Chapter 4: Filtering, Join, and Subquery

Analytics Tensor

Mahesh KC

mahesh.kc@analyticstensor.com

https://analyticstensor.com

Introduction

- Filtering
- Evaluate Condition
- Types of Condition
  - Equality condition
  - Inequality condition
  - Range condition
  - Membership condition
  - Matching condition
- Join
- Subquery
- Wrap-up

Filtering

Filtering is used to filter dataset based on business logic. They are used for multiple purpose such as:-

Creating subset of data.

Removing unwanted or duplicate records.

Updating/Deleting records based on certain condition.

Filtering always refers to *where clause* in SQL statement. For grouping it refer s to *having clause*.

Evaluate Condition

Condition is composed of one or more *expressions* with multiple *operators*. The expression can be:
- Number, like 1, 100
- Column in a table or view
- String literal, like 'Virginia', 'Engineer'
- Built-in function, upper('Java')
- Subquery
- List of expression, like ('C#', 'PHP', 'Javascript')

The operators are:
Comparison Operators, like =, !, <, >, !=, LIKE, IN, BETWEEN
Arithmetic Operators, like +, - , *, /
Boolean Operators, like AND, OR, NOT

Evaluate Condition (cont..)

Table 2: **AND**, **OR** operator with three conditions

Condition	Result
WHERE true AND (true OR true)	True
WHERE true AND (true OR false)	True
WHERE true AND (false OR true)	True
WHERE true AND (false OR false)	False
WHERE false AND (true OR true)	False
WHERE false AND (true OR false)	False
WHERE false AND (false OR true)	False
WHERE false AND (false OR false)	False

Table 1: **OR** operator with two conditions

Condition	Result
WHERE true OR true	True
WHERE true OR false	True
WHERE false OR true	True
WHERE false OR false	False

For example:
WHERE state = 'VA' AND purchase_date > '2019-07-12' -- both condition must be true.
WHERE state = 'VA' OR purchase_date > '2019-07-12' -- either one condition must be true.

Evaluate Condition (cont..)

Table 3: **AND**, **OR** and **NOT** operator with three conditions

Condition	Result
WHERE true AND NOT (true OR true)	False
WHERE true AND NOT (true OR false)	False
WHERE true AND NOT (false OR true)	False
WHERE true AND NOT (false OR false)	True
WHERE false AND NOT (true OR true)	False
WHERE false AND NOT (true OR false)	False
WHERE false AND NOT (false OR true)	False
WHERE false AND NOT (false OR false)	False

NOT is the negation operator. It can also be written as ! operator. For example:

WHERE state != 'VA'

If there are more than three conditions then you should use parentheses to make sql statement clear which also helps to make your code easier to read.
For example:
WHERE total_quantity is not NULL AND (state = 'VA' AND purchase_date > '2019-07-12') -- all condition must be true.

Types of Condition

There are various type of conditions to filter data. Such as:

- Equality condition
- Inequality condition
- Range condition
- Membership condition
- Matching condition

All the condition above have their own usages.

Equality Condition

Equality condition equate one expression to another. It has a form of
column_name = expression

Both value must match to satisfy the equality condition.
For example:
position = 'Developer'
id = 10115
price= 25

Inequality Condition

Inequality condition is the negation of equality condition. It ensure that two expression are not equal. It has a form of *column_name != expression* or *column_name <> expression*

The value must not match with each other.

For example:

city != 'Richmond'

dept_no <> 10

amount != 100

Range Condition

Range condition is used to evaluate an expression which satisfies the data within a certain range. It is basically used with numeric and temporal data types.

BETWEEN operator is used to filter data between two range. It requires both lower limit and upper limit of range. For example:

select student_id, fname, lname from student where start_date between '2015-01-01' AND '2018-01-01' ;

Select * from employees where dept_no between 10 and 30;

Membership Condition

Member condition is used to choose a set of finites values rather than just a single values or range of values.

IN operators is used to specify the set of values. This operators is the combination of multiple OR operators.

NOT IN is the reverse of IN operator.

For example:
select * from employees where dept_no in (10, 30, 50)
which can also be written as
select * from employees where dept_no = 10 OR dept_no = 30 or dept_no = 50

Matching Condition

Matching condition is used for partial string matching. It uses wildcard character to match string. Wildcards are used to search character or substring in a string at any positing either beginning, ending, or anywhere within the string.

Table 4: Wildcard characters

Wildcard character	Matches
_	Exactly one character
%	Any number of characters (including 0)

Matching Condition (cont..)

Table 5: String Search Expression

Search expression	Interpretation
S%	Strings beginning with S
%a	Strings ending with a
%jan%	Strings containing the substring 'jan'
_ _i_	Four-character strings with a i in the third position
_ _ _-_ _-_ _ _	11-character strings with dashes in the fourth and seventh positions

JOIN

Join is used to retrieve data from multiple tables in relational structure.

Types of Join
- Inner Join
- Left outer Join
- Right outer Join
- Full outer Join
- Self Join
- Cross Join

Subquery

Subquery is a query contained within a SQL statement. It is a query within a query. Subquery is always enclosed within parentheses. Subquery return dataset that contains:

- Single row with a single column
- Multiple rows with a single column
- Multiple rows and column

Types of Subquery
- Noncorrelated subquery
- Correlated subquery

Subquery (cont..)

Noncorrelated Subquery: The query that does not reference anything from the containing (outer) statement is known as noncorrelated subquery. These queries are independent of their containing statements.

Correlated Subquery: Correlated subquery is dependent on its containing statement from which it references one or more columns. A correlated subquery is executed once for each candidate row.

Exists operator

Wrap-up

- Important points.
- Q & A.

Assignment-5

TBD