is true if and only if $s$ is equal to no tuple in $R$.

  • $s$ can be a list of attributes and the entire tuple is compared
Example with IN

SELECT name
FROM
Student
WHERE country IN
(SELECT countryName
FROM EnglishSpeakingCountries)

<table>
<thead>
<tr>
<th>Name</th>
<th>Country</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>Canada</td>
<td>3</td>
</tr>
<tr>
<td>John</td>
<td>Britain</td>
<td>3</td>
</tr>
<tr>
<td>Tom</td>
<td>Canada</td>
<td>3.5</td>
</tr>
<tr>
<td>Maria</td>
<td>Mexico</td>
<td>4</td>
</tr>
</tbody>
</table>
Exercise

Suppose we have tables R(a, b) and S(b, c).

1. What does this query do?
   
   ```sql
   SELECT a
   FROM R
   WHERE b IN (SELECT b FROM S);
   ```

2. Can we express this query without using IN?
EXISTS

- For subquery R:
  - **EXISTS R** is a condition that is true if R is not empty.
  - Read it as “exists at least one row in the subquery result”
Correlated subqueries

• EXISTS (NOT EXISTS) are used with correlated subqueries

• The EXISTS operator checks if the inner query returns at least one row, and it returns TRUE or FALSE

• If a subquery refers only to names defined inside it, it can be evaluated once and used repeatedly in the outer query.

• If it refers to any name defined outside of itself, it must be evaluated once for each tuple in the outer query. These are called correlated subqueries.
Example 1: EXISTS

```
SELECT Teacher.Name
FROM Teacher outer
WHERE EXISTS
  (SELECT '1'
   FROM Student
   WHERE name = outer.name);
```

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tom</td>
<td>4</td>
</tr>
<tr>
<td>Kim</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student</th>
<th>Country</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob</td>
<td>Canada</td>
<td>3</td>
</tr>
<tr>
<td>John</td>
<td>Britain</td>
<td>3</td>
</tr>
<tr>
<td>Tom</td>
<td>Canada</td>
<td>3.5</td>
</tr>
<tr>
<td>Maria</td>
<td>Mexico</td>
<td>4</td>
</tr>
</tbody>
</table>
Example 2: EXISTS

SELECT name, gpa
FROM Student
WHERE EXISTS ( 
    SELECT *
    FROM Took
    WHERE Student.name = Took.name
    AND grade > 85
);
Example 3: EXISTS

```
SELECT DISTINCT Course
FROM Took
WHERE EXISTS (
    SELECT *
    FROM Took t, Offering o
    WHERE
        t.course = o.course AND
        t.course <> Took.course AND
        o.dept = 'CSC' AND
        took.name = t.name );
```
SELECT clause
Expressions in SELECT clauses

• Instead of a simple projection, you can use an expression in a SELECT clause.

• Operands: attributes, constants
  Operators: arithmetic ops, string ops

• Examples:
  ```sql
  SELECT name, grade+10 AS adjusted
  FROM Took;
  ```

  ```sql
  SELECT dept||course
  FROM Offering;
  ```
Operations involving NULL

• If we operate with arithmetic operators on two values: $a + b$ – and $a$ is NULL, the result is NULL
Substituting NULL’s in SELECT

• The Postgre `coalesce` function converts a NULL value to an actual value supplied as an argument
  `coalesce (column, value)`

• Coalesce evaluates the arguments in order and returns the current value of the first expression that initially does not evaluate to NULL

Examples:
  `coalesce (comission,0)`
  `coalesce (prerequisites, ‘None’)`
Current date

SELECT CURRENT_DATE;

• Example: computing age (Approximate)
SELECT (CURRENT_DATE – birthdate)/365.25
FROM student

Function $AGE$ computes number of years and months between 2 dates, if 1 argument – default is the current date

SELECT $age$ (birthdate)
FROM student;
DISTINCT
Relations can have duplicates in SQL

• A table can have duplicate tuples, unless this would violate an integrity constraint.

• And SELECT-FROM-WHERE statements leave duplicates in unless you say not to.

• Why?
  • Getting rid of duplicates is expensive!
  • We may want the duplicates because they tell us how many times something occurred.

• To eliminate duplicates need to explicitly use DISTINCT:
  
  SELECT DISTINCT *
  FROM R;
Bags

• SQL treats tables as “bags” (or “multisets”) rather than sets.
• Bags are just like sets, but duplicates are allowed.
• \{6, 2, 7, 1, 9\} is a set (and a bag)
  \{6, 2, 2, 7, 1, 9\} is not a set, but is a bag.
• Like with sets, order doesn’t matter.
  \{6, 2, 7, 1, 9\} = \{1, 2, 6, 7, 9\}
Impact of null values on DISTINCT

• Does SELECT DISTINCT treat two NULLs as the same?

create table X(a int, b int);
insert into X values (1, 2), (null, 3), (null, 4);
select * from X

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>
Impact of null values on DISTINCT

• Does SELECT DISTINCT treat two NULLs as the same?

create table X(a int, b int);
insert into X values (1, 2), (null, 3), (null, 4);
select * from X

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>
Impact of null values on DISTINCT

• If we ask for distinct values, the two NULLs are collapsed to one - SELECT DISTINCT has considered the two NULL values to be the same.

```
select distinct a from x;
 a
---
  1
```

(2 rows)

• This behavior is DBMS-dependent
ORDER BY clause
ORDER BY

• To put the tuples in order, add this as the final clause: ORDER BY «attribute list» [DESC]

• The default is ascending order; DESC overrides it to force descending order.

• The attribute list can include expressions: e.g., ORDER BY sales+rentals

• The ordering is the last thing done before the SELECT, so all attributes are still available.
Bonus: TOP-N analysis

• Top-N queries are used to sort rows in a table and then to find the first-N largest (smallest) values

• In PostgreSQL and in SQLite:

```
SELECT gpa, name FROM Student
ORDER BY gpa DESC
LIMIT 5
```
Example 1: TOP-4 largest rooms

SELECT Building, RoomNo, Capacity
FROM location
ORDER BY Capacity DESC
LIMIT 4;
Example 2: TOP-3 lowest salaries

SELECT Lname, Fname, Salary
FROM employee
ORDER BY Salary
LIMIT 3;