

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

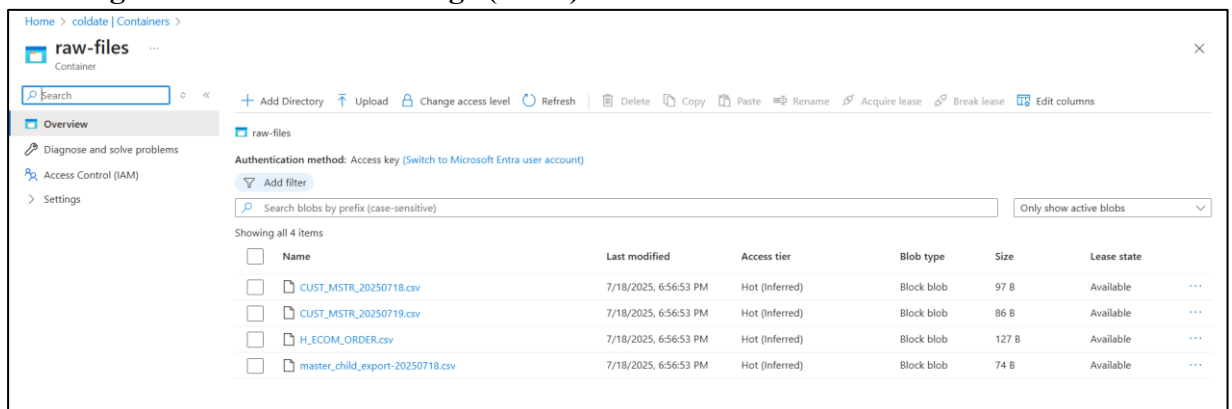
1. For the "CUST_MSTR" starting name of the file You have to create an additional column for a date that will fetch the data value from the filename and put it into an additional column Date format: 2019-11-12 and load it into the "CUST_MSTR" table.
2. For the "master_child_export" starting name of the 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

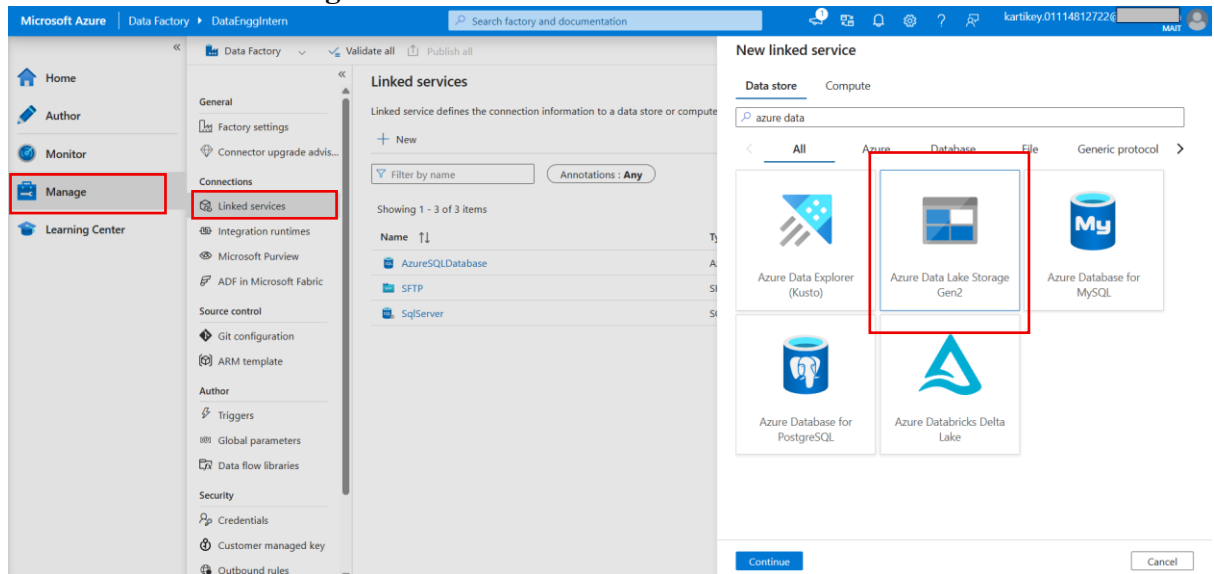
A. Create Linked Services

In Azure Data Factory:

Go to Manage > Linked Services.

Create:

Azure Data Lake Storage Gen2 linked service



New linked service
Azure Data Lake Storage Gen2 [Learn more](#)

Name *
AzureDataLakeStorage

Description

Connect via integration runtime *
☒ AutoResolveIntegrationRuntime

Authentication type
Account key

Account selection method
☒ From Azure subscription ☐ Enter manually

Azure subscription
Azure for Students

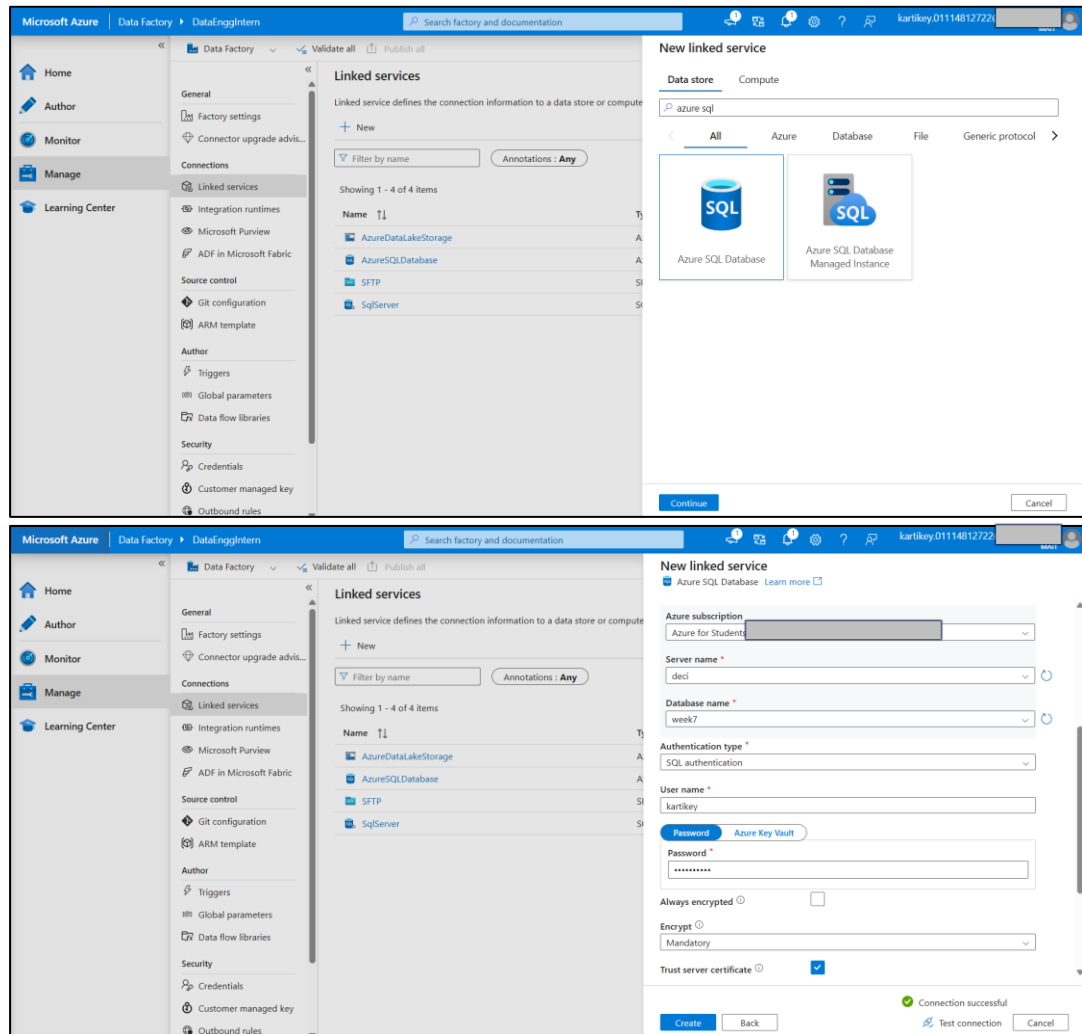
Storage account name *
coldate

Test connection
☒ To linked service ☐ To file path

Buttons: Create, Back, Test connection, Cancel

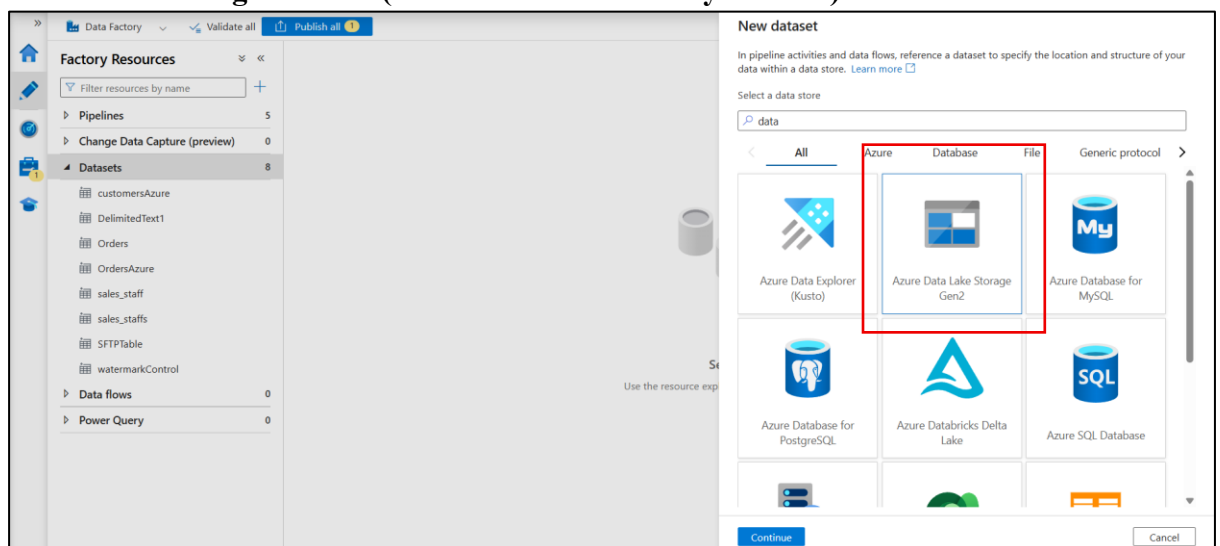
Status: Connection successful

Azure SQL Database linked service



3. Create new datasets

Data Lake Storage Dataset (CSV format and Binary Format)



01 Binary
custData

Connection Parameters

Linked service * AzureDataLakeStorage Test connection Edit + New Learn more

File path raw-files / Directory / File name Browse

Compression type No compression

DelimitedText
CUST_MSTR

Connection Schema Parameters

Linked service * AzureDataLakeStorage Test connection Edit + New Learn more

File path @dataset().FolderPath / Directory / @dataset().FileName Browse

Compression type No compression

Column delimiter Comma (,)

Row delimiter Default (\r,\n, or \r\n)

Encoding Default(UTF-8)

Quote character Double quote (")

Escape character Backslash (\)

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

DelimitedText
CUST_MSTR

Connection Schema Parameters

+ New Delete

Name	Type	Default value
FolderPath	String	Value
FileName	String	Value

This ensures dynamic file handling during pipeline execution.

Azure SQL dataset

Create new datasets for Azure SQL in Data Factory

The screenshot shows the Azure Data Factory portal interface. On the left, the 'Factory Resources' pane is expanded to 'Datasets', where 'custCHF' is selected. The main pane displays the configuration for the 'custCHF' dataset, which is an Azure SQL Database. The 'Connection' tab is active, showing the 'Linked service' as 'AzureSqlDatabaseW7' and the 'Table' as 'cust.mstch'. There are buttons for 'Test connection', 'Edit', 'New', and 'Learn more'. Below the table selection, there is a checkbox for 'Enter manually'.

The screenshot shows the Azure Data Factory portal interface. On the left, the 'Factory Resources' pane is expanded to 'Datasets', where 'custMSTR' is selected. The main pane displays the configuration for the 'custMSTR' dataset, which is an Azure SQL Database. The 'Connection' tab is active, showing the 'Linked service' as 'AzureSqlDatabaseW7' and the 'Table' as 'cust.mstr'. There are buttons for 'Test connection', 'Edit', 'New', and 'Learn more'. Below the table selection, there is a checkbox for 'Enter manually'.

The screenshot shows the Azure Data Factory portal interface. On the left, the 'Factory Resources' pane is expanded to 'Datasets', where 'CUSTECOMF' is selected. The main pane displays the configuration for the 'CUSTECOMF' dataset, which is an Azure SQL Database. The 'Connection' tab is active, showing the 'Linked service' as 'AzureSqlDatabaseW7' and the 'Table' as 'cust.ecom'. There are buttons for 'Test connection', 'Edit', 'New', and 'Learn more'. Below the table selection, there is a checkbox for 'Enter manually'.

For creating SQL tables in Azure SQL:

CREATE SCHEMA cust;

1. cust.mstr

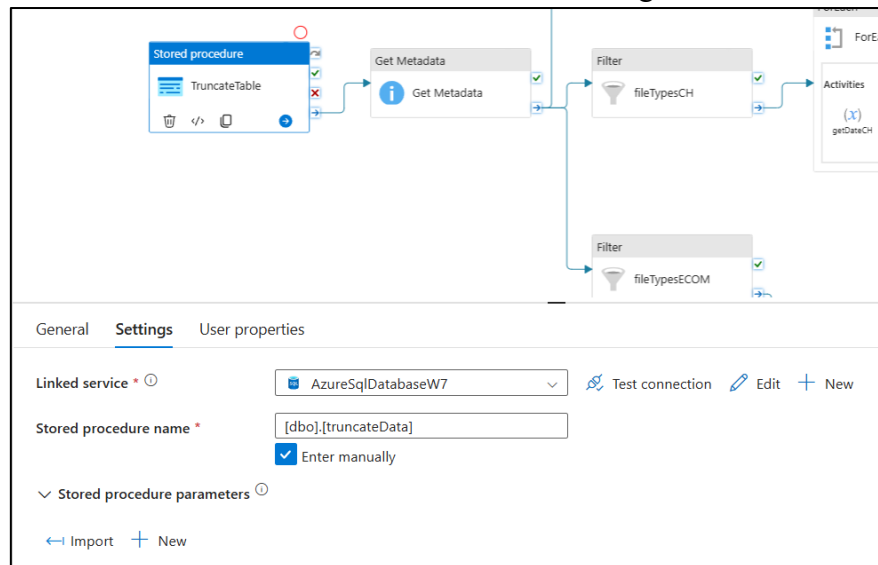
```
CREATE TABLE cust.mstr (  
    CustomerID VARCHAR(10),  
    CustomerName VARCHAR(100),  
    Country VARCHAR(100),  
    dateInserted date
```

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

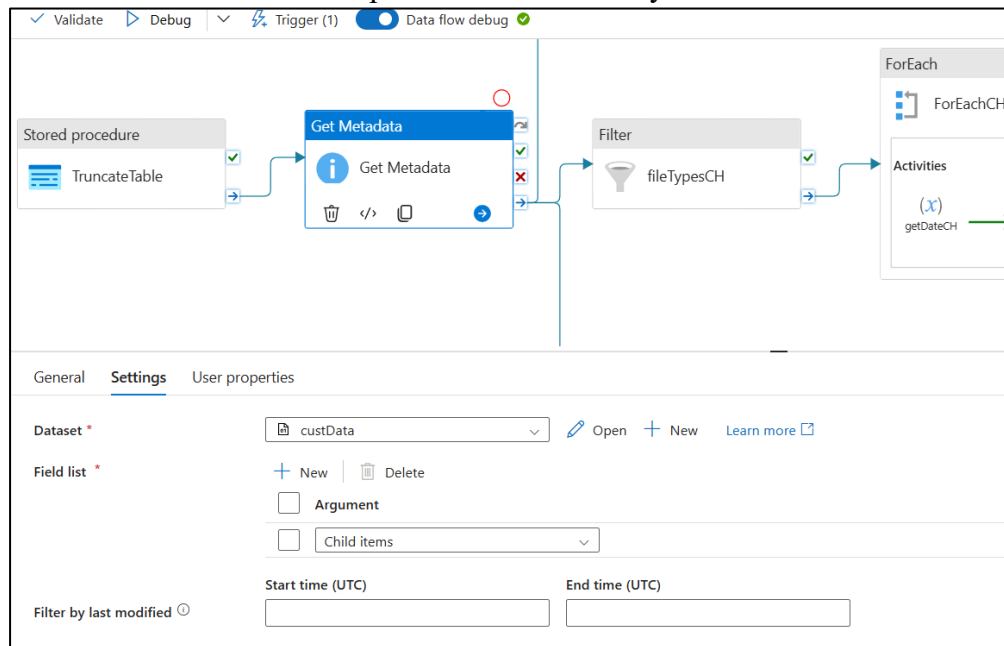


Where, stored procedure:

```
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.

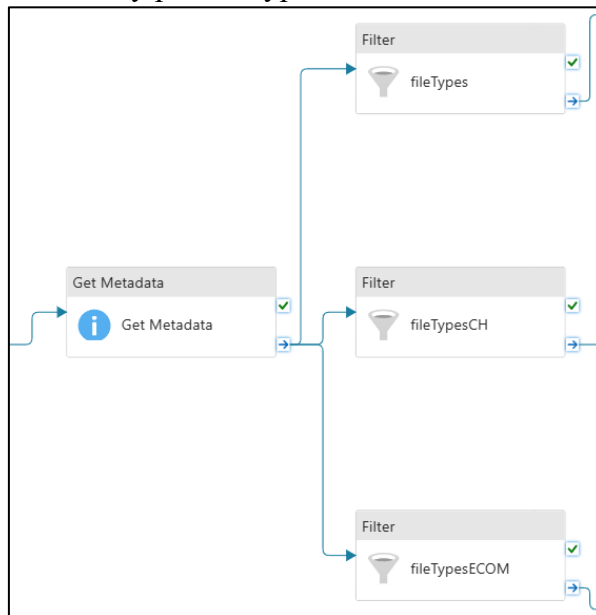


Dataset: custData (binary format dataset created earlier)

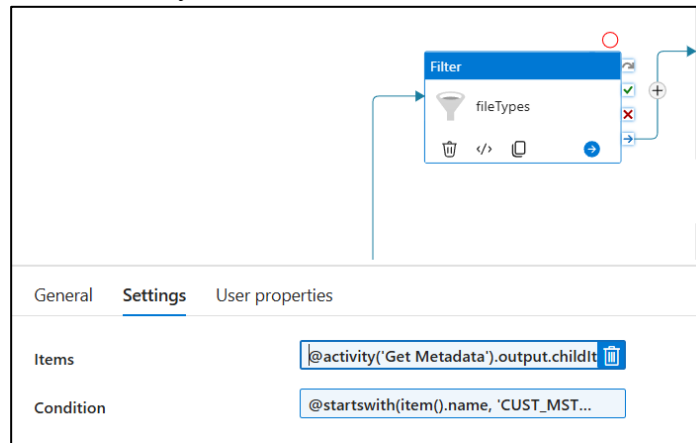
Set Field list: Child Items

c. Creating Filter Activity

One filter activity per file type



In each Filter Activity

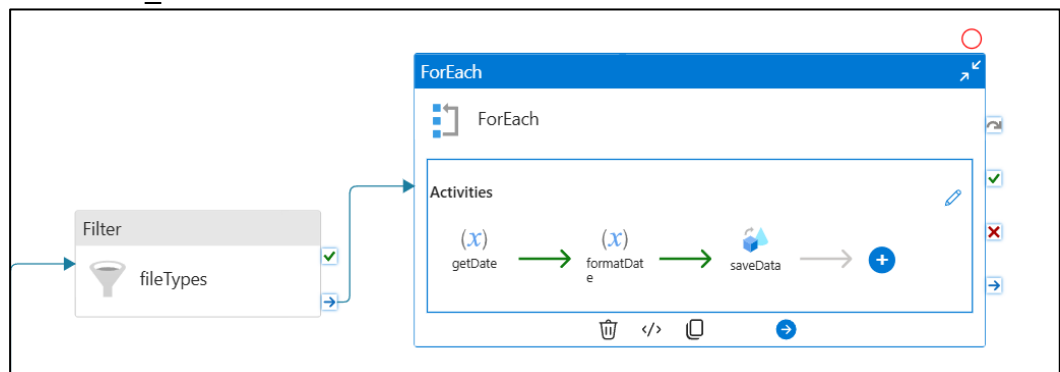


1. FileTypes (for CUST_MASTER)
Items: @activity('Get Metadata').output.childItems
Condition: @startswith(item().name, 'CUST_MSTR_')
2. FileTypesCH (for master_child)
Items: @activity('Get Metadata').output.childItems
Condition: @startswith(item().name, ' master_child_export-')
3. FileTypesECOM (ECOM)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

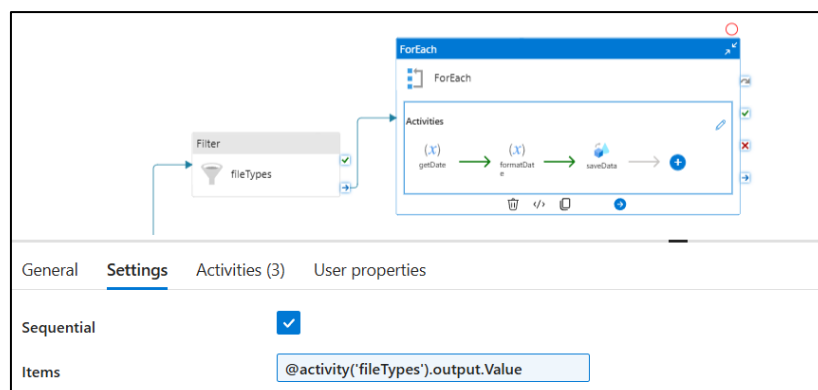
d. ForEach Activity

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

1. For CUST_MASTER

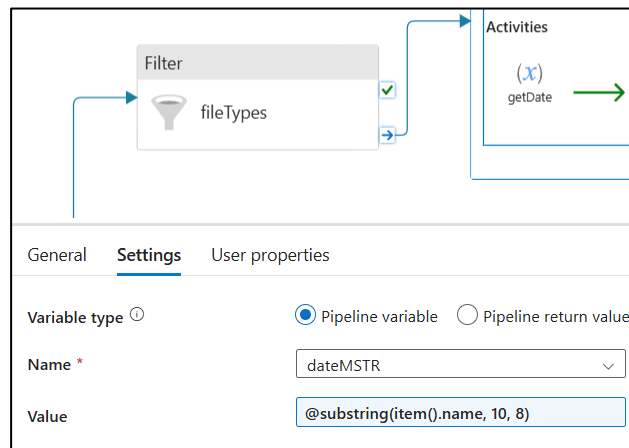


Configure the ForEach activity to take the name of files from Filter activity:



Activities to be inserted:

I. Set Variable



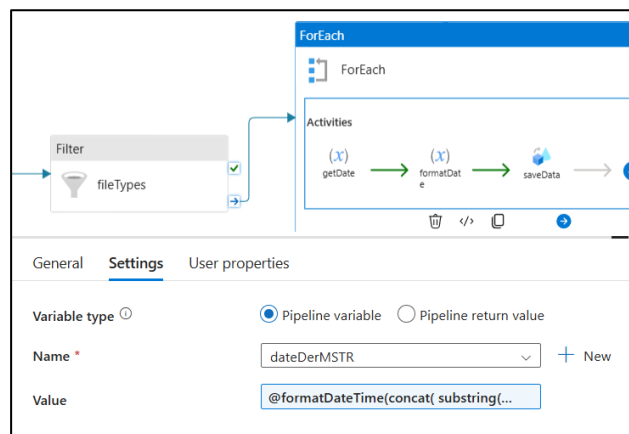
Value: @substring(item().name, 10, 8)

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

Visual workflow diagram showing a Filter activity (fileTypes) connected to a ForEach loop. The ForEach loop contains a sequence of activities: getDate, formatDate, and saveData.

Settings for Data flow * CUSTMSTR:

- source parameters:

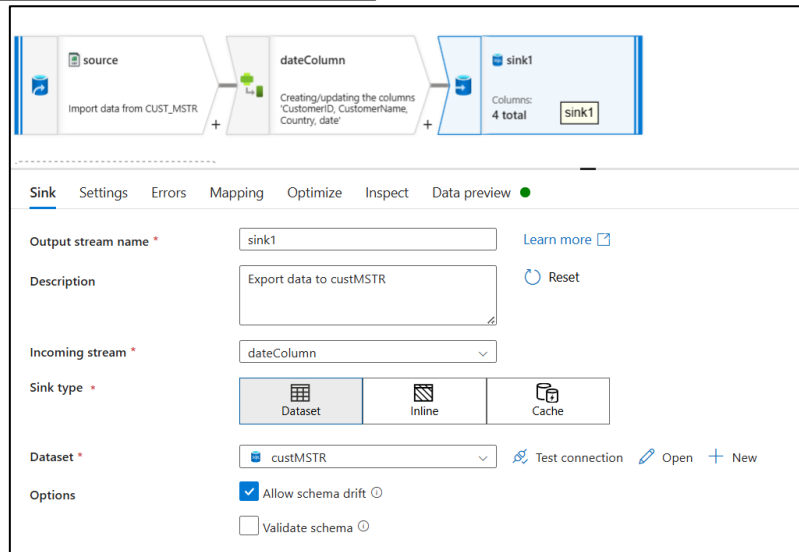
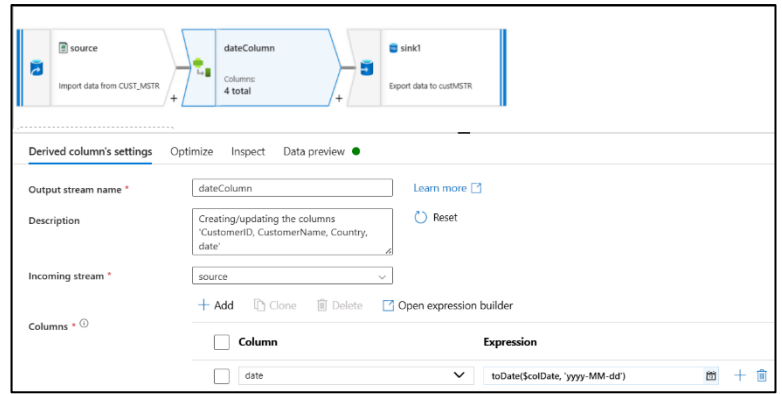
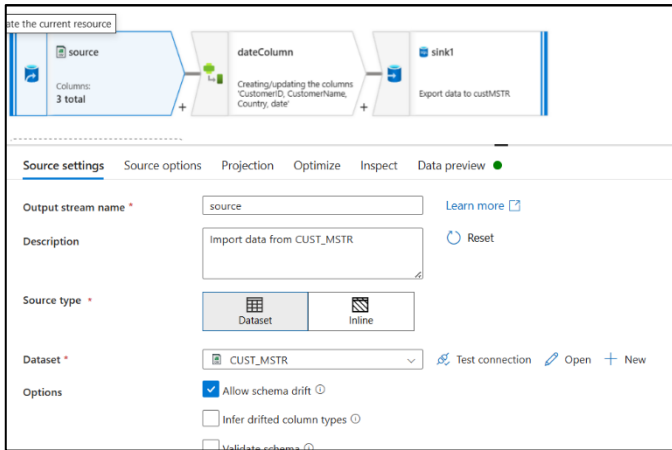
Name	Value	Type
FolderPath	raw-files	object
FileName	@item().name	string

- Run on (Azure IR): AutoResolveIntegrationRuntime
- Compute size: Small

Visual workflow diagram showing a source activity (Import data from CUST_MSTR) connected to a dateColumn activity (Creating/updating the columns 'CustomerID, CustomerName, Country, date') connected to a sink1 activity (Export data to custMSTR).

Parameters for CUSTMSTR:

Name	Type	Default value
colDate	abc string	'2000-01-01'

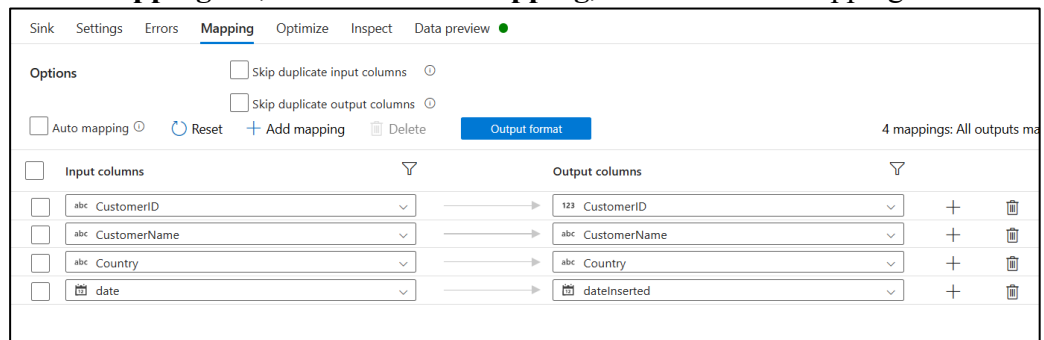


Where,

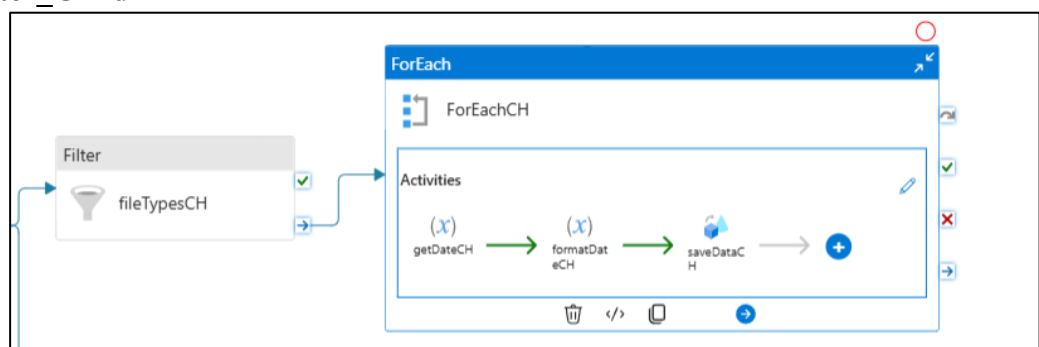
- a. **Source Dataset:** CSV file of Data Lake Storage created earlier.
 - a. Go to the **Projection** tab and import projection, provide the parameter values and import schema.
- b. **DerivedColumn:** For adding the extra date column
 - a. Create **date** column, where in expression, create new parameter: **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:

General Settings Parameters User properties			
Data flow parameters			
Name	Value	Type	Expression
colDate	@variables("dateDerMSTR")	string	<input type="checkbox"/>

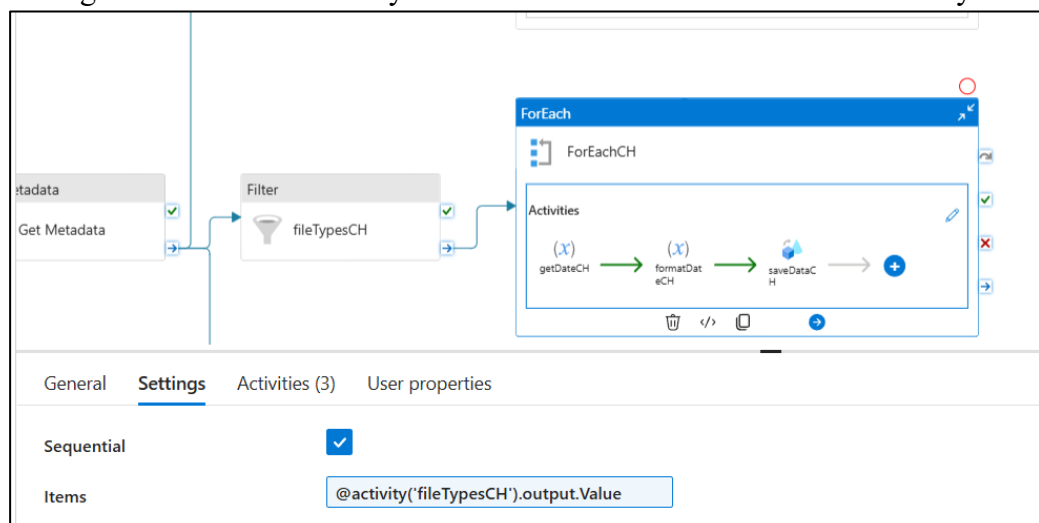
- c. **Sink:** custMSTR table created in Azure SQL and imported in data factory
- a. Go to **Mapping** tab, uncheck **Auto Mapping**, create manual mappings



2. For Master_Child

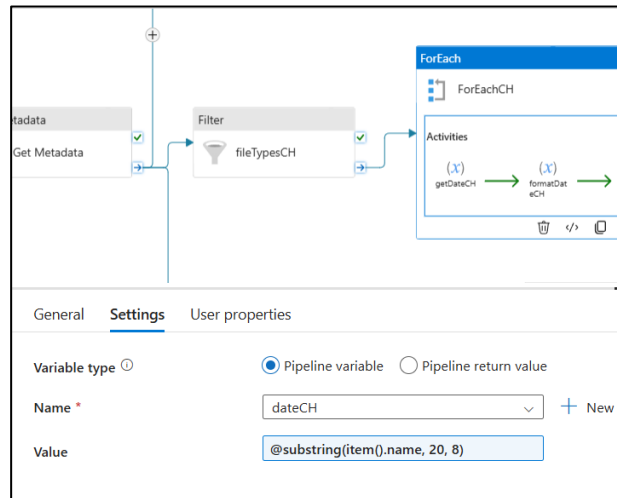


Configure the ForEach activity to take the name of files from Filter activity:



Activities to be inserted:

I. Set Variable



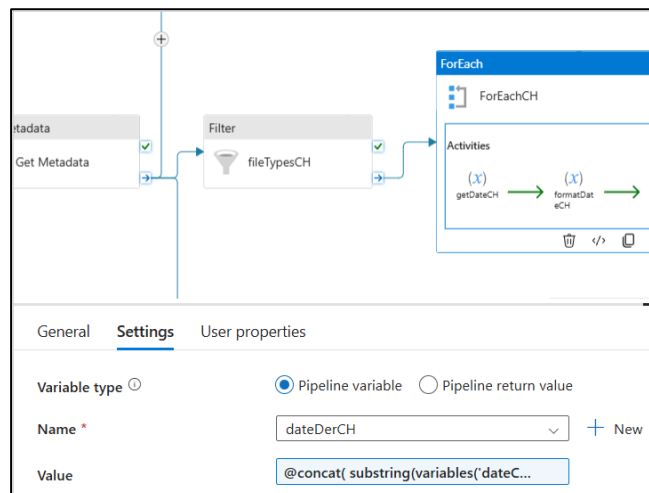
Value: @substring(item().name, 20, 8)

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

The screenshot displays the Azure Data Factory (ADF) interface for creating a new data flow. The top section shows a visual pipeline diagram with the following components:

- Get Metadata** activity.
- Filter** activity with the filter expression `fileTypesCH`.
- ForEach** loop activity containing a **ForEachCH** sub-loop.
- Inside the **ForEachCH** loop, the **Activities** pane shows a sequence: `getDatesCH` → `formatDateCH` → `copyDataCH`.

The bottom section shows the **Settings** tab for the data flow:

- Data flow**: CUSTCH
- source parameters**:

Name	Value	Type
FolderPath	raw-files	object
FileName	@item().name	string
- Run on (Azure IR)**: AutoResolveIntegrationRuntime
- Compute size**: Small

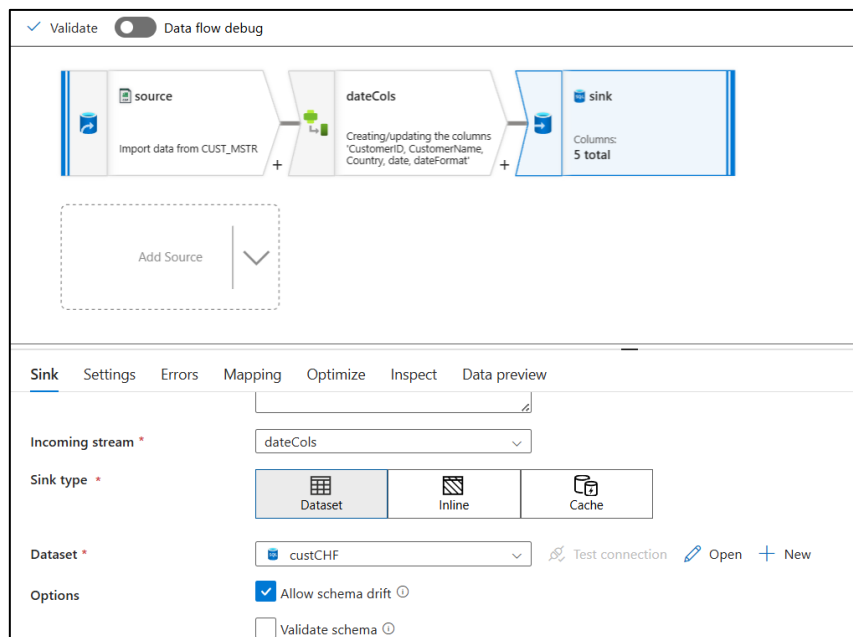
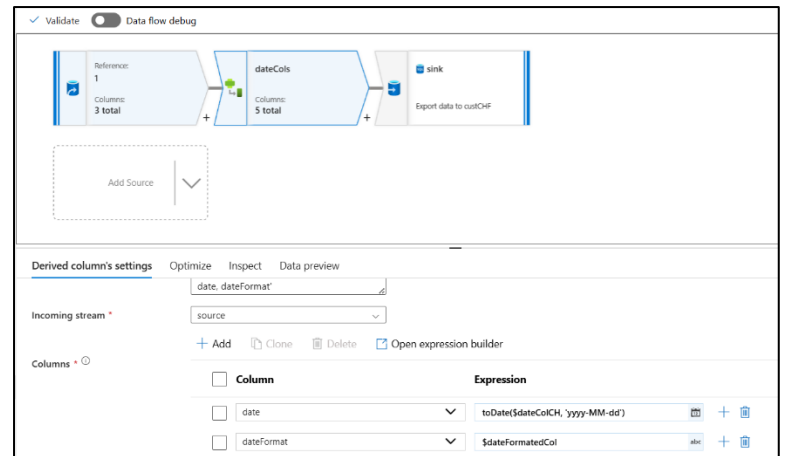
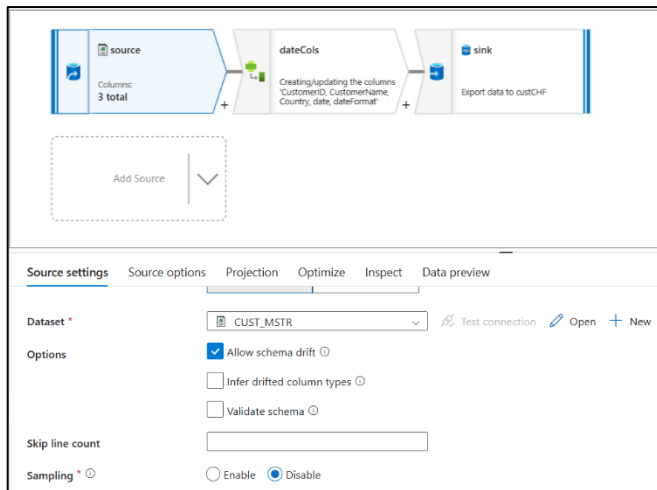
The middle section shows the **Validate** tab with a visual representation of the data flow:

- source**: Import data from CUST_MSTR
- Reference**: 1, Columns: 5 total
- sink**: Export data to custCHF

The bottom section shows the **Parameters** tab:

- Parameters** and **Settings** tabs are visible.
- Parameters** list:

Name	Type	Default value
dateColCH	abc string	Enter expression... ANY
dateFormattedCol	abc string	Enter expression... ANY



Where,

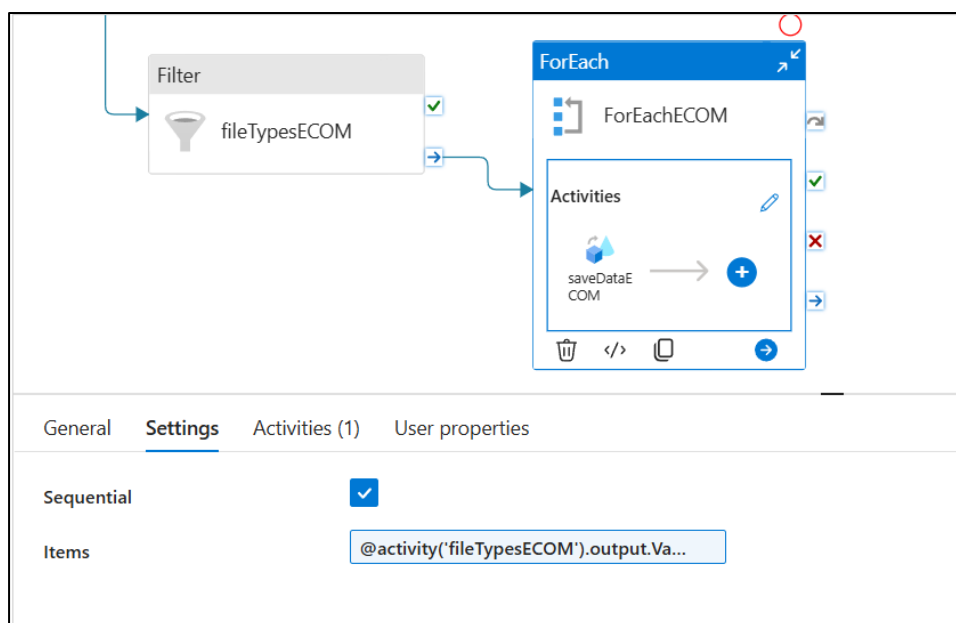
- a. **Source Dataset:** CSV file of Data Lake Storage created earlier.
 - a. Go to the **Projection** tab and import projection, provide the parameter values and import schema.
- b. **DerivedColumn:** For adding the extra date column
 - a. Create **date** column, where in expression, create new parameter: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

General Settings Parameters User properties			
Data flow parameters			
Name	Value	Type	Expression
dateColCH	@variables('dateDerCH')	string	
dateFormattedCol	@variables('dateCH')	string	

- c. **Sink:** custCH table created in Azure SQL and imported in data factory
- a. Go to **Mapping** tab, uncheck **Auto Mapping**, create manual mappings

Sink	Settings	Errors	Mapping	Optimize	Inspect	Data preview
<input type="checkbox"/> Auto mapping	<input checked="" type="checkbox"/> Skip duplicate output columns	Reset	+ Add mapping	Delete	Output format	5 mappings: All outputs mapped
Input columns		Output columns				
<input type="checkbox"/> abc CustomerID	→	abc CustomerID	+	🗑		
<input type="checkbox"/> abc CustomerName	→	abc CustomerName	+	🗑		
<input type="checkbox"/> abc Country	→	abc Country	+	🗑		
<input type="checkbox"/> date	→	dateInserted	+	🗑		
<input type="checkbox"/> abc dateFormat	→	abc dateFormat	+	🗑		

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.

GeneralSettingsParametersUser properties

Data flow *

CUSTECOM

Open

New

source parameters ⓘ

Name	Value	Type
FolderPath	raw-files	object
FileName	@item().name	string

Run on (Azure IR) * ⓘ

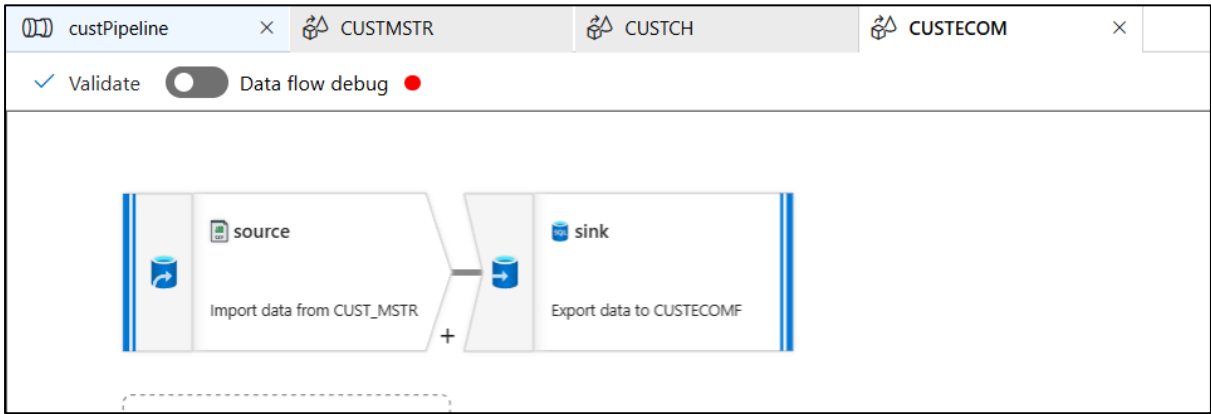
AutoResolveIntegrationRuntime

Compute size * ⓘ

Small

Advanced

Create a new data flow



Validate

Data flow debug

source

Columns: 4 total

+

sink

Export data to CUSTECOMF

Source settings

Source options

Projection

Optimize

Inspect

Data preview

Description

Import data from CUST_MSTR

Reset

Source type *

Dataset

Inline

Dataset *

CUST_MSTR

Test connection

Open

Options

Allow schema drift ⓘ

Infer drifted column types ⓘ

Validate schema ⓘ

Skip line count

Sampling * ⓘ

Enable

Disable

Sink

Settings

Errors

Mapping

Optimize

Inspect

Data preview

Output stream name *

sink

Learn more ⓘ

Description

Export data to CUSTECOMF

Reset

Incoming stream *

source

Sink type *

Dataset

Inline

Cache

Dataset *

CUSTECOMF

Test connection

Open

New

Options

Allow schema drift ⓘ

Validate schema ⓘ

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

Sink Settings Errors **Mapping** Optimize Inspect Data preview

Options ☒ Skip duplicate input columns ☒ Skip duplicate output columns

☐ Auto mapping 4 mappings: All outputs mapped

Input columns	Output columns
<input type="checkbox"/> abc CustomerID	<input type="checkbox"/> abc CustomerID
<input type="checkbox"/> abc OrderID	<input type="checkbox"/> 123 OrderID
<input type="checkbox"/> OrderDate	<input type="checkbox"/> OrderDate
<input type="checkbox"/> 1.2 Amount	<input type="checkbox"/> 1.2 Amount

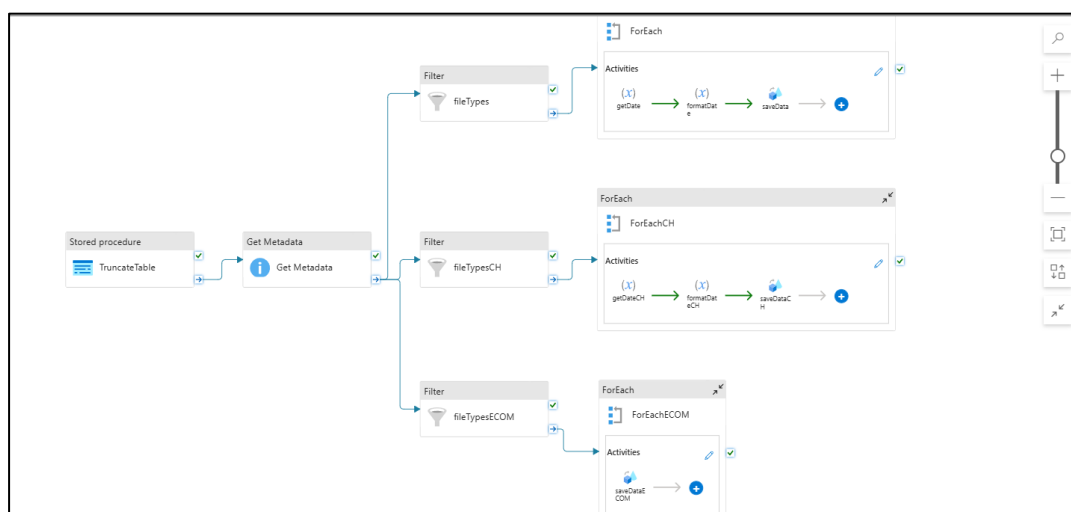
All ForEach Activities have been created.

Variables used in entire pipeline:

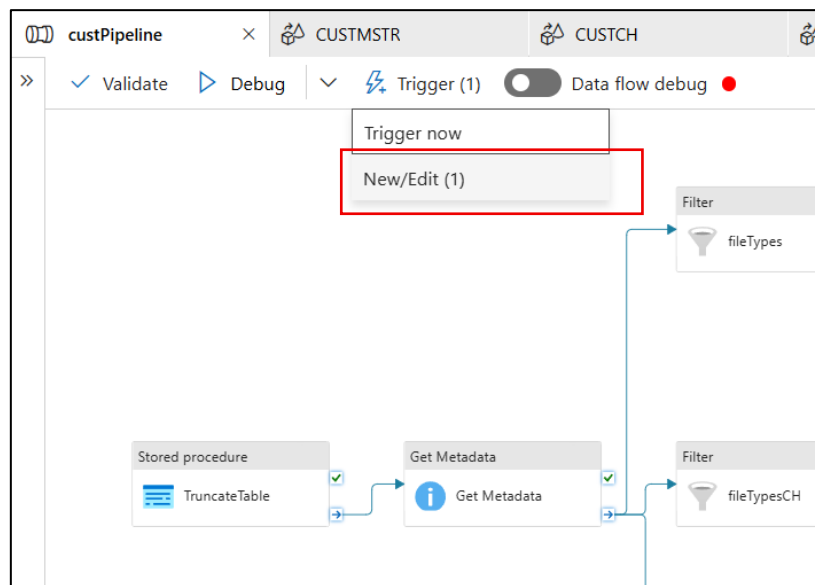
Parameters **Variables** Settings Output

Name	Type	Default value
dateMSTR	String	Value
dateDerMSTR	String	Value
dateCH	String	Value
dateDerCH	String	Value
dateECOM	String	Value
dateDerECOM	String	Value

Final pipeline:



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



Edit trigger

Name *
DailyUpdate

Description

Type *
ScheduleTrigger

Start date * ⓘ
7/18/2025, 2:20:00 PM

Time zone * ⓘ
Chennai, Kolkata, Mumbai, New Delhi (UTC+5:30)

Recurrence * ⓘ
Every Day(s)

✓ **Advanced recurrence options**

Execute at these times ⓘ

Hours
Minutes

Schedule execution times
14:20

Sample CSV Files:

CUST_MSTR_20250718.csv X

CUST_MSTR_20250718.csv

1

CustomerID, CustomerName, Country

2

C001, John Doe, USA

3

C002, Jane Smith, Canada

4

C003, Alice Brown, UK

5

CUST_MSTR_20250719.csv X

CUST_MSTR_20250719.csv

1

CustomerID, CustomerName, Country

2

C004, Rahul, India

3

C005, Prerna, Russia

4

C006, Aditi, UK

5

master_child_export-20250718.csv X

CUST_MSTR_20250718.csv

master_child_export-20250718.csv

1

CustomerID, CustomerName, Country

2

C001, John Doe, USA

3

C002, Jane Smith, Canada

4

C003, Alice Brown, UK

5

H_ECOM_ORDER.csv X

master_child_export-20250718.csv

H_ECOM_ORDER.csv

1

OrderID, CustomerID, OrderDate, Amount

2

01001, C001, 2025-07-18, 150.00

3

01002, C002, 2025-07-18, 200.00

4

01003, C003, 2025-07-18, 300.00

5

Pipeline Triggered Execution:

Pipeline runs

Triggered

Debug

Rerun

Cancel options

Refresh

Edit columns

List

Gantt

Filter by run ID or name

Chennai, Kolkata, Mu... : Last 24 hours

Pipeline name : All

Status : All

Runs : Latest runs

Copy filters

Export to CSV

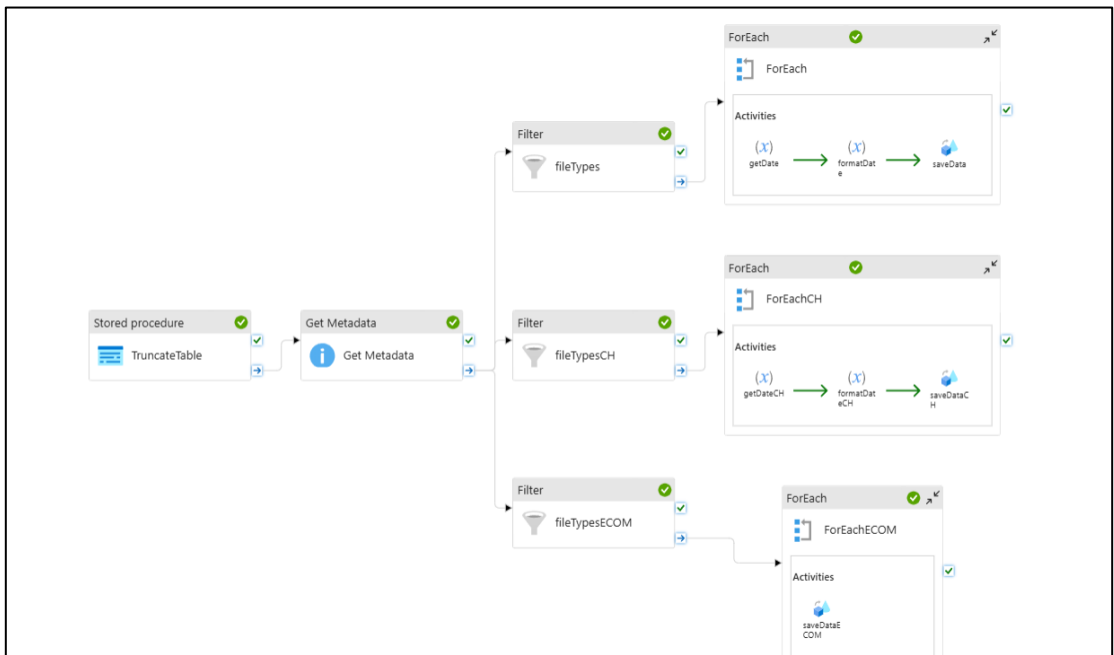
Triggered by : All

Add filter

Showing 1 - 2 items

Last refreshed 0 minutes a

<input type="checkbox"/>	Pipeline name ↑↓	Run start ↑↓	Run end ↑↓	Duration	Triggered by	Status ↑↓	Run	Parameters
<input type="checkbox"/>	custPipeline	7/19/2025, 2:20:00 PM	7/19/2025, 2:24:38 PM	4m 39s	DailyUpdate	Succeeded	Original	
<input type="checkbox"/>	custPipeline	7/19/2025, 1:25:01 PM	7/19/2025, 1:31:03 PM	6m 2s	DailyUpdate	Failed	Original	



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:

▼

Tables

>

cust.ecom

...

>

cust.mstch

...

>

cust.mstr

...

>

Views

▼

Stored Procedures

>

System Stored Procedures

>

dbo.truncateData

...

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	dateFormatted
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