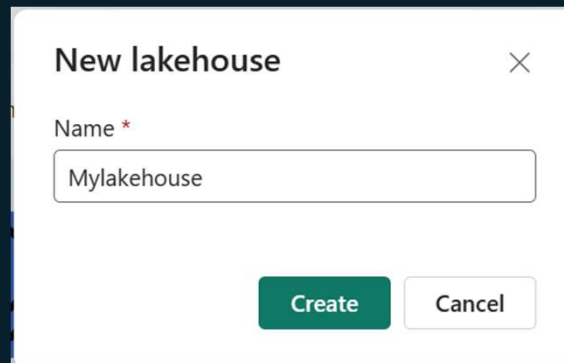# Lab – Build a Direct Lake dataset

Content in Lab

1. Build lakehouse
2. Ingest data
3. Build data model
4. Create report

Prerequisites

- Fabric workspace is created – i.e. workspace is in Trial, Premium capacity or Fabric capacity

Build lakehouse

1. In workspace, select New > Show all and select Lakehouse under Data Engineering,
   or
   In bottom left-hand corner, select Data Engineering and select Lakehouse on this screen

2. Provide a name for your lakehouse
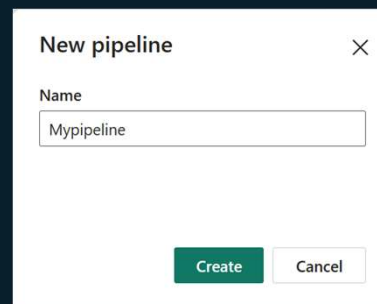
Ingest data

1. In the lakehouse,
   select New data pipeline
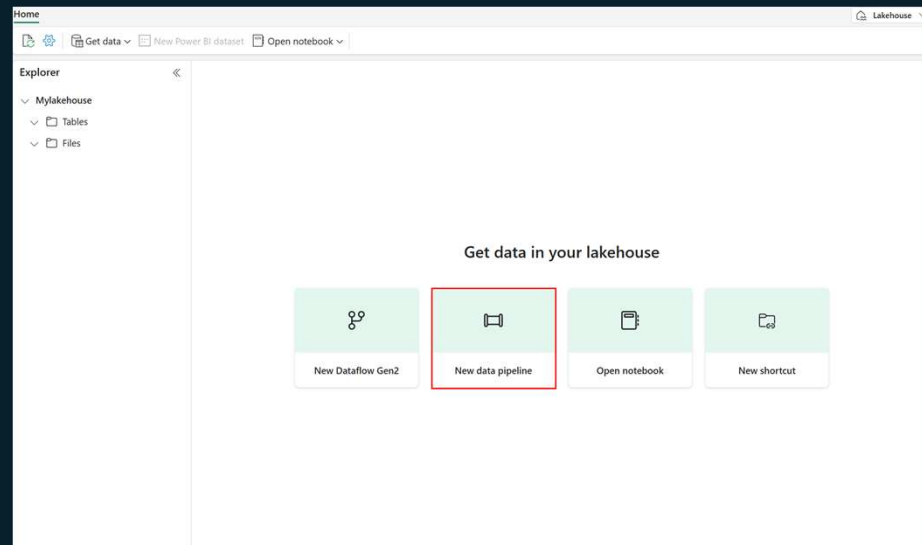
2. Provide a name for your pipeline

Ingest data

3. In the data pipeline experience, "Copy data" activity has been pre-selected.
    Scroll to "Data sources" and select "Azure SQL Database"
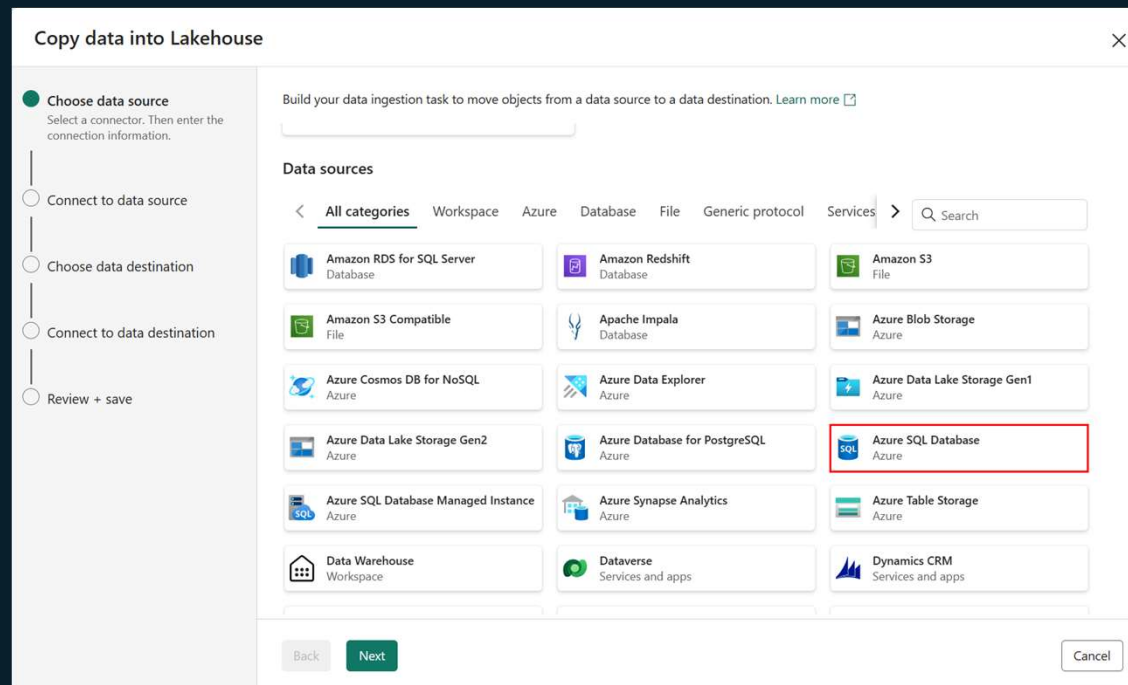
Ingest data

4. Use following connection settings:
   Server: laander.database.windows.net
   Database: DemoDW
   Authentication kind: Basic
   Username: fabric5sep
   Password: fabric.12345

5. Select all 4 tables

Ingest data

6. Make sure your newly created lakehouse is set as your destination

7. Make sure all 4 tables are in the list of "Connect to data destination"
   preferably set "Table action" to Overwrite for each table, if you need to re-run

Ingest data

7. Save + Run your pipeline

Ingest data

8. Now your lakehouse has 4 tables

Build data model

1. In lakehouse view, click on "New Power BI dataset" to build the Direct Lake model

2. Include all 4 tables
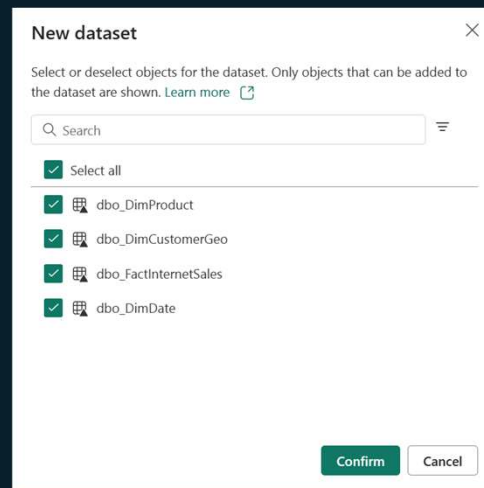
Build data model

3. You are now in the web modeling experience and need to configure relationships between fact and dimension tables

Build data model

4. Create the following relationships – drag column from/to dimension to/from fact:
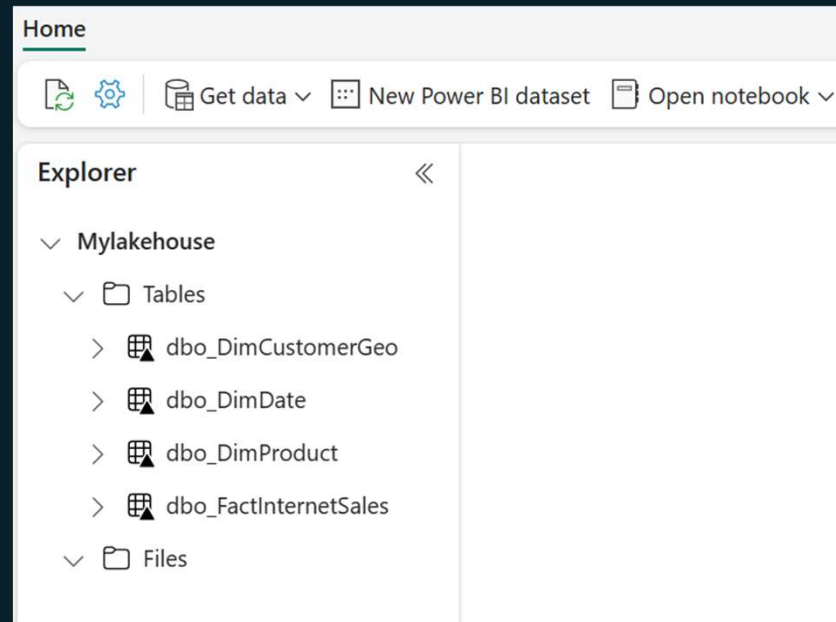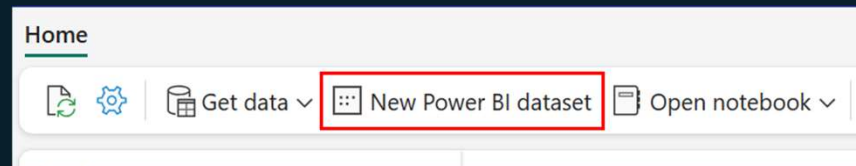
    dbo_DimCustomerGeo.CustomerKey = dbo_FactInternetSales.CustomerKey

    dbo_DimProduct.ProductKey = dbo_FactInternetSales.ProductKey

    dbo_DimDate.DateKey = dbo_FactInternetSales.OrderDateKey
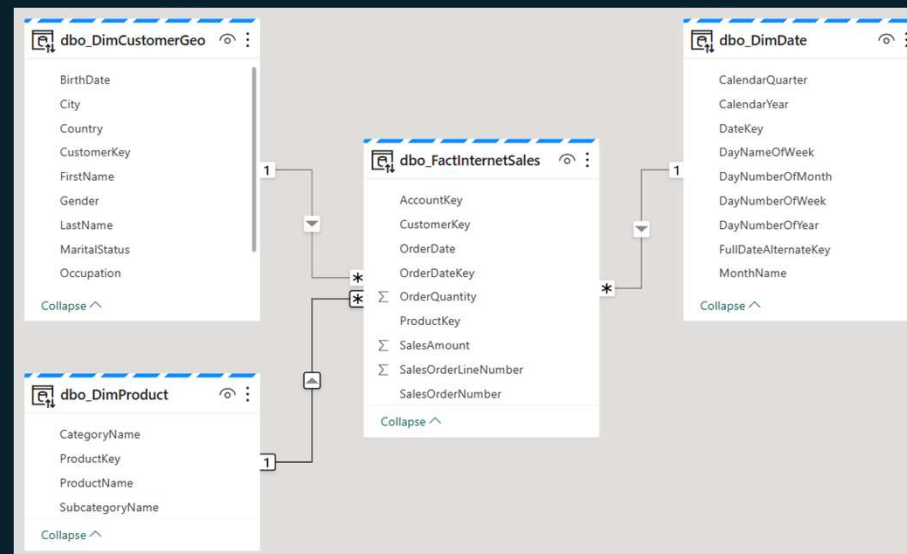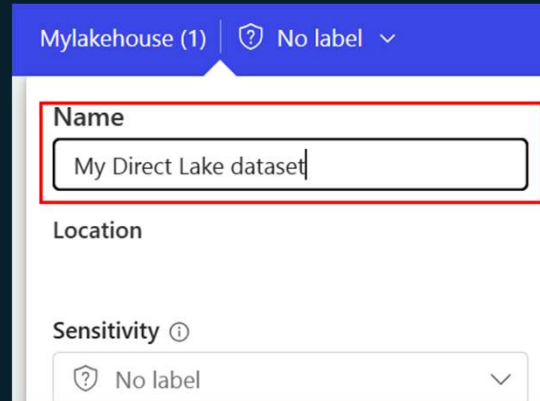
Build data model

5. Refine the model *(just a little bit, in real-world scenarios you will do more ☺)*
    > Hide all columns in dbo_FactInternetSales
    > Create two measures in dbo_FactInternetSales
    - Total Sales = SUM(dbo_FactInternetSales[SalesAmount])
    - Total Orders = SUM(dbo_FactInternetSales[OrderQuantity])
    > In dbo_DimDate select the column MonthName. In Properties > Advanced, change "Sort by column" to MonthNumberOfYear
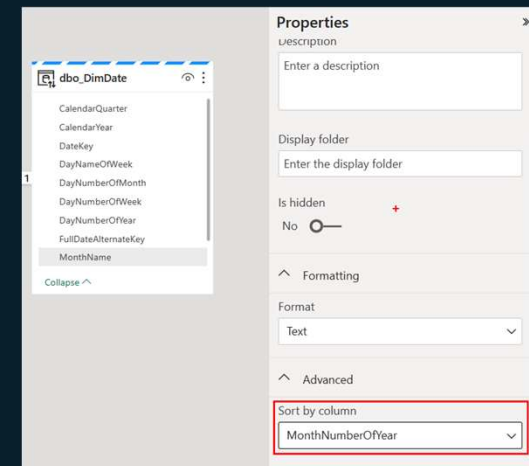
Build data model

6. Optionally, rename your Direct Lake dataset – click name in upper left corner

Create report

1. If you are in the web modeling experience, select "New report"
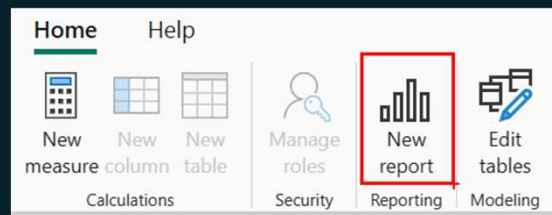
Else navigate to workspace and select "Create report"
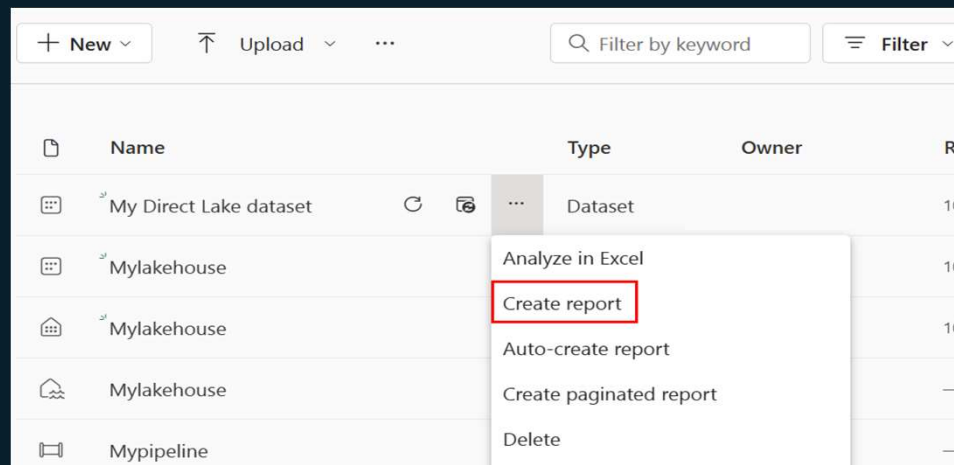
Create report

2. Look at the data model and build your own report, it could look something like this

Extra – Refreshing your Direct Lake dataset

1. You can configure refresh on your Direct Lake dataset by going to the setting of the dataset

Extra – Refreshing your Direct Lake dataset

2. Initially, you must enter credentials for your dataset. Select "Edit credentials" under "Data source credentials"
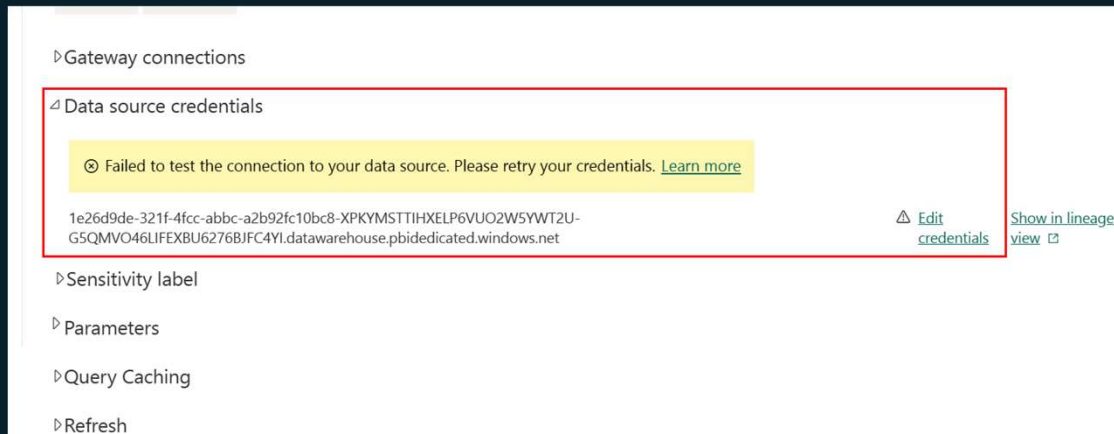
Use "Authentication method" OAuth2 and sign in with your Fabric credentials
Notice: this is credentials for your lakehouse and not the Azure SQL database
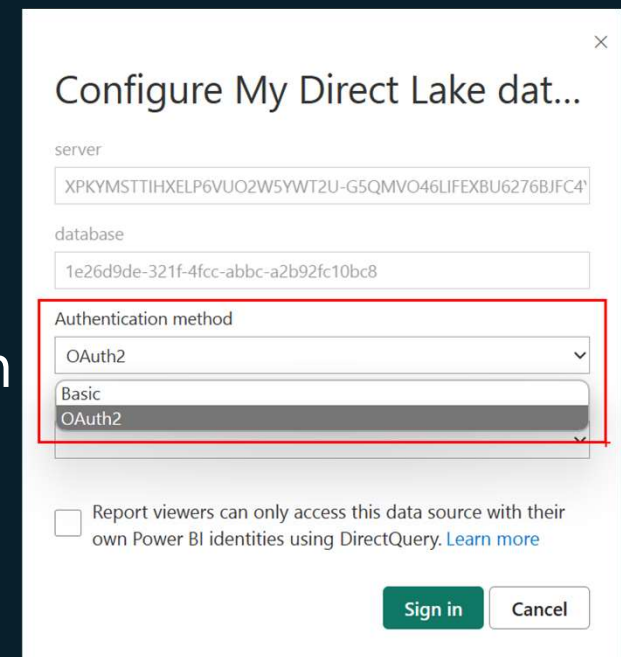
Extra – Refreshing your Direct Lake dataset

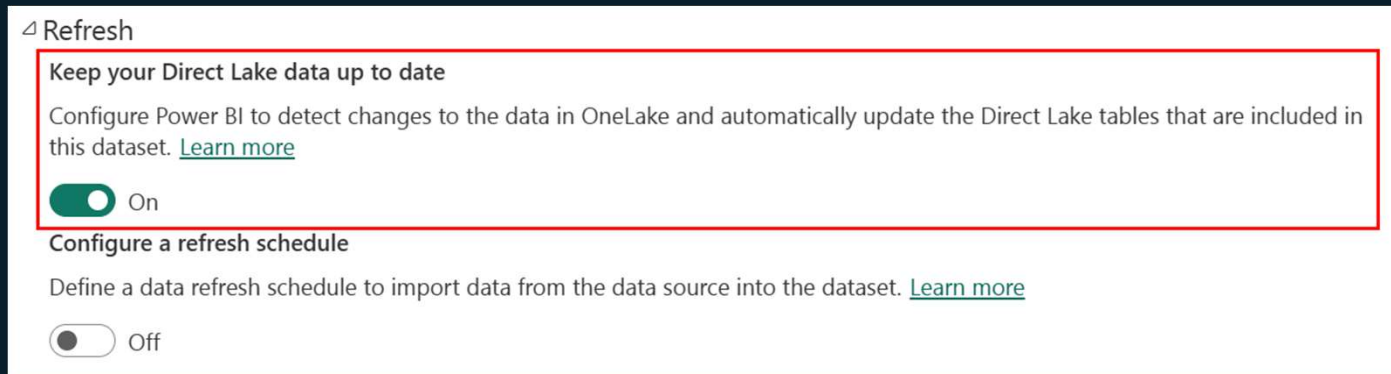2. Under Refresh a Direct Lake dataset has a new setting "Keep you Direct Lake data up to date" which is turned on by default

Having this setting turned on mean that whenever new data is added to your table in lakehouse, it will automatically be available in your Direct Lake dataset. Turning the setting off mean that you must control the refresh of your dataset either by "Configure a refresh schedule" or trigger refresh using API

Direct Lake documentation

[Learn about Direct Lake in Power BI and Microsoft Fabric](https://learn.microsoft.com/en-us/power-bi/enterprise/directlake-overview)
(https://learn.microsoft.com/en-us/power-bi/enterprise/directlake-overview)