

Fabric Dataflow gen 1

Explore the dataflow gen1 Capabilities

First, we need to create a pipeline

Workspace -> New Item -> Data pipeline

The screenshot shows the Fabric workspace interface. On the left, there's a sidebar with icons for Home, Workspaces, OneLake catalog, Monitor, Real-Time, Workloads, and fabricworkspace. The main area is titled 'New item' under 'Get data'. It lists several options:

- Copy job**: Makes it easy to copy data in Fabric. Includes full copy, incremental copy, and event-based copy modes.
- Data pipeline**: Ingest data at scale and schedule data workflows.
- Dataflow Gen1**: Prep, clean, and transform data.
- Dataflow Gen2**: Prep, clean, and transform data.
- Eventstream**: Capture, transform, and route real-time event stream to various destinations in desired format with no-code experience.
- Mirrored Azure Cosmos DB (pre...)**: Easily replicate data from an existing source into an analytics-friendly format.
- Mirrored Azure Database for Po...**: Easily replicate data from an existing source into an analytics-friendly format.
- Mirrored Azure SQL Database**: Easily replicate data from an existing source into an analytics-friendly format.
- Mirrored Azure SQL Managed In...**: Easily replicate data from an existing source into an analytics-friendly format.

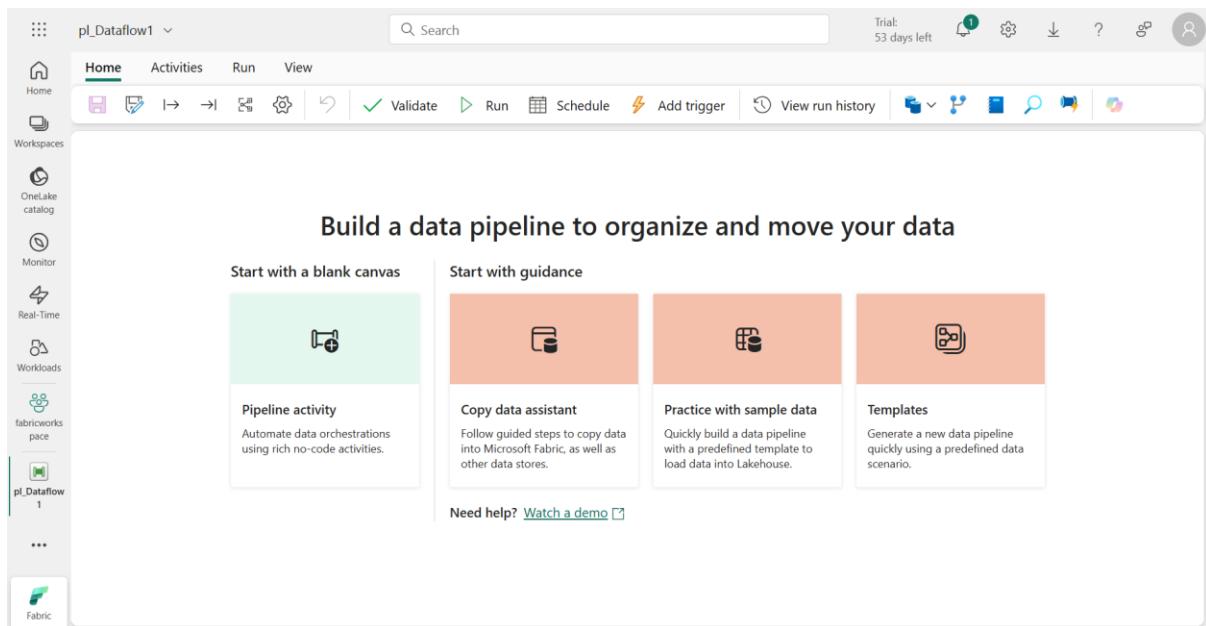
Click on Data pipeline

The screenshot shows a 'New pipeline' dialog box overlaid on a list of existing pipelines. The dialog has a 'Name' input field containing 'pl_Dataflow1'. At the bottom are 'Create' and 'Cancel' buttons. The background table lists the following pipelines:

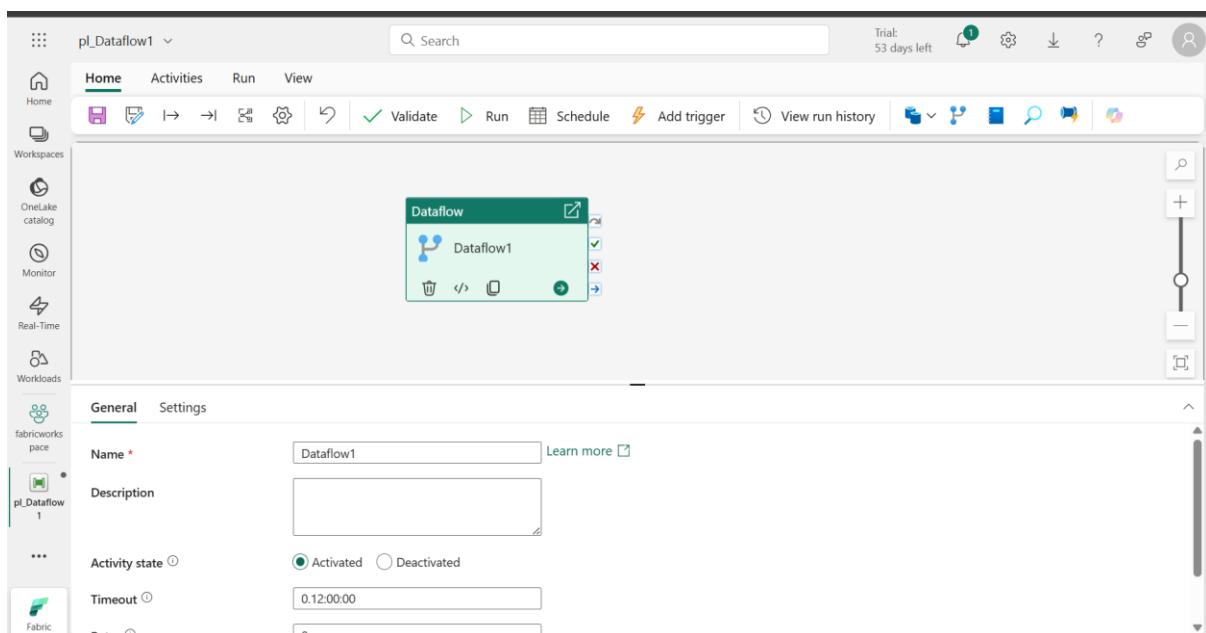
Name	Type	Owner	Last refresh	Next refresh	Endorsements	Sensitivity	Included in app
Dataflow	Dataflow	deefabric	N/A	N/A	—	—	—
Fabric_lakehouse	Lakehouse	deefabric	—	—	—	—	—
Fabric_lakehouse	Semantic	deefabric	—	—	—	—	—
Fabric_lakehouse	SQL analysis	deefabric	—	—	—	—	—
pl_dataflow	Data pipeline	deefabric	—	—	—	—	—
pl_loaddata	Data pipeline	deefabric	—	—	—	—	—
pl_warehousetolakehouse	Data pipeline	deefabric	—	—	—	—	—

Create pipeline

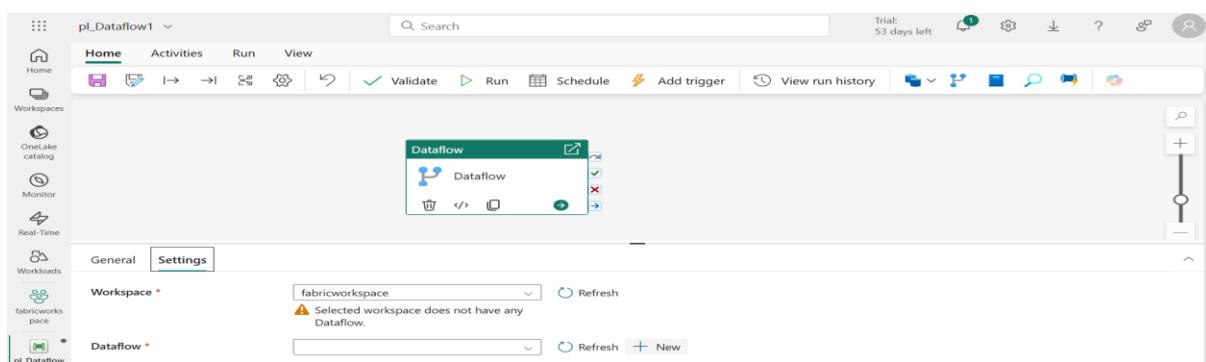
Pipeline created.

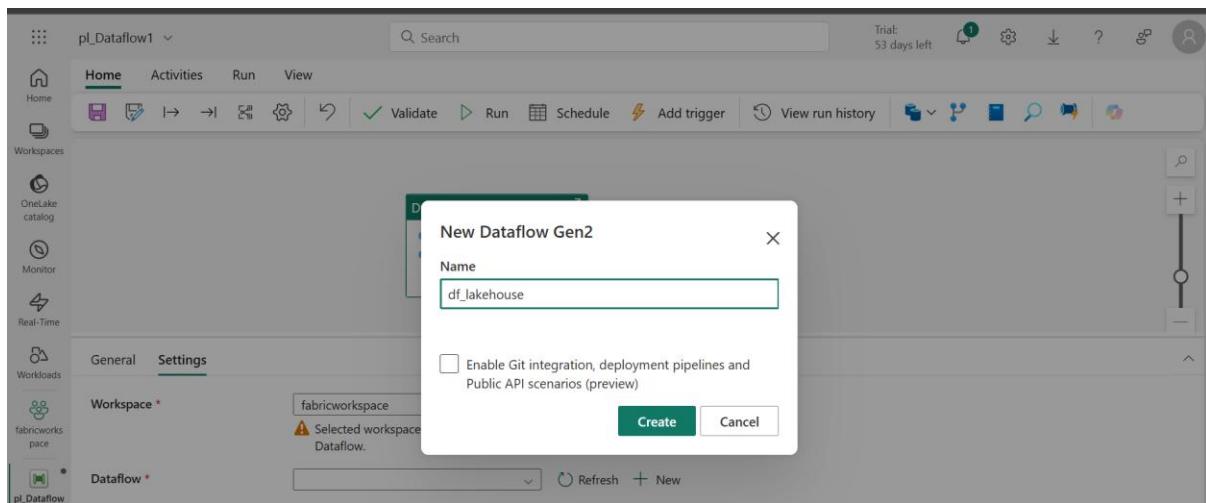


Now select Dataflow from the activities

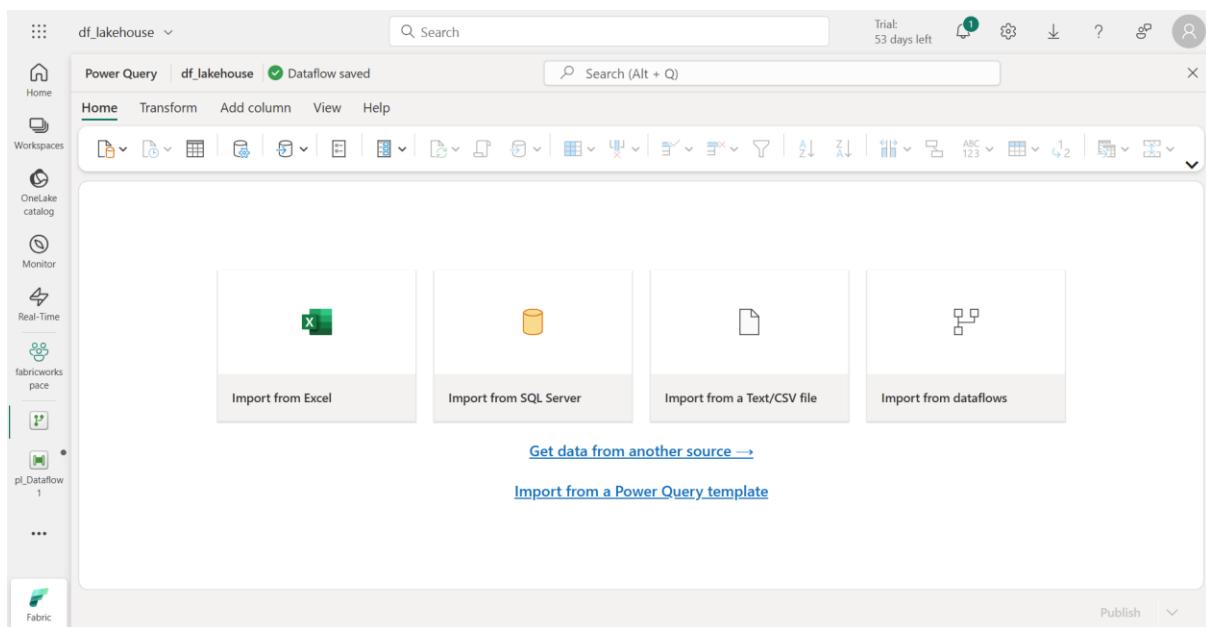


Now we need to load the data for that we need to create a connection, here our source is Lake house so we need to build a connection between dataflow and lakehouse. For that go to settings -> workspace is auto selected -> Dataflow -> New



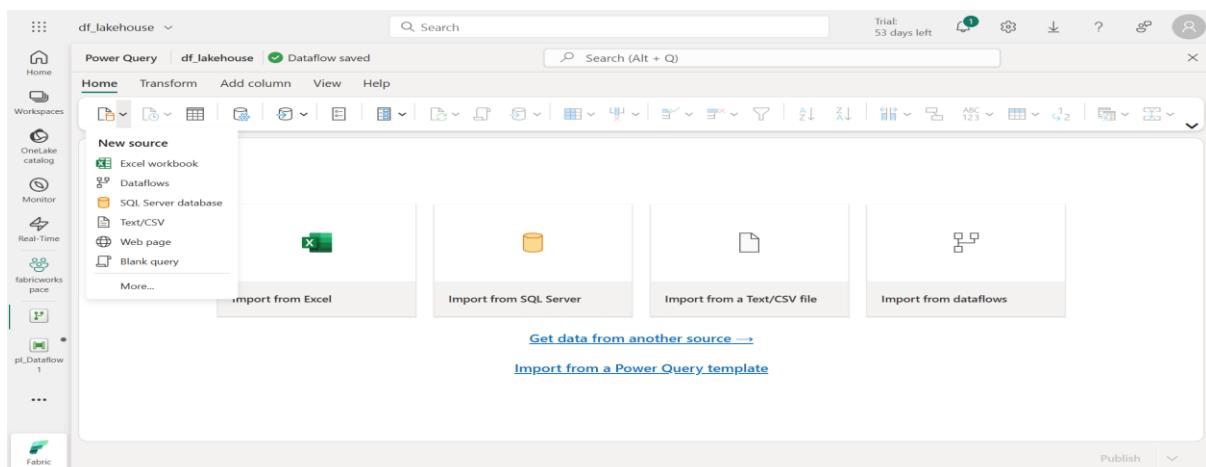


Create Dataflow

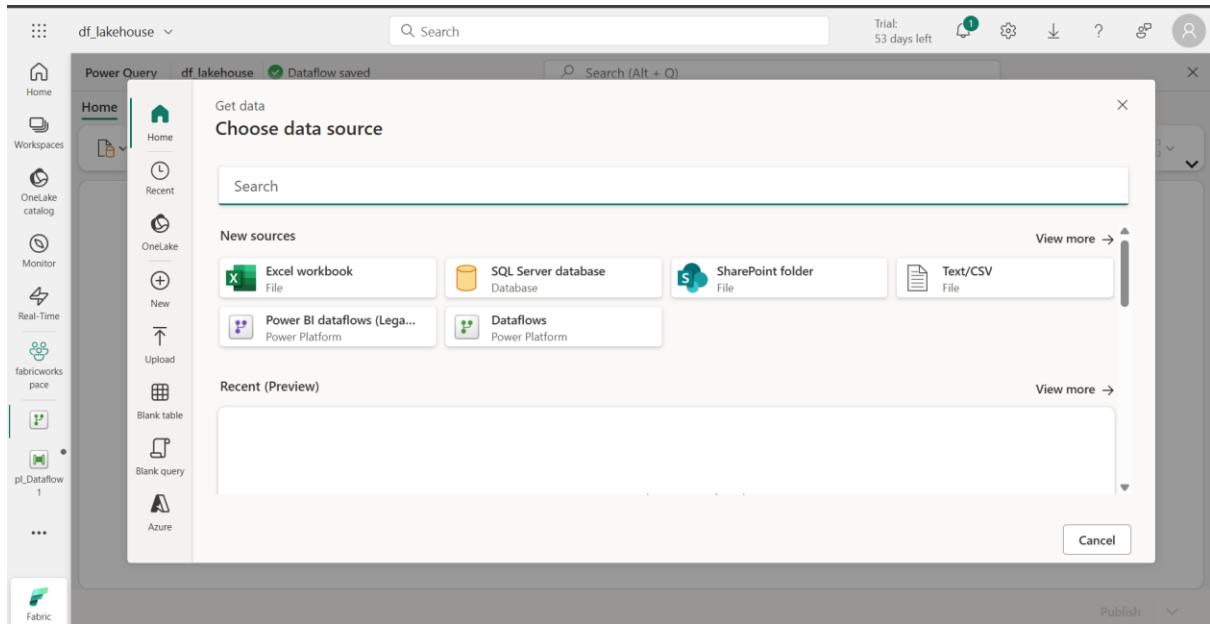


Data flow is created.

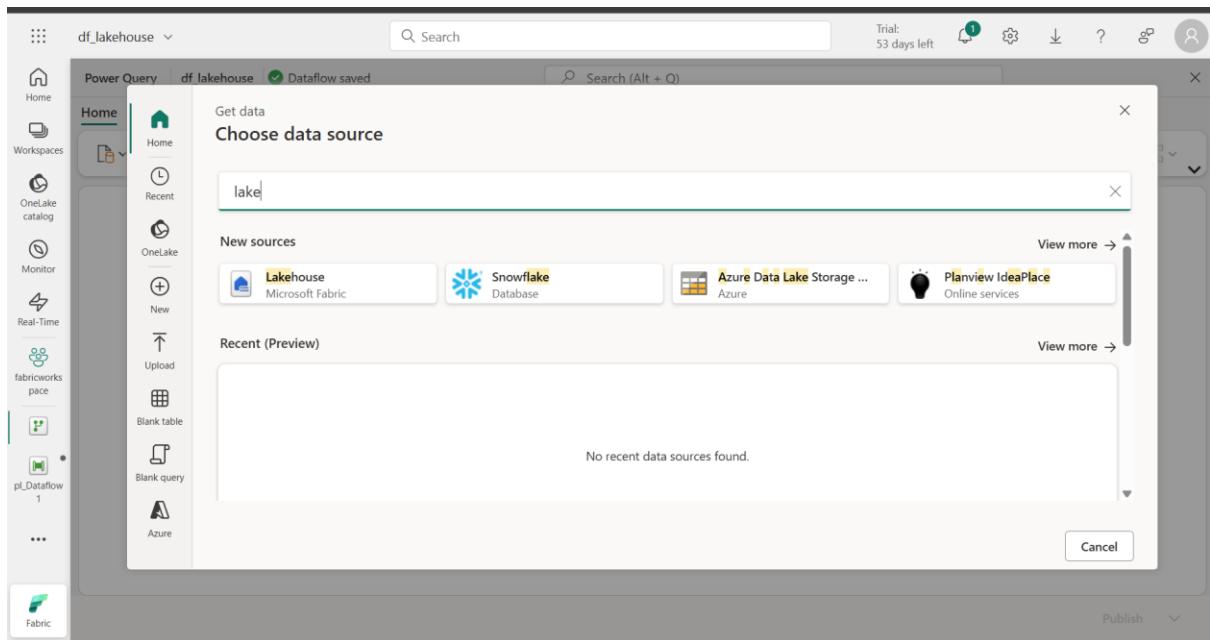
Now to get source data we need to create a connection. For that click on New source -> Choose source.



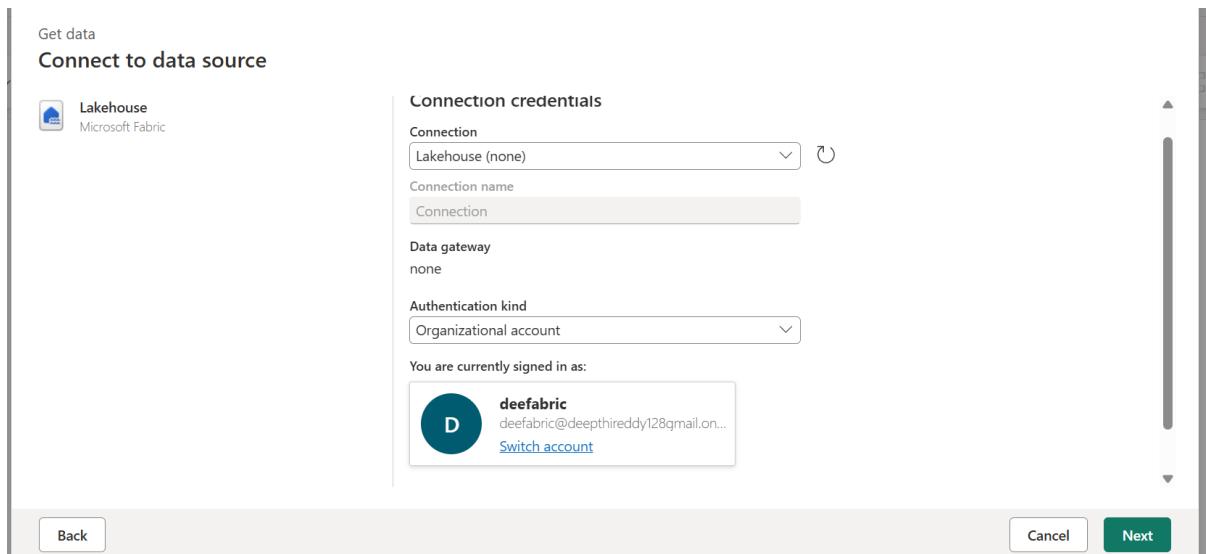
To view more source options, click on more.



Here we can search for any source which we want, here our source is lakehouse so we will search for lakehouse.



Once we select the lakehouse automatically it will create connection to all the workspaces which are available for us, so we can select any lakehouse within any workspace we want.



Next, we have to select which workspace lake house we want and what is the source from the lakehouse a file or a table.

ID	Name	City	PhoneNo
1	Sam	Brampton	4567824567
2	Ieo	Sudbury	4567824567
3	Naho	Paris	4567824567
4	Robert	Toronto	4567824567
5	John	New York	4567824567

Here, we see a folder to get the data we need to click on the link Table, then we will get the data.

Here my folder has subfolder, so I will have to click on 3 times on table to get the data.

Connection to lake house is created.

On the right hand side we have query setting as a option, here we have applied steps options, which will store what ever steps we are doing in the dataset, if we want to undo any steps we can do it from there.

The screenshot shows the Power Query Editor interface. On the left, there's a 'Queries [1]' pane with a single item named '2025'. The main area displays a table with columns: ID, Name, City, and Phoneno. The 'Applied steps' pane on the right lists several steps, including 'Navigation' steps and a 'Promote' step. The 'Data destination' section is currently set to 'No data destination'.

To view this in diagrammatic view we have option on right side bottom.

This screenshot shows the same Power Query Editor interface, but the 'Data destination' pane has been expanded to show the 'Diagram view' option, which is highlighted with a red border.

Click on Diagram view

This screenshot shows the Power Query Editor with the 'Diagram view' selected. The 'Applied steps' pane now displays a visual representation of the data flow, showing a sequence of steps: Source, Navigation 1, Navigation 2, Navigation 3, Navigation 4, Navigation 5, Navigation 6, Navigation 7, Imported CSV, and Promoted. The 'Query settings' pane is also visible on the right.

We get this view; to collapse these steps we can click on collapse option.

The screenshot shows the Microsoft Power Query Editor interface. On the left, there's a sidebar with icons for Home, Workspaces, OneLake catalog, Monitor, Real-Time, and fabricworks pace. The main area is titled 'df_lakehouse' and shows a 'Power Query' tab. Below the tabs are 'Home', 'Transform', 'Add column', 'View', and 'Help'. The 'Queries [1]' pane shows a single step named '2025' with 12 steps. The step content is a table transformation with the formula: Table.TransformColumnTypes(#"Promoted headers", {"ID", Int64.Type}, {"Name", type text}, {"City", type text}, {"Phoneno", Int64.Type}}). The table has columns 'ID', 'Name', 'City', and 'Phoneno', with rows for Sam, leo, Naho, Robert, and John. The table is styled with alternating row colors. At the bottom of the editor, it says 'Completed (0.39 s) Columns: 4 Rows: 5' and 'Add default destination...'. There are also 'Step', 'Stop', and 'Run' buttons.

Now we can more source either from same workspace lakehouse or different workspace lakehouse, as well we can also add source from warehouse or any other data source.

Go to new source and select where we want our source from.

My second source is also a lakehouse from same workspace, but this time I am choosing a table. As we already created a connection with lake house it will look like below.

The screenshot shows the 'Get data' dialog in Microsoft Power Query. On the left, it says 'Get data' and 'Connect to data source'. Below that is a 'Lakehouse Microsoft Fabric' icon. The right side is titled 'Connection credentials' with a dropdown 'Connection' set to 'Lakehouse (none)' and 'Authentication kind: Organizational account' with a link 'Edit connection'. At the bottom are 'Back', 'Cancel', and 'Next' buttons.

After this we need to choose the path from where we want our data to be taken.

The screenshot shows the 'Choose data' dialog in Microsoft Power Query. On the left, it says 'Get data' and 'Choose data'. Below that is a 'Display options' dropdown with 'customer_table' selected. A tree view shows 'customer_table' expanded, with 'sys.managed_delta_ta...' and 'customer_table' checked. Other items like 'credit_table' and 'newwws' are also listed. To the right is a preview of the 'customer_table' data with columns 'ID', 'Name', 'City', and 'Phoneno', and rows for Sam, leo, Naho, Robert, and John. At the bottom are 'Back', 'Cancel', and 'Create' buttons.

Here our second source is a Table.

Second source is added, as we opted for diagram view second source is also showing in diagram view as well.

The screenshot shows the Power Query Editor interface. At the top, there's a navigation bar with 'df_lakehouse' selected, a search bar, and a trial status. Below the bar are tabs for 'Power Query', 'df_lakehouse', and 'Dataflow saved'. The main area is titled 'Queries [2]' and contains a dataflow diagram. The first step is a table named 'customer_table' with four columns: 'ID', 'Name', 'City', and 'Phoneno'. This step has three 'Navigation' steps connected to it. A preview pane below the table shows five rows of data:

ID	Name	City	Phoneno
1	Sam	Brampton	4567824567
2	Ieo	Sudbury	4567824567
3	Naho	Paris	4567824567
4	Robert	Toronto	4567824567
5	John	New York	4567824567

Now, we will add 3rd source, from different workspace, lake house.

The screenshot shows the 'Get data' wizard. The current step is 'Connect to data source'. On the left, there's a sidebar with 'Get data' and 'Recent'. The main area shows a list of connection options under 'Microsoft Fabric': 'Lakehouse'. To the right, there's a 'Connection credentials' panel with a 'Connection' dropdown set to 'Lakehouse (none)' and an 'Edit connection' link. At the bottom, there are 'Back', 'Cancel', and 'Next' buttons.

Here we see our workspace is newws.

Get data

Choose data

Newcustomerdata.csv

CSV

onelake.dfs.fabric.microsoft.com
123 bytes

Back Cancel Create

df_lakehouse

Power Query df_lakehouse Dataflow saved

Home Transform Add column View Help

Queries [3]

customer_table

Newcustomerdata.csv

Csv.Document("Navigation 4", [Delimiter = ",", Columns = 4, QuoteStyle = QuoteStyle.None])

ID	Name	City	Phoneno
1	Anna	Florida	6784563782
2	James	Toronto	9876785467
3	Lily	NY	2456783487
4	Will	Italy	8796578256
5			

Completed (0.78 s) Columns: 4 Rows: 5 Add default destination... Step Publish

Three sources are added here.

Now we can append all these sources.

Right click on any of the source -> Append queries as new

df_lakehouse

Power Query df_lakehouse Dataflow saved

Home Transform Add column View

Queries [3]

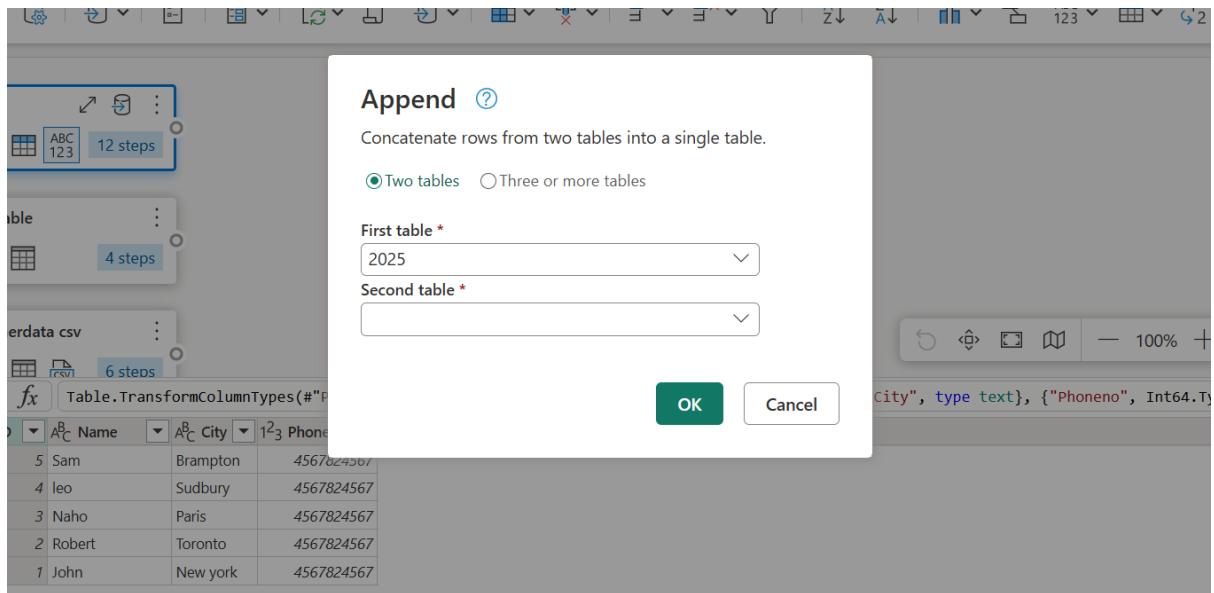
customer_table

Newcustomerdata.csv

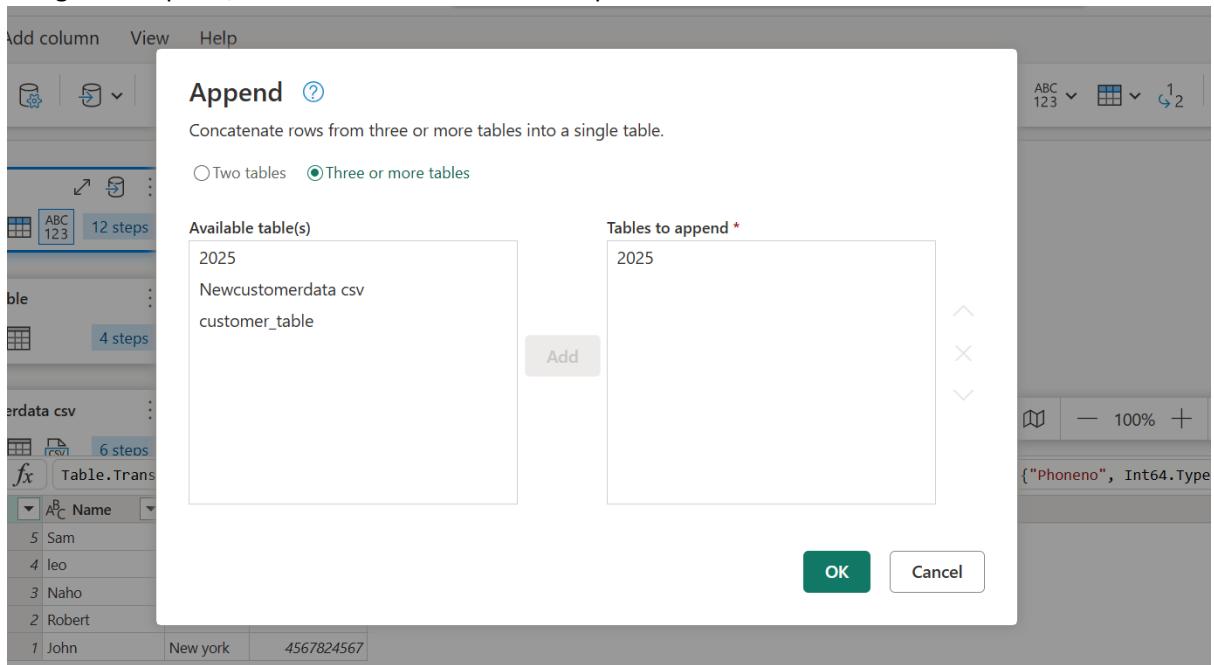
Table.Transform

Columns: 4 Rows: 5 Add default...

Append queries as new



We get this option, as we have 3 tables need to opt for # or more tables



Now we have to choose all the tables.

Added all three tables, and if we want to change the order we can just click on X mark on right side and change the order or remove any table which we don't want.

Append ?

Concatenate rows from three or more tables into a single table.

Two tables Three or more tables

Available table(s)

2025
Newcustomerdata csv
customer_table

Tables to append *

2025
Newcustomerdata csv
customer_table

Add



OK

Cancel

The screenshot shows the Power Query Editor interface. On the left, the 'Queries [4]' pane lists four queries: '2025', 'customer_table', 'Newcustomerdata', and 'Append'. The 'customer_table' and 'Newcustomerdata' queries are connected by an 'Append' step. The main workspace displays the 'Table.TransformColumnTypes' step for the 'Append' query, which has 15 rows of sample data. The data includes columns like 'ID', 'Name', 'City', 'PhoneNo', and several unnamed columns labeled 'Column1' through 'Column4'. The 'Query settings' pane on the right shows the 'Name' is set to 'Append'. The 'Applied steps' section shows the 'Source' step and the 'Changed' step. The bottom status bar indicates the process completed in 4.16 seconds with 15 rows.

After append we see there are null values and different columns this is because of column names are different so we need to rename the column names.

Column names are different in Newcustomerdata file, so click on that file, go to column and right click-> Rename.

Or else we can double click on the column then we can rename it.

The screenshot shows the Power Query Editor interface with the 'df_lakehouse' dataflow selected. The 'Queries [4]' pane lists four queries: '2025', 'customer_table', 'Newcustomer...', and 'Imported C...'. A context menu is open over the 'Newcustomer...' query, showing options like 'Remove columns', 'Remove other columns', 'Duplicate column', 'Add column from examples...', 'Remove duplicates', 'Remove errors', 'Split column', 'Replace values...', 'Replace errors...', 'Change type', 'Transform column', 'Group by...', 'Unpivot columns', 'Unpivot other columns', 'Unpivot only selected columns', 'Rename...', 'Resize...', 'Move', 'Drill down', and 'Add as new query'. The main workspace displays the 'Imported CSV' table with four columns: ID, Name, City, and Phoneno. The 'Query settings' pane on the right shows the name 'Newcustomerdata csv' and applied steps for 'Source', 'Navigation 1' through 'Navigation 4', and 'Imported C...'. The 'Applied steps' pane also lists these steps. The bottom right corner has a 'Publish' button.

This screenshot shows the Power Query Editor with the 'df_lakehouse' dataflow selected. The 'Queries [4]' pane shows the 'Append' step connecting 'customer_table' and 'Imported CSV'. The 'Imported CSV' table has four columns: ID, Name, City, and Phoneno. The 'Query settings' pane shows the name 'Newcustomerdata csv' and applied steps for 'Source', 'Navigation 1' through 'Navigation 4', and 'Imported C...'. The 'Applied steps' pane also lists these steps. The 'Data destination' section indicates 'No data destination'. The bottom right corner has a 'Publish' button.

All the column names have changed , but not the datatypes, so hover over the data type and click on it.

This screenshot shows the Power Query Editor with the 'df_lakehouse' dataflow selected. The 'Queries [4]' pane shows the 'Imported CSV' table with four columns: ID, Name, City, and Phoneno. The 'Query settings' pane shows the name 'Newcustomerdata csv' and applied steps for 'Source', 'Navigation 1' through 'Navigation 4', and 'Imported C...'. The 'Applied steps' pane also lists these steps. The 'Data destination' section indicates 'No data destination'. The bottom right corner has a 'Publish' button. A detailed view of the 'Phoneno' column's datatype dropdown is shown, listing options: '1.2 Decimal number', '\$ Currency', '123 Whole number', '% Percentage', '2 Date/Time', '3 Date', '4 Time', '5 Date/Time/Zone', '6 Duration', '7 Text', '8 True/False', '9 Binary', and '10 Using locale...'. The current selection is '123 Whole number'. To the right of the dropdown, a preview of the 'Imported CSV' table shows the 'Phoneno' column with values: 6784563782, 9876785467, 2456783487, and 8796578256.

From here we can change the datatype.

Table.TransformColumnTypes(#"Renamed columns", {{"ID", type number}, {"Phoneno", type number}})

	1.2 ID	A ^B _C Name	A ^B _C City	1.2 Phoneno
1	[Error]	Name	City	[Error]
2	6	Anna	Florida	6784563782
3	7	James	Toronto	9876785467
4	8	Lilly	NY	2456783487
5	9	Will	Italy	8796578256

We see here first row is throwing error, so we can delete this row, click in table icon.

The screenshot shows the Power Query Editor interface. The main area displays a table with five rows. The first row contains errors in the 'Name' and 'City' columns. The 'Applied steps' pane on the right lists several steps, including 'Renamed columns'. A context menu is open over the first row, with the 'Remove top rows' option highlighted.

Choose Remove top rows.

The screenshot shows the 'Remove top rows' dialog box. The 'Number of rows' input field is set to 1. The 'OK' button is highlighted. The background shows the Power Query Editor with the table and the 'Applied steps' pane.

Give how many rows we have to remove.

The screenshot shows the Power Query Editor with the transformed table. The first row is now empty and has been removed, leaving four rows of valid data.

First row is removed.

Now check the append.

The screenshot shows the Power Query Editor interface. In the center, there's a table titled "Table.Combine({#"2025", customer_table, #"Newcustomerdata csv"})" with columns "1.2 ID", "Name", "City", and "1.2 Phoneno". The table has 11 rows. The right side of the screen displays the "Query settings" pane with "Properties" set to "Name: Append" and the "Applied steps" pane which includes a step named "Append".

All the rows have appended correctly into one query.

Now, as we see we have duplicate values, we should remove those, for that, click on table icon and opt for remove duplicates option.

The screenshot shows the Power Query Editor interface. A context menu is open over the "1.2 Phoneno" column header, with the "Remove duplicates" option highlighted. The right side of the screen displays the "Query settings" pane with "Properties" set to "Name: Append" and the "Applied steps" pane which includes a step named "Append".

We get this pop up

The screenshot shows a "Insert step" dialog box. It contains the text "Are you sure you want to insert a step? Inserting an intermediate step may affect subsequent steps, which could cause your query to break." Below this is a checkbox labeled "Remember my choice and don't show this dialog again." At the bottom are two buttons: "Insert" and "Cancel". The background shows the Power Query Editor interface with a table and the "Append" step in the Applied steps pane.

Click on insert option.

	ID	Name	City	Phoneno
1	5	Sam	Brampton	4567824567
2	4	leo	Sudbury	4567824567
3	3	Naho	Paris	4567824567
4	2	Robert	Toronto	4567824567
5	1	John	New York	4567824567
6	6	Anna	Florida	6784563782
7	7	James	Toronto	9876785467
8	8	Lilly	NY	2456783487
9	9	Will	Italy	8796578256

Duplicates are removed from the query.

ID	Name	City	Phoneno

This is power query option

If we want to see query view we need to click on right side bottom -> step -> Query Script.

ID	Name	City	Phoneno
4	Robert	Toronto	4567824567
5	John	New York	4567824567
6	Anna	Florida	6784563782
7	James	Toronto	9876785467
8	Lilly	NY	2456783487

ID	Name	City	Phoneno
1	Sam	Brampton	4567824567
2	leo	Sudbury	4567824567
3	Naho	Paris	4567824567

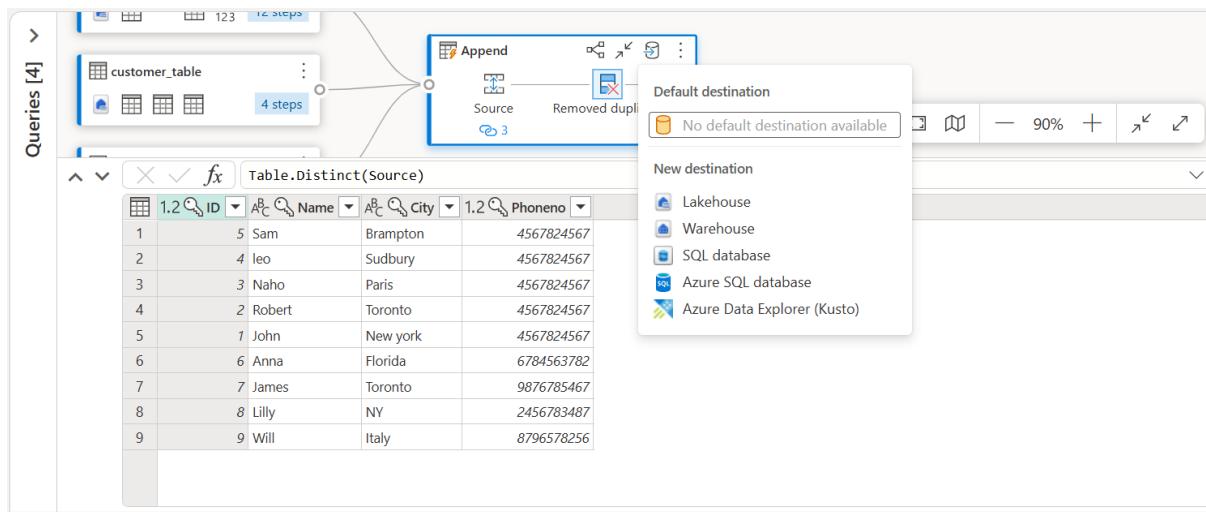
We got power query for all the steps we performed.

```
1 let
2     Source = Table.Combine({#"2025", customer_table, #"Newcustomerdata csv"}),
3     #"Removed duplicates" = Table.Distinct(Source),
4     #"Changed column type" = Table.TransformColumnTypes(#"Removed duplicates", {{"Column1", type text}, {"Column2", type text},
5     {"Column3", type text}, {"Column4", type text}})
6 in
7     #"Changed column type"
```

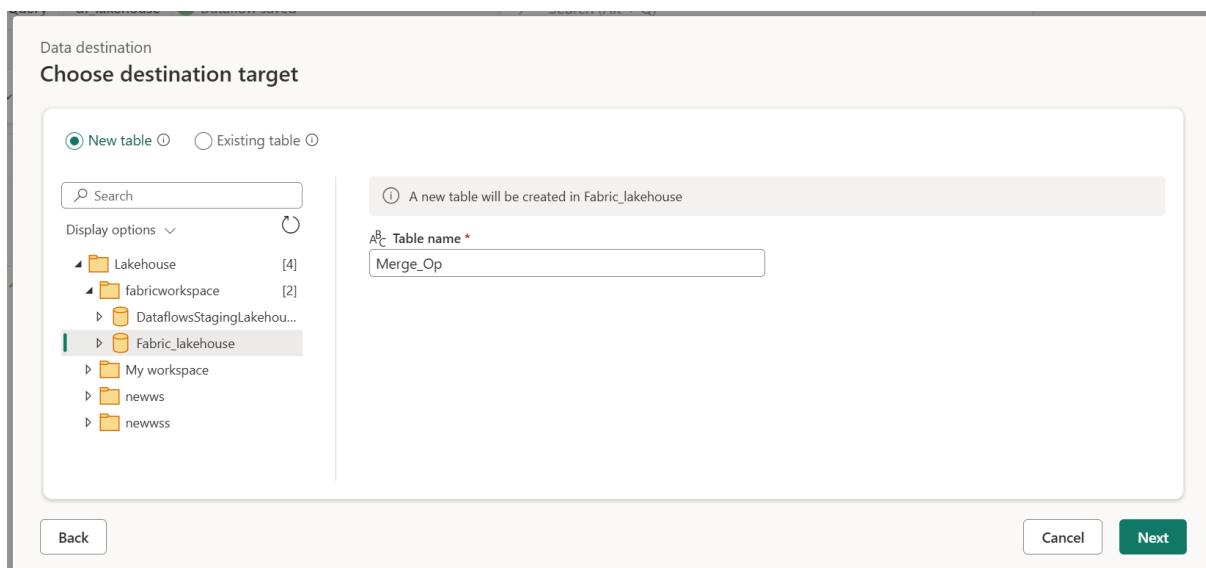
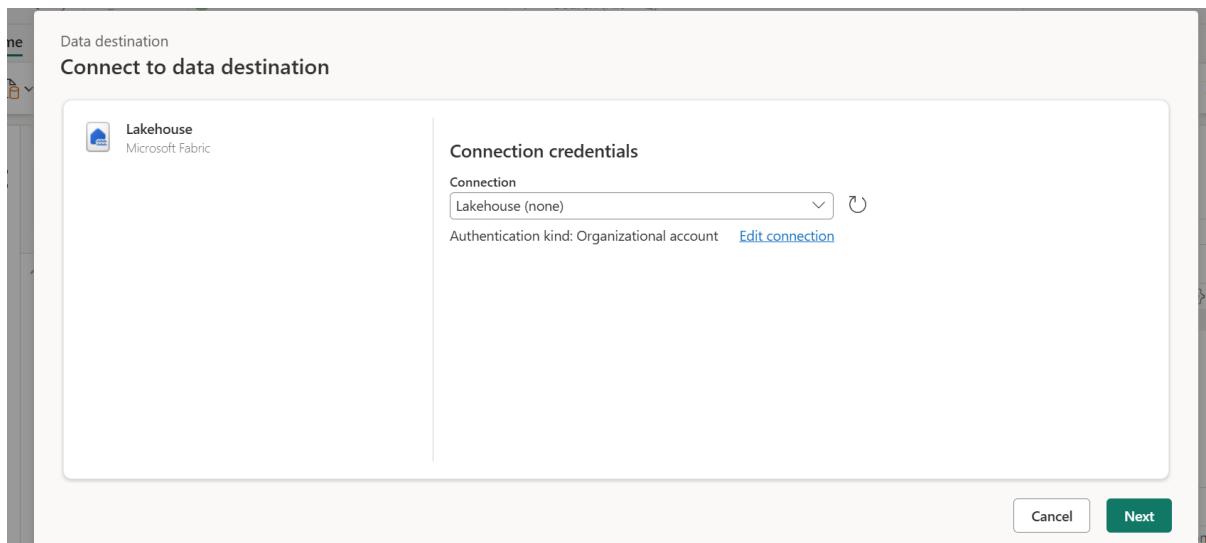
We can add our source directly in query as well.

Now we can add destination.

Now click on destination option on Append.



We can opt our destination from here, I want to store this data in a lake house.



Select which workspace lakehouse you want to store and also give the table name.

Data destination

Choose destination settings

To improve the performance of the data load into the destination, we are going to disable staging for the source query.

Use automatic settings

Column mapping

Source	Source type	Destination	Destination type
Name	Text	Name	Text
City	Text	City	Text
Phoneno	1.2 Decimal number	Phoneno	Decimal number

Back Cancel Save settings

If want to change any setting, turn off the Use automatic settings options.

Data destination

Choose destination settings

Use automatic settings

Update method

Append Replace

Schema options on publish

Dynamic schema Fixed schema

Column mapping

Back Cancel Save settings

Then we can change settings.

We can change update method, schema and also data types.

The screenshot shows the Power Query Editor interface with a data flow named 'df_lakehouse'. The flow consists of four steps: 'customer_table', 'Table.Distinct(Source)', 'Append', and 'Lakehouse'. The 'Append' step has a context menu open, displaying 'Update method' (set to 'Replace') and 'Column mapping' (described as 'Manual mapping and allow for schema change'). The 'Lakehouse' destination is selected in the 'Data destination' pane on the right. The bottom right corner features a large green 'Publish' button.

Destination is added, now publish dataflow by click on below right corner publish option.

The screenshot shows a dataflow preview window with a table containing 9 rows and 4 columns. The columns are labeled 8, 9, Name, and ID. The rows show data for 'Lilly' (NY, 2456783487) and 'Will' (Italy, 8796578256). Below the table, it says 'Columns: 4 Rows: 9' and 'Add default destination...'. To the right, there's a 'Publish' button with a dropdown menu. The menu has two options: 'Publish now' (highlighted in blue) and 'Publish later'.

We have 2 options, publish now or later.

Let's choose publish now.

The screenshot shows the 'fabricworkspace' list view. It displays two items: 'Dataflow' and 'df_lakehouse'. Both items are listed under the 'Type' column as 'Dataflow ...'. The 'Owner' column shows 'deefabric' for both. The 'Refreshed' column shows '—' for both. The 'Next refresh' column shows 'N/A' for both. The 'Included in app' column shows '—' for both. The 'Endorsement' and 'Sensitivity' columns also show '—' for both.

Dataflow is publishing.

This screenshot is similar to the previous one, but the 'df_lakehouse' item now has a different 'Last refresh' timestamp: '4/8/2025, 8:20 AM'. All other details remain the same.

Dataflow published.

Now go to lakehouse and check if data is loaded or not.

The screenshot shows the OneLake catalog interface. On the left, the 'Explorer' sidebar shows a tree structure with 'Fabric_lakehouse' expanded, showing 'Tables', 'customer_table', 'credit_table', and 'Merge_Op' selected. On the right, a table titled 'Merge_Op' is displayed with 9 rows. The columns are 'ID', 'Name', 'City', and 'PhoneNo'. The data matches the original table: Sam (Brampton, 4567824567), Robert (Toronto, 4567824567), Leo (Sudbury, 4567824567), John (New York, 4567824567), Naho (Paris, 4567824567), Will (Italy, 8796578256), Lilly (NY, 2456783487), Anna (Florida, 6784563782), and James (Toronto, 9876785467).

Data is updated correctly.

To generate a report for this automatically, we need to go to SQL analytical endpoint -> reporting -> Manage default semantic model -> choose which table we want.

Fabric_Lakehouse

Reporting

New report | New semantic model | Manage default semantic model | Automatically update semantic model

Explorer

+ Warehouses

Fabric_Lakehouse

Schemas

dbo

INFORMATION...

queryinsights

sys

Security

Views

My queries

Shared queries

Model layouts

Fabric

Manage default semantic model

Select or deselect tables for the semantic model.

Search

Select all

customer_table

Merge_Op

Views

Confirm

Cancel

New SQL query

Click confirm.

Fabric_Lakehouse

Reporting

New report | New semantic model | Manage default semantic model | Automatically update semantic model

Explorer

+ Warehouses

Fabric_Lakehouse

Schemas

dbo

INFORMATION...

queryinsights

sys

Security

Views

My queries

Shared queries

Model layouts

Fabric

Manage default semantic model

Select or deselect tables for the semantic model.

Search

Select all

credit_table

customer_table

Merge_Op

Working on it...

Confirm

New SQL query

Now go to semantic model -> 3 dots -> Auto- create report.

Fabric

File | Export | Save | Explore | Set alert | Edit | Show data table

Quick summary

Merge_Op

Sum of ID by City

City

Sum of ID

Italy 9.0

Toronto 9.0

NY 8.0

Florida 6.0

Brampton 5.0

Sudbury 4.0

Paris 3.0

New York 1.0

Sum of ID by Name

City

Sum of ID

Will 9.0

Lili 7.0

James 6.0

Anna 5.0

Sum of Phoneno by City

City

Sum of Phoneno

Toronto 14bn

Italy 9bn

Florida 5bn

Brampton 5bn

New York 5bn

NY 2bn

Sum of ID by Phoneno

Phoneno

Sum of ID

4,567,824 15

8,796,578 9

2,456,783 8

9,876,785 7

6,784,563 6

Your data

Search

Merge_Op

Count of rows

City

ID

Name

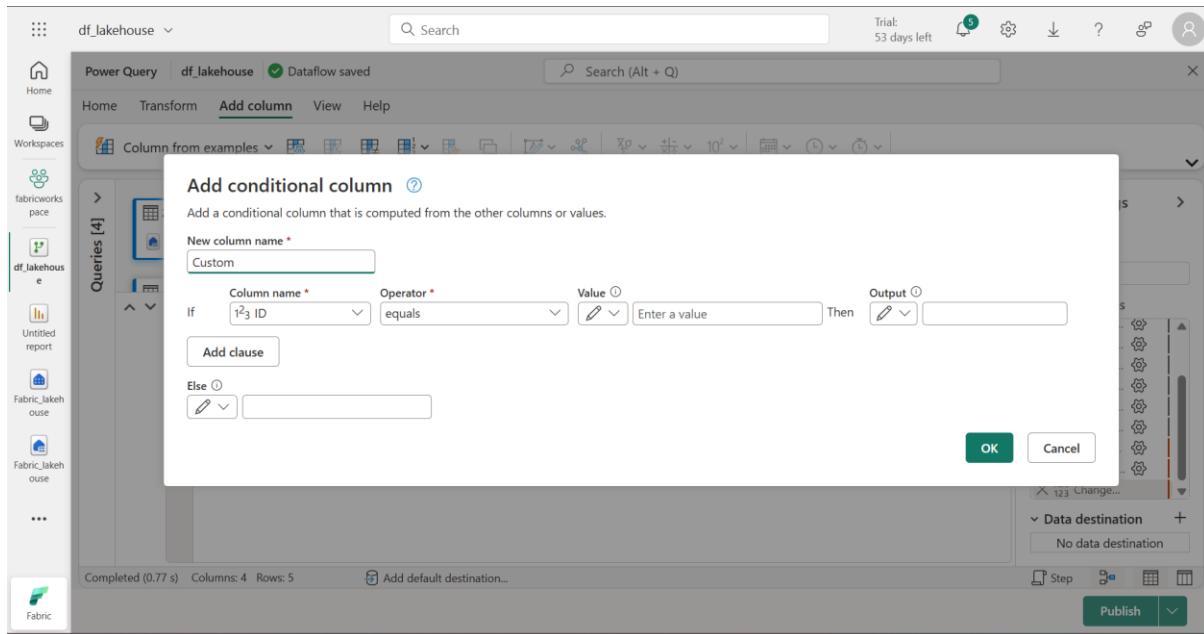
Phoneno

Report is generated.

If we need this, we can save this report.

In dataflow we can add a new column.

Go to Add column -> select what type of column we need to add -> custom or condition column



This is how we create a conditional column, this will be used full in applying SCD Type 1 logic.

Dataflow Gen 1 VS Gen 2.

Feature Area	Dataflow Gen1	Dataflow Gen2 (Fabric)
Engine	Power Query	Apache Spark
Storage	ADLS Gen2, Power BI	One Lake, Lakehouse
Performance	Moderate	High (distributed processing)
Orchestration	Limited	Full (via Data Pipelines)
Governance	Basic	Enterprise-grade (Fabric-native)