JOHNS HOPKINS
CAREY BUSINESS SCHOOL

Database Management

BU.330.770

Session 7 (Part II)

Instructor: Changmi Jung, Ph.D.

Subqueries

Session Objectives (1/2)

- Determine when using a subquery is appropriate
- Identify which clauses can contain subqueries
- Distinguish between an outer query and a subquery
- Distinguish between single-row and multiple-row comparison operators
- Use a single-row subquery and multiple-row subquery in
  - a WHERE clause
  - a HAVING clause
  - a SELECT clause

Session Objectives (2/2)

- Create an inline view using a multiple-column subquery in a FROM clause
- Distinguish between correlated and uncorrelated subqueries
- Nest a subquery inside another subquery
- Use a subquery in a DML action

Subqueries and Their Uses

- Subquery – a query nested inside another query
- Used when a query is based on an unknown value
- Requires SELECT and FROM clauses
- Must be enclosed in parentheses
- Place on the right side of a comparison operator (if used in WHERE or HAVING clauses)

Types of Subqueries

Subquery	Description
Single-row subquery	Returns to the outer query **one row** of results that consists of **one column**
Multiple-row subquery	Returns to the outer query **more than one row** of results
Multiple-column subquery	Returns to the outer query **more than one column** of results
Correlated subquery	References a column in the outer query, and executes the subquery once for every row in the outer query
Uncorrelated subquery	Executes the subquery first and passes the value to the outer query

Single-Row Subqueries

- Can only return one result to the outer query
- Operators include =, >, <, >=, <=, < >

Single-Row Subquery in a WHERE Clause

- Used for comparison against individual data

Worksheet | Query Builder

```
SELECT title, cost, category
FROM books
WHERE cost > (SELECT cost
              FROM books
              WHERE title = 'DATABASE IMPLEMENTATION')
and category = 'COMPUTER';
```

Script Output | Query Result

SQL | All Rows Fetched: 2 in 0.074 seconds

	TITLE	COST	CATEGORY
1	HOLY GRAIL OF ORACLE	47.25	COMPUTER
2	E-BUSINESS THE EASY WAY	37.9	COMPUTER

Let's display all computer books at a higher cost than the 'Database Implementation.'

We don't know the cost of 'Database Implementation,' so we need to retrieve it first

Single-Row Subquery in a HAVING Clause

- Required when returned value is compared to grouped data

Worksheet | Query Builder

```
SELECT category, AVG(retail-cost) "Average Profit"
FROM books
GROUP BY category
HAVING AVG(retail-cost) > (SELECT AVG(retail-cost)
                           FROM books
                           WHERE category = 'LITERATURE');
```

Query Result

SQL | All Rows Fetched: 2 in 0.018 seconds

	CATEGORY	Average Profit
1	FAMILY LIFE	24.875
2	COMPUTER	18.2625

Let's display all book categories with a higher average profit than the Literature category.

Need to query the average profit of the books in the Literature category.

Single-Row Subquery in a SELECT Clause

- Replicates subquery value for each row displayed

Worksheet | Query Builder

```
SELECT title, retail,
       (SELECT ROUND(AVG(retail), 2) FROM books) "Overall Avg Retail",
        retail - (SELECT TO_CHAR(AVG(retail), '999.99') FROM books) "Difference"
FROM books;
```

Query Result

SQL | All Rows Fetched: 14 in 0.024 seconds

	TITLE	RETAIL	Overall Avg Retail	Difference
1	BODYBUILD IN 10 MINUTES A DAY	30.95	40.98	-10.03
2	REVENGE OF MICKEY	22	40.98	-18.98
3	BUILDING A CAR WITH TOOTHPICKS	59.95	40.98	18.97
4	DATABASE IMPLEMENTATION	55.95	40.98	14.97
5	COOKING WITH MUSHROOMS	19.95	40.98	-21.03
6	HOLY GRAIL OF ORACLE	75.95	40.98	34.97
7	HANDCRANKED COMPUTERS	25	40.98	-15.98
8	E-BUSINESS THE EASY WAY	54.5	40.98	13.52
9	PAINLESS CHILD-REARING	89.95	40.98	48.97
10	THE WOK WAY TO COOK	28.75	40.98	-12.23
11	BIG BEAR AND LITTLE DOVE	8.95	40.98	-32.03
12	HOW TO GET FASTER PIZZA	29.95	40.98	-11.03
13	HOW TO MANAGE THE MANAGER	31.95	40.98	-9.03
14	SHORTEST POEMS	39.95	40.98	-1.03

Multiple-Row Subqueries

- Return more than one row of results
- Require use of IN, ANY, ALL, or EXISTS operators in the clause

good point

Since it returns multiple rows, you cannot use a single-value operator like =, >, etc.

Multiple-Row Subquery with IN

Worksheet | Query Builder

```
SELECT title, retail, category
FROM books
WHERE retail IN (SELECT MAX(retail)
                 FROM books
                 GROUP BY category)
ORDER BY category;
```

Query Result

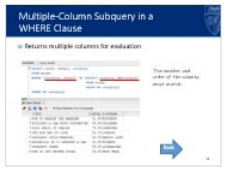
SQL | All Rows Fetched: 8 in 0.023 seconds

	TITLE	RETAIL	CATEGORY
1	HOW TO MANAGE THE MANAGER	31.95	BUSINESS
2	BUILDING A CAR WITH TOOTHPICKS	59.95	CHILDREN
3	HOLY GRAIL OF ORACLE	75.95	COMPUTER
4	THE WOK WAY TO COOK	28.75	COOKING
5	PAINLESS CHILD-REARING	89.95	FAMILY LIFE
6	BODYBUILD IN 10 MINUTES A DAY	30.95	FITNESS
7	SHORTEST POEMS	39.95	LITERATURE
8	HOW TO GET FASTER PIZZA	29.95	SELF HELP

Can we identify books with a retail value matching the highest retail value for any book category?

Need to identify the highest retail price in each category first

Can use **IN** operator or **=ANY** they are same.



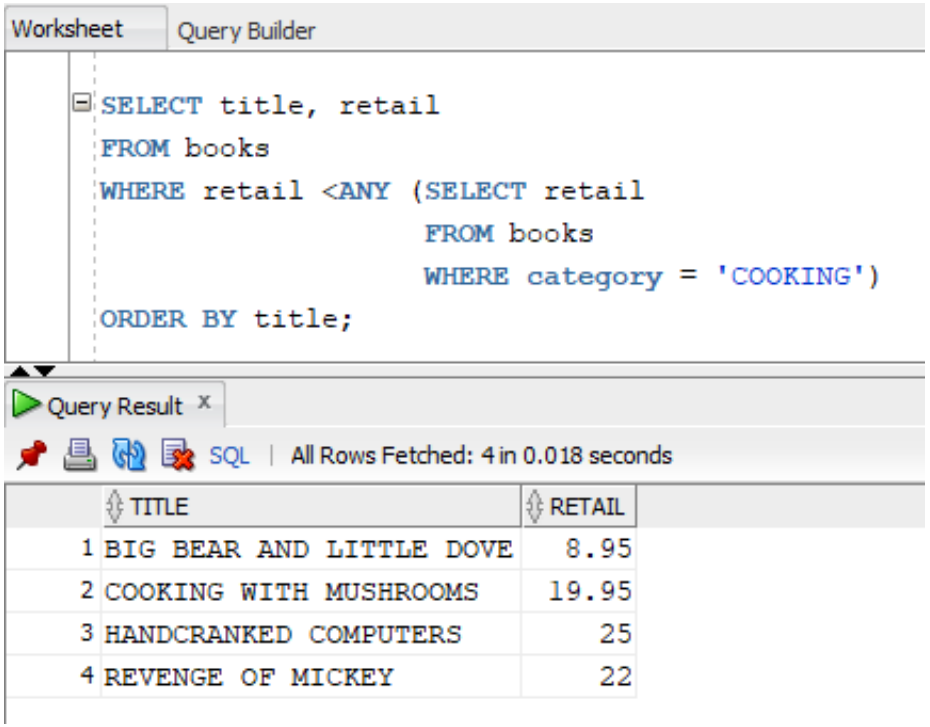
ANY and ALL Operators

- Combine with arithmetic operators to treat a subquery's results as a set of values instead of single values.

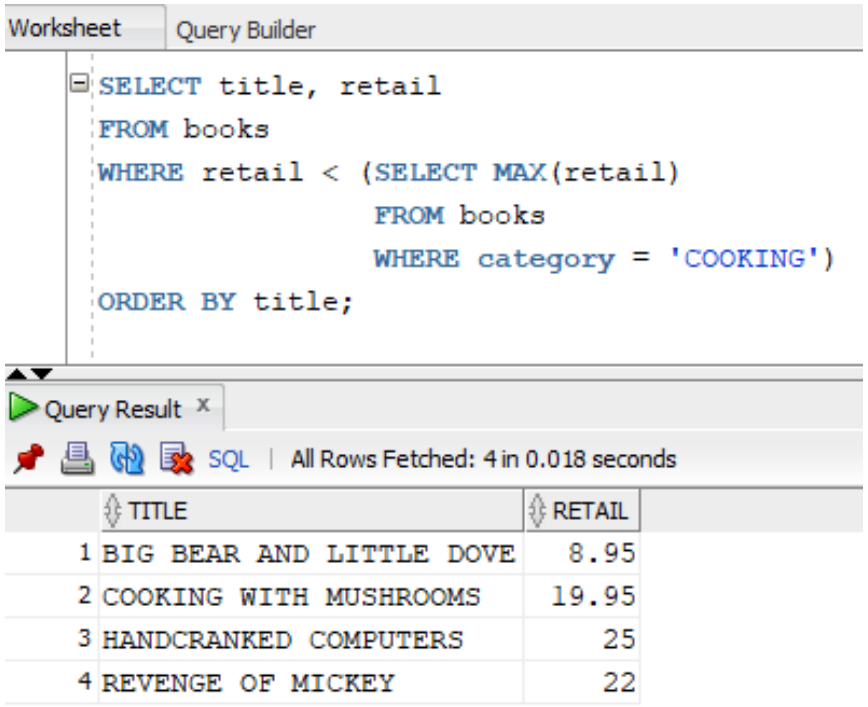
Operator	Description
>ALL	More than the highest value returned by the subquery
<ALL	Less than the lowest value returned by the subquery
>ANY	More than the lowest value returned by the subquery
<ANY	Less than the highest value returned by the subquery
=ANY	Equal to any value returned by the subquery (same as IN)

Imp

Multiple-Row Subquery with ANY



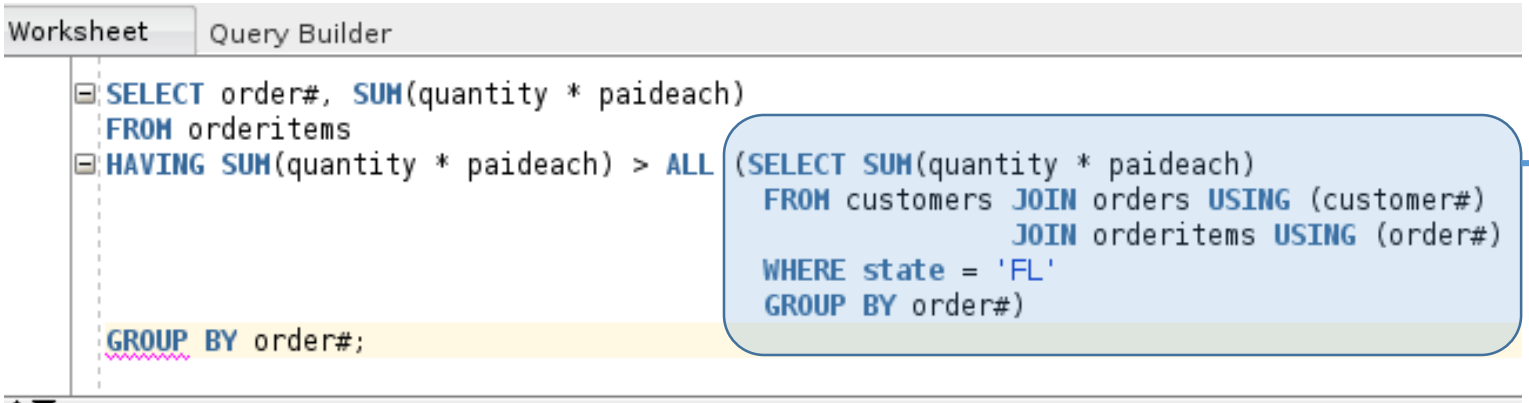
	TITLE	RETAIL
1	BIG BEAR AND LITTLE DOVE	8.95
2	COOKING WITH MUSHROOMS	19.95
3	HANDCRANKED COMPUTERS	25
4	REVENGE OF MICKEY	22



	TITLE	RETAIL
1	BIG BEAR AND LITTLE DOVE	8.95
2	COOKING WITH MUSHROOMS	19.95
3	HANDCRANKED COMPUTERS	25
4	REVENGE OF MICKEY	22

BTW, the max (retail) from the Cooking category is $28.75.

Multiple-Row Subquery with ALL in a HAVING Clause



Total sales amount for each order from Florida-based customers.

Query Result

SQL | All Rows Fetched: 5 in 0.061 seconds

	ORDER#	SUM(QUANTITY*PAIDEACH)
1	1001	117.4
2	1002	111.9
3	1004	170.9
4	1007	335.85
5	1012	166.4

Find all orders with amounts greater than every order amount from Florida-based customers.

Multiple-Column Subqueries

- Return more than one column to the outer query
- Can return more than one row
- When used in WHERE or HAVING clause:
  - Use the **IN** operator for WHERE and HAVING clauses
  - Syntax: **WHERE (col1, col2, …) IN (Select … Subquery)**
  - The column list on the left side of an operator must be in parentheses
  - Column names listed in the WHERE clause must be in the same order as they're listed in the subquery's SELECT clause

Multiple-Column Subquery in a FROM Clause

- Creates a temporary table that can be referenced by other clauses of the outer query
- The temporary table is called an **inline view**

Worksheet | Query Builder

```
SELECT b.title, b.retail, a.category, a.cat_average
FROM books b, (SELECT category, AVG(retail) cat_average
               FROM books
               GROUP BY category) a
WHERE b.category = a.category
AND b.retail > a.cat_average
ORDER BY b.category, b.title;
```

Query Result

SQL | All Rows Fetched: 6 in 0.041 seconds

	TITLE	RETAIL	CATEGORY	CAT_AVERAGE
1	BUILDING A CAR WITH TOOTHPICKS	59.95	CHILDREN	34.45
2	DATABASE IMPLEMENTATION	55.95	COMPUTER	52.85
3	E-BUSINESS THE EASY WAY	54.5	COMPUTER	52.85
4	HOLY GRAIL OF ORACLE	75.95	COMPUTER	52.85
5	THE WOK WAY TO COOK	28.75	COOKING	24.35
6	PAINLESS CHILD-REARING	89.95	FAMILY LIFE	55.975

Suppose we need a list of all books from the BOOKS table with a selling price higher than the average selling price of books within the same category.

Multiple-Column Subquery in a WHERE Clause

- Returns multiple columns for evaluation

Worksheet | Query Builder

```
SELECT title, retail, category
FROM books
WHERE (category, retail) IN (SELECT category, MAX(retail)
                             FROM books
                             GROUP BY category)
ORDER BY category;
```

Query Result

SQL | All Rows Fetched: 8 in 0.03 seconds

	TITLE	RETAIL	CATEGORY
1	HOW TO MANAGE THE MANAGER	31.95	BUSINESS
2	BUILDING A CAR WITH TOOTHPICKS	59.95	CHILDREN
3	HOLY GRAIL OF ORACLE	75.95	COMPUTER
4	THE WOK WAY TO COOK	28.75	COOKING
5	PAINLESS CHILD-REARING	89.95	FAMILY LIFE
6	BODYBUILD IN 10 MINUTES A DAY	30.95	FITNESS
7	SHORTEST POEMS	39.95	LITERATURE
8	HOW TO GET FASTER PIZZA	29.95	SELF HELP

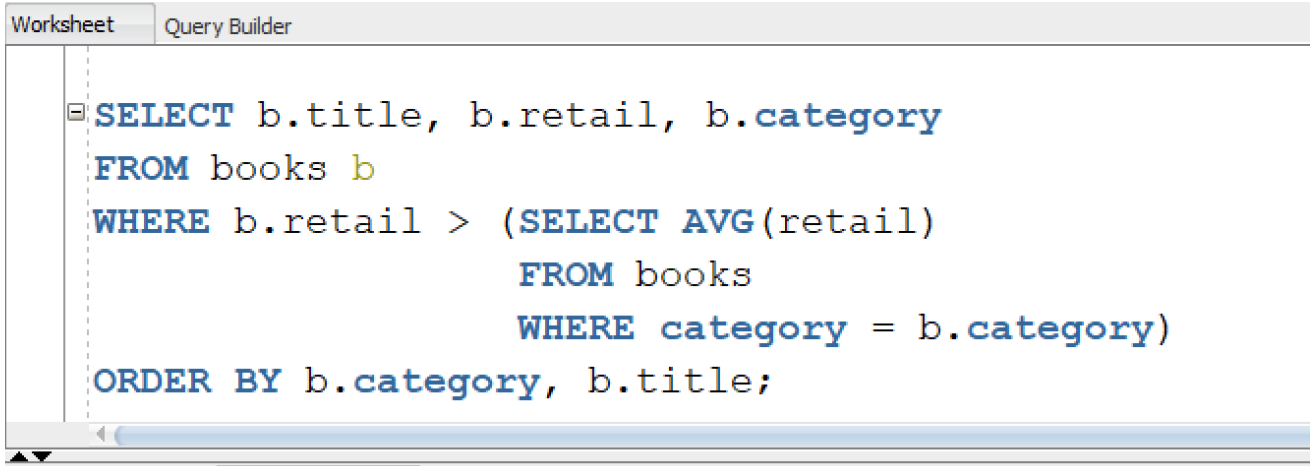
The number and order of the columns must match.

Back

Uncorrelated vs. Correlated Subqueries

- Uncorrelated
  - Inner query is executed first
  - The result is passed to the outer query
  - Outer query is executed
- Correlated
  - Inner query is executed once for each row processed by the outer query
  - Inner query references the row contained in the outer query

Correlated Subqueries Example 1



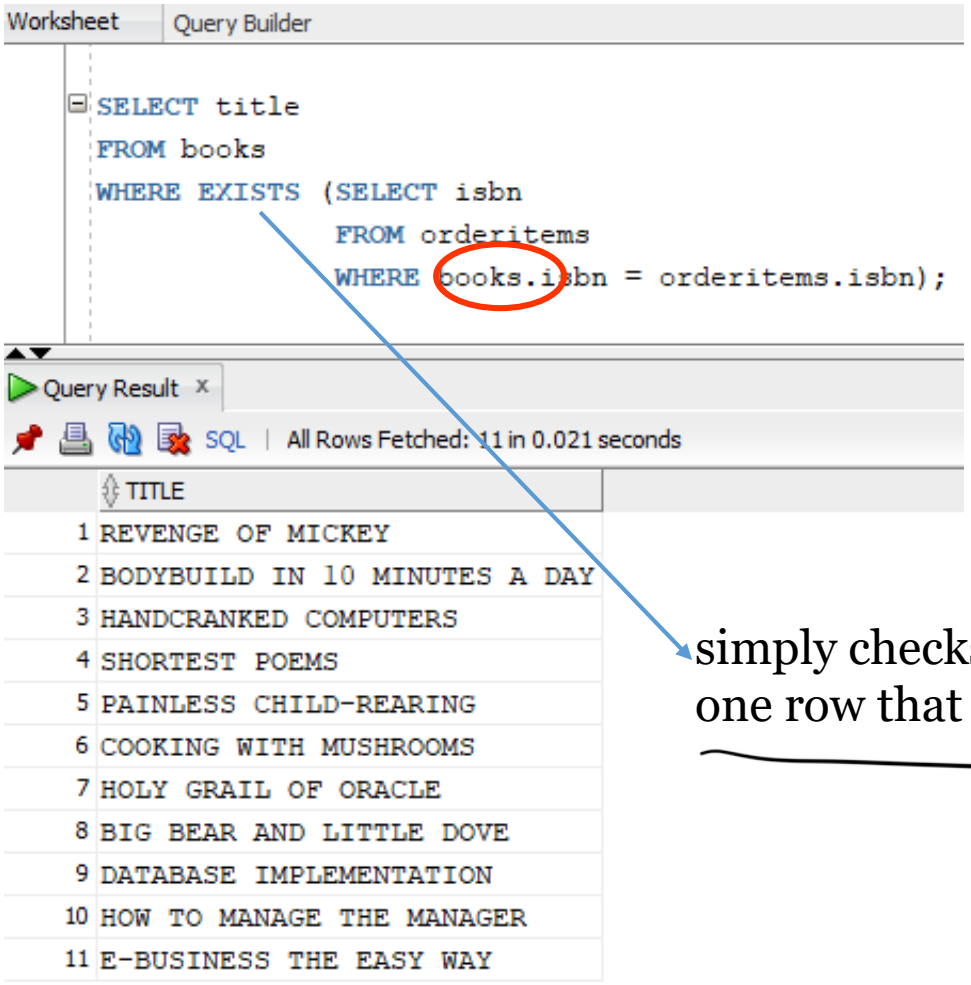
Script Output | Query Result

SQL | All Rows Fetched: 6 in 0.034 seconds

	TITLE	RETAIL	CATEGORY
1	BUILDING A CAR WITH TOOTHPICKS	59.95	CHILDREN
2	DATABASE IMPLEMENTATION	55.95	COMPUTER
3	E-BUSINESS THE EASY WAY	54.5	COMPUTER
4	HOLY GRAIL OF ORACLE	75.95	COMPUTER
5	THE WOK WAY TO COOK	28.75	COOKING
6	PAINLESS CHILD-REARING	89.95	FAMILY LIFE

Convert the previous uncorrelated subquery to a correlated one.

Correlated Subqueries Example 2



Suppose that we need a list of the books that have been ordered.

simply checks for the existence of at least one row that matches the condition.

Nested Subqueries

- Maximum of 255 subqueries if nested in the WHERE clause
- No limit if nested in the FROM clause
- Innermost subquery is resolved first, then the next level, etc.

Nested Subqueries: Example

- Innermost is resolved first (A), then the second level (B), then the outer query (C)

Worksheet | Query Builder

```
SELECT customer#, c.lastname, c.firstname
FROM customers c JOIN orders o USING(customer#)
WHERE o.order# IN (SELECT order#
                   FROM orderitems
                   GROUP BY order#
                   HAVING COUNT(*) = (SELECT MAX(COUNT(*))
                                      FROM orderitems
                                      GROUP BY order#));
```

C B A

Query Result

SQL | All Rows Fetched: 2 in 0.025 seconds

	CUSTOMER#	LASTNAME	FIRSTNAME
1	1007	GIANA	TAMMY
2	1017	NELSON	BECCA

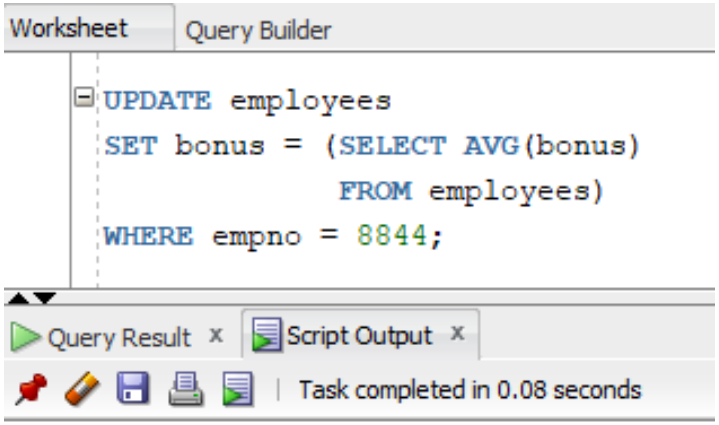
1) Nested Subquery A Identifies the most items in one order.

2) The value of the highest count of items ordered is passed to the subquery B

3) Subquery B identifies which orders have the same number of items as the highest number of items found by A

4) After B identifies the order number, it's passed to the outer query, C, which determines the customer number and name of the person who placed the orders. In this case, two customers tied for placing an order with the most items.

Subquery in a DML action




```
1 row updated.
```

Let's update Sue Stuart's bonus to the average employee bonus amount.

Recollect we used a subquery for INSERT INTO statement as well!

Summary (1/2)

- A subquery is a complete query nested in the SELECT, FROM, HAVING, or WHERE clause of another query
  - The subquery must be enclosed in parentheses and have a SELECT and a FROM clause, at a minimum
- Subqueries are completed first; the result of the subquery is used as input for the outer query
- A single-row subquery can return a maximum of one value
- Single-row operators include =, >, <, >=, <=, and <>
- Multiple-row subqueries return more than one row of results

Summary (2/2)

- Operators that can be used with multiple-row subqueries include IN, ALL, ANY, and EXISTS
- Multiple-column subqueries return more than one column to the outer query
- Correlated subqueries reference a column contained in the outer query
- With nested subqueries, the innermost subquery is executed first, then the next highest level subquery is executed, and so on, until the outermost query is reached

Knowledge Check

Join at

slido.com

#716 354