TASK

Let's suppose you have 3 different types of file

1. CUST_MSTR_20191112.csv
2. master_child_export-20191112.csv
3. H_ECOM_ORDER.csv

All these files will be in the data lake container You have to fetch all three types of files into their respective folders. Note: There could be multiple files on all 3 types for different dates for example CUST_MSTR_20191112.csv and CUST_MSTR_20191113.csv

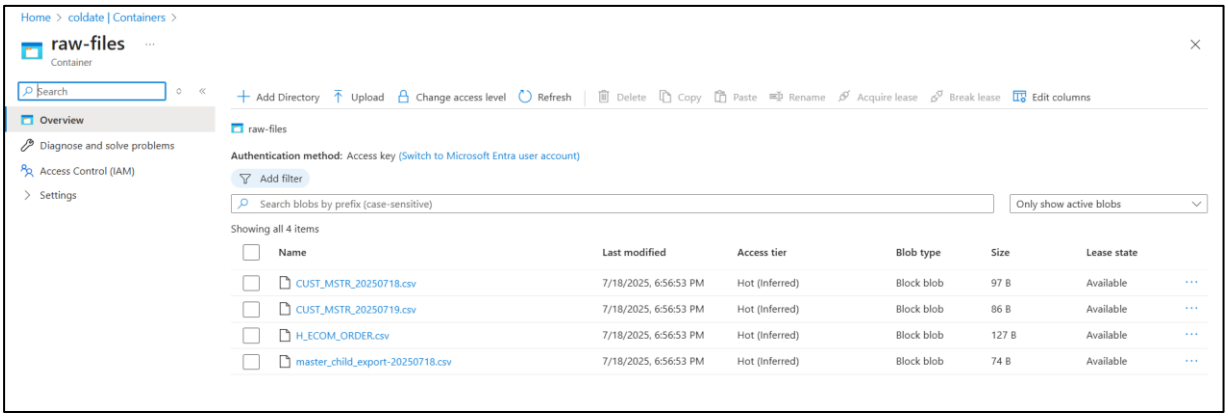
1. For the "CUST_MSTR" starting name of the file You have to create an additional column for a date that will fetch the data value from the filename and put it into an additional column Date format: 2019-11-12 and load it into the "CUST_MSTR" table.
2. For the "master_child_export" starting name of the file You have to create two additional columns date and date key which will fetch the data from the filename and put it into the additional columns. Date format: 2019-11-12 DateKey format: 20191112 and load it into the "master_child" table.
3. For the "H_ECOM_ORDER" type of file you have to load it into the database as it is. and load it into "H_ECOM_Orders" table

Note: This process will work on truncate load on a daily basis

Step by Step Guide:

(The steps use the variable names from the pipeline creation process. If you used different variable names, adjust the steps accordingly.)

1. **Creating Azure Data Lake Storage (Gen2) with the files in a container.**



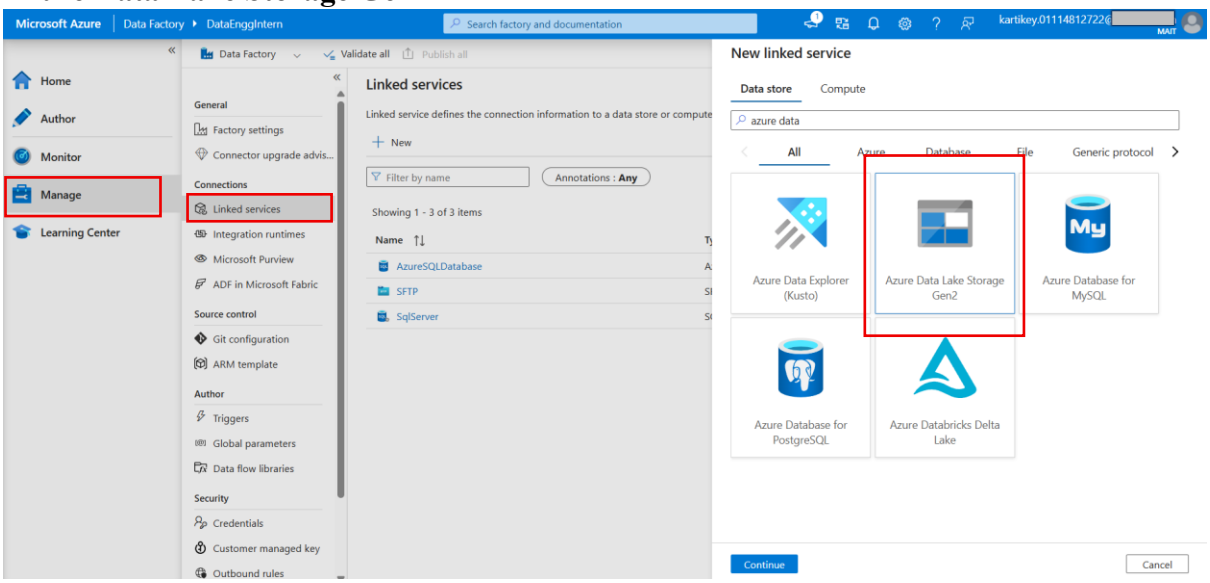
2. **Azure Data Factory instance**
   A. Create Linked Services

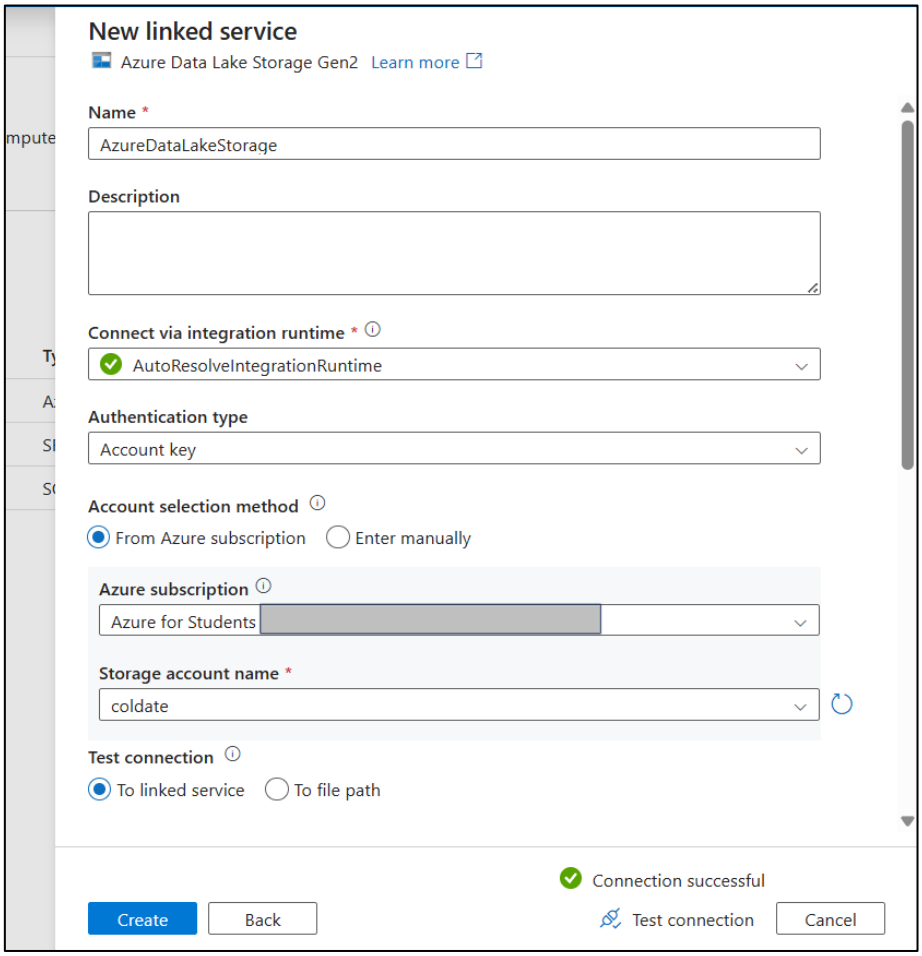
   In **Azure Data Factory**:

   Go to **Manage > Linked Services**.

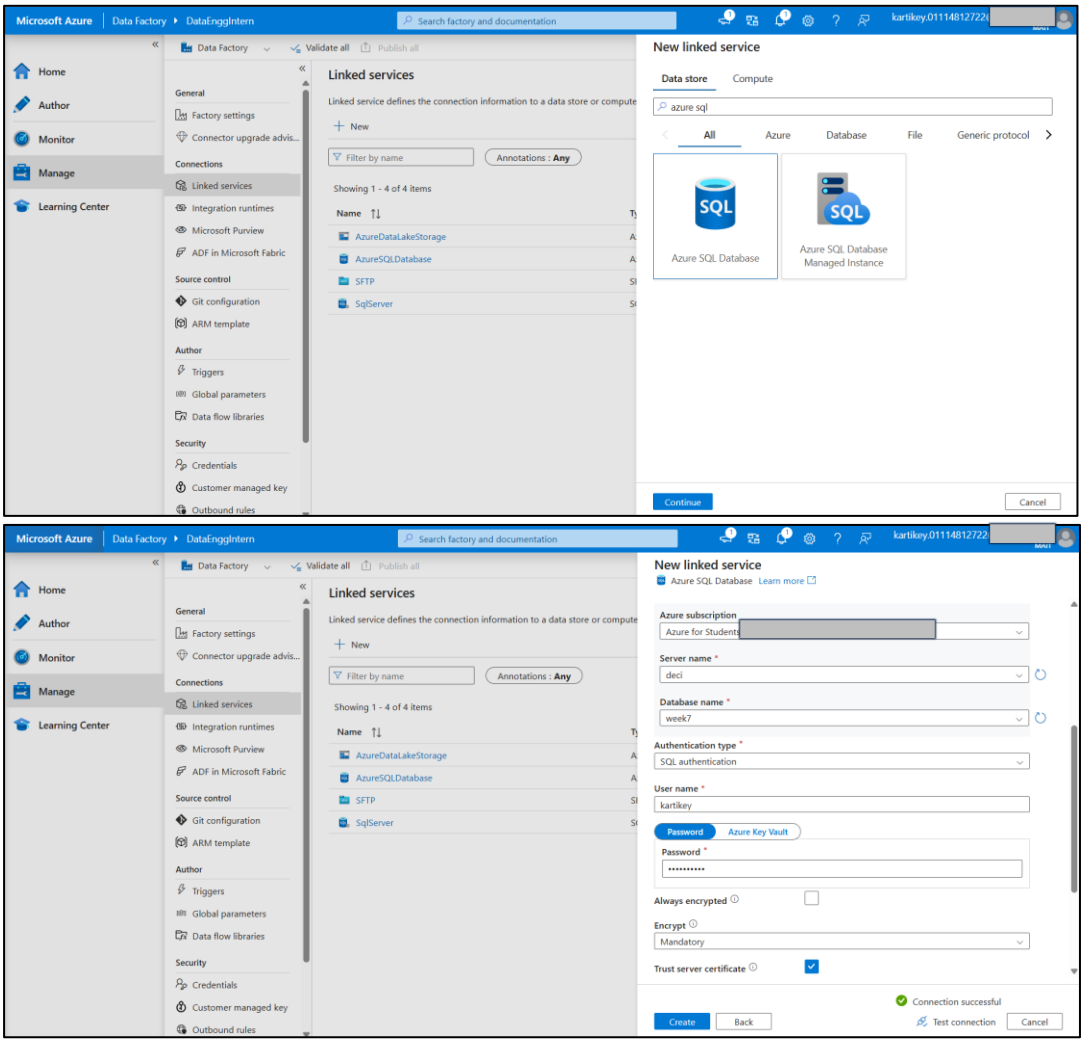
Create:

Azure Data Lake Storage Gen2 linked service



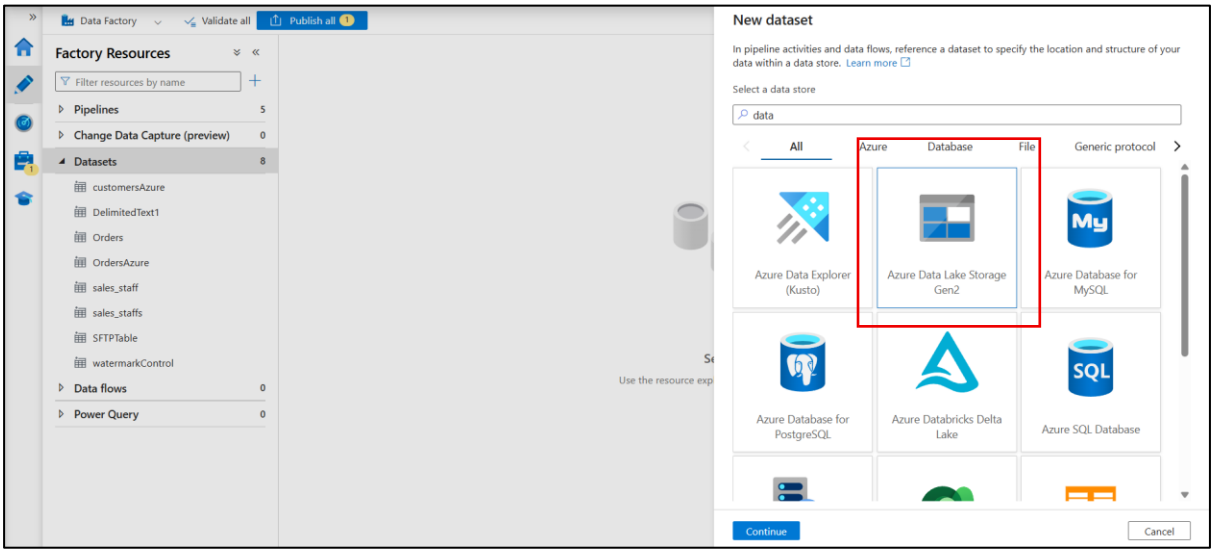


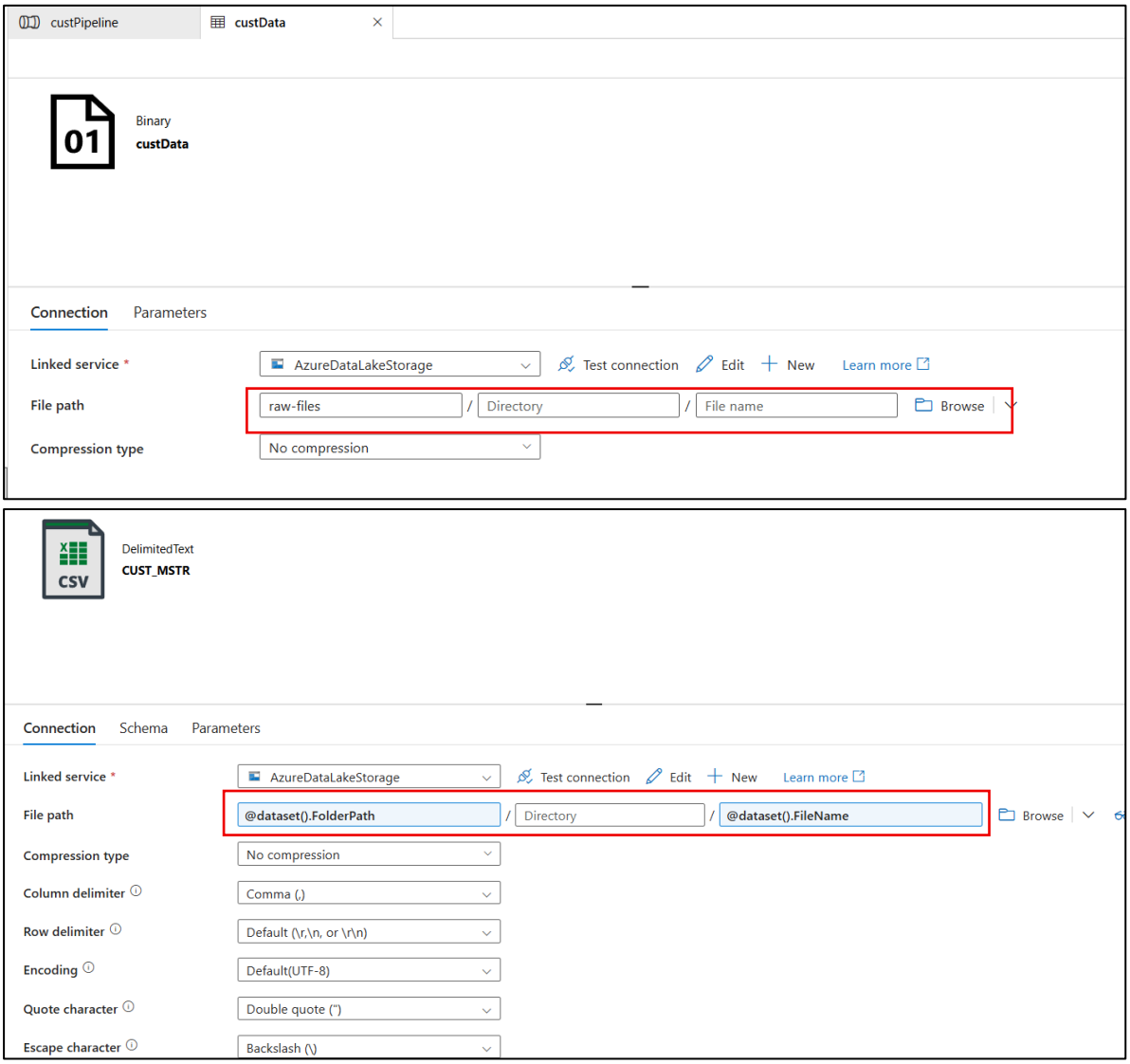
Azure SQL Database linked service



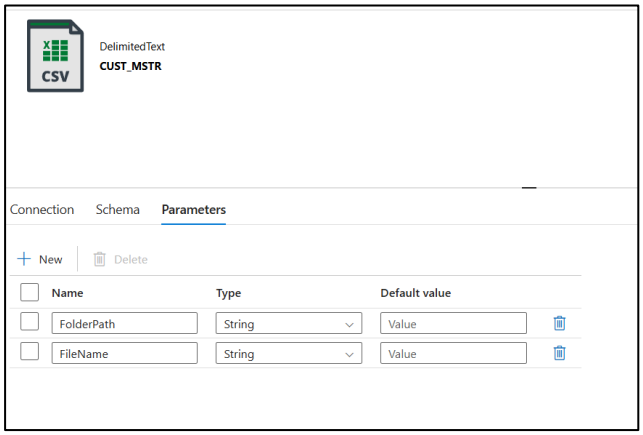
3. Create new datasets

Data Lake Storage Dataset (CSV format and Binary Format)





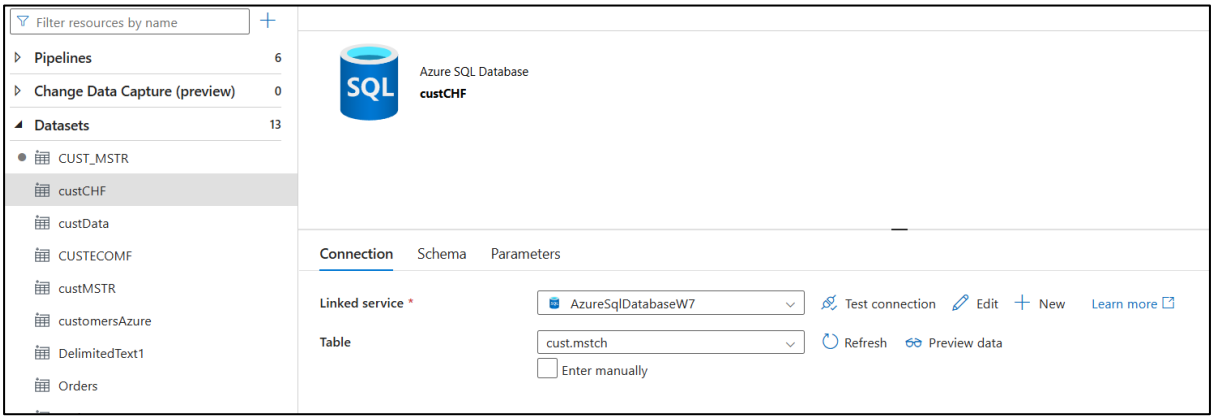
Insert parameters and update the file path as shown in the above image

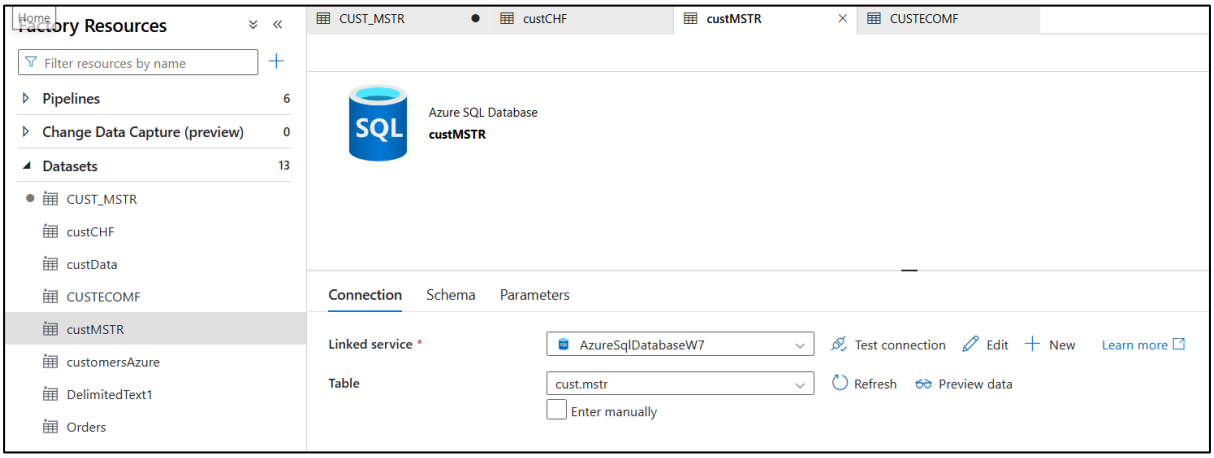


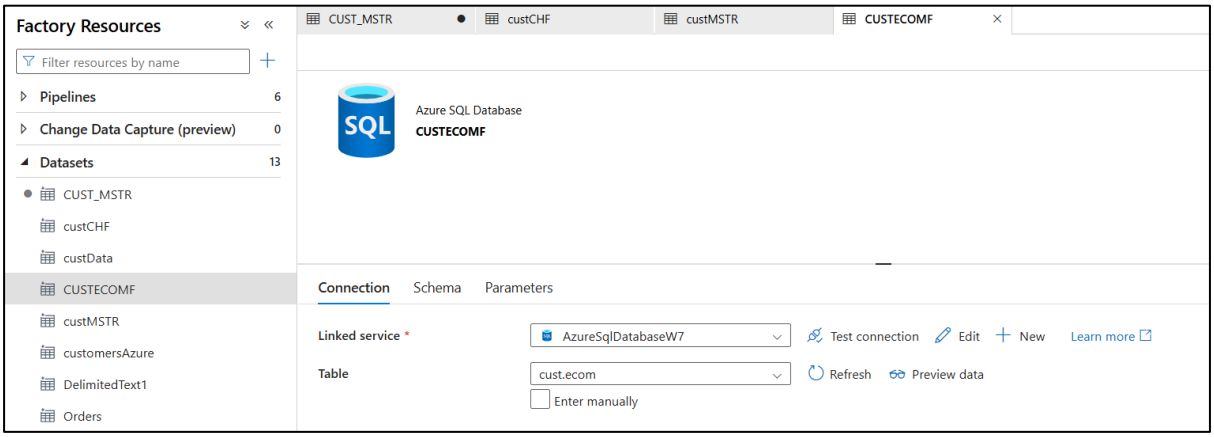
This ensures dynamic file handling during pipeline execution.

Azure SQL dataset

Create new datasets for Azure SQL in Data Factory







For creating SQL tables in Azure SQL:

CREATE SCHEMA cust;

1. cust.mstr

```
CREATE TABLE cust.mstr (
    CustomerID VARCHAR(10),
    CustomerName VARCHAR(100),
    Country VARCHAR(100),
    dateInserted date
```

)

2. cust.mstch

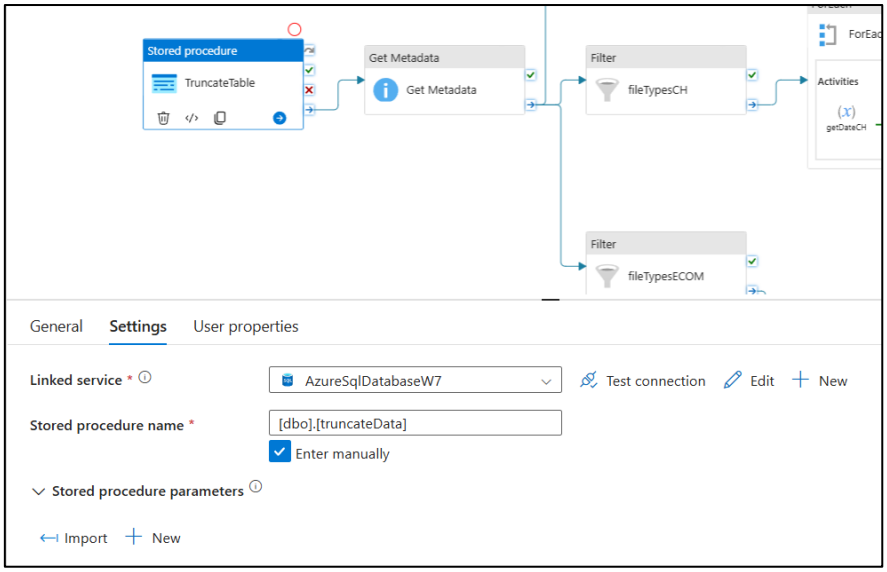
```
CREATE TABLE cust.mstch (
    CustomerID VARCHAR(10),
    CustomerName VARCHAR(100),
    Country VARCHAR(100),
    dateInserted date,
    dateFormated VARCHAR(10)
)
```

3. cust.ecom

```
CREATE TABLE cust.ecom (
    CustomerID VARCHAR(10),
    OrderID VARCHAR(10),
    OrderDate date,
    Amount float
)
```

4. **Create ADF pipeline**
   a. **Stored Procedure Activity**

As specified we need to truncate the table before insertng the data.

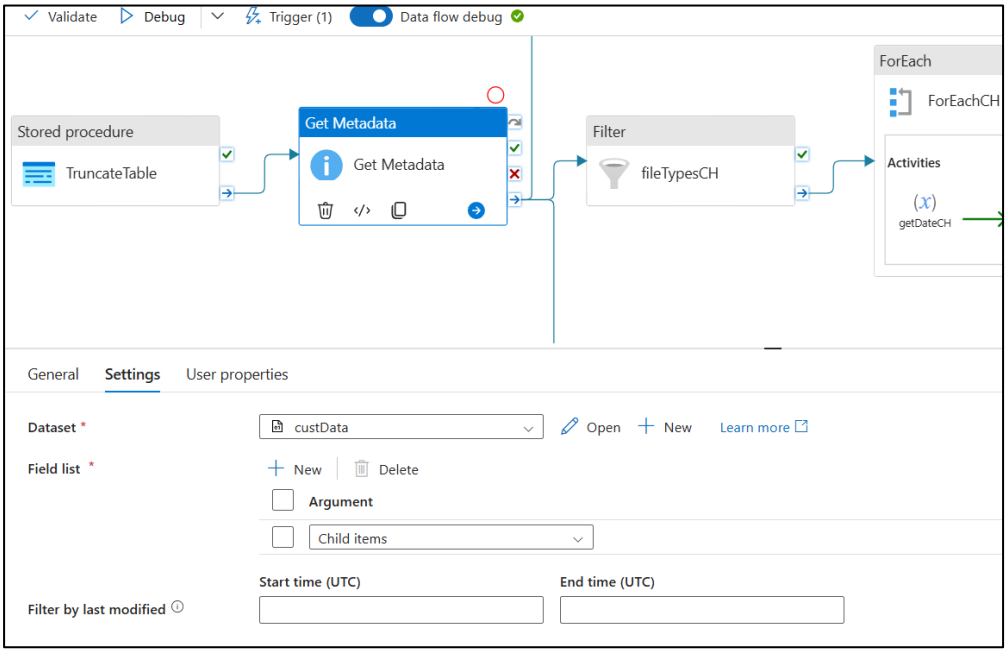


Where, stored procedure:

```
CREATE PROCEDURE truncateData
AS
BEGIN
   TRUNCATE TABLE cust.mstr;
   TRUNCATE TABLE cust.mstch;
   TRUNCATE TABLE cust.ecom;
END;
```

b. Create **Get Metadata** Activity

It extracts all the file names present in the directory.

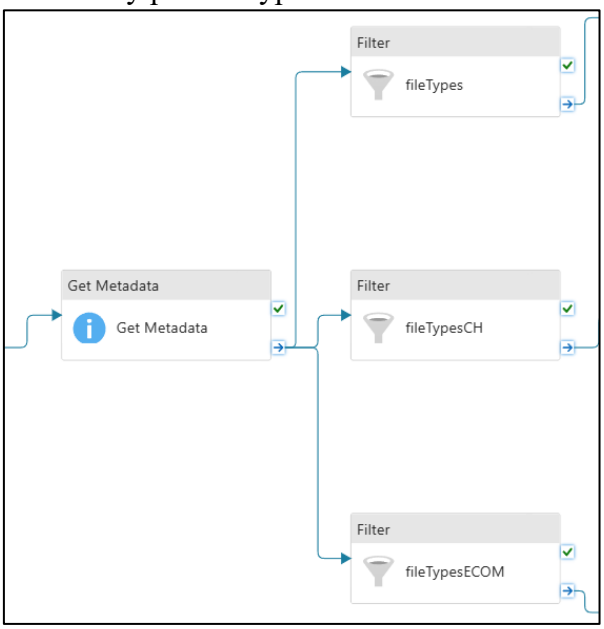


Dataset: custData (binary format dataset created earlier)

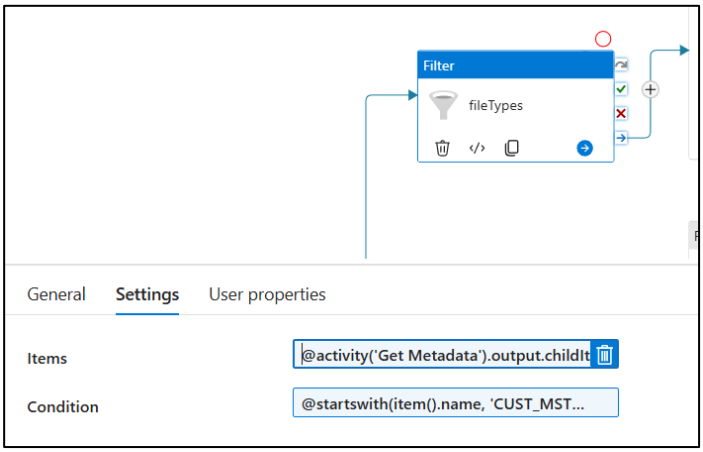
Set Field list: Child Items

c. Creating **Filter** Activity

One filter activity per file type



In each Filter Activity

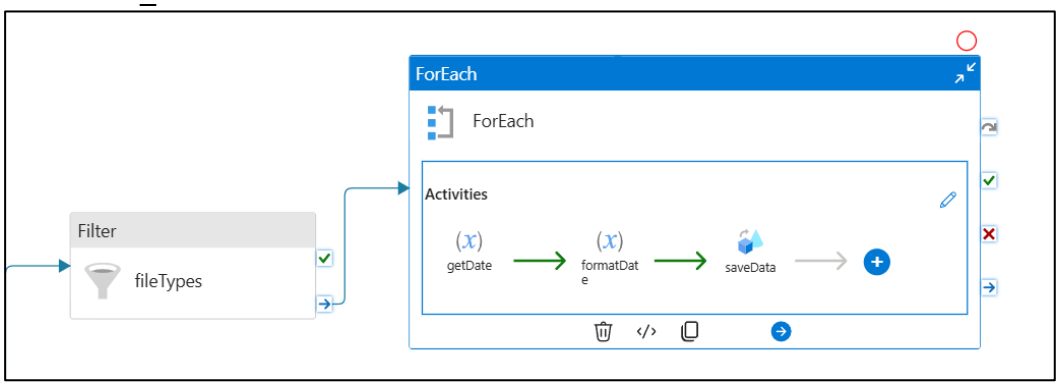


1. FileTypes (for CUST_MASTER)
   Items: @activity('Get Metadata').output.childItems
   Condition: @startswith(item().name, 'CUST_MSTR_')
2. FileTypesCH (for master_child)
   Items: @activity('Get Metadata').output.childItems
   Condition: @startswith(item().name, ' master_child_export-')
3. FileTypesECOM (ECOM)
   Items: @activity('Get Metadata').output.childItems
   Condition: @startswith(item().name, ' H_ECOM_ORDER')

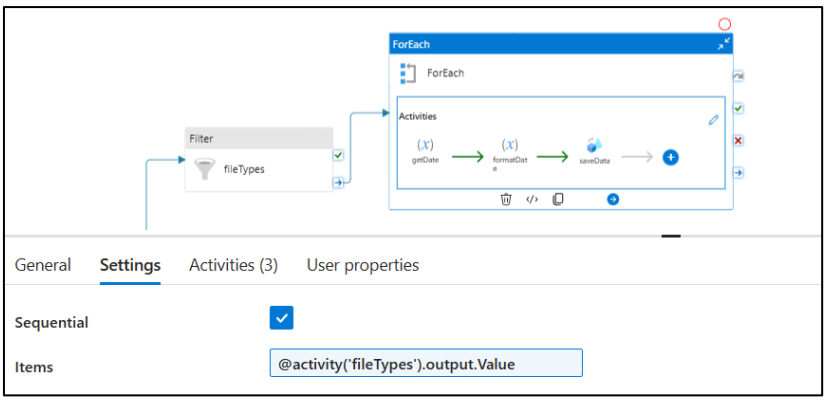
d. ForEach Activity

For each file type, we will create ForEach acitivity to loop over all the files of same type and update their data into Azure SQL.

1. For **CUST_MASTER**

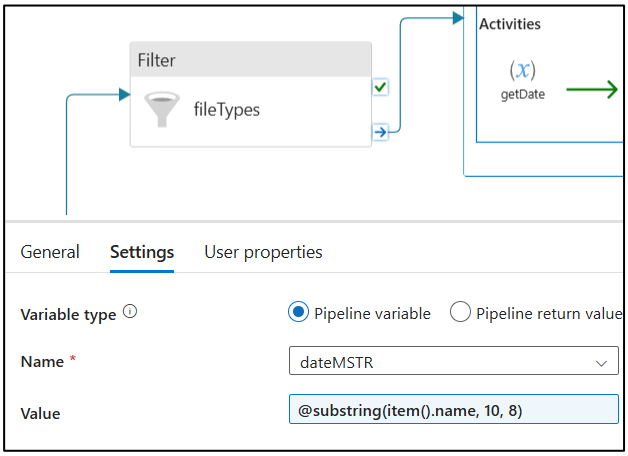


Configure the ForEach activity to take the name of files from Filter activity:



Activities to be inserted:

I. **Set Variable**



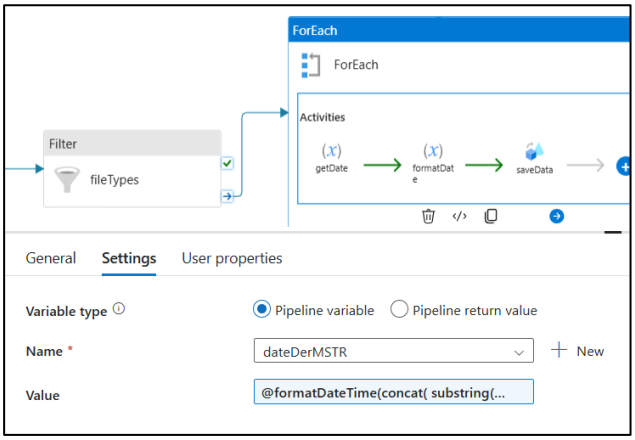
Value: @substring(item().name, 10, 8)

This variable extracts the date from the file

Eg. If file name is: **CUST_MSTR_20250718.csv**

The date starts from 10th letter and spans next 8 letters.

II. **Set Variable**

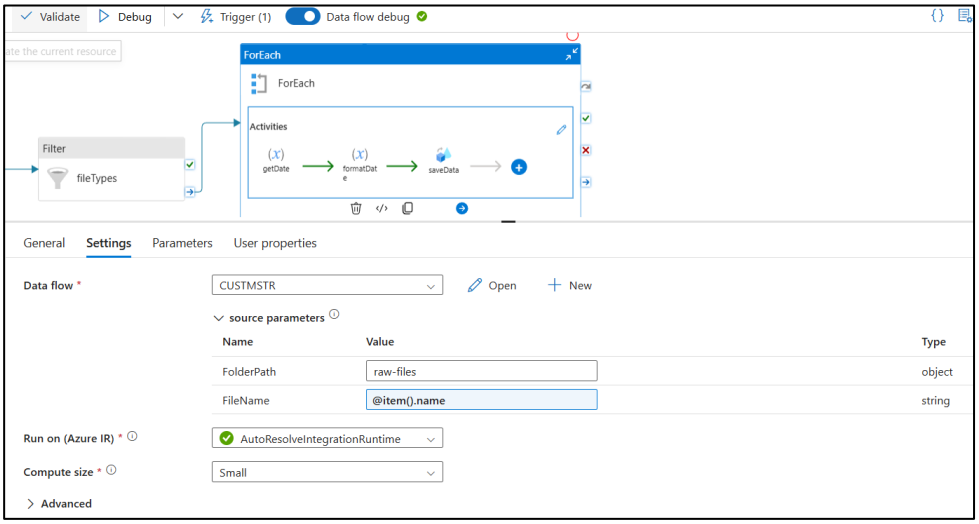


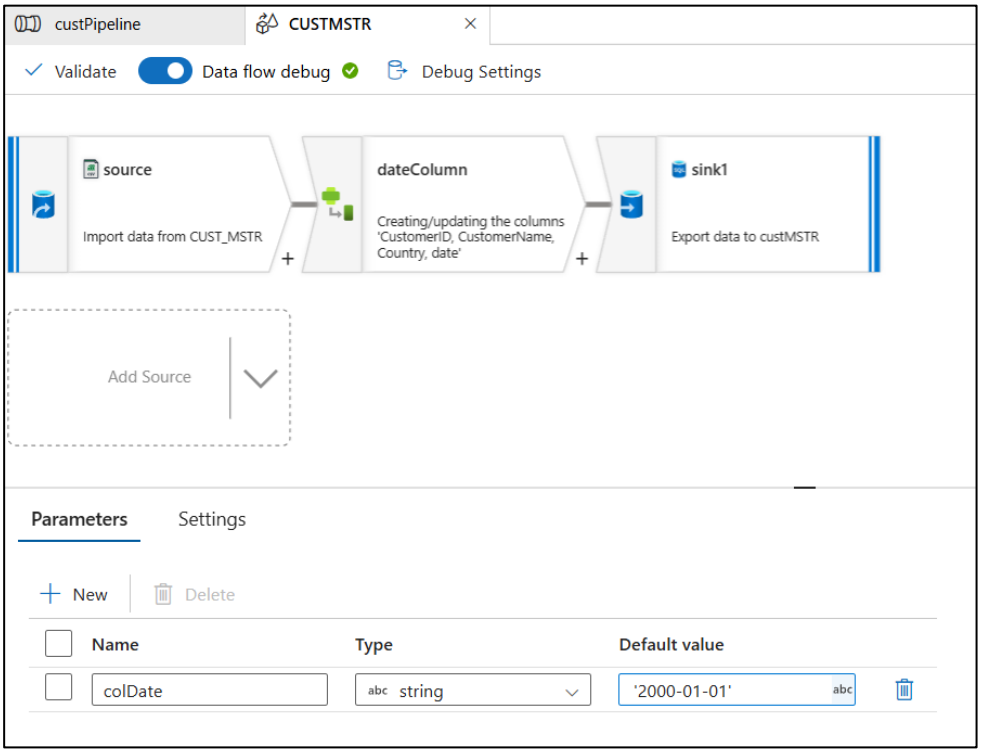
Value: @formatDateTime(concat(
substring(variables('dateMSTR'), 0, 4), '-',
substring(variables('dateMSTR'), 4, 2), '-',
substring(variables('dateMSTR'), 6, 2)
), 'yyyy-MM-dd')

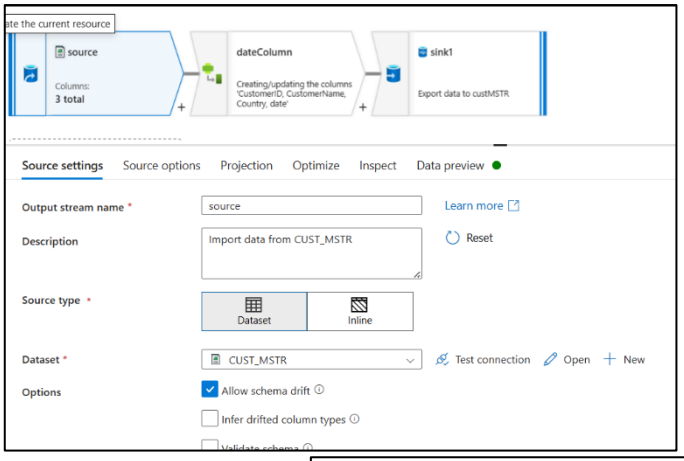
This variable converts the date into ISO format.

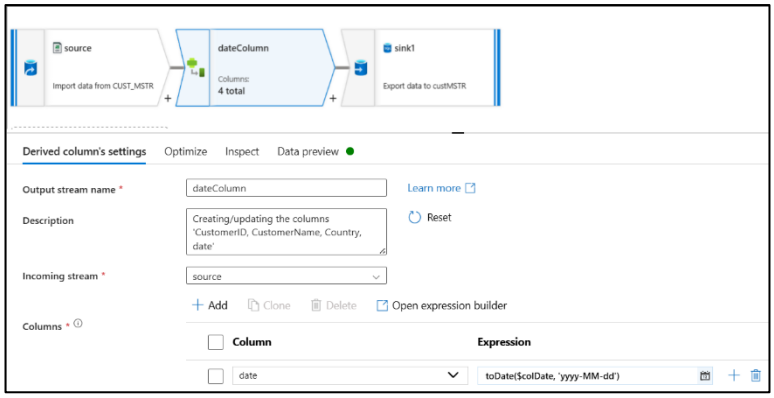
III. Data Flow

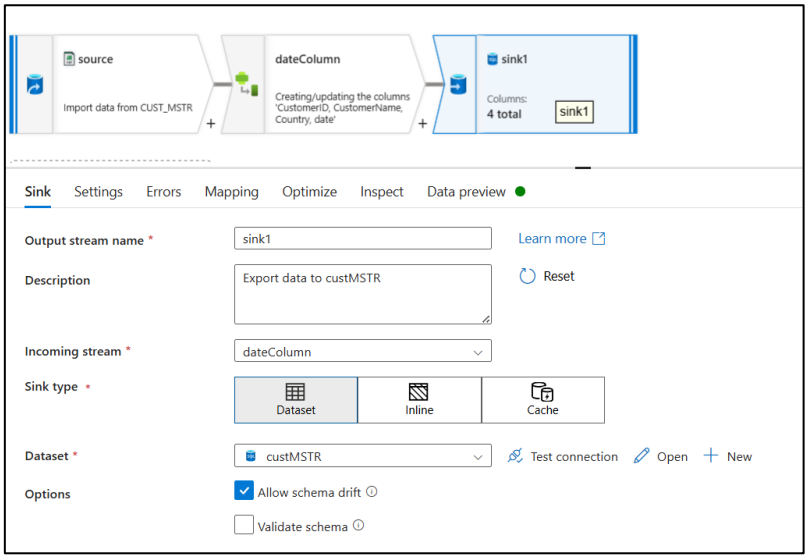
Create a new data flow





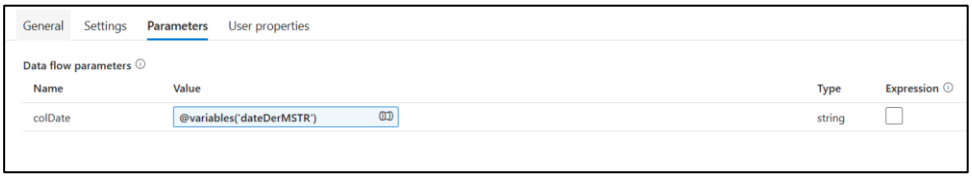






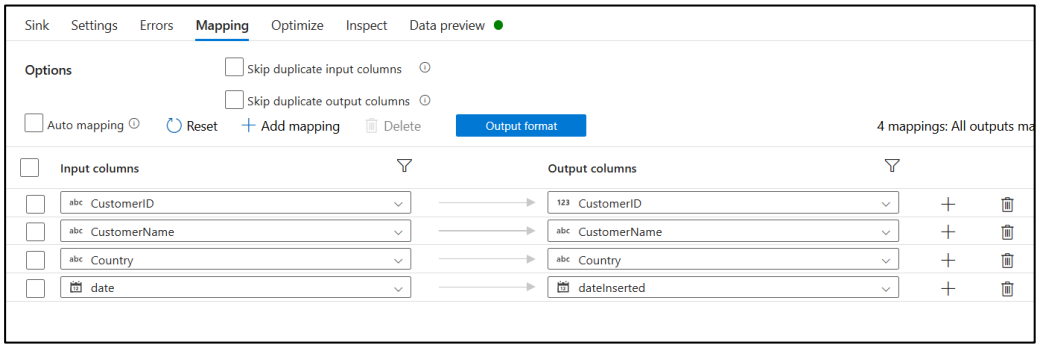
Where,

a. **Source Dataset:** CSV file of Data Lake Storage created earlier.
    a. Go to the **Projection** tab and import projection, provide the parameter values and import schema.
b. **DerivedColumn:** For adding the extra date column
    a. Create **date** column, where in expression, create new parameter: **colDate** of type string and set the expression as: **toDate($colDate, 'yyyy-MM-dd')**
    b. Go to the Data Flow Activity > Parameters
        i. Set colDate parameter:

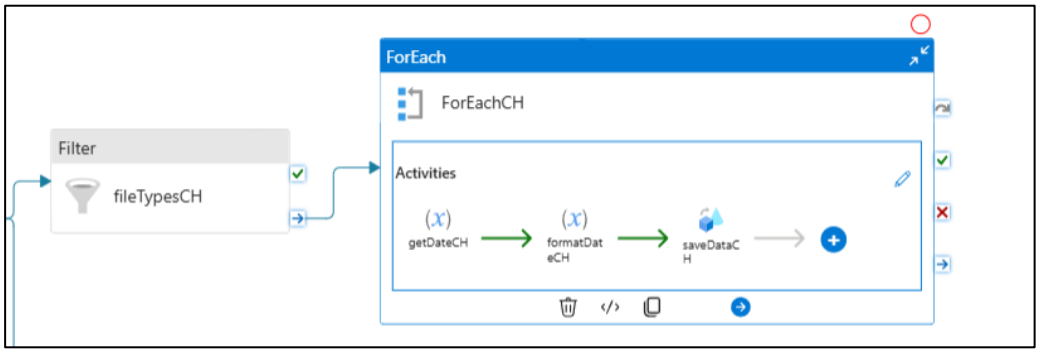


c. **Sink:** custMSTR table created in Azure SQL and imported in data factory

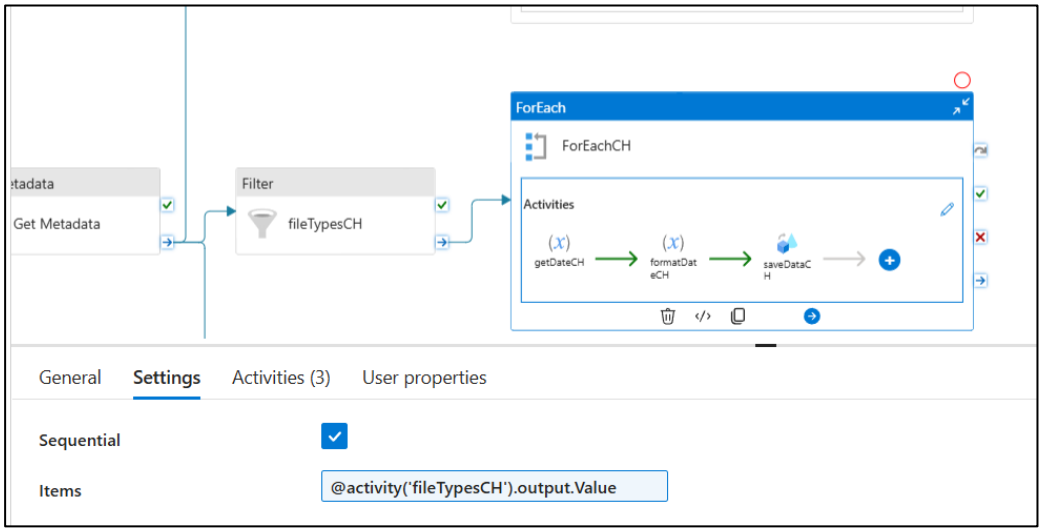
a. Go to **Mapping** tab, uncheck **Auto Mapping**, create manual mappings



2. For **Master_Child**

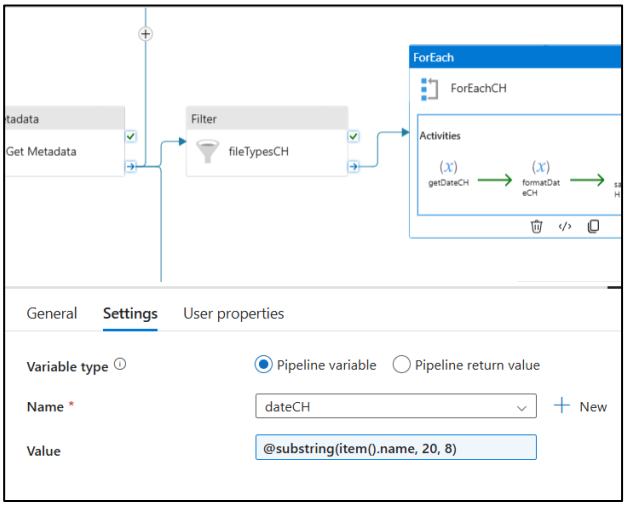


Configure the ForEach activity to take the name of files from Filter activity:



Activities to be inserted:

I. **Set Variable**



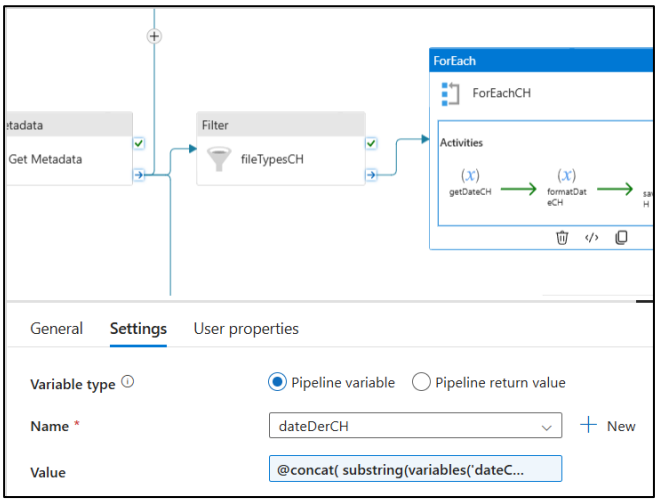
Value: @substring(item().name, 20, 8)

This variable extracts the date from the file

Eg. If file name is: **master_child_export-20250718.csv**

The date starts from 20th letter and spans next 8 letters.

II. **Set Variable**



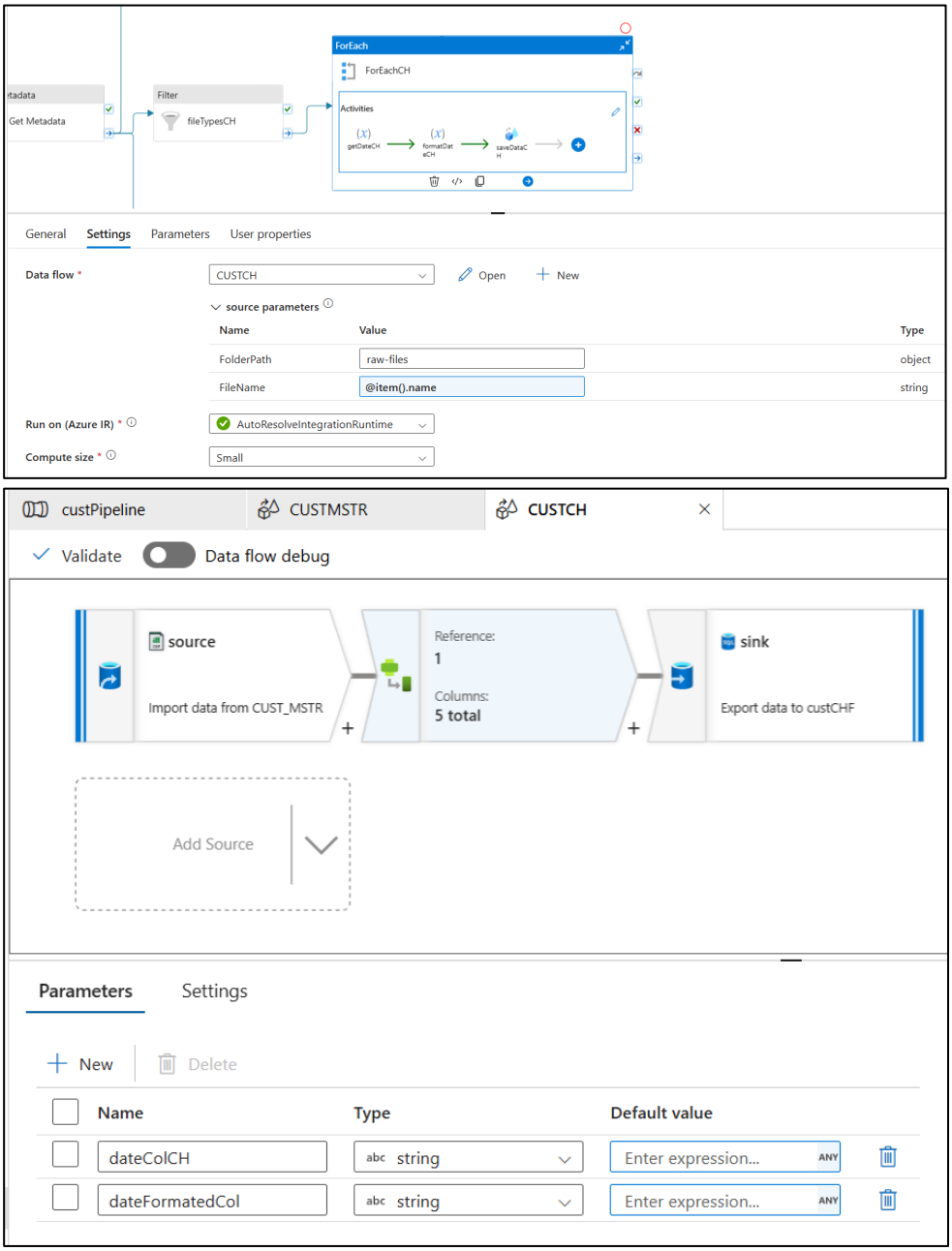
Value: @concat(

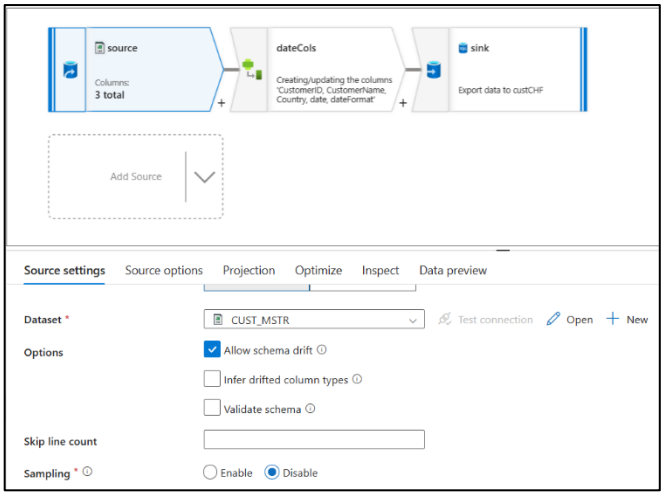
substring(variables('dateCH'), 0, 4), '-',
substring(variables('dateCH'), 4, 2), '-',
substring(variables('dateCH'), 6, 2)
)

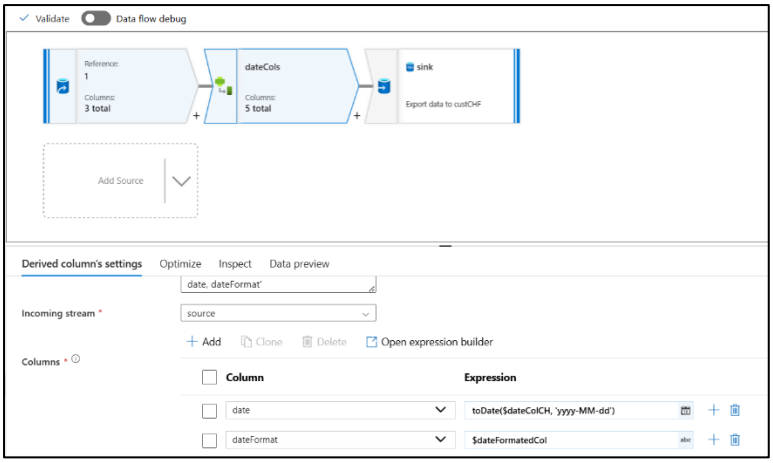
This variable converts the date into ISO format.

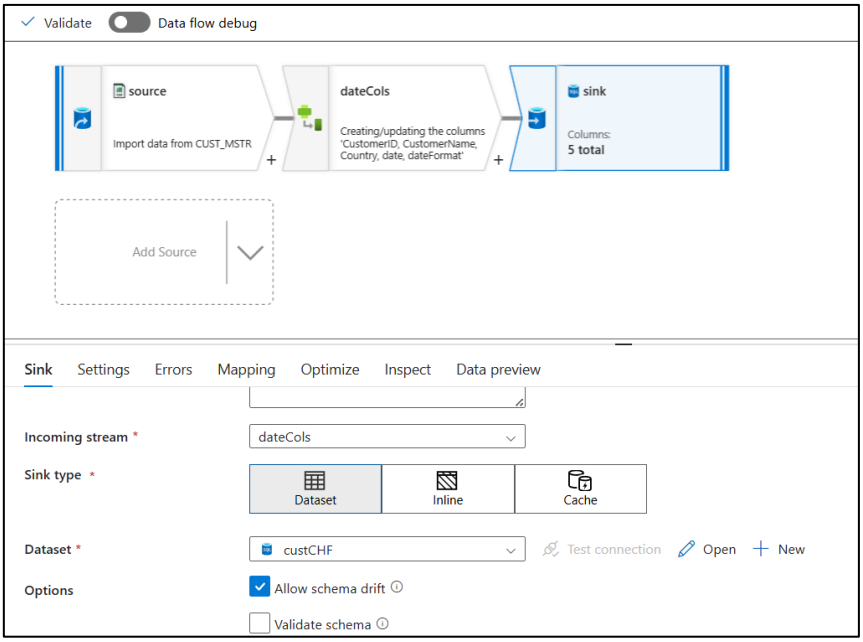
III. Data Flow

Create a new data flow







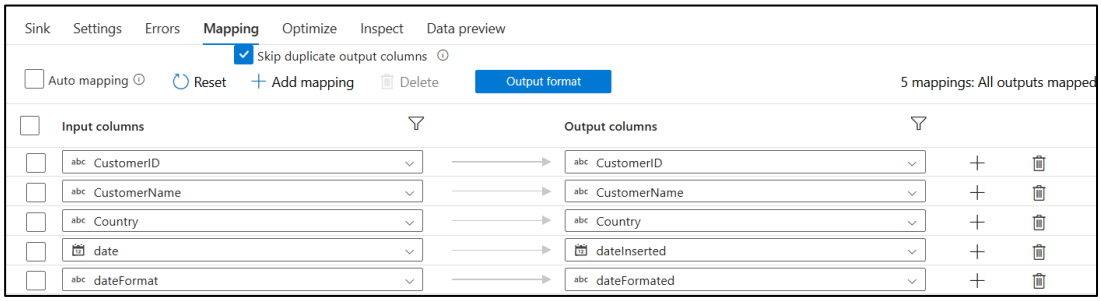


Where,

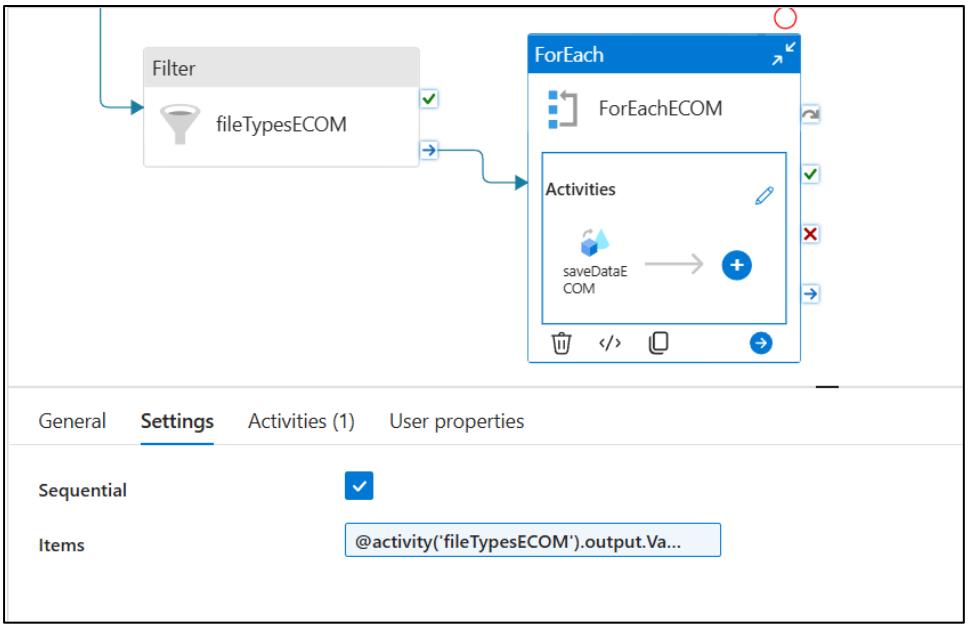
a. **Source Dataset:** CSV file of Data Lake Storage created earlier.
    a. Go to the **Projection** tab and import projection, provide the parameter values and import schema.
b. **DerivedColumn:** For adding the extra date column
    a. Create **date** column, where in expression, create new parameter: **colDate** of type string and set expression: **toDate($dateColCH, 'yyyy-MM-dd')**
    b. Create **dateFormat** column, where in expression, create new parameter: dateFormatedCol of type string and set expression: **$dateFormatedCol**
    c. Go to the Data Flow Activity > Parameters
        i. Set parameters:

c. **Sink:** custCH table created in Azure SQL and imported in data factory

a. Go to **Mapping** tab, uncheck **Auto Mapping**, create manual mappings



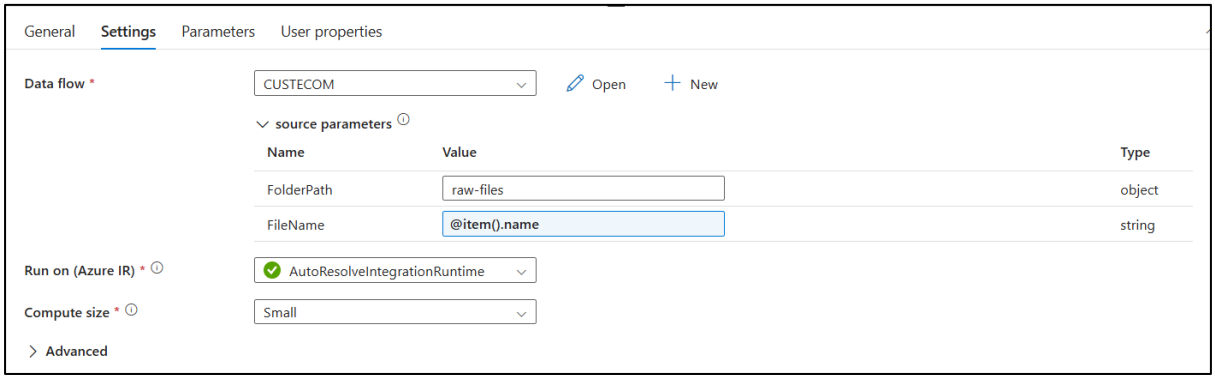
3. **ECOM_ORDER**



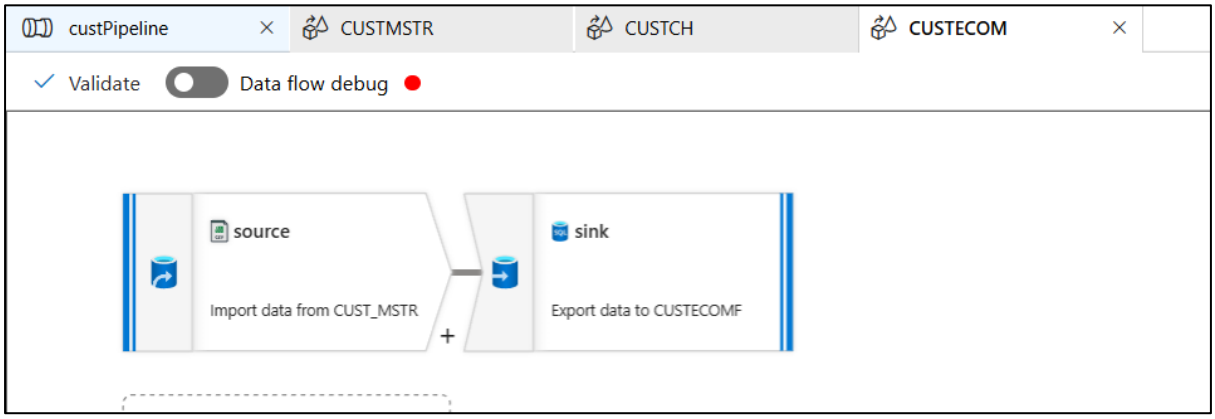
Items: @activity('fileTypesECOM').output.Value

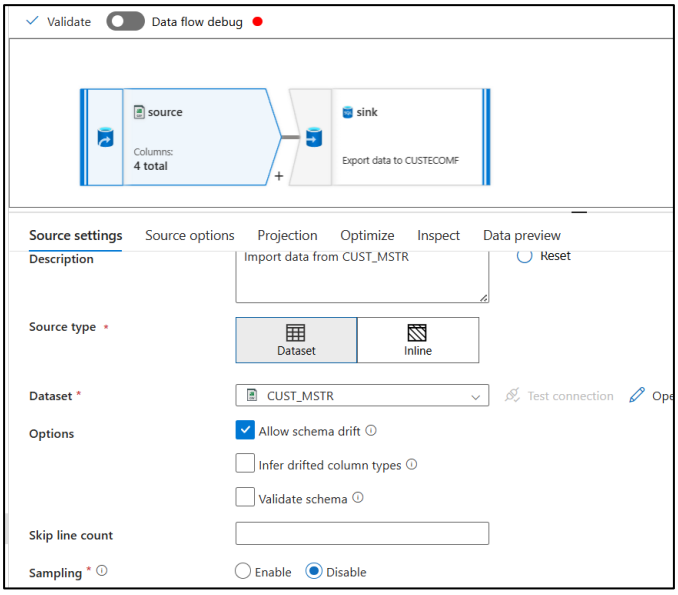
We only need to transfer data from file to the table, no additional columns are required.

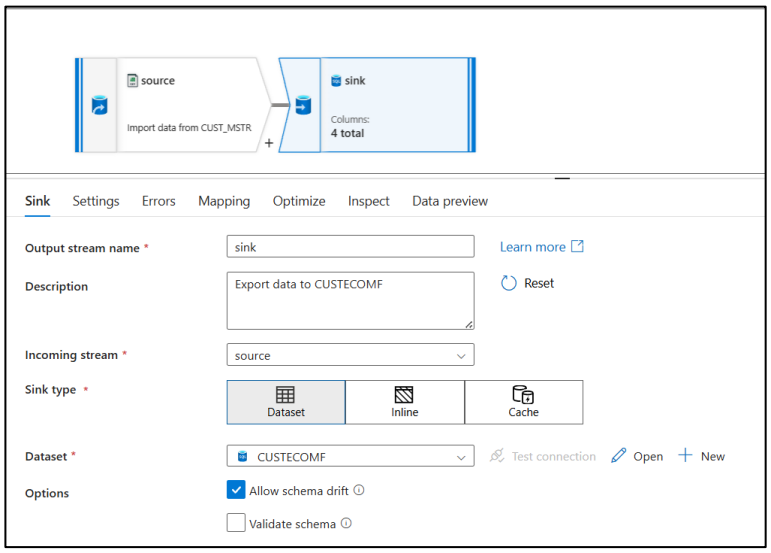
Therefore, only one activity: **Data Flow** is needed.



Create a new data flow





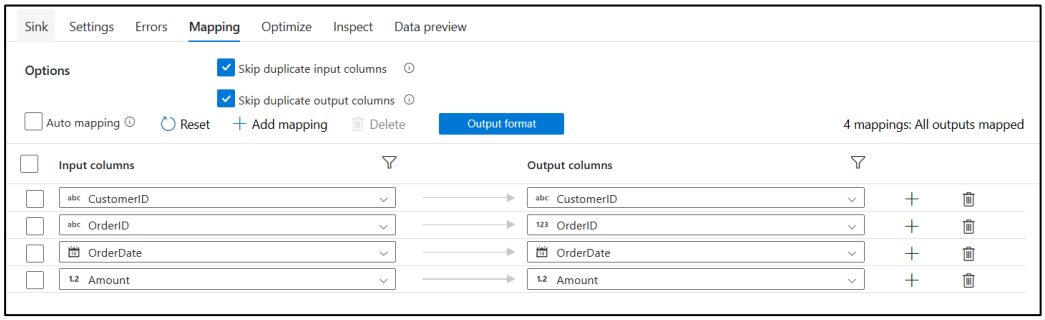


Where,

a. **Source Dataset:** CSV file of Data Lake Storage created earlier.
    a. Go to the **Projection** tab and import projection, provide the parameter values and import schema.

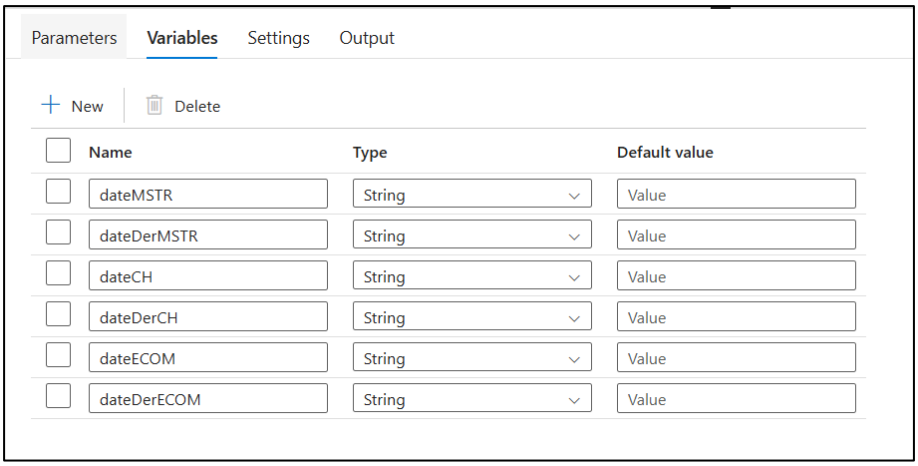
b. **Sink:** ECOM table created in Azure SQL and imported in data factory

a. Go to **Mapping** tab, uncheck **Auto Mapping**, create manual mappings

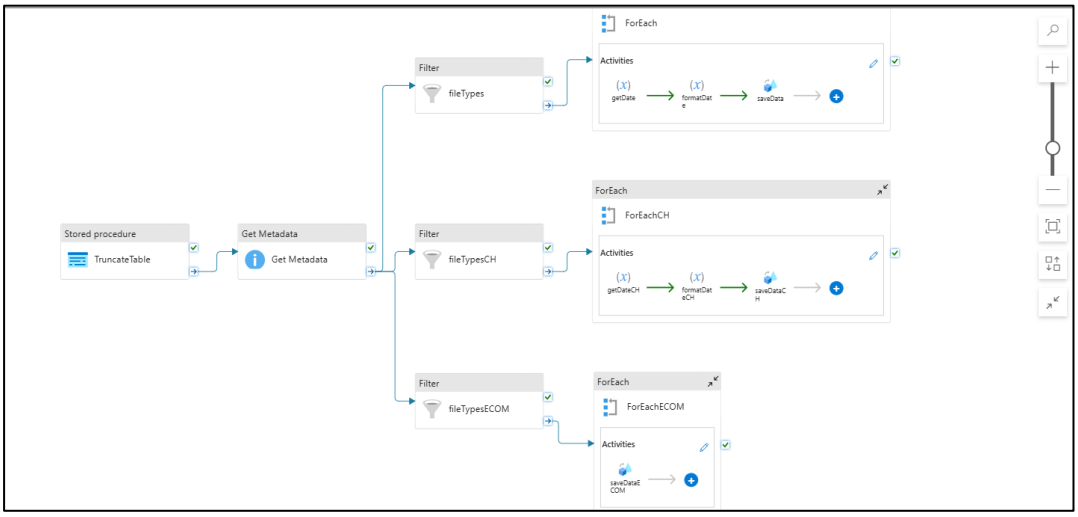


All ForEach Activities have been created.

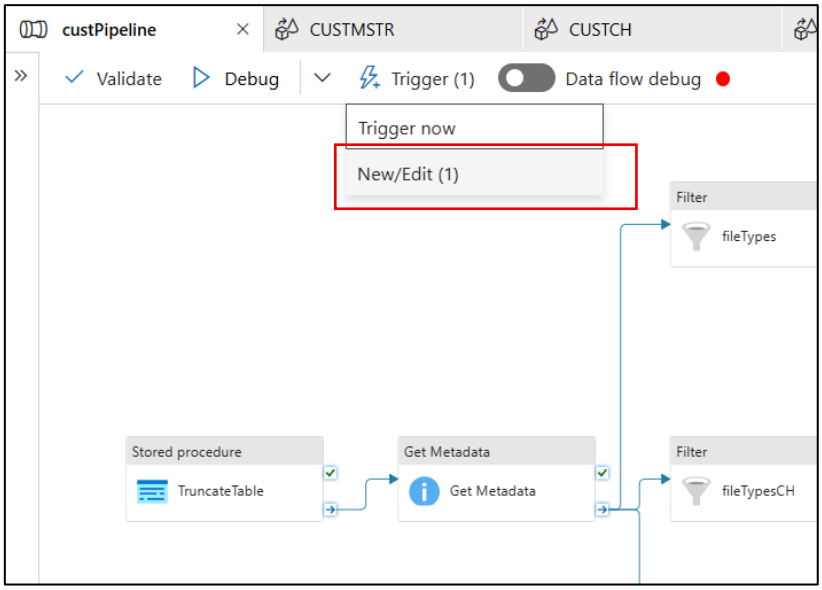
Variables used in entire pipeline:

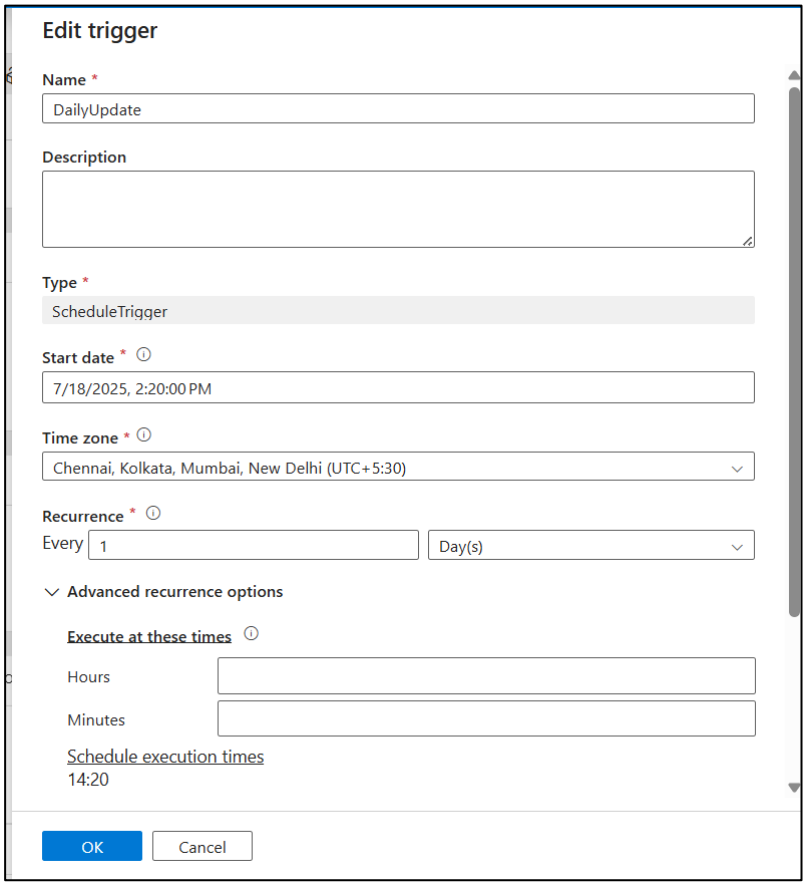


Final pipeline:



Since, the pipeline has to be executed on daily basis, we need to add **trigger.**





Sample CSV Files:

CUST_MSTR_20250718.csv

```
CustomerID,CustomerName,Country
C001,John Doe,USA
C002,Jane Smith,Canada
C003,Alice Brown,UK
```

CUST_MSTR_20250719.csv

```
CustomerID,CustomerName,Country
C004,Rahul,India
C005,Prerna,Russia
C006,Aditi,UK
```

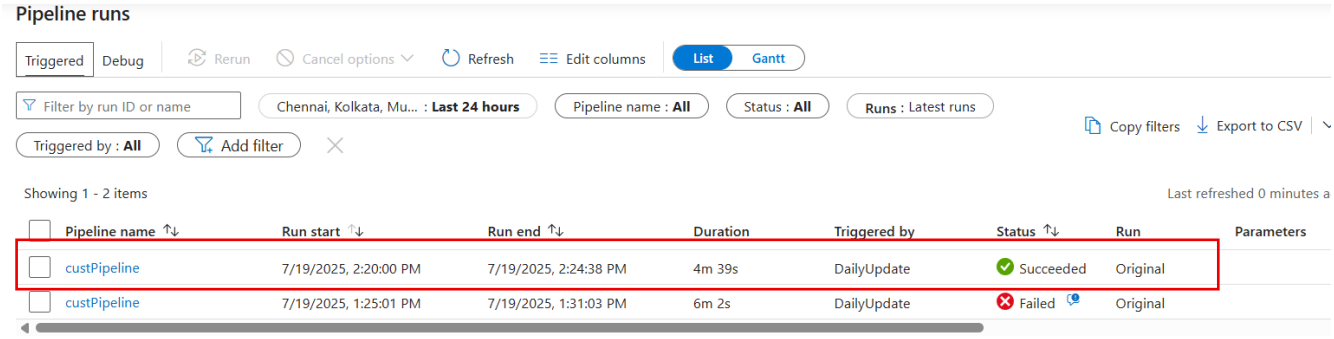
master_child_export-20250718.csv

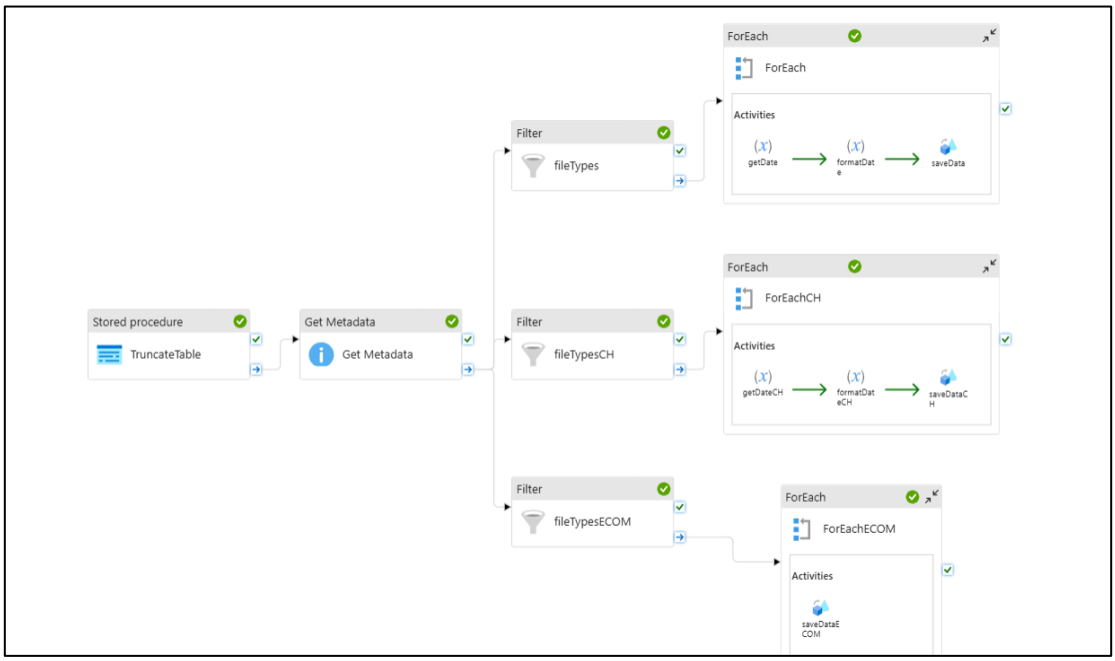
```
CustomerID,CustomerName,Country
C001,John Doe,USA
C002,Jane Smith,Canada
C003,Alice Brown,UK
```

H_ECOM_ORDER.csv

```
OrderID,CustomerID,OrderDate,Amount
O1001,C001,2025-07-18,150.00
O1002,C002,2025-07-18,200.00
O1003,C003,2025-07-18,300.00
```

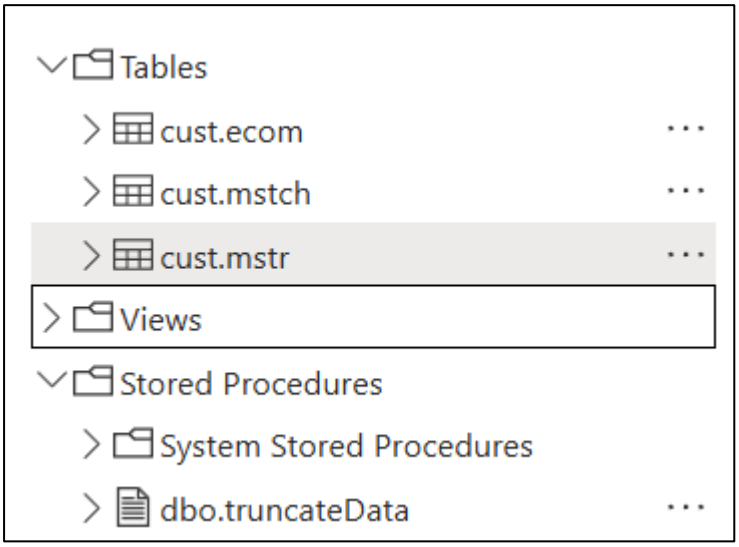
Pipeline Triggered Execution:





TruncateTable	Succeeded	Stored procedu	7/19/2025, 2:20:04 PM	4s	AutoResolveIntegrationRuntime (Central India)
Get Metadata	Succeeded	Get Metadata	7/19/2025, 2:20:09 PM	3s	AutoResolveIntegrationRuntime (Central India)
fileTypesCH	Succeeded	Filter	7/19/2025, 2:20:13 PM	Less than 1s	
fileTypesECOM	Succeeded	Filter	7/19/2025, 2:20:13 PM	Less than 1s	
fileTypes	Succeeded	Filter	7/19/2025, 2:20:13 PM	Less than 1s	
ForEachECOM	Succeeded	ForEach	7/19/2025, 2:20:14 PM	4m 0s	
ForEach	Succeeded	ForEach	7/19/2025, 2:20:14 PM	4m 22s	
ForEachCH	Succeeded	ForEach	7/19/2025, 2:20:14 PM	4m 2s	
getDate	Succeeded	Set variable	7/19/2025, 2:20:15 PM	Less than 1s	
saveDataECOM	Succeeded	Data flow	7/19/2025, 2:20:15 PM	3m 57s	AutoResolveIntegrationRuntime (Central India)
getDateCH	Succeeded	Set variable	7/19/2025, 2:20:15 PM	Less than 1s	
formatDate	Succeeded	Set variable	7/19/2025, 2:20:16 PM	Less than 1s	
formatDateCH	Succeeded	Set variable	7/19/2025, 2:20:16 PM	Less than 1s	
saveDataCH	Succeeded	Data flow	7/19/2025, 2:20:17 PM	3m 56s	AutoResolveIntegrationRuntime (Central India)
saveData	Succeeded	Data flow	7/19/2025, 2:20:17 PM	3m 57s	AutoResolveIntegrationRuntime (Central India)
getDate	Succeeded	Set variable	7/19/2025, 2:24:16 PM	Less than 1s	
formatDate	Succeeded	Set variable	7/19/2025, 2:24:17 PM	Less than 1s	
saveData	Succeeded	Data flow	7/19/2025, 2:24:18 PM	14s	AutoResolveIntegrationRuntime (Central India)

Updated Azure SQL Tables:



1. **CUST_MSTR Table**

Results | Messages

Search to filter items...

CustomerID	CustomerName	Country	dateInserted
C001	John Doe	USA	2025-07-18
C002	Jane Smith	Canada	2025-07-18
C003	Alice Brown	UK	2025-07-18
C004	Rahul	India	2025-07-19
C005	Prerna	Russia	2025-07-19
C006	Aditi	UK	2025-07-19

2. **Master_Child Table**

Results | Messages

Search to filter items...

CustomerID	CustomerName	Country	dateInserted	dateFormated
C001	John Doe	USA	2025-07-18	20250718
C002	Jane Smith	Canada	2025-07-18	20250718
C003	Alice Brown	UK	2025-07-18	20250718

3. ECOM Table

Results Messages

Search to filter items...

CustomerID	OrderID	OrderDate	Amount
C001	O1001	2025-07-18	150
C002	O1002	2025-07-18	200
C003	O1003	2025-07-18	300