Nawatech

Beyond Engineering

nawatech.co

Dataflow Gen2 in Microsoft Fabric

Hands-on: Dataflow Gen2

📝 Step 1: Connect to the Data Source

1. Open **Microsoft Fabric** and go to **Dataflows**.
2. Click **New Dataflow** → **Add a Table**.
3. Choose **Lakehouse** as the data source.
4. Select the **Lakehouse ID** and find your file (e.g., `flight.csv` or ‘flights.txt).
5. Click **Create**

📌 Step 2: Remove blank rows (For txt format)

Problem: The data have blank rows.

Solution:

In the ribbon, go to **Home** → **Reduce rows** → **Remove rows** → **Remove blank rows**.

📌 Step 3: Process first row as header

Problem: first row as header

Solution:

In ribbon **Transform** → **Use first row as headers** → **Use first row as headers**

Hands-on: Dataflow Gen2

📌 Step 4: Add new column ‘Date_Parsed’

Problem: ‘Date_of_Journey’ have string type

Solution:

1. In ribbon **Add column** → select **Custom column**.
2. In **New column name**, fill ‘Data_Parsed’
3. In **Costum column formula**, fill from ‘Hands-on Day 2 Dataflow’ in GitHub

📌 Step 5: Remove ‘Date_of_Journey’ and replace ‘Date_Parsed’ with ‘Date_of_Journey’

1. Right-click the old `Date_of_Journey` column.
2. Click **Remove columns**.
3. Right-click the ‘Date_Parsed’ column.
4. Click **Rename**.

Hands-on: Dataflow Gen2

📌 Step 6: Remove columns

1. Block column like ‘Route’, ‘Dep_Time’, ‘Arrival_Time’, ‘Duration’, ‘Total_Stops’, ‘Additional_Info’
2. Right-click then select **Remove columns**

📌 Step 7: Group ‘Airline’ and aggregate ‘Price’

1. In ribbon **Transform** → select **Group by**.
2. Select **Basic**.
3. In Group by, select ‘Airline’ column.
4. In **New column name**, fill ‘Sum_Price’.
5. In **Operation**, select ‘Sum’,
6. In **Column**, select ‘Price’
7. Click ‘**OK**’

Hands-on: Dataflow Gen2

📌 Step 8: Save the result to lakehouse

1. In Right Side Bar, Click '+' on **Data Destination** → select **Lakehouse**.
2. Select **New table**.
3. Click drop-down workspace 'fabric-data-kai' → select 'your lakehouse'
4. In **Table name**, fill the table name.
5. Click **Next**
6. Click **Save settings**

📌 Step 9: Publish the dataflow

1. In bottom right corner, select drop-down **Publish**
2. Select **Publish now**, if want to run now or
3. Select **Publish later**, if want to exit without run

Data Factory and Spark Notebook

Hands-on: Data Factory and Spark Notebook

📌 Step 1: Create Pipeline and Add Copy Data

1. Open **Microsoft Fabric** and go to **Factory Pipeline**.
2. In ribbon **Home** → **Copy data** → **Add to canvas**.
3. Click **Copy data**.
4. In **General** → **Name** → fill name

📌 Step 2: Fill Source in Copy Data

1. In **Connection** → Select **More**
2. In New sources, select ‘View more’ → **Azure Blobs**
3. In Account name or URL, fill URL from Readme in GitHub
4. In Authentication, select ‘Account key’
5. In Account key, fill key from Readme in GitHub → click **Next**
6. In **File path**, fill container with ‘workshop-kai’
7. In **File format**, select **DelimitedText**

Hands-on: Data Factory and Spark Notebook

📌 Step 3: Fill Destination in Copy Data

1. In **Connection** → Select ‘your lakehouse’
2. In **File format**, select **DelimitedText**
3. Click **Copy data**.
4. In **General** → **Name** → fill name

📌 Step 4: Add Notebook

1. In ribbon **Home**, select **Notebook**.
2. Click on **Notebook** activities
3. In **General**, fill your notebook name in **Name** field
4. In **Settings**, select ‘yout workspace’ in **Workspace** field (e.g. ‘fabric-data-kai’)
5. In **Notebook**, select ‘your notebook’

Hands-on: Data Factory and Spark Notebook

📌 Step 5: Add new Paramater Pipeline for looping copy data

1. Click on blank space.
2. In bottom, Click **Parameter**.
3. In **Name** field, fill ‘your parameter name’
4. In **Type** field, select **Array**.
5. In **Value** field, fill with [“flights”,”bookings”,”passengers”,”payments”,”airports”]

📌 Step 6: Add **ForEach** activities

1. In ribbon **Activities**, select **ForEach**.
2. Click on **ForEach** activities.
3. In **General**, fill your ForEach name in **Name** field
4. In **Settings**, click text bar in **Items** field → click **Add dynamic content**
5. Click **Parameter** → select ‘your parameter’ → **OK**

Hands-on: Data Factory and Spark Notebook

📌 Step 7: Add Copy activities on ForEach activities

1. Double click on ForEach activities.
2. In ribbon **Home** → **Copy data** → **Add to canvas**
3. Click on Copy activities.
4. In **General**, fill ‘your copy name’ on **Name** field

📌 Step 8: Fill source on Copy data in ForEach activities

1. Click Copy activities → click **Source**
2. In **Connection**, select ‘your lakehouse’
3. In **Root folder**, select **Tables**
4. In **Table**, click text bar → click **Add dynamic content**
5. Select **ForEach iterator** → select ‘your ForEach name’ → **OK**

Hands-on: Data Factory and Spark Notebook

📌 Step 9: Fill destination on Copy data in ForEach activities

1. Click Copy activities → click **Destination**
2. In **Connection**, select **More**
3. In **New Fabric item**, select **Warehouse**
4. In **Name**, fill ‘your warehouse name’
5. Click **Create and connect**
6. In **Table option** → **Auto create table**
7. In **Table**, click on ‘table name’ bar → click **Add dynamic content**
8. In **ForEach iterator**, click ‘your

📌 Step 10: Run the pipeline

1. In pipeline space, click **Main canvas**
2. In ribbon **Home**, click **Validate** to check the pipeline is work and can to run
3. After pipeline has been validated, in ribbon **Home** → click **Run**