Handling Null and Blank Values

Monday, July 22, 2024 8:19 PM

It's crucial to discern that NULL is distinctly different from zero or a blank space; it illustrates a field that doesn't have any value. Therefore, NULL value doesn't simply pose an aesthetic issue in your table – it can lead to significant complications when setting up your data for examination.

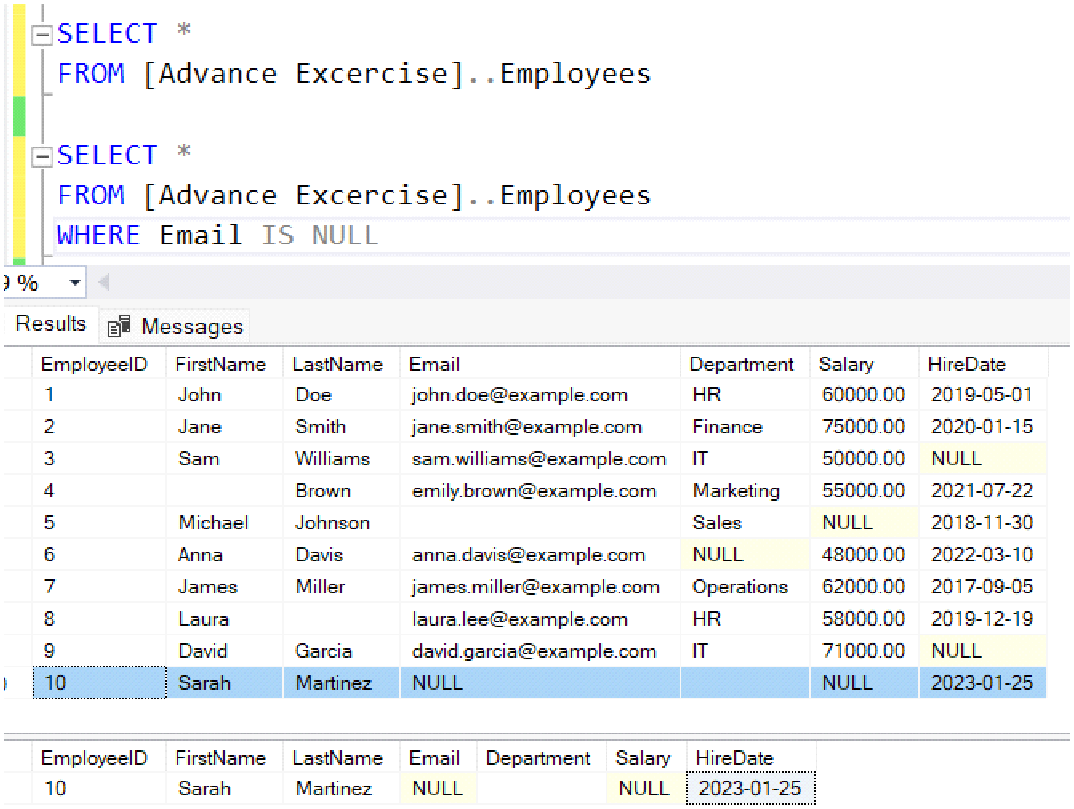
Why are NULL values troublesome?

The problem emerges when you select your data with the aim of analyzing it. As an example, when we draft a query to compare a known value (in our case, NULL) with another known value, the result is typically unknown and isn't included in the result set, leading to a potential loss of data.

How do we resolve this NULL value issue?

The solution to handling NULL values is in utilizing ISNULL or ISNOTNULL operators to verify if your table has a NULL value in a particular field. Let's apply this theory to a real table to grasp how it functions.

The requirement is to check if any of the columns contain NULL values. How do we proceed? We start by crafting a simple SELECT statement and then executing it.



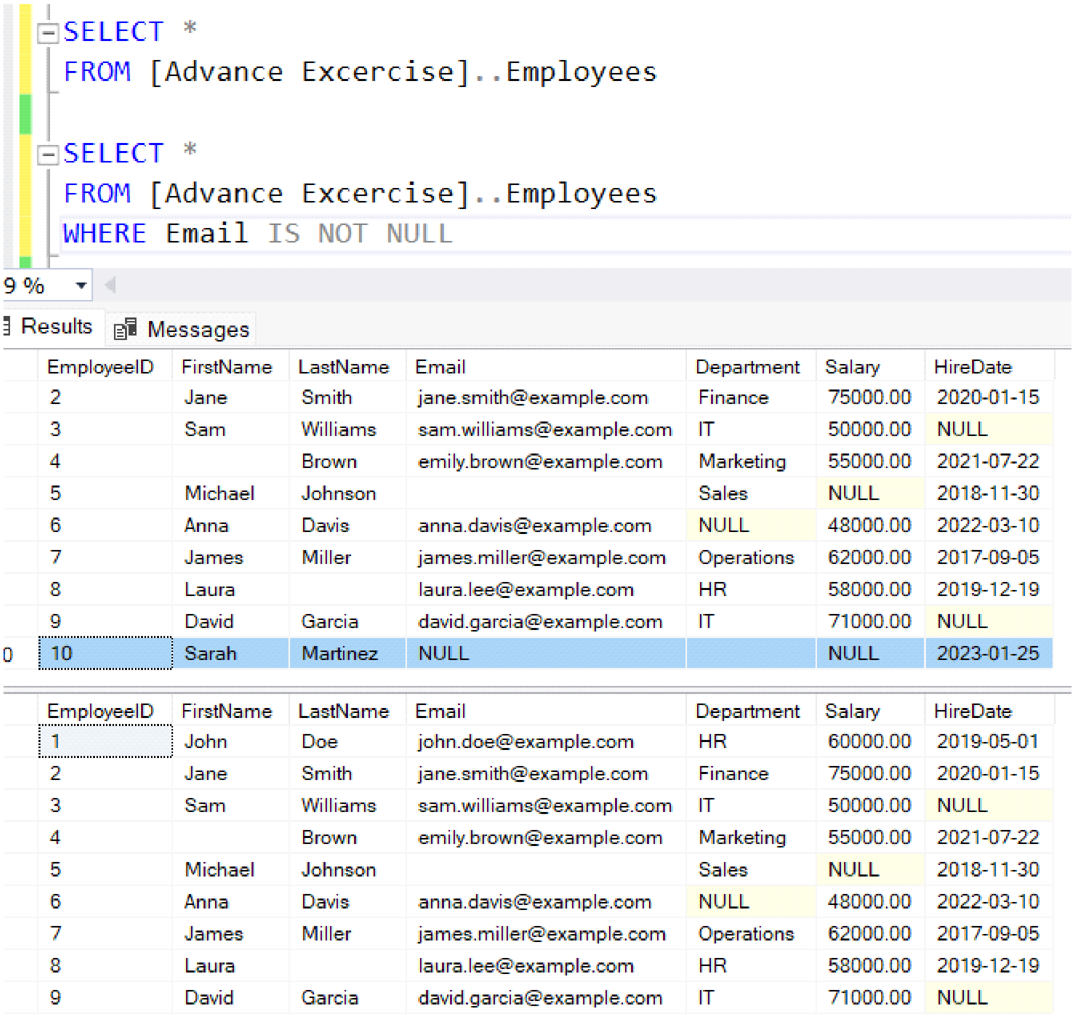
```
SELECT *
FROM [Advance Excercise]..Employees

SELECT *
FROM [Advance Excercise]..Employees
WHERE Email IS NULL
```

EmployeeID	FirstName	LastName	Email	Department	Salary	HireDate
1	John	Doe	john.doe@example.com	HR	60000.00	2019-05-01
2	Jane	Smith	jane.smith@example.com	Finance	75000.00	2020-01-15
3	Sam	Williams	sam.williams@example.com	IT	50000.00	NULL
4		Brown	emily.brown@example.com	Marketing	55000.00	2021-07-22
5	Michael	Johnson		Sales	NULL	2018-11-30
6	Anna	Davis	anna.davis@example.com	NULL	48000.00	2022-03-10
7	James	Miller	james.miller@example.com	Operations	62000.00	2017-09-05
8	Laura		laura.lee@example.com	HR	58000.00	2019-12-19
9	David	Garcia	david.garcia@example.com	IT	71000.00	NULL
10	Sarah	Martinez	NULL		NULL	2023-01-25

EmployeeID	FirstName	LastName	Email	Department	Salary	HireDate
10	Sarah	Martinez	NULL		NULL	2023-01-25

How do we manage these NULL values while gathering data for our analysis?

The answer is simple - we should omit these NULL records. If we require data that doesn't contain any NULL value in a specific column, we should employ the IS NOT NULL operator in our query.



```sql
SELECT *
FROM [Advance Excercise]..Employees

SELECT *
FROM [Advance Excercise]..Employees
WHERE Email IS NOT NULL
```

EmployeeID	FirstName	LastName	Email	Department	Salary	HireDate
2	Jane	Smith	jane.smith@example.com	Finance	75000.00	2020-01-15
3	Sam	Williams	sam.williams@example.com	IT	50000.00	NULL
4		Brown	emily.brown@example.com	Marketing	55000.00	2021-07-22
5	Michael	Johnson		Sales	NULL	2018-11-30
6	Anna	Davis	anna.davis@example.com	NULL	48000.00	2022-03-10
7	James	Miller	james.miller@example.com	Operations	62000.00	2017-09-05
8	Laura		laura.lee@example.com	HR	58000.00	2019-12-19
9	David	Garcia	david.garcia@example.com	IT	71000.00	NULL
10	Sarah	Martinez	NULL		NULL	2023-01-25

EmployeeID	FirstName	LastName	Email	Department	Salary	HireDate
1	John	Doe	john.doe@example.com	HR	60000.00	2019-05-01
2	Jane	Smith	jane.smith@example.com	Finance	75000.00	2020-01-15
3	Sam	Williams	sam.williams@example.com	IT	50000.00	NULL
4		Brown	emily.brown@example.com	Marketing	55000.00	2021-07-22
5	Michael	Johnson		Sales	NULL	2018-11-30
6	Anna	Davis	anna.davis@example.com	NULL	48000.00	2022-03-10
7	James	Miller	james.miller@example.com	Operations	62000.00	2017-09-05
8	Laura		laura.lee@example.com	HR	58000.00	2019-12-19
9	David	Garcia	david.garcia@example.com	IT	71000.00	NULL

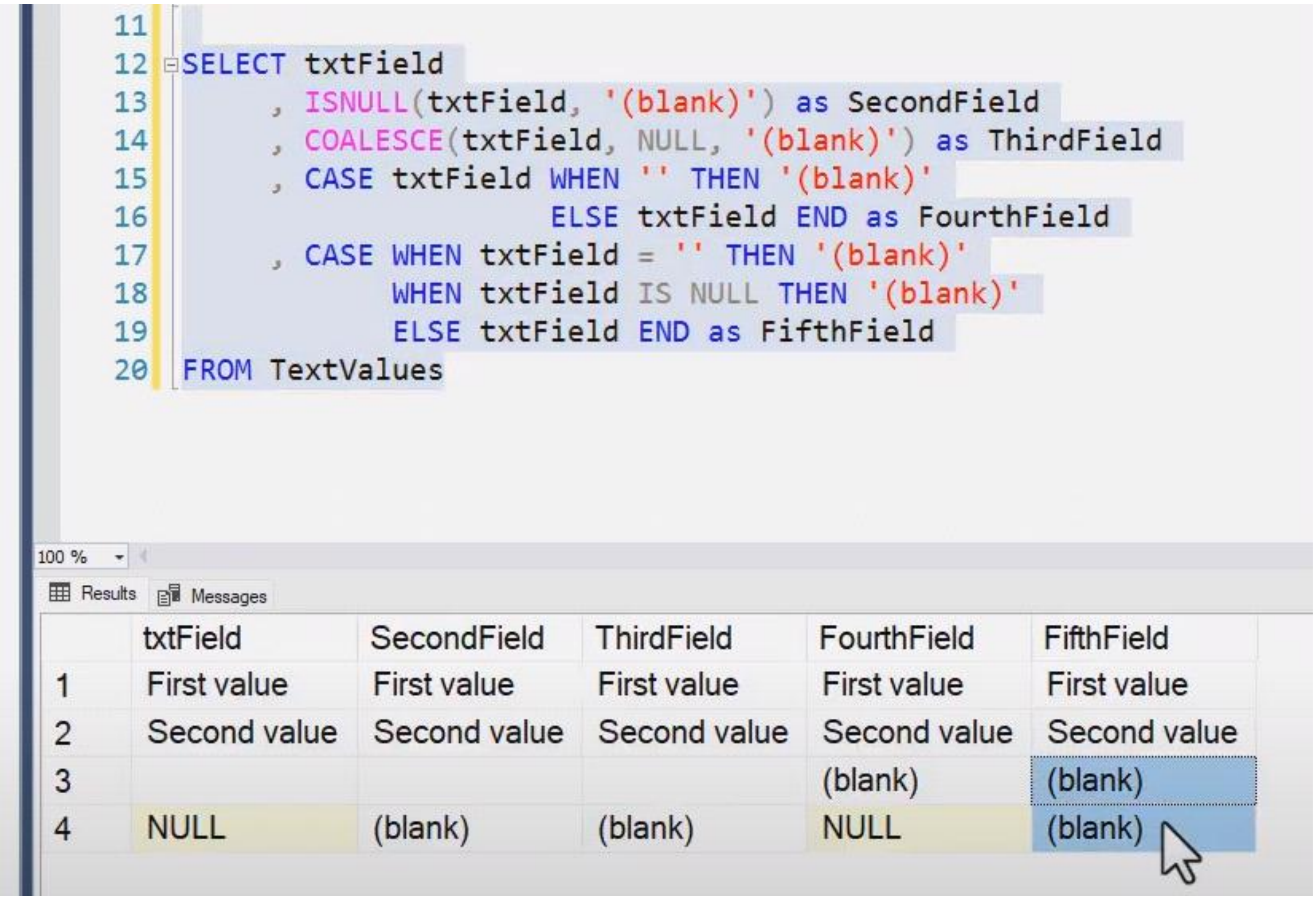
Excluding rows with null values.

From this in-depth exploration, it's clear that mastering SQL NULL values is crucial for precise and successful data analytics. By understanding how to implement the IS NULL and IS NOT NULL operators, we can overcome potential hurdles that NULL values pose. Remember, each learning curve, including resolving SQL NULL values, only equips us better for our data analysis, transforming us into superior analysts.

Credits: https://ai2sql.io/tackling-sql-null-values-step-by-step-guide-optimised-data-analysis

Handling NULL and Blank values using different methods

Credits: Practice Activity: Replacing NULL and blank values in Microsoft SQL Server



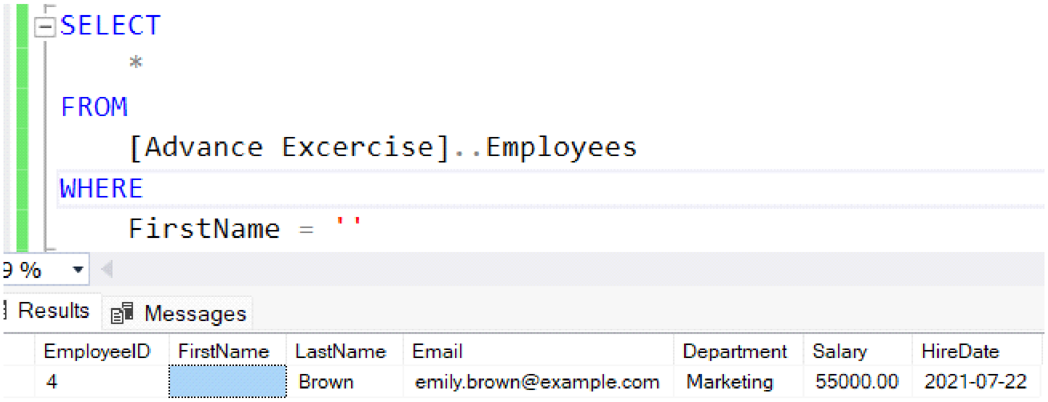
```
SELECT txtField
     , ISNULL(txtField, '(blank)') as SecondField
     , COALESCE(txtField, NULL, '(blank)') as ThirdField
     , CASE txtField WHEN '' THEN '(blank)'
                     ELSE txtField END as FourthField
     , CASE WHEN txtField = '' THEN '(blank)'
            WHEN txtField IS NULL THEN '(blank)'
            ELSE txtField END as FifthField
FROM TextValues
```

	txtField	SecondField	ThirdField	FourthField	FifthField
1	First value	First value	First value	First value	First value
2	Second value	Second value	Second value	Second value	Second value
3				(blank)	(blank)
4	NULL	(blank)	(blank)	NULL	(blank)

The first methods that mentioned before are IS NULL and IS NOT NULL in WHERE clause. But these methods are not enough when handling complex data. For example a rows with blank value. This is not null. It has value. It simply a blank space but this data may lead to inaccuracy.

Handling Blank values using WHERE clause

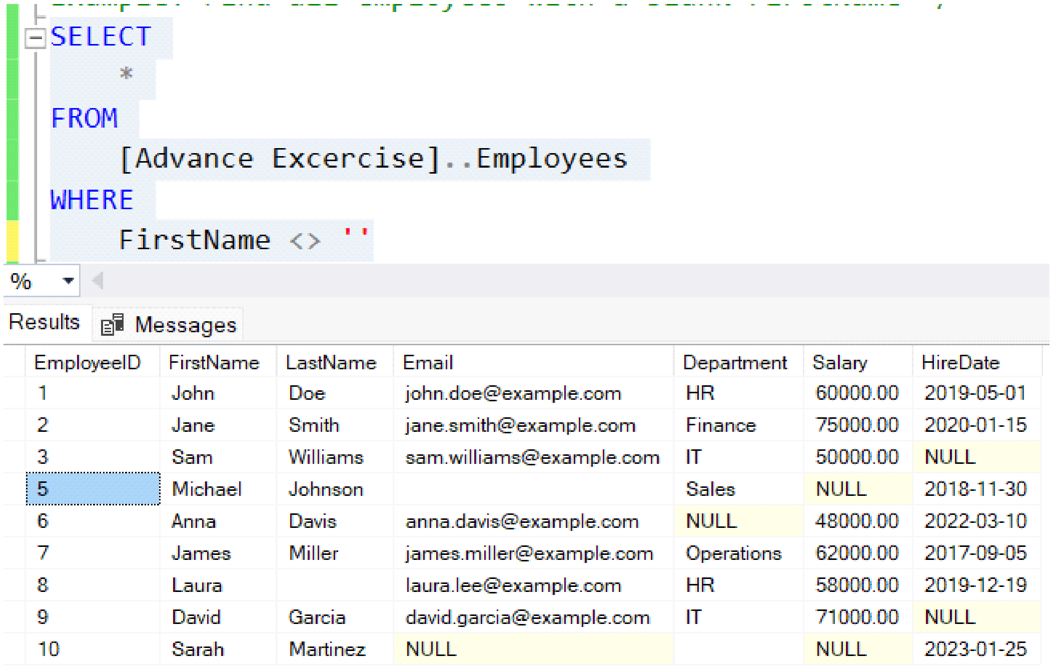
Similar to IS NULL and IS NOT NULL before we implement this condition in WHERE clause.



```
SELECT
    *
FROM
    [Advance Excercise]..Employees
WHERE
    FirstName = ''
```

	EmployeeID	FirstName	LastName	Email	Department	Salary	HireDate
	4		Brown	emily.brown@example.com	Marketing	55000.00	2021-07-22

Selecting FirstName column with blank space rows. We simply implement a condition in WHERE clause that select all the rows equal to blank. The employee with EmployeeID = 4 were selected because he met the condition.



```sql
SELECT
    *
FROM
    [Advance Excercise]..Employees
WHERE
    FirstName <> ''
```

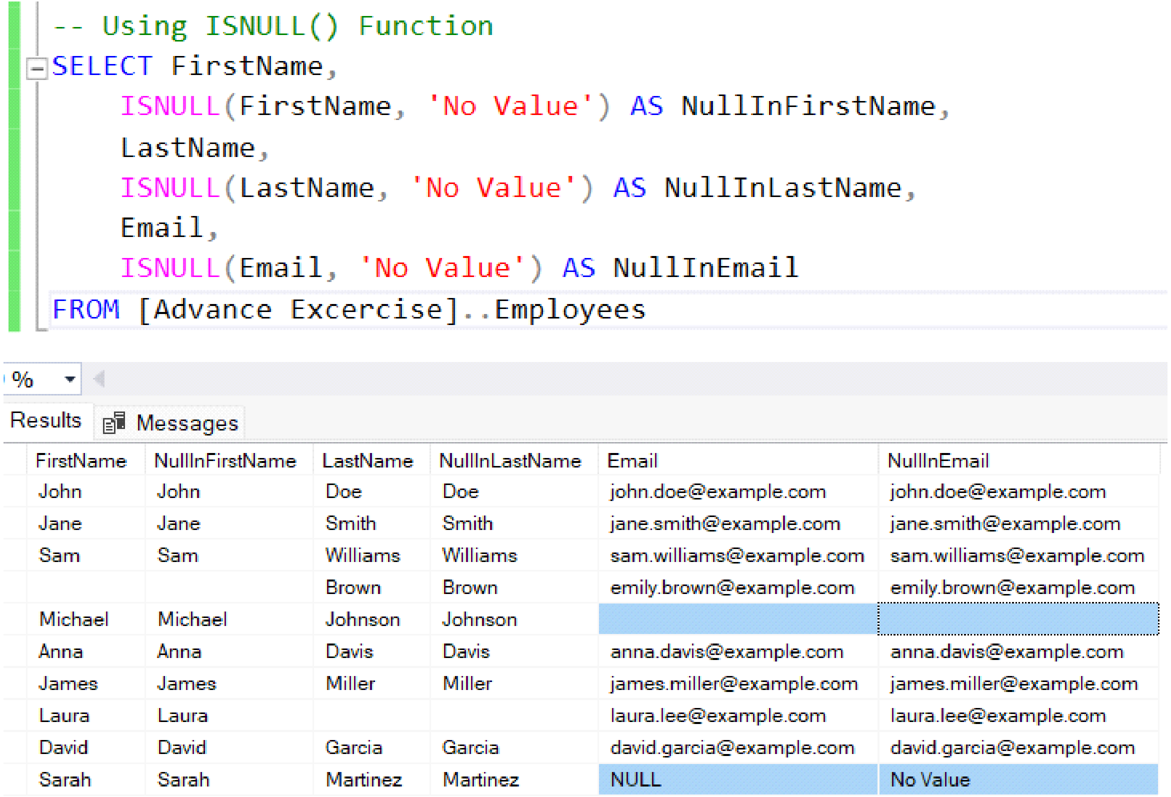
EmployeeID	FirstName	LastName	Email	Department	Salary	HireDate
1	John	Doe	john.doe@example.com	HR	60000.00	2019-05-01
2	Jane	Smith	jane.smith@example.com	Finance	75000.00	2020-01-15
3	Sam	Williams	sam.williams@example.com	IT	50000.00	NULL
5	Michael	Johnson		Sales	NULL	2018-11-30
6	Anna	Davis	anna.davis@example.com	NULL	48000.00	2022-03-10
7	James	Miller	james.miller@example.com	Operations	62000.00	2017-09-05
8	Laura		laura.lee@example.com	HR	58000.00	2019-12-19
9	David	Garcia	david.garcia@example.com	IT	71000.00	NULL
10	Sarah	Martinez	NULL		NULL	2023-01-25

In this query, we simply make a condition in WHERE clause that all the rows with blank value should exclude in select statement. This time the employee with EmployeeID = 4 is excluded because it holds blank value.

Using ISNULL() function

Return the specified value IF the expression is NULL, otherwise return the expression

From <https://www.w3schools.com/sql/func_sqlserver_isnull.asp>



```sql
-- Using ISNULL() Function
SELECT FirstName,
    ISNULL(FirstName, 'No Value') AS NullInFirstName,
    LastName,
    ISNULL(LastName, 'No Value') AS NullInLastName,
    Email,
    ISNULL(Email, 'No Value') AS NullInEmail
FROM [Advance Excercise]..Employees
```

FirstName	NullInFirstName	LastName	NullInLastName	Email	NullInEmail
John	John	Doe	Doe	john.doe@example.com	john.doe@example.com
Jane	Jane	Smith	Smith	jane.smith@example.com	jane.smith@example.com
Sam	Sam	Williams	Williams	sam.williams@example.com	sam.williams@example.com
		Brown	Brown	emily.brown@example.com	emily.brown@example.com
Michael	Michael	Johnson	Johnson		
Anna	Anna	Davis	Davis	anna.davis@example.com	anna.davis@example.com
James	James	Miller	Miller	james.miller@example.com	james.miller@example.com
Laura	Laura			laura.lee@example.com	laura.lee@example.com
David	David	Garcia	Garcia	david.garcia@example.com	david.garcia@example.com
Sarah	Sarah	Martinez	Martinez	NULL	No Value

The syntax is very simple: ISNULL(ColumnName, 'Changed value'). But it only works with NULL values. See in row 5 it holds blank value. Blank does not mean null. It has value, blank.

Using COALESCE() function

The COALESCE() function in SQL is used to return the first non-NULL value from a list of arguments. It is a very useful function for handling NULL values and providing default values when NULL values are encountered.

> Syntax:
> COALESCE(expression1, expression2, ..., expressionN)

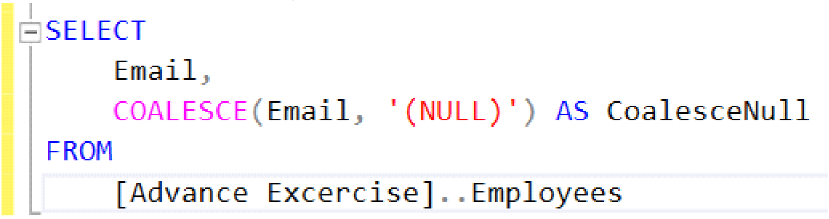
expression1, expression2, ..., expressionN: These are the values or expressions to evaluate. COALESCE returns the first non-NULL value among these expressions. If all the expressions evaluate to NULL, then COALESCE returns NULL.

How It Works

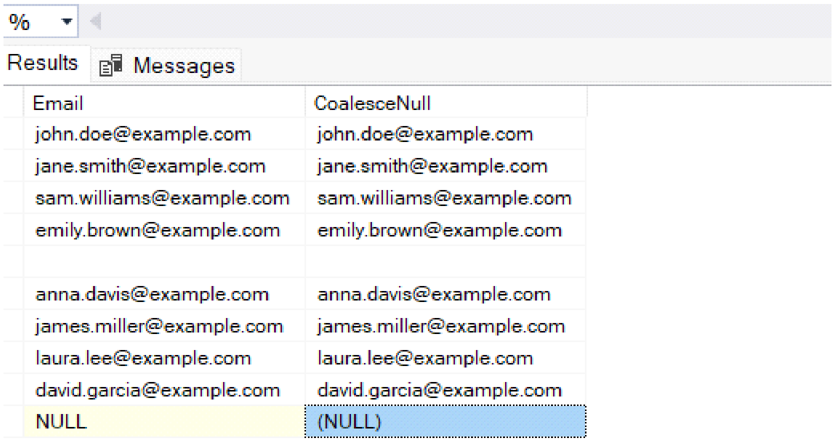
The COALESCE() function evaluates the arguments in the order they are provided and returns the first non-NULL value. If all arguments are NULL, it returns NULL.

Simple Case: Replacing NULL with a Default Value

Suppose you have a table Employees with a column Department that might contain NULL values. You want to replace NULL values with a default value 'Unassigned'.



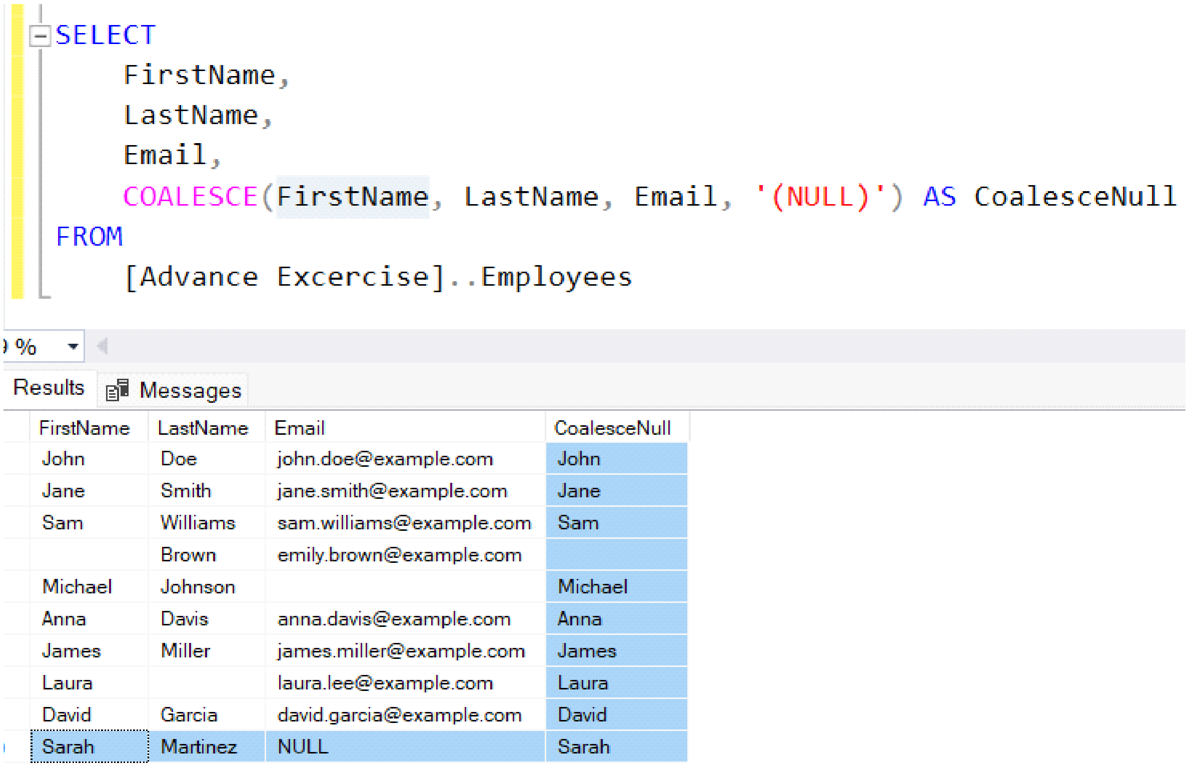
```
SELECT
    Email,
    COALESCE(Email, '(NULL)') AS CoalesceNull
FROM
    [Advance Excercise]..Employees
```



Email	CoalesceNull
john.doe@example.com	john.doe@example.com
jane.smith@example.com	jane.smith@example.com
sam.williams@example.com	sam.williams@example.com
emily.brown@example.com	emily.brown@example.com
anna.davis@example.com	anna.davis@example.com
james.miller@example.com	james.miller@example.com
laura.lee@example.com	laura.lee@example.com
david.garcia@example.com	david.garcia@example.com
NULL	(NULL)

Multiple Columns: Return First Non-NULL Value

You can use COALESCE to return the first non-NULL value from multiple columns.
For example, if you have columns FirstName, LastName, and Email and you want to get the first non-NULL email:



```sql
SELECT
    FirstName,
    LastName,
    Email,
    COALESCE(FirstName, LastName, Email, '(NULL)') AS CoalesceNull
FROM
    [Advance Excercise]..Employees
```

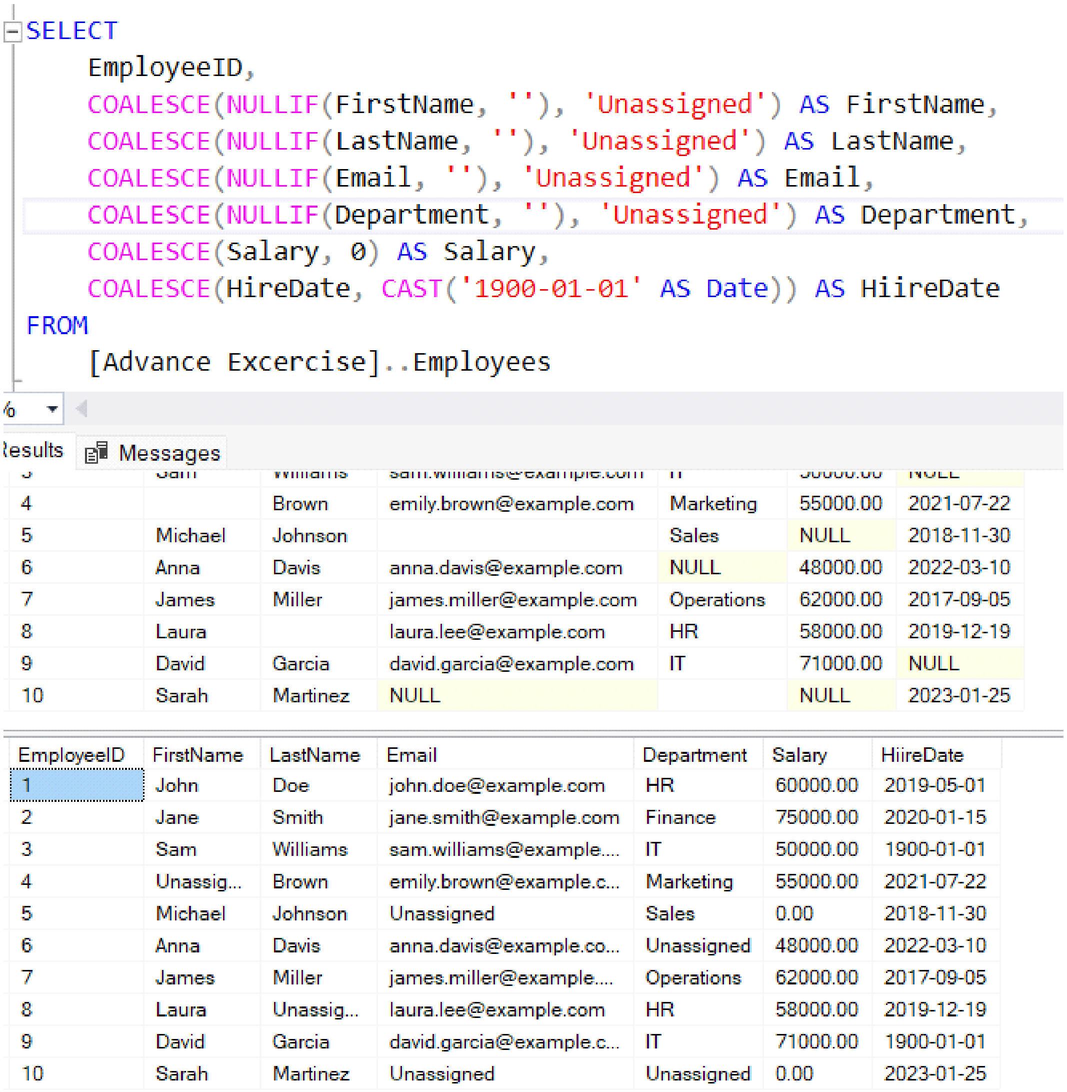
FirstName	LastName	Email	CoalesceNull
John	Doe	john.doe@example.com	John
Jane	Smith	jane.smith@example.com	Jane
Sam	Williams	sam.williams@example.com	Sam
	Brown	emily.brown@example.com	
Michael	Johnson		Michael
Anna	Davis	anna.davis@example.com	Anna
James	Miller	james.miller@example.com	James
Laura		laura.lee@example.com	Laura
David	Garcia	david.garcia@example.com	David
Sarah	Martinez	NULL	Sarah

But similar to ISNULL() function, it only deal with null but in much complex way.

Using COALESCE() NULLIF() function to deal with NULL and BLANK rows

The NULLIF() function returns NULL if two expressions are equal, otherwise it returns the first expression.

Inside the COALESCE() function we used NULLIF() function to compare the two parameters. If the first parameter met the condition in second parameter (which is blank) it then converts the value into NULL but since we used COALESCE() function (which we knew deals with null rows) it then converts the converted NULL value into 'Unassigned' string value.



```sql
SELECT
    EmployeeID,
    COALESCE(NULLIF(FirstName, ''), 'Unassigned') AS FirstName,
    COALESCE(NULLIF(LastName, ''), 'Unassigned') AS LastName,
    COALESCE(NULLIF(Email, ''), 'Unassigned') AS Email,
    COALESCE(NULLIF(Department, ''), 'Unassigned') AS Department,
    COALESCE(Salary, 0) AS Salary,
    COALESCE(HireDate, CAST('1900-01-01' AS Date)) AS HiireDate
FROM
    [Advance Excercise]..Employees
```

3	Sam	Williams	sam.williams@example.com	IT	50000.00	NULL
4		Brown	emily.brown@example.com	Marketing	55000.00	2021-07-22
5	Michael	Johnson		Sales	NULL	2018-11-30
6	Anna	Davis	anna.davis@example.com	NULL	48000.00	2022-03-10
7	James	Miller	james.miller@example.com	Operations	62000.00	2017-09-05
8	Laura		laura.lee@example.com	HR	58000.00	2019-12-19
9	David	Garcia	david.garcia@example.com	IT	71000.00	NULL
10	Sarah	Martinez	NULL		NULL	2023-01-25

EmployeeID	FirstName	LastName	Email	Department	Salary	HiireDate
1	John	Doe	john.doe@example.com	HR	60000.00	2019-05-01
2	Jane	Smith	jane.smith@example.com	Finance	75000.00	2020-01-15
3	Sam	Williams	sam.williams@example....	IT	50000.00	1900-01-01
4	Unassig...	Brown	emily.brown@example.c...	Marketing	55000.00	2021-07-22
5	Michael	Johnson	Unassigned	Sales	0.00	2018-11-30
6	Anna	Davis	anna.davis@example.co...	Unassigned	48000.00	2022-03-10
7	James	Miller	james.miller@example....	Operations	62000.00	2017-09-05
8	Laura	Unassig...	laura.lee@example.com	HR	58000.00	2019-12-19
9	David	Garcia	david.garcia@example.c...	IT	71000.00	1900-01-01
10	Sarah	Martinez	Unassigned	Unassigned	0.00	2023-01-25

We can see that COALESCE() and NULLIF() function deals both with null and blank by converting blank into null using NULLIF() function and converts null into 'Unassigned' by using COALESCE() function.