A correlated subquery is a subquery that references columns from the outer query. The subquery is executed once for each row processed by the outer query. This makes correlated subqueries powerful for certain types of queries but also potentially slower because the subquery runs multiple times.

Let's create an example to help students understand correlated subqueries. We'll use a simple dataset of students and their grades.

keyboard\_arrow\_down

Creating the Dataset

students Table

CREATE TABLE students (  
    student\_id INT PRIMARY KEY,  
    student\_name VARCHAR(100)  
);  
  
INSERT INTO students (student\_id, student\_name) VALUES  
(1, 'Alice Johnson'),  
(2, 'Bob Smith'),  
(3, 'Charlie Brown'),  
(4, 'David Wilson'),  
(5, 'Eva Green');

grades Table

CREATE TABLE grades (  
    grade\_id INT PRIMARY KEY,  
    student\_id INT,  
    subject VARCHAR(100),  
    grade DECIMAL(5, 2),  
    FOREIGN KEY (student\_id) REFERENCES students(student\_id)  
);  
  
INSERT INTO grades (grade\_id, student\_id, subject, grade) VALUES  
(1, 1, 'Math', 85.00),  
(2, 1, 'English', 92.00),  
(3, 2, 'Math', 78.00),  
(4, 2, 'English', 81.00),  
(5, 3, 'Math', 89.00),  
(6, 3, 'English', 95.00),  
(7, 4, 'Math', 70.00),  
(8, 4, 'English', 75.00),  
(9, 5, 'Math', 95.00),  
(10, 5, 'English', 98.00);

Example Correlated Subquery

**Question**: Find the students who have scored above the average grade in any subject.

Query with Correlated Subquery

SELECT   
    student\_name,  
    subject,  
    grade  
FROM   
    grades g1  
WHERE   
    grade > (  
        SELECT   
            AVG(grade)  
        FROM   
            grades g2  
        WHERE   
            g1.subject = g2.subject  
    );

**Explanation**:

1. The outer query selects student\_name, subject, and grade from the grades table aliased as g1.
2. The subquery calculates the average grade for the same subject (g1.subject = g2.subject) from the grades table aliased as g2.
3. The WHERE clause of the outer query checks if the student's grade is greater than the average grade for that subject.

Understanding the Correlated Subquery

* The subquery (SELECT AVG(grade) FROM grades g2 WHERE g1.subject = g2.subject) is executed once for each row of the outer query.
* For each grade in the outer query, the subquery calculates the average grade of the corresponding subject.
* If a student's grade is higher than the average grade of the subject, that student's details are included in the result set.

Example Output

Given the dataset, the result of the query would be:

| **student\_name** | **subject** | **grade** |
| --- | --- | --- |
| Alice Johnson | English | 92.00 |
| Charlie Brown | Math | 89.00 |
| Charlie Brown | English | 95.00 |
| Eva Green | Math | 95.00 |
| Eva Green | English | 98.00 |

These students scored above the average grade in the respective subjects.

Summary

* **Correlated Subqueries**: A subquery that uses values from the outer query.
* **Execution**: The subquery is run once for each row in the outer query.
* **Use Cases**: Useful for comparing each row against a set of aggregated data, like finding rows that meet certain criteria relative to the average, maximum, or minimum values in a group.

This example helps students understand how correlated subqueries work by using a practical and easy-to-follow scenario.

<https://www.scaler.com/topics/sql/correlated-subquery/>