

Solutions:

Solution: Assignment 8: ADF Data Flow 5 Question

00:45:52

Assignment 8: ADF Data Flow 5 Question

1. Create a pipeline to copy the customer data from csv file to SQL where the customer id is an even number.(Attaching the Customer File)
File: SalesLT.Customer-xRyuaEu2p7-WdCbzuWlIo-80V7BfCQtP
2. Create a pipeline to join the two files (Customer, Customer Address) based on customer id and save the result as a JSON file.
3. Create a pipeline to read the Customer table data from SQL and CustomerAddress data from CSV, join both of them,
and then save the result where customer id > 1000 & Customer id < 2000 in ascending order as a Parquet file.
4. create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.
Ensure that product shouldn't be of blue in color and save the result as CSV file inside ProductResult folder.
5. create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.
Ensure that product shouldn't be of blue in color and save the result as a SINGLE CSV file inside ProductSingleResult folder.

Solutions:

Solutions:

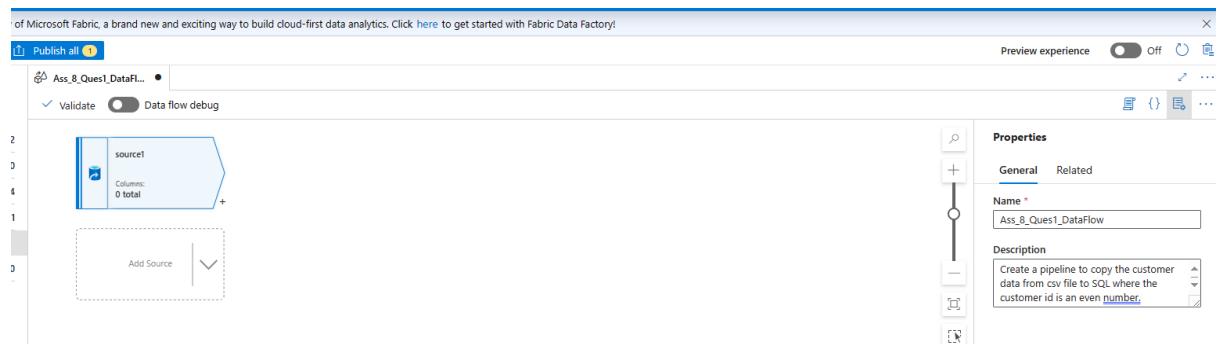
Assignment 8: ADF Data Flow 5 Question

1. Create a pipeline to copy the customer data from csv file to SQL where the customer id is an even number.(Attaching the Customer File)
File: SalesLT.Customer-xRyuaEu2p7-WdCbzuWlIo-80V7BfCQtP

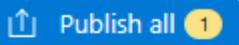
Solution :

Problem is input file is in CSV format,(but not in SQL Table), Therefore I have to use data flow concept using “Filter” or so

The screenshot shows the Microsoft Azure Data Factory interface. At the top, there's a navigation bar with a back arrow, a refresh icon, a URL (<https://adf.azure.com/en/authoring/dataflow/dataflow1?factory=%2Fsubscriptions%2F10d4af9c-1612-4d63-bed9->), and a search bar. Below the navigation is a message about the public preview of Microsoft Fabric. The main area is titled "Factory Resources" and contains a sidebar with icons for Home, Pipelines, Change Data Capture (preview), Datasets, Data flows, Power Query, and Copy Data tool. Under "Data flows", "dataflow1" is selected. A modal window is open, showing a list of resources: Pipeline, Change Data Capture (preview), Dataset, Data flow, Power Query, and Flowlet. The "Data flow" item is highlighted with a dashed border.



view of Microsoft Fabric, a brand new and exciting way to build cloud-first data

all  Publish all 1

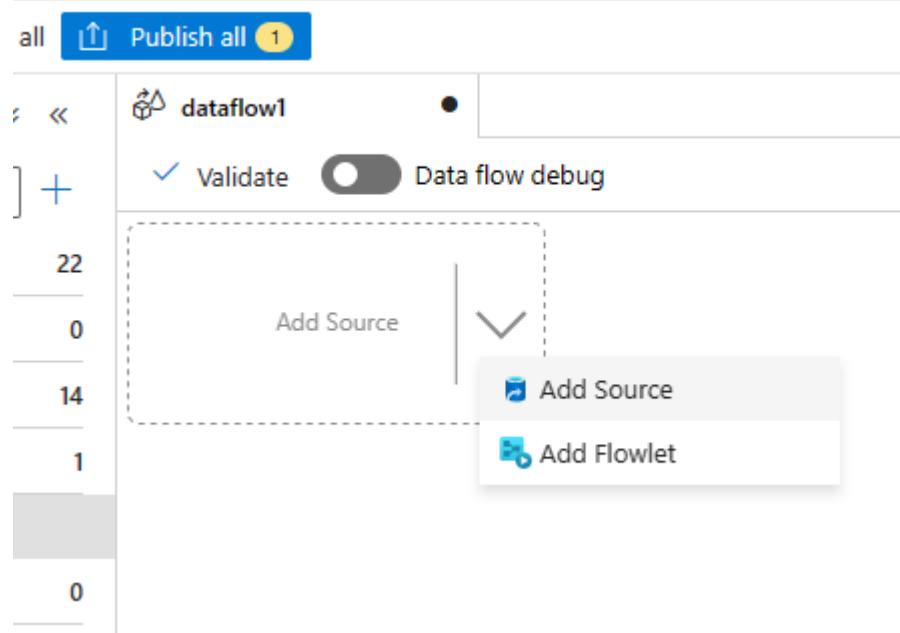
dataflow1

Validate Data flow debug

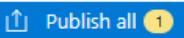
Add Source

Add Source

Add Flowlet



Lic preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Flow.

Validate all  Publish all 1

dataflow1

Validate Data flow debug

source1
Columns: 0 total

Add Source

Source settings Source options Projection Optimize Inspect Data preview

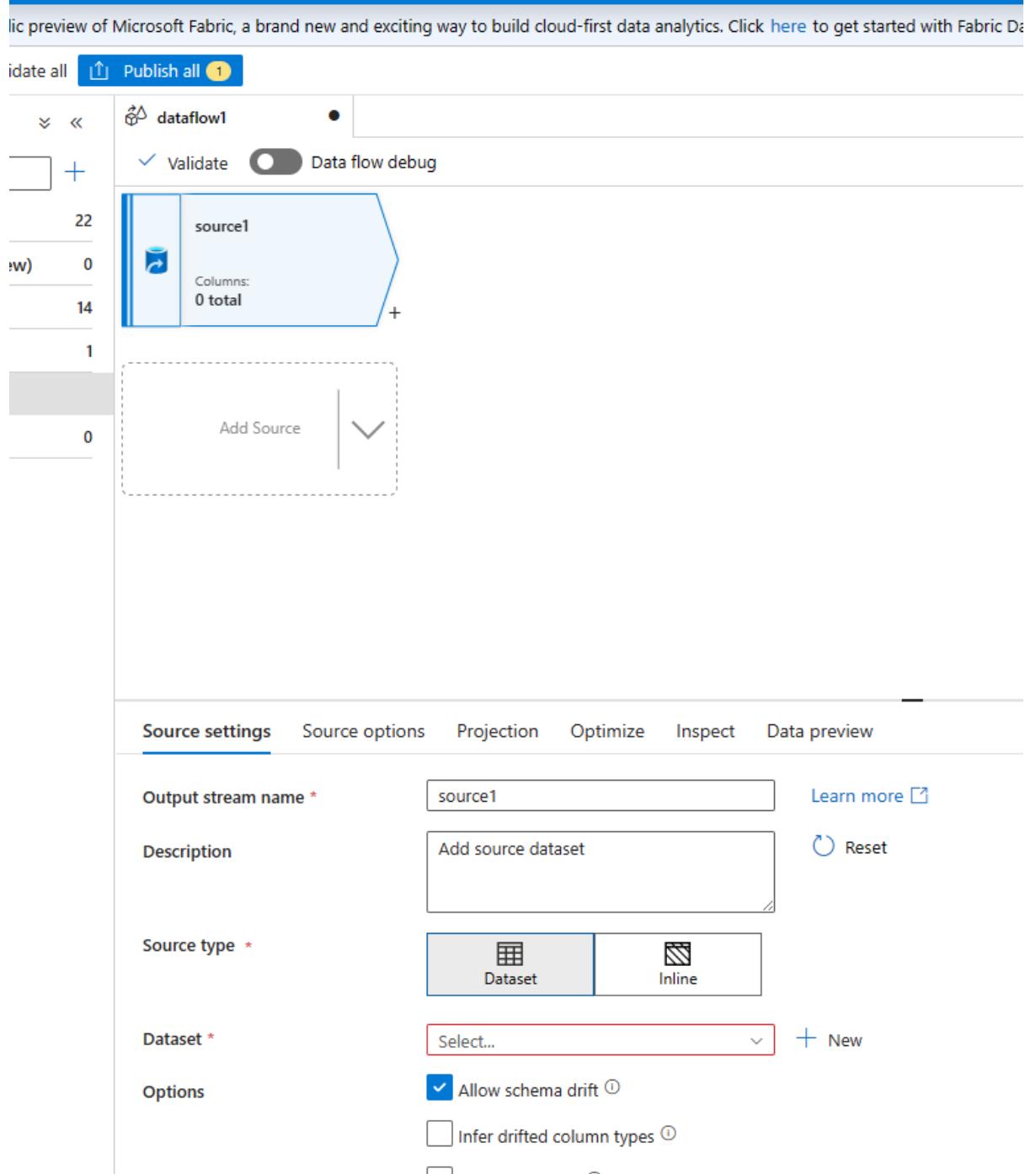
Output stream name * source1 Learn more 

Description Add source dataset 

Source type *  Dataset  Inline

Dataset * Select... 

Options Allow schema drift 
 Infer drifted column types 



Since CSV is input file, It is not available , we need to upload to ADLS storage container account first .

gopinathma09@outlook.com
DEFAULT DIRECTORY (ANIL24)

Add Directory

Name *

Objects

Size

Save Give feedback

https://portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2F...%2FresourceGroups%2Fgopi01012024adls1%2Fcontainers%2Flanding

Microsoft Azure Upgrade Search resources, services, and docs

Home > Storage accounts > gopi01012024adls1 | Containers >

landing Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: landing

Search blobs by prefix (case-sensitive)

Name	Modified
<input type="checkbox"/> Ass_7_4_Table_HighWaterMarkFile	
<input type="checkbox"/> Ass_8_Ques_1_ADF_Dataflow	
<input type="checkbox"/> CSV_Pipe	

Upgrade

Search resources, services, and docs (G+/)

pi01012024adls1 | Containers >

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: landing / Ass_8_Ques_1_ADF_Dataflow

Upload blob

2 file(s) selected: SalesLT.CustomerAddress-nnCI751651 (1)-0A4ckSV5g...
Drag and drop files here or Browse for files

Overwrite if files already exist

Advanced

Upload

Give feedback

Name	Modified	Access tier	Archive status
<input type="checkbox"/> SalesLT.CustomerAddress-nnCI751651 (1)-0A4ckSV5g...			

Microsoft Azure

Search resources, services, and docs (G+/)

Storage accounts > gopi01012024adls1 | Containers >

landing

Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: landing / Ass_8_Ques_1_ADF_Dataflow

Overview

Diagnose and solve problems

Access Control (IAM)

Search blobs by prefix (case-sensitive)

Name	Modified	Access tier	Archive status
<input checked="" type="checkbox"/> SalesLT.Customer-xRyuEu2p7-WdCbzuWllo-80V7BfCQtP.txt	2/27/2024, 4:49:22 PM	Hot (Inferred)	
<input checked="" type="checkbox"/> SalesLT.CustomerAddress-nnCI751651 (1)-0A4ckSV5gM-jBPnfa8HWQ.txt	2/27/2024, 4:48:53 PM	Hot (Inferred)	
<input checked="" type="checkbox"/> SalesLT.Product-760RKVH4kL-x8c0uQ9qBG-pu11LFOEY0.txt	2/27/2024, 4:48:54 PM	Hot (Inferred)	

- Shared access tokens
- Manage ACL
- Access policy
- Properties
- Metadata

« [Upload](#) [Add Directory](#) [Refresh](#) | [Rename](#) [Delete](#) [Change](#)

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: [landing / Ass_8_input_files](#)

Search blobs by prefix (case-sensitive)

Name
<input type="checkbox"/> SalesLT.Customer-xRyuaEu2p7-WdCbzuWllo-80V7BfCQtP.txt
<input type="checkbox"/> SalesLT.CustomerAddress-nnCl751651 (1)-0A4ckSV5gM-jBPnfa8HWQ.txt
<input type="checkbox"/> SalesLT.Product-760RKVH4kL-x8c0uQ9qBG-pul1LFOEyO.txt

Now from DataFlow – load this Dataset of CSV file from ADLS container

The screenshot shows the Azure Data Flow pipeline editor. On the left, there's a visual representation of the pipeline with a source activity labeled "source1" and a sink activity below it. The source is connected to a dashed box labeled "Add Source". On the right, there's a detailed configuration pane for the source activity.

Source settings tab (selected):

- Description: Add source dataset
- Source type: Dataset (selected)
- Dataset: Select... (dropdown menu)
- Options:
 - Allow schema drift: checked
 - Infer drifted column types: unchecked
 - Preserve headers: unchecked

Source options tab (disabled):

- Projection: None
- Optimize: None
- Inspect: None

Projection tab (disabled):

- None

Optimize tab (disabled):

- None

Inspect tab (disabled):

- None

Data preview tab (disabled):

- None

On the right side of the editor, there's a grid of icons representing various data stores:

All	Azure	Database	File	Generic protocol	NoSQL	Services and apps

gopinathmca09@DESKTOP-1HJL9D9

Select format

Choose the format type of your data

 Avro	 DelimitedText	 Excel
 JSON	 ORC	 Parquet
 XML	 Binary	

[Continue](#) [Back](#)

DEFAULT DIRECTORY

Set properties

Name
Ass_8_Q1_Customer_CSV

Linked service *
ADLS_Azure_DataLakeStorage1_Linked_service

File path
landing / Ass_8_input_files / SalesLT.Customer-xRyu...

First row as header

Import schema
 From connection/store From sample file None

> Advanced

I cannot “Data Preview” my data , until I enable the “Data flow debug” button

Source settings Source options Projection Optimize Inspect Data preview

⚠ Please turn on debug mode and wait until the cluster is ready to preview data...

↑ Publish all 2

Ass_8_Ques1_DataFl... ●

✓ Validate Data flow debug ↵

source1
Columns: 15 total

Add Source ✓

Source settings Source options Projection Optimize Inspect Data preview

Description Import data from Ass_8_Q1_Customer_CSV ⏪ Reset

This screenshot shows the Microsoft Fabric Data Flow interface. At the top, there's a 'Publish all' button with a count of 2. Below it is a list item for 'Ass_8_Ques1_DataFl...', which includes a 'Validate' checkbox and a 'Data flow debug' toggle switch, both of which are circled in red. A red arrow points to the 'Data flow debug' switch. The main workspace shows a single source component named 'source1' with 15 columns. Below the workspace is a dashed box labeled 'Add Source' with a checkmark icon. At the bottom, there's a navigation bar with tabs: 'Source settings' (which is active and underlined), 'Source options', 'Projection', 'Optimize', 'Inspect', and 'Data preview'. The 'Data preview' tab is also circled in red. Underneath this bar, there's a 'Description' field containing the text 'Import data from Ass_8_Q1_Customer_CSV' and a 'Reset' button.

↑ Publish all 2

Ass_8_Ques1_DataFl... ●

✓ Validate Data flow debug ⏪ Debug Settings

22

This screenshot shows the Microsoft Fabric Data Flow interface, similar to the one above but with some differences. It features a 'Publish all' button with a count of 2. Below it is a list item for 'Ass_8_Ques1_DataFl...', with a 'Validate' checkbox and a 'Data flow debug' toggle switch, both of which are circled in red. A red arrow points to the 'Data flow debug' switch. The main workspace shows a single source component named 'source1' with 15 columns. Below the workspace is a dashed box labeled 'Add Source' with a checkmark icon. At the bottom, there's a navigation bar with tabs: 'Source settings' (active), 'Source options', 'Projection', 'Optimize', 'Inspect', and 'Data preview'. The 'Data preview' tab is also circled in red. Underneath this bar, there's a 'Description' field containing the text 'Import data from Ass_8_Q1_Customer_CSV' and a 'Reset' button. Additionally, there's a 'Debug Settings' link next to the 'Data flow debug' switch. The bottom of the screen shows a progress bar with the number '22'.

DEFAULT DIRECTORY

Turn on data flow debug

Integration runtime

AutoResolveIntegrationRuntime

AutoResolveIntegrationRuntime

Region	AutoResolve
Compute size	Small

Debug time to live ⓘ

1 hour

[Search] [New] [Delete]

Turn on data flow debug

Integration runtime

AutoResolveIntegrationRuntime

AutoResolveIntegrationRuntime

Region	AutoResolve
Compute size	Small

Debug time to live ⓘ

1 hour

Until 2 hrs

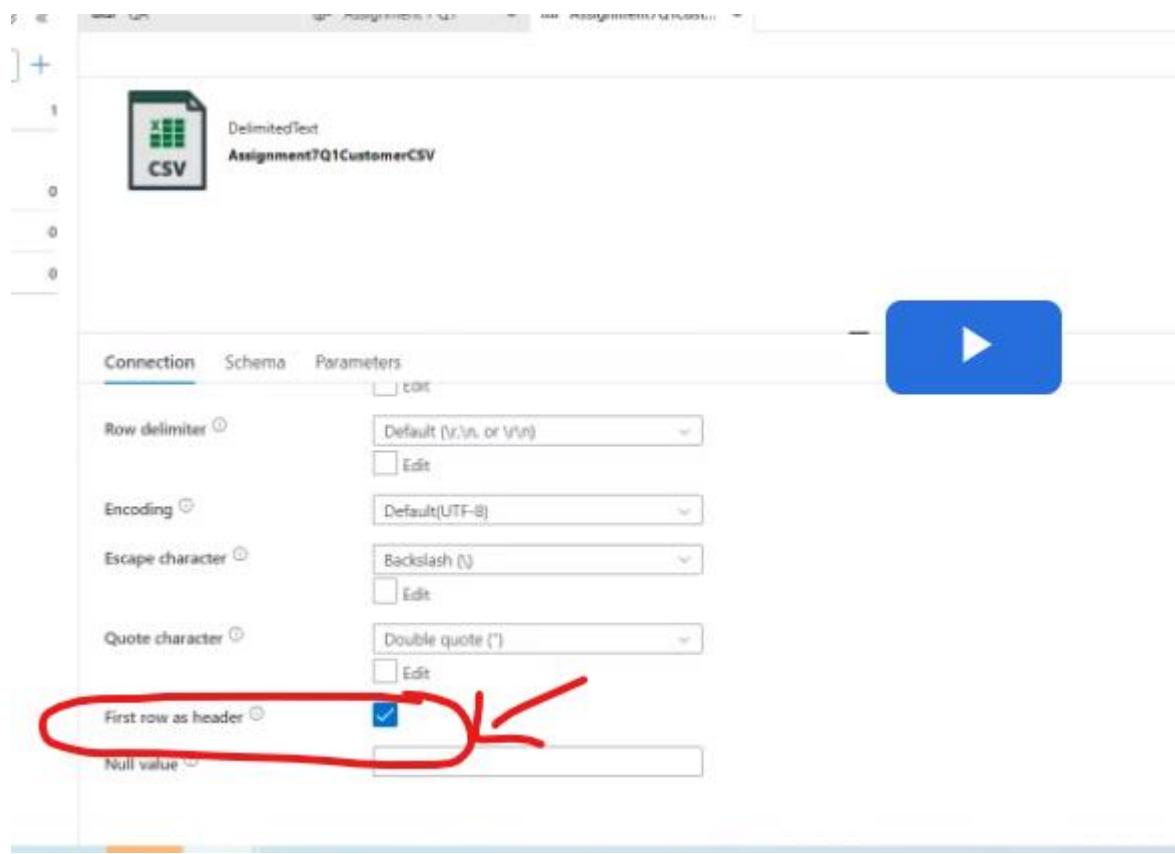
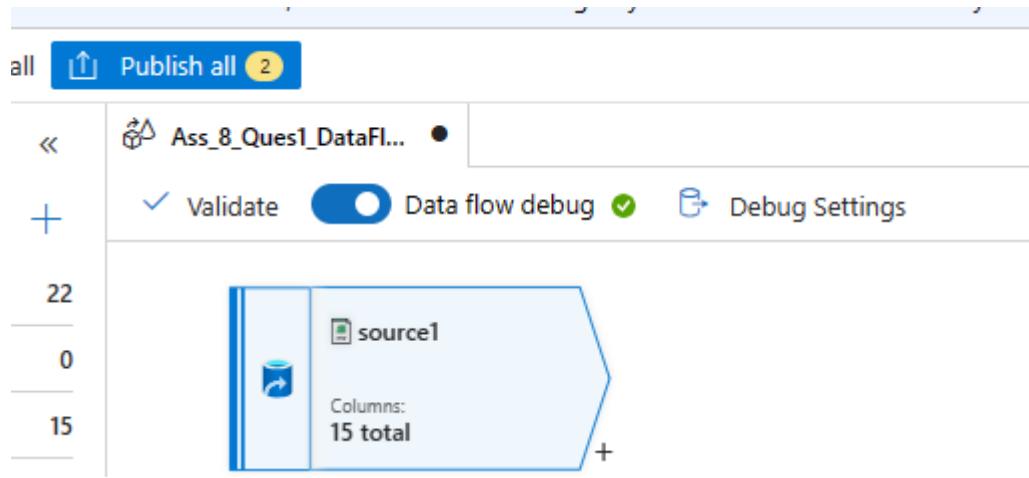
preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click

ate all Publish all 2

Ass_8_Ques1_DataFl... ●

+ Validate Data flow debug Debug Settings

22



Source settings Source options **Projection** Optimize Inspect Data preview ●

Define default format Detect data type Import projection Reset schema

Column name	Type	Format
CustomerID	short	Specify format
NameStyle	boolean	Specify format
Title	string	Specify format
FirstName	string	Specify format
MiddleName	string	Specify format
LastName	string	Specify format
Suffix	string	Specify format
CompanyName	string	Specify format

Source settings Source options Projection Optimize **Inspect** Data preview ●

Number of columns **Total 15**

Order ↑↓	Column ↑↓
1	CustomerID
2	NameStyle
3	Title
4	FirstName
5	MiddleName
6	LastName
7	Suffix
...	CompanyName

Source settings Source options Projection Optimize Inspect **Data preview** ●

Number of rows **INSERT 100** **UPDATE 0** **DELETE 0** **UPSERT 0**

Custom...	NameSt...	Title	FirstName	Middle...	LastName	Suffix	Compan...	SalesPe...
1	False	Mr.	Orlando	N.	Gee	NULL	A Bike St...	adventu...
2	False	Mr.	Keith	NULL	Harris	NULL	Progress...	adventu...
3	False	Ms.	Donna	F.	Carreras	NULL	Advance...	adventu...
4	False	Ms.	Janet	M.	Gates	NULL	Modular...	adventu...
5	False	Mr.	Lucy	NULL	Harringt...	NULL	Metropo...	adventu...
6	False	Ms.	Rosmarie	J.	Carroll	NULL	Aerobic ...	adventu...
7	False	Mr.	Dominic	P.	Gash	NULL	Associat...	adventu...

Next step is to pick customer ID == Even number? SO I need to “Filter” it for even number--

Create a pipeline to copy the customer data from csv file to SQL where the customer id is an even number.
(Attaching the Customer File)

The screenshot shows the Azure Data Factory Pipeline Designer interface. On the left, there is a source component named "source1" with 15 total columns. A red circle highlights the "+" button next to the "Derived Column" action in the context menu. Another red circle highlights the "Filter" action under the "Row modifier" section of the context menu.

source1
Columns: 15 total

+ Derived Column

- Select
- Aggregate
- Surrogate Key
- Pivot
- Unpivot
- Window
- Rank
- External Call
- Cast

Source options

Columns **Total 15**

Row modifier

- Filter
- Sort
- Alter Row

preview

New to Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started.

[Publish all](#) 2

« [Ass_8_Ques1_DataFl... ●](#) | [Ass_8_Q1_Customer... ●](#)

Validate Data flow debug Debug Settings

22
0
15

The diagram illustrates a data flow transformation. A source stream (Reference: 1, Columns: 15 total) merges into a filter stream (filter1, Columns: 15 total). A dashed box labeled "Add Source" is present.

Filter settings Optimize Inspect Data preview ●

Output stream name * [Learn more](#)

Description [Reset](#)

Incoming stream *

Filter on * ANY

Filter settings Optimize Inspect Data preview ●

Output stream name * [Learn more](#)

Description [Reset](#)

Incoming stream *

Filter on * ANY

 [Open expression builder](#)

Dataflow expression builder

filter1

Expression

CustomerID%2 == 0

+ - * / || && !

Expression elements

- All
- Functions
- Input schema
- Parameters

Expression values ↗

Filter by keyword

Create new

CustomerID

NameStyle

The screenshot shows the Dataflow expression builder interface. At the top, there's a breadcrumb navigation: 'Dataflow expression builder' and 'filter1'. Below that is a section titled 'Expression' containing the expression 'CustomerID%2 == 0', which is circled in red. Underneath the expression are standard arithmetic operators (+, -, *, /, ||, &&, !). To the left is a sidebar with 'Expression elements' and a list of items: All, Functions, Input schema, and Parameters. To the right is a section titled 'Expression values' with a search bar ('Filter by keyword') and a 'Create new' button. Two items are listed: 'CustomerID' (with a circled '123' icon) and 'NameStyle'. A red arrow points from the 'CustomerID' entry in the 'Expression values' list towards the circled 'CustomerID%2' in the expression editor.

The screenshot shows the Microsoft Azure Data Factory Dataflow expression builder. At the top, there is a toolbar with operators: +, -, *, /, ||, &&, !, and ^. Below the toolbar are two panes: 'Expression elements' on the left and 'Expression values' on the right.

Expression elements

- All
- Functions
- Input schema
- Parameters
- Cached lookup
- Data flow library functions

Expression values

- CustomerID
- NameStyle
- Title
- FirstName
- MiddleName

At the bottom left, there is a 'Data preview' button with a refresh icon, which is circled in red with a handwritten note pointing to it.

The screenshot shows the Microsoft Azure Data Factory Dataflow expression builder. At the top, there is a toolbar with operators: +, -, *, /, ||, &&, !, and ^.

Expression

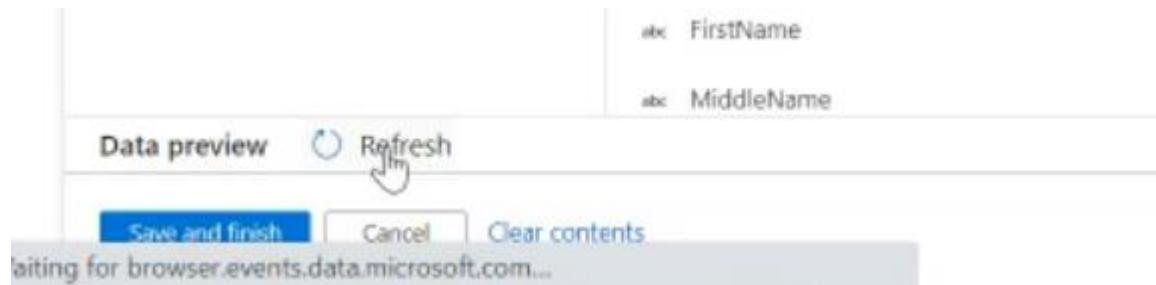
```
CustomerID%2 == 0
```

Expression elements

- All
- Functions
- Input schema
- Parameters
- Cached lookup

Expression values

- CustomerID
- NameStyle



Refresh the Data Preview shows the preview data:

Data preview Refresh

Output: ✓

	CustomerID 123
x	1
✓	2
x	3
✓	4
x	5
✓	6
x	7
✓	10
x	11

Searched lookup

Data preview Refresh

Output: ✓

	CustomerID 123
x	1
✓	2
x	3
✓	4

x

✓

x

✓

Save and finish Cancel Clear contents

https://en/authoring/dataflow/Assignment%207%20Q1?factory=%2Fsubscriptions%2F7dce91bb-c16e-47d8-a73e-c8d57748ac2a%...

Search

Deepakgoyal11@hotmail.com

Set properties

Name: SQLDbCustomerDB

Linked service: SQL_DB_LS

Select from existing table (radio button)

Schema and table name: saleslt - CustomerClone

Output stream name: sink1

Description: Add sink dataset

Incoming stream: filter1

Sink type: Dataset

Dataset: Select... (dropdown menu)

Options: Allow schema drift (checked), Validate schema (unchecked)

Validate all 2

Assignment 7 Q1

Assignment7Q1Customer...

source1

Import data from Assignment7Q1CustomerCSV

filter1

Filtrating rows using expressions on column customerID

Columns: 15 total

sink1

CustomerClone

Advanced

```
graph LR; source1 --> filter1; filter1 --> sink1[sink1]; sink1 --> CustomerClone[CustomerClone]
```

Ass_8_Ques1_DataFl... ● Ass_8_Q1_Customer_...

Validate Data flow debug Debug Settings

source1

Import data from Ass_8_Q1_Customer_CSV

filter1

Columns: 15 total

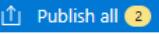
sink

Destination

Sink

Add Source

```
graph LR; source1 --> filter1; filter1 --> sink
```

all  Publish all 2

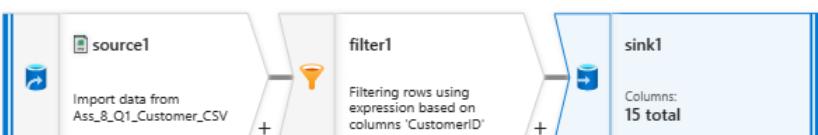
« Ass_8_Ques1_DataFl... ● Ass_8_Q1_Customer_... ●

+ Validate  Data flow debug ✓ 

22
0
15

asset

1
0



```
graph LR; source1[source1<br/>Import data from Ass_8_Q1_Customer_CS] --> filter1[filter1<br/>Filtering rows using expression based on columns 'CustomerID']; filter1 --> sink1[sink1<br/>Columns: 15 total]
```

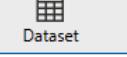
Add Source

Sink Settings Errors Mapping Optimize Inspect Data preview

Output stream name * sink1 [Learn more](#)

Description Add sink dataset [Reset](#)

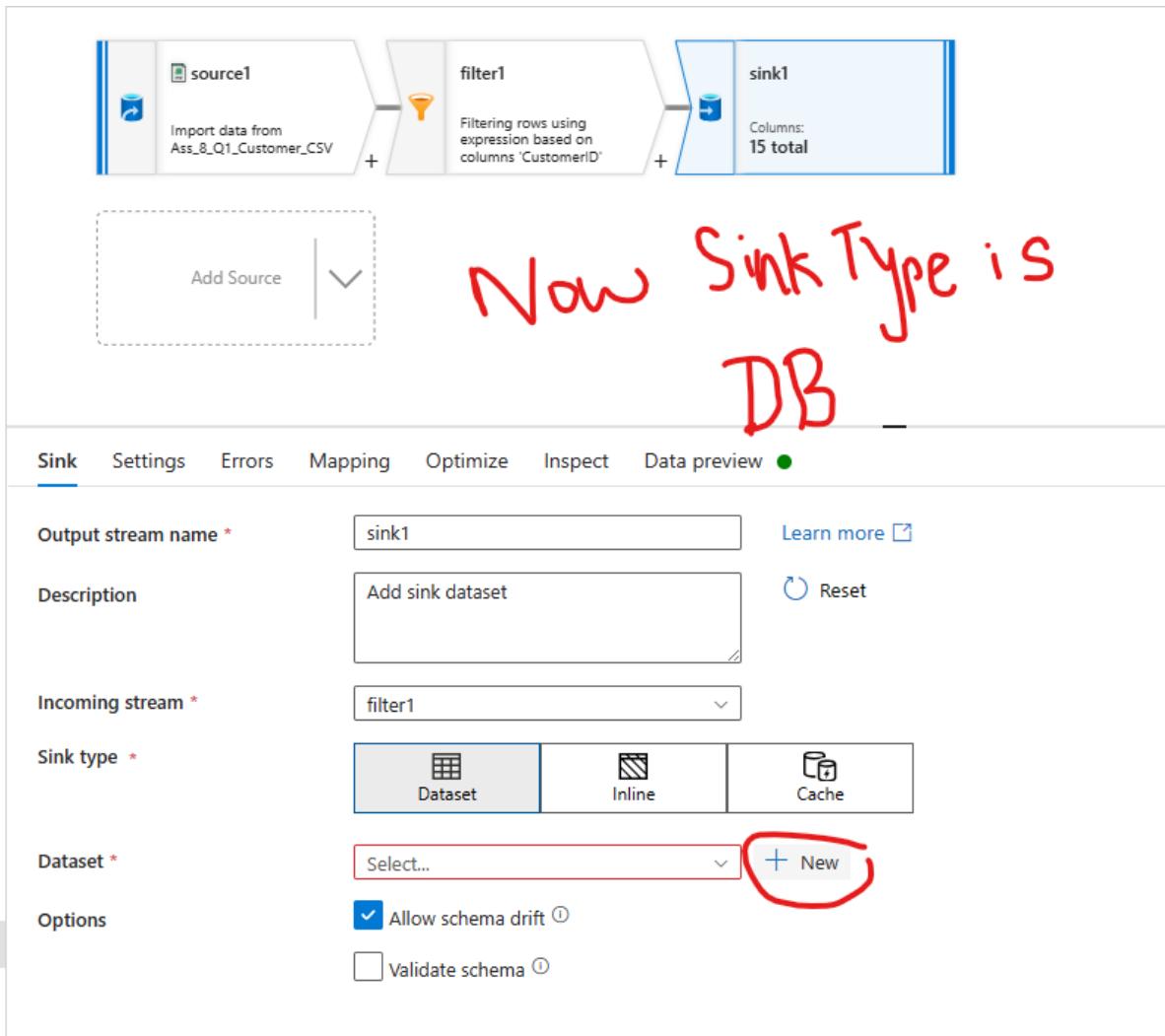
Incoming stream * filter1

Sink type *  Dataset   Cache

Dataset * Select... [New](#)

Options Allow schema drift [①](#) Validate schema [①](#)

Diagram description: The screenshot shows the Azure Data Factory Data Flow blade. At the top, there are tabs for 'Sink' (selected), 'Settings', 'Errors', 'Mapping', 'Optimize', 'Inspect', and 'Data preview'. Below these are configuration fields: 'Output stream name' set to 'sink1', 'Description' set to 'Add sink dataset', 'Incoming stream' set to 'filter1', 'Sink type' set to 'Dataset', and a dropdown for 'Dataset' currently showing 'Select...'. Under 'Options', the 'Allow schema drift' checkbox is checked, while 'Validate schema' is unchecked. On the left side, there's a vertical sidebar with navigation links like 'all', 'Ass_8_Ques1_DataFl...', 'Ass_8_Q1_Customer...', and a search bar. A status bar at the bottom shows '22', '0', '15', 'asset', '1', and '0'.



Create a pipeline to copy the customer data from csv file to SQL where the customer id is an even number.(Attaching the Customer File)
 File: SalesLT.Customer-xRyuaEu2p7-WdCbzuWIlo-80V7BfCQtP

DEFAULT

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure data within a data store. [Learn more](#)

Select a data store



sql

All

Azure

Database

File

Generic protocol

NoSQL

Services and ai



Azure Cosmos DB for
NoSQL



Azure Database for
MySQL



Azure Database for
PostgreSQL



Azure SQL Database



Azure SQL Database
Managed Instance



SQL server



Amazon RDS for SQL
Server



MySQL



PostgreSQL

Continue



gopinathma09@outlook.co
DEFAULT DIRECTO

Set properties

Name

SQLDB_CustomerDB

Linked service *

SQL_DB_Linked_service1

X



Select from existing table

New table

Schema and table name

SalesLt

. CustomerClone

> Advanced

OK

Back

Cancel



gopinathma09@outlook.co
DEFAULT DIRECTO

Set properties

Name

SQLDB_CustomerDB

Linked service *

SQL_DB_Linked_service1

X



Select from existing table New table

Schema and table name

SalesLt

. CustomerClone

> Advanced

OK

Back

Cancel

Microsoft Azure | Search resources, services, and docs (G+)

Home > mission100adebatch3 (mission100adebatch31/mission100adebatch3)

mission100adebatch3 (mission100adebatch31/mission100adebatch3) | Query editor (preview) ...

Search Login New Query Open query Feedback

Overview Activity log Tags Diagnose and solve problems Getting started Query editor (preview) Settings Compute + storage Connection strings Properties Locks Data management Replicas Sync to other databases Integrations

For full capability please click here to open Azure Data Studio.

Tables

- > dbo.BuildVersion
- > dbo.Employee_Type1
- > dbo.Employee_Type2
- > dbo.ErrorLog
- > dbo.HighWaterMarkTable
- > dbo.HWM
- > SalesLT.Address
- > SalesLT.Customer
- > SalesLT.CustomerAddress
- > SalesLT.Product
- > SalesLT.ProductCategory
- > SalesLT.ProductDescription
- > SalesLT.ProductModel
- > SalesLT.ProductModelProductD...
- > SalesLT.SalesOrderDetail
- > SalesLT.SalesOrderHeader

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 select * from sa
```

abc [SalesLT].[Address]
abc [SalesLT].[Customer]
abc [SalesLT].[CustomerAddress]
abc [SalesLT].[Product]
abc [SalesLT].[ProductCategory]
abc [SalesLT].[ProductDescription]
abc [SalesLT].[ProductModel]
abc [SalesLT].[ProductModelProductDescription]
abc [SalesLT].[SalesOrderDetail]
abc [SalesLT].[SalesOrderHeader]

Results Messages

Search to filter items...

Ready

This screenshot shows the Microsoft Azure Query editor interface. The left sidebar contains navigation links for Overview, Activity log, Tags, Diagnose and solve problems, Getting started, Query editor (preview), Settings, Compute + storage, Connection strings, Properties, Locks, Data management, Replicas, Sync to other databases, and Integrations. The main area displays a list of tables in the database, including dbo.BuildVersion, dbo.Employee_Type1, dbo.Employee_Type2, dbo.ErrorLog, dbo.HighWaterMarkTable, dbo.HWM, SalesLT.Address, SalesLT.Customer, SalesLT.CustomerAddress, SalesLT.Product, SalesLT.ProductCategory, SalesLT.ProductDescription, SalesLT.ProductModel, SalesLT.ProductModelProductD..., SalesLT.SalesOrderDetail, and SalesLT.SalesOrderHeader. A query window titled 'Query 1' is open with the command 'select * from sa'. A dropdown menu lists various schema objects starting with 'abc [SalesLT]'. The 'Results' tab is selected, and the output area shows the results of the query. The status bar at the bottom indicates 'Ready'.

Microsoft Azure | Search resources, services, and docs (G+)

Home > mission100adebatch3 (mission100adebatch31/mission100adebatch3)

mission100adebatch3 (mission100adebatch31/mission100adebatch3) | Query editor (preview) ...

Search Login New Query Open query Feedback

Overview Activity log Tags Diagnose and solve problems Getting started Query editor (preview) Settings Compute + storage Connection strings Properties Locks Data management Replicas Sync to other databases Integrations

For full capability please click here to open Azure Data Studio.

Tables

- > dbo.BuildVersion
- > dbo.Employee_Type1
- > dbo.Employee_Type2
- > dbo.ErrorLog
- > dbo.HighWaterMarkTable
- > dbo.HWM
- > SalesLT.Address
- > SalesLT.Customer
- > SalesLT.CustomerAddress
- > SalesLT.Product
- > SalesLT.ProductCategory
- > SalesLT.ProductDescription
- > SalesLT.ProductModel
- > SalesLT.ProductModelProductD...
- > SalesLT.SalesOrderDetail

Query 1

Run Cancel query Save query Export data as Show only Editor

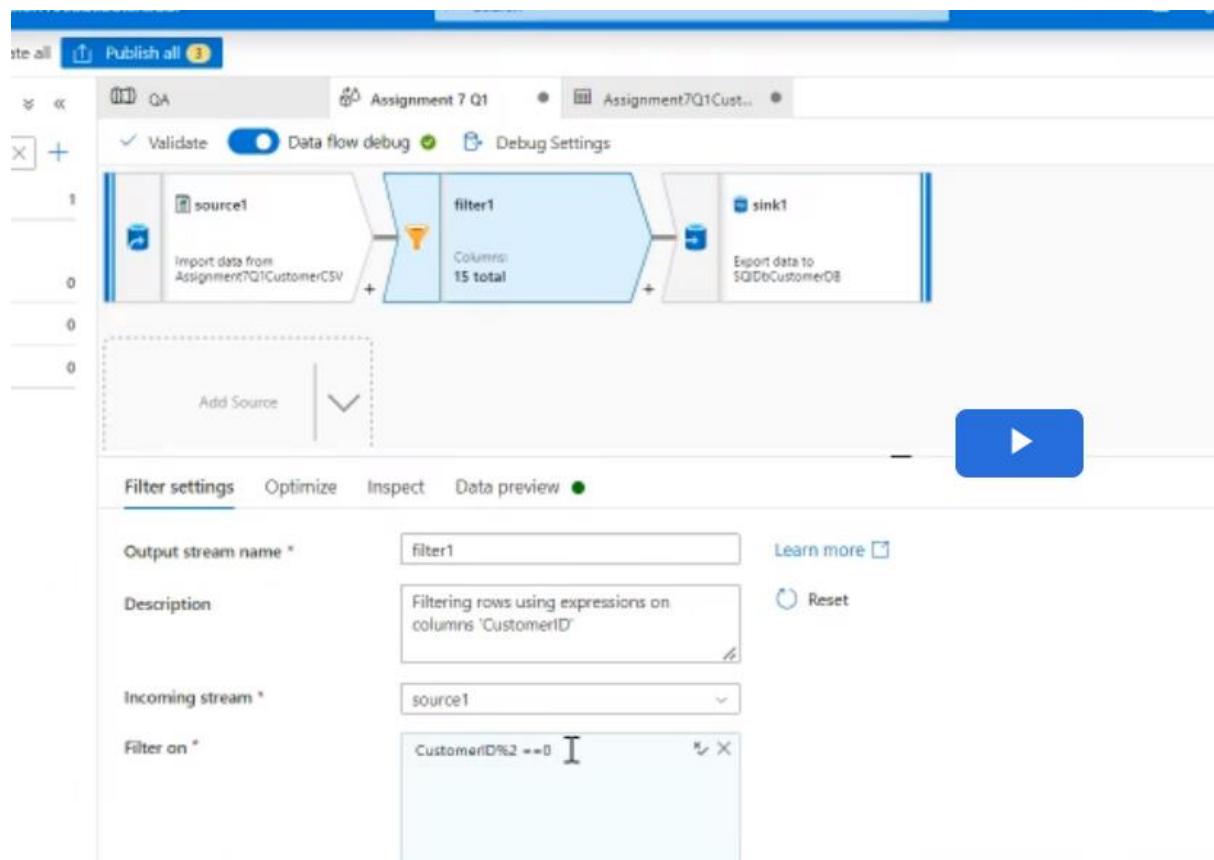
```
1 select * from [SalesLT].[CustomerClone]
```

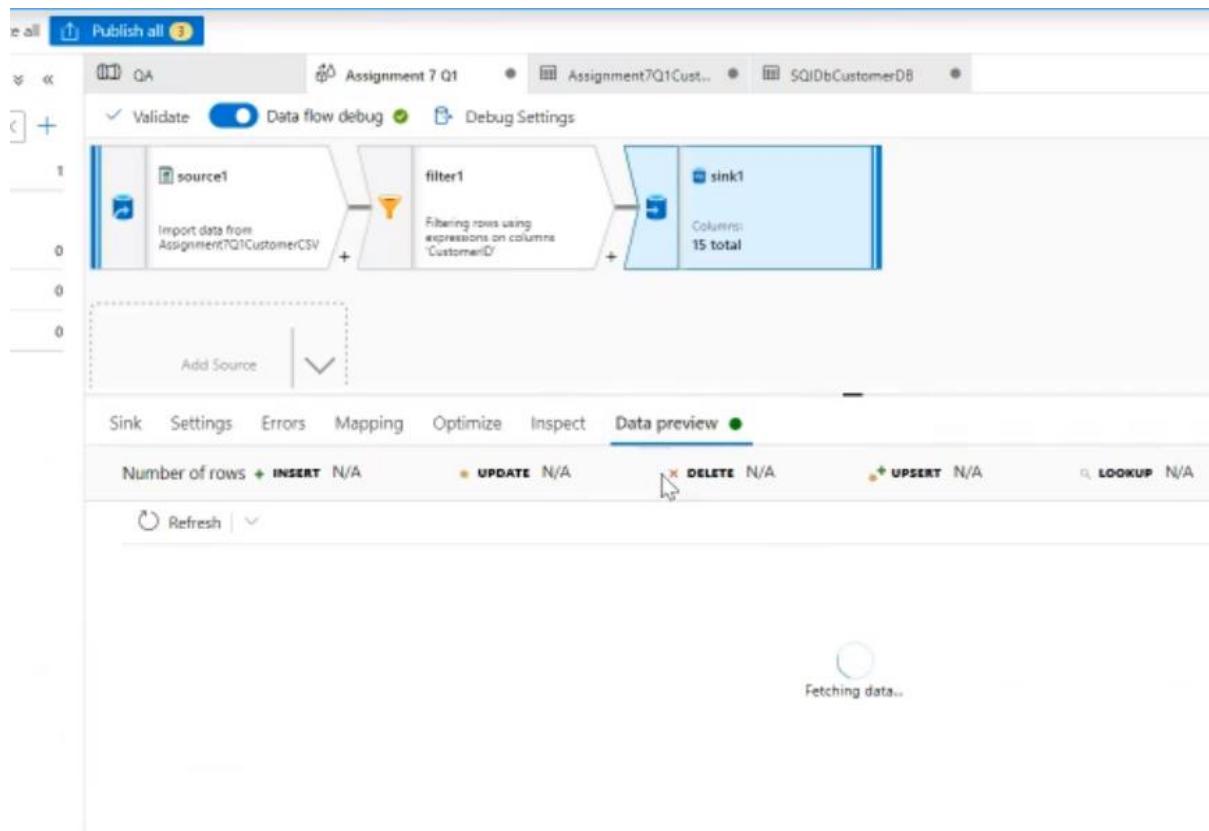
No Table

Results Messages

Failed to execute query. Error: Invalid object name 'salesLT.CustomerClone'.

This screenshot shows the Microsoft Azure Query editor interface, similar to the one above. The left sidebar contains the same navigation links. The main area displays a list of tables, including dbo.BuildVersion, dbo.Employee_Type1, dbo.Employee_Type2, dbo.ErrorLog, dbo.HighWaterMarkTable, dbo.HWM, SalesLT.Address, SalesLT.Customer, SalesLT.CustomerAddress, SalesLT.Product, SalesLT.ProductCategory, SalesLT.ProductDescription, SalesLT.ProductModel, SalesLT.ProductModelProductD..., and SalesLT.SalesOrderDetail. A query window titled 'Query 1' is open with the command 'select * from [SalesLT].[CustomerClone]'. The 'Messages' tab is selected, and an error message is displayed: 'Failed to execute query. Error: Invalid object name 'salesLT.CustomerClone''. A red circle highlights this error message, and a red arrow points from the text 'No Table' above it towards the error message. The status bar at the bottom indicates 'Failed to execute query'.





Assignment 7 Q1

Validate Data flow debug Debug Settings

```

graph LR
    source1[source1] --> filter1[filter1]
    filter1 --> sink1[sink1]
    
```

source1: Import data from Assignment7Q1CustomerCSV

filter1: Reference 1
Columns 15 total

sink1: Columns 15 total

Add Source

Sink Settings Errors Mapping Optimize Inspect Data preview

Number of rows + INSERT 0 UPDATE 0 DELETE 0 UPSERT 0 LOOKUP 0 ERROR 0 TOTAL 440

Refresh | Statistics | Export to CSV |

	CustomerID	NameStyle	Title	FirstName	MiddleName	LastName	Suffix	CompanyName	SalesPerson
+	2	X	Mr.	Keith	NULL	Harris	NULL	Progress...	advent
+	4	X	Ms.	Janet	M.	Gates	NULL	Modular...	advent
+	6	X	Ms.	Rosmarie	J.	Carroll	NULL	Aerobic...	advent
+	10	X	Ms.	Kathleen	M.	Garza	NULL	Rural Cy...	advent
+	12	X	Mr.	Johnny	A.	Caprio	Mr.	Bikes an...	advent
+	16	X	Mr.	Christopher	R.	Beck	Mr.	Bulk Dis...	advent
+	18	X	Mr.	David	J.	Liu	NULL	Catalog ...	advent

11:06

(mission100adebatch31/mission100adebatch3) | Query editor (preview) ⋮

Login + New Query ⚡ Open query ⚡ Feedback

For full capability please click here to open Azure Data Studio.

Tables

- > dbo.BuildVersion ...
- > dbo.Employee_Type1 ...
- > dbo.Employee_Type2 ...
- > dbo.ErrorLog ...
- > dbo.HighWaterMarkTable ...
- > dbo.HWM ...
- > SalesLT.Address ...
- > SalesLT.Customer ...
- > SalesLT.CustomerAddress ...
- > SalesLT.Product ...

Query 1 ×

Run Cancel query Save query Export data as Show only Editor

```
1 select * from [SalesLT].[CustomerClone]
```

Results Messages

SQL database

Search Login + New Query ⚡ Open query ⚡ Feedback

For full capability please click here to open Azure Data Studio.

Tables

- > dbo.BuildVersion ...
- > dbo.Employee_Type1 ...
- > dbo.Employee_Type2 ...
- > dbo.ErrorLog ...
- > dbo.HighWaterMarkTable ...
- > dbo.HWM ...
- > SalesLT.Address ...
- > SalesLT.Customer ...
- > SalesLT.CustomerAddress ...
- > SalesLT.Product ...
- > SalesLT.ProductCategory ...
- > SalesLT.ProductDescription ...
- > SalesLT.ProductModel ...
- > SalesLT.ProductModelProductD ...
- > SalesLT.SalesOrderDetail ...
- > SalesLT.SalesOrderHeader ...

Query 1 ×

Run Cancel query Save query Export data as Show only Editor

```
1 select * from [SalesLT].[CustomerClone]
```

Results Messages

Ready

SQL database

Search

Login + New Query Open query Feedback

Overview

Activity log

Tags

Diagnose and solve problems

Getting started

Query editor (preview) **(circled)**

Tables

- dbo.BuildVersion
- dbo.Employee_Type1
- dbo.Employee_Type2
- dbo.ErrorLog
- dbo.HighWaterMarkTable
- dbo.HWM
- SalesLT.Address
- SalesLT.Customer
- SalesLT.CustomerAddress
- SalesLT.Product
- SalesLT.ProductCategory
- SalesLT.ProductDescription
- SalesLT.ProductModel
- SalesLT.ProductModelProductID
- SalesLT.SalesOrderDetail
- SalesLT.SalesOrderHeader

Run Cancel query Save query Export data as Show only Editor

1 select * from [SalesLT].[CustomerClone]

Results Messages

No Clone Cust Exist

Query 1 ×

Run Cancel query Save query Export data as Show only Editor

1 select * from [SalesLT].[CustomerClone]

↑

Results Messages

Failed to execute query. Error: Invalid object name 'SalesLT.CustomerClone'.

mission100adebatch31/mission100adebatch3)

batch3 (mission100adebatch31/mission100adebatch3) | Query editor (preview) ...

[Login](#) [New Query](#) [Open query](#) [Feedback](#)

For full capability please click here to open Azure Data Studio.

Tables

- > dbo.BuildVersion
- > dbo.Employee_Type1
- > dbo.Employee_Type2
- > dbo.ErrorLog
- > dbo.HighWaterMarkTable
- > dbo.HWM
- > SalesLT.Address
- > SalesLT.Customer
- > SalesLT.CustomerAddress
- > SalesLT.Product
- > SalesLT.ProductCategory
- > SalesLT.ProductDescription
- > SalesLT.ProductModel

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 select * from [SalesLT].[CustomerClone]
```

Results **Messages**

Search to filter items...

CustomerID	NameStyle	Title
2	False	Mr.

view of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Ass_8_Ques1_DataFl... ● Ass_8_Q1_Customer... ●

Validate Data flow debug   Debug Settings

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

Output stream name * sink1 [Learn more](#)

Description Export data to SQLDB_CustomerDB [Reset](#)

Incoming stream * filter1

Sink type *

- Dataset
- Inline
- Cache

Dataset * SQLDB_CustomerDB [Test connection](#) [Open](#) [New](#)

Options

- Allow schema drift 
- Validate schema 

Ass_8_Ques1_DataFl... ● | Ass_8_Q1_Customer_... ●

Validate Data flow debug Debug Settings

```
graph LR; source1[source1<br/>Import data from Ass_8_Q1_Customer_CSV] --> filter1[filter1<br/>Filtering rows using expression based on columns 'CustomerID']; filter1 --> sink1[sink1<br/>Columns: 15 total]
```

Add Source

Sink Settings Errors Mapping Optimize Inspect Data preview

Number of rows + INSERT 0 * UPDATE 0 × DELETE 0

Refresh | Statistics Export to CSV |

Custom...	NameSt...	Title	FirstName	Middle...	LastName	Suffix	C
2	X	Mr.	Keith	NULL	Harris	NULL	P
4	X	Ms.	Janet	M.	Gates	NULL	N
6	X	Ms.	Rosmarie	J.	Carroll	NULL	A
10	X	Ms.	Kathleen	M.	Garza	NULL	R
12	X	Mr.	Johnny	A.	Caprio	Jr.	B
16	X	Mr.	Christop...	R.	Beck	Jr.	B
18	X	Mr.	David	J.	Liu	NULL	C
20	X	Ms.	Jean	P.	Handlev	NULL	C

The screenshot shows the Azure Data Studio interface. At the top, there's a navigation bar with 'Login', 'New Query', 'Open query', and 'Feedback'. Below it is a message: 'For full capability please click here to open Azure Data Studio.' On the left, a sidebar titled 'Tables' lists various tables from the SalesLT database, such as 'dbo.BuildVersion', 'dbo.Employee_Type1', 'dbo.Employee_Type2', etc. The main area is titled 'Query 1' with the following SQL code:

```
1 select * from [SalesLT].[CustomerClone]
```

Below the code is a blue play button icon. To the right, there are tabs for 'Results' and 'Messages', with 'Results' being selected. The results table shows one row of data:

CustomerID	NameStyle	Title
2	False	Mr.

Data flow does not run its by own –so u need to create a pipeline

This screenshot shows the Microsoft Fabric Data Factory interface. The left sidebar lists 'Factory Resources' including 'Activities', 'Datasets', and 'Pipelines'. Under 'Activities', 'Pipeline' is selected, and a new pipeline named 'Ass_8_Q1' is shown. The properties pane on the right is configured with the following details:

- Name:** Ass_8_Q1
- Description:** Create a pipeline to copy the customer data from csv file to SQL where the customer id is an even number.

This screenshot shows the Microsoft Fabric Data Factory interface. The left sidebar lists 'Factory Resources' including 'Activities', 'Datasets', and 'Pipelines'. Under 'Activities', 'Data flow' is selected, and a new data flow named 'Assignment_8_Ques_1' is shown. The properties pane on the right is configured with the following details:

- Name:** Assignment_8_Ques_1
- Description:** Create a pipeline to copy the customer data from csv file to SQL where the customer id is an even number.

Ass_8_Q1_Customer... ● Assignment_8_Que... ●

Validate Debug Add trigger Data flow debug ✓

Data flow Data Flow1

General Settings Parameters User properties

Name * Data Flow1 Learn more ↗

Description Call DataFlow Ass_8_Ques1_DataFlow

Activity state Activated Deactivated

Timeout 0.12:00:00

Retry 0

Retry interval (sec) 30

Secure output

Azure | Data Factory > mission100Gopt01012024dat-backup1 Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory < Validate all Publish all 4

Activity Resources < Filter resources by name + Ass_8_Ques1_DataFlow Ass_8_Q1_Customer... Assignment_8_Que...

Activities < Validate Debug Add trigger Data flow debug

Ass_8_Ques1_DataFlow

Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query

Data flow Data Flow Ass 8 Ques1 Dat...

General Settings Parameters User properties

Name * Data Flow Ass_8_Ques1_DataFlow Learn more

Description Call DataFlow Ass_8_Ques1_DataFlow

Activity state Activated Deactivated

Timeout 0:12:00:00

Retry 0

Retry interval (sec) 30

The screenshot shows the Azure Data Factory interface. On the left, there's a sidebar with 'Activity Resources' containing a list of various activities like 'Move and transform', 'Synapse', etc., and a specific 'Data flow' item which is highlighted. The main area is titled 'Activities' and shows a single 'Data flow' activity named 'Ass_8_Ques1_DataFlow'. Below the activity title, there are tabs for 'General', 'Settings', 'Parameters', and 'User properties'. Under 'General', the 'Name' is set to 'Data Flow Ass_8_Ques1_DataFlow', and the 'Description' field contains the text 'Call DataFlow Ass_8_Ques1_DataFlow'. The 'Activity state' is set to 'Activated'. Under 'Timeout', the value is '0:12:00:00'. Under 'Retry', the value is '0'. Under 'Retry interval (sec)', the value is '30'. There are also icons for deleting the activity and for opening its details.

Ass_8_Q1_Customer... • Assignment_8_Que... •

Validate Debug Add trigger Data flow debug ✓

Data flow Data Flow1

General Settings Parameters¹ User properties

Data flow * Ass_8_Ques1_DataFlow

Run on (Azure IR) * Ass_8_Ques1_DataFlow

Compute size *

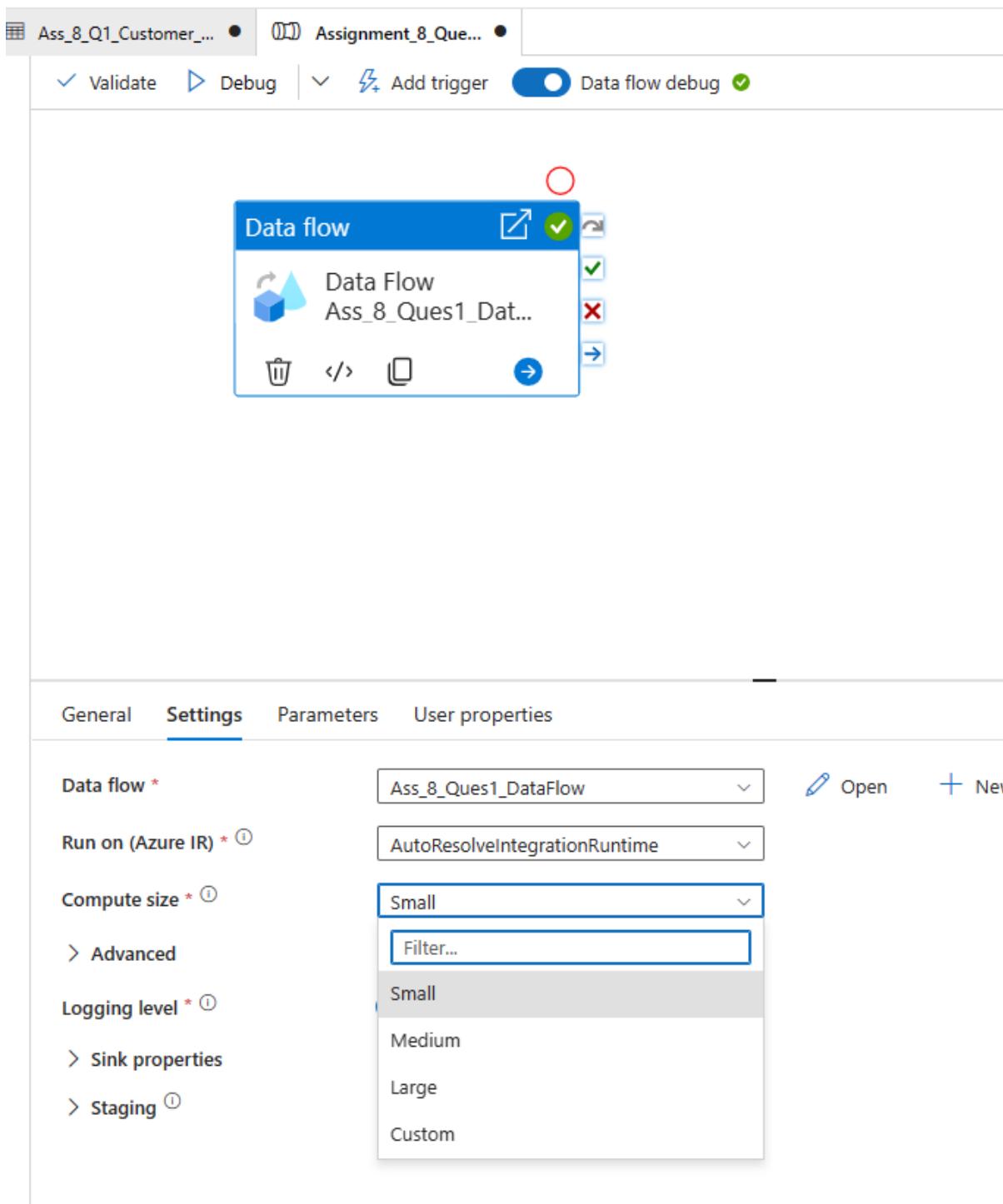
> Advanced

Logging level * Verbose Basic None

> Sink properties

> Staging

The screenshot shows the Azure Data Factory interface for managing a data flow. The top navigation bar includes tabs for 'Ass_8_Q1_Customer...' and 'Assignment_8_Que...'. Below the tabs are buttons for 'Validate', 'Debug', 'Add trigger', and a toggle for 'Data flow debug' which is turned on. The main workspace contains a 'Data flow' section with a blue header labeled 'Data flow' and 'Data Flow1'. To the right of this section are icons for edit, checkmark, error, and refresh. A red circle is drawn around the edit icon. On the far left, there's a vertical toolbar with various icons. The bottom half of the screen shows the 'Settings' tab of the 'Data flow' configuration pane. It includes fields for 'Data flow' (set to 'Ass_8_Ques1_DataFlow'), 'Run on (Azure IR)' (set to 'Ass_8_Ques1_DataFlow'), 'Compute size' (set to 'Small'), and 'Logging level' (set to 'Verbose'). There are also sections for 'Advanced', 'Sink properties', and 'Staging' which are currently collapsed.



Announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

[Validate all](#) [Publish all 4](#)

Sources

- + [Ass_8_Ques1_DataFlow](#)
- [int_5_Ingestion_Product...](#)
- [int_6_Foreach_Example](#)
- [int_6_part6_threshold_file](#)
- [Customer_JSON_To_Folder](#)
- [Customer_Table_To_CSVC](#)
- [Customer_To_CSVPipe](#)
- [Customer_To_JSON](#)
- [ductTable_To_CSVC](#)
- [Customer_SQLDB_ADLS](#)
- [Customer_SQLDB_ADLS_F...](#)
- [Product_To_JSON](#)
- [int_8_Ques_1](#)
- [Foreach_Example_Pipeline](#)
- [Capture \(preview\)](#) 0
- 16
- [Customer_CSVDataset](#)
- [Customer_CSVPipe](#)
- [DataLakeStorage_CSVDataset](#)
- [Generic_CSVDs](#)
- [Generic_JSON_DS](#)

Activities

Trigger test runs of the current pipeline without publishing your changes to the service.

[Move and transform](#)

- [Copy data](#)
- [Data flow](#)

[Azure Data Explorer](#)

- [Azure Data Explorer C...](#)

[Databricks](#)

- [Notebook](#)
- [Jar](#)
- [Python](#)

[Data Lake Analytics](#)

- [U-SQL](#)

[General](#)

- [Get Metadata](#)

General **Settings** **Parameters** **User properties**

Data flow * [Ass_8_Ques1_DataFlow](#) [Open](#) [New](#)

Run on (Azure IR) * [AutoResolveIntegrationRuntime](#)

Compute size * [Small](#)

Advanced

Logging level * Verbose Basic None

Preview experience

[Data flow](#) [Data Flow1](#)

Output

Pipeline run ID: fd76553c-fbc8-4eed-8ef3-b09354dc1fd4 [Logs](#) [Run history](#)

Pipeline status: [In progress](#) [Monitor in Azure Metrics](#)

All status

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity run ID
Data Flow1	In progress	Data flow	2/27/2024, 5:42:51 PM	24s			071518a2-...

Parameters	Variables	Settings	Output
Pipeline run ID: fd76553c-fbc8-4eed-8ef3-b09354dc1fd4			
All status			Pipeline status Succeeded

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity run ID
Data Flow1	Succeeded	Data flow	2/27/2024, 5:42:51 PM	50s	debugpool-8Cores-Gen1		071518a2-2

ced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started.

« All pipeline runs > Assignment_8_Ques_1 - Activity runs > Data Flow1

Pipeline was modified after this run. The current pipeline configuration is shown.

Data Flow1

Cluster startup time: 1s 124ms Number of transformations: 3 Data flow status: Success

Refresh Auto refresh On Edit dataflow

Sinks All streams

Sink	Status	Processing time ↑↓
sink1	Succeeded	1s 954ms

The screenshot shows the Azure Data Studio interface for a SQL database named 'mission100Gopi01012024db'. The left sidebar has a red box around the 'Query editor (preview)' section. The main area shows a query window with a red box around the query 'select * from [SalesLT].[CustomerClone]'. Below the query is a results grid with a red box around the first column 'CustomerID'. The results grid displays the following data:

CustomerID	NameStyle	Title	FirstName	MiddleName
2	False	Mr.	Keith	
4	False	Ms.	Janet	M.
6	False	Ms.	Rosmarie	J.
10	False	Ms.	Kathleen	M.
12	False	Mr.	Johnny	A.
16	False	Mr.	Christopher	R.
18	False	Mr.	David	J.
20	False	Ms.	Jean	P.

Only Even Customer ID are displayed, hence our pipeline is correct

The screenshot shows the Azure Data Studio interface for the same database. The left sidebar has a red box around the 'Query editor (preview)' section. The main area shows a query window with the same query 'select * from [SalesLT].[CustomerClone]'. Below the query is a results grid with a red box around the first column 'CustomerID'. The results grid displays the following data:

CustomerID	NameStyle	Title	FirstName	MiddleName	LastName
2	False	Mr.	Keith		Harris
4	False	Ms.	Janet	M.	Gates
6	False	Ms.	Rosmarie	J.	Carroll
10	False	Ms.	Kathleen	M.	Garza
12	False	Mr.	Johnny	A.	Caprio
16	False	Mr.	Christopher	R.	Beck
18	False	Mr.	David	J.	Liu
20	False	Ms.	Jean	P.	Handley

Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (4)

NAME	CHANGE	EXISTING
▼ Pipelines		
Assignment_8_Ques_1	(New)	-
▼ Datasets		
Ass_8_Q1_Customer_CSV	(New)	-
SQLDB_CustomerDB	(New)	-
▼ Data flows		
Ass_8_Ques1_DataFlow	(New)	-

Publish

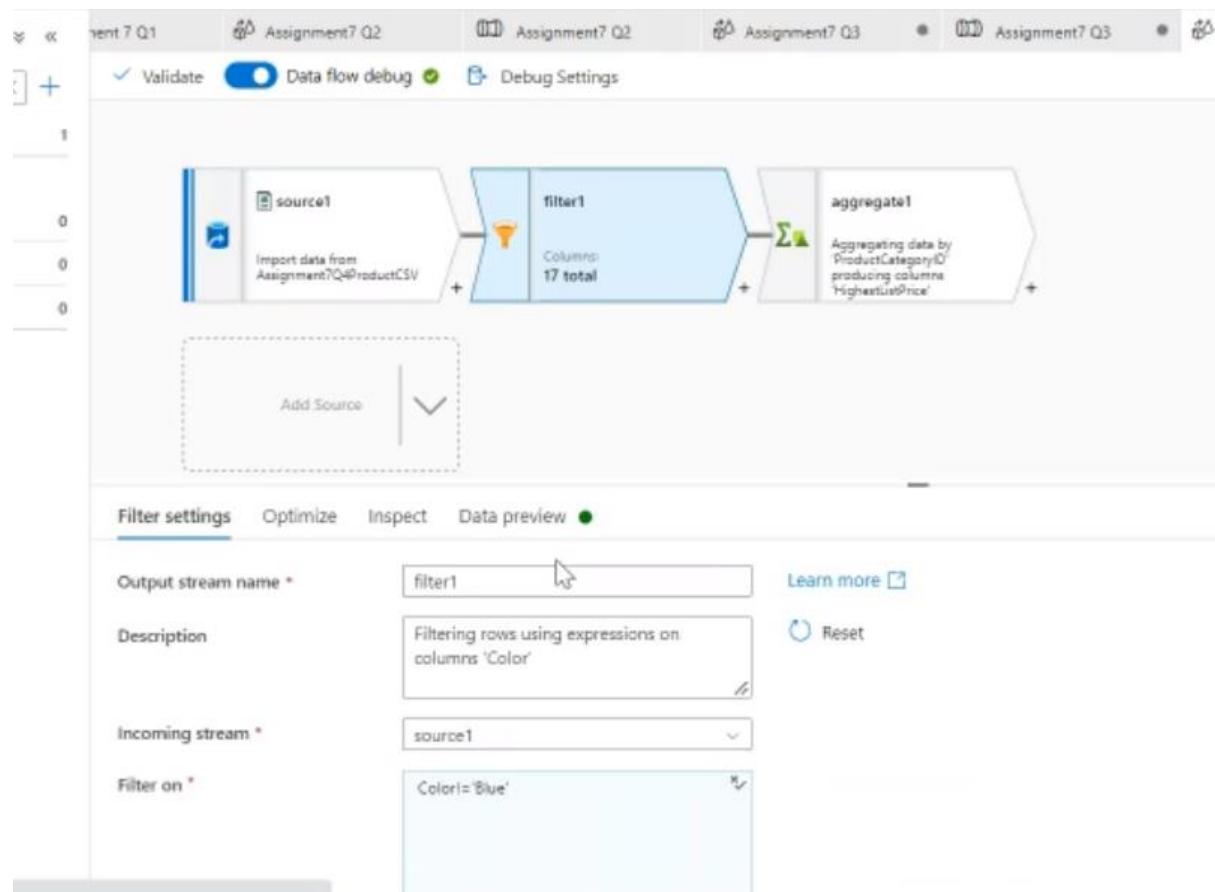
Cancel

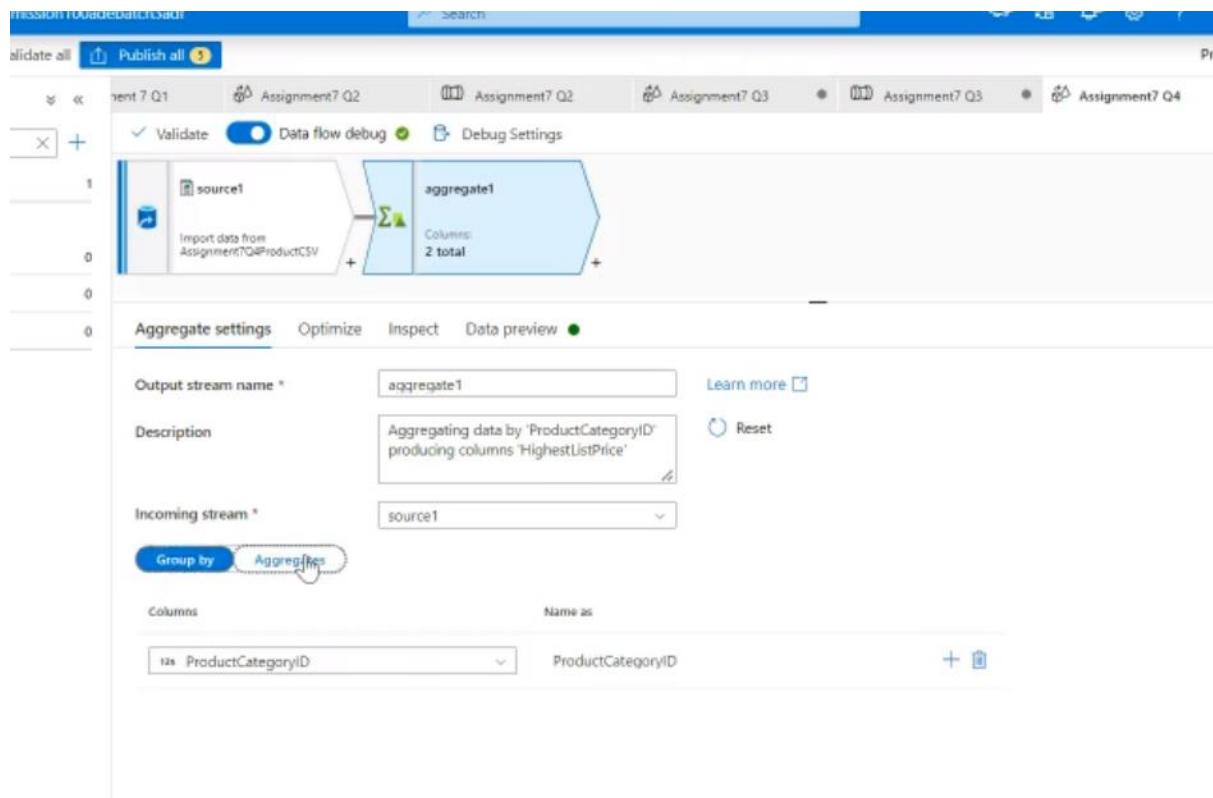
Question 4:

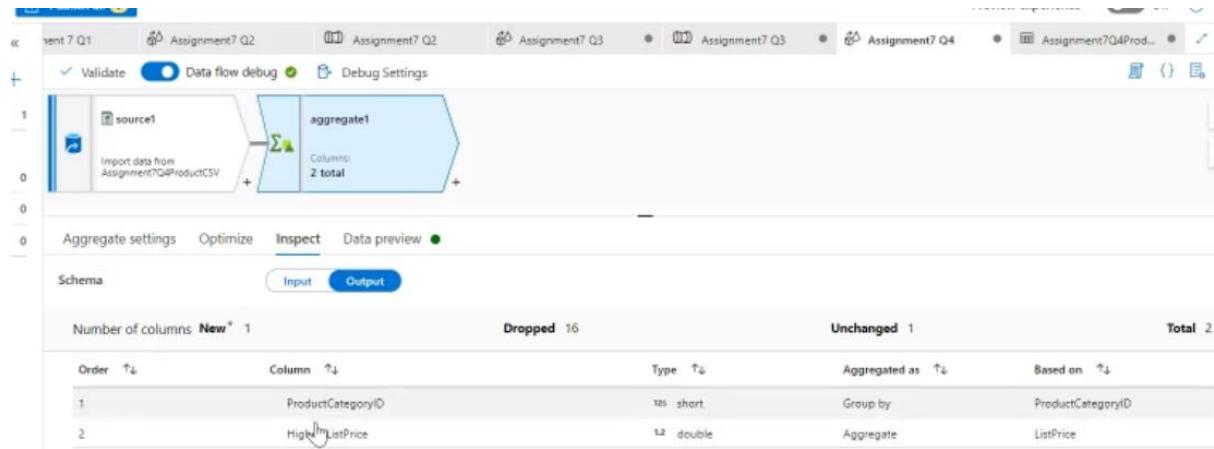
create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.

Ensure that product shouldn't be of blue in color and save the result as CSV file inside ProductResult folder.

Solution:







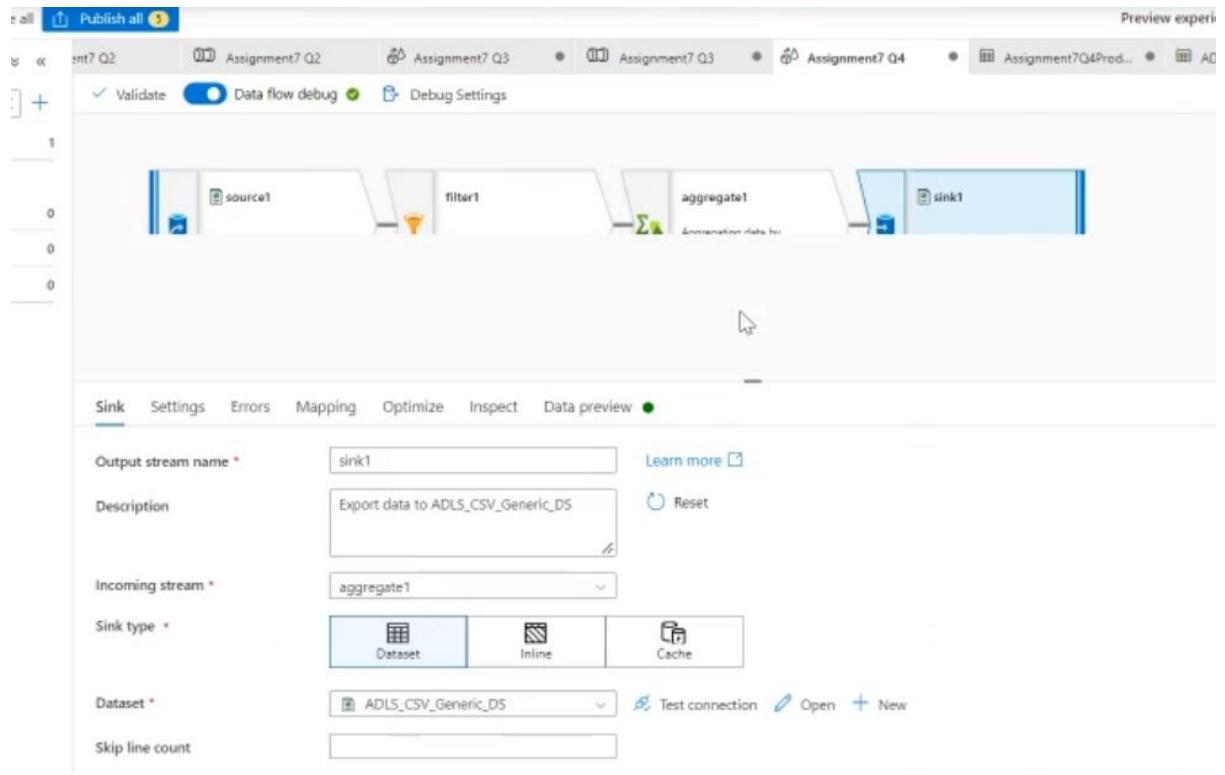
The screenshot shows the 'Aggregate settings' configuration for the 'aggregate1' component. At the top, there are buttons for 'Validate', 'Data flow debug' (selected), and 'Debug Settings'. Below these, the pipeline diagram is shown again.

The 'Aggregate settings' tab is selected. It contains the following fields:

- Output stream name:** aggregate1
- Description:** Aggregating data by 'ProductCategoryID' producing columns 'HighestListPrice'
- Incoming stream:** source1

Below these fields, there are two tabs: 'Group by' (selected) and 'Aggregates'. Under 'Group by', it says 'Grouped by: ProductCategoryID'. There are buttons for '+ Add', 'Clone', 'Delete', and 'Open expression builder'. Under 'Aggregates', there is a table with one row:

Column	Expression
HighestListPrice	max(ListPrice)



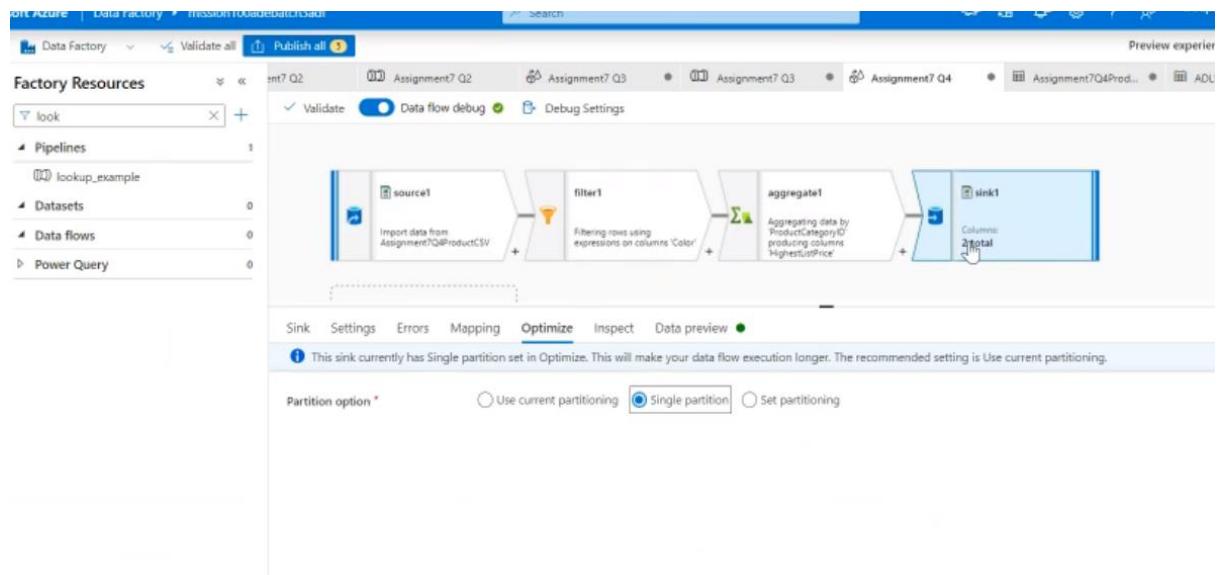
Question 5:

5. create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.
Ensure that product shouldn't be of blue in color and save the result as a SINGLE CSV file inside ProductSingleResult folder

Solution Q5:

Choose “**Single Partition**” in optimize tab (so there wont be multiple partition)

Rather than using parallel processng/ paralleising multiple files, it willconvert/ merge into single file/one file only instead of multiple files



Assignment7 Q1 Assignment7 Q2 Assignment7 Q2 Assignment7 Q3

+ ✓ Validate Data flow debug Debug Settings

source1 aggregate1

Import data from Assignment7Q4ProductCSV

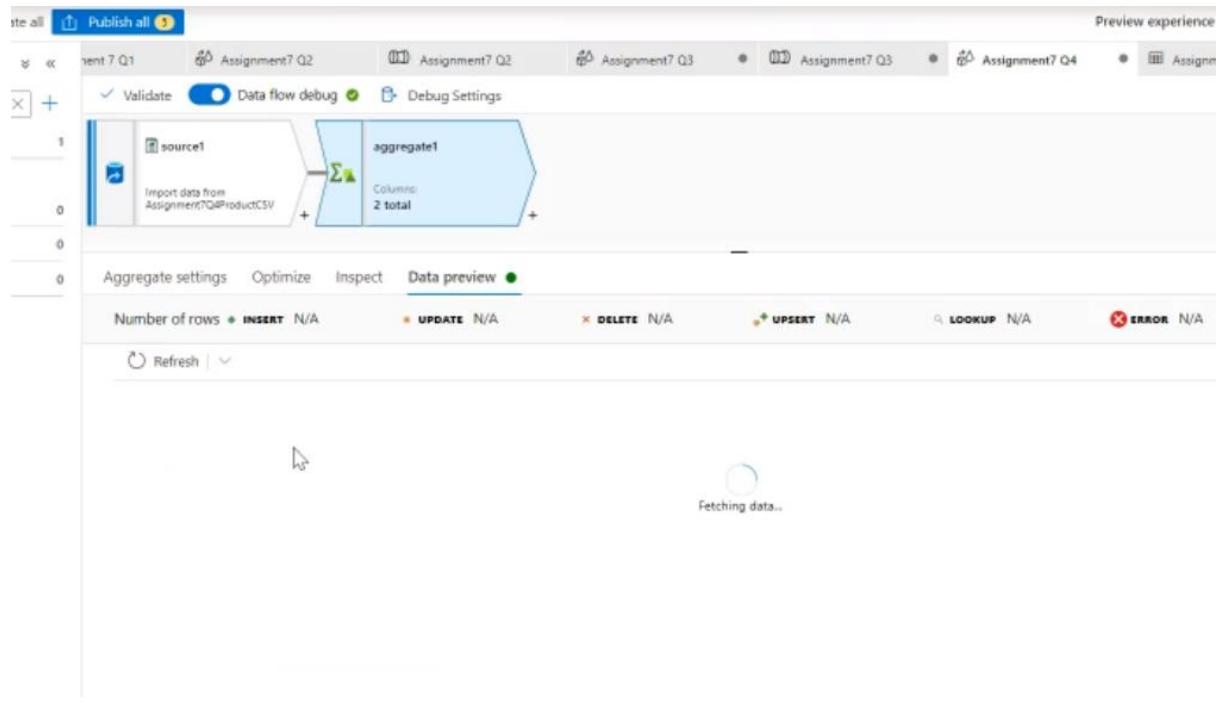
Columns: 2 total

Aggregate settings Optimize Inspect **Data preview**

Number of rows: **INSERT** 37 **UPDATE** 0 **DELETE** 0

Refresh Typecast Modify Map drifted Statistics Remove

ProductCategoryID	HighestListPrice
14	229.99
15	124.73
9	120.27
21	357.06
26	69.99
39	125.0
38	25.0
40	24.99
37	44.99



Dataflow expression builder

filter1

Expression

```
Color != 'Red'
```

+, -, *, /, ||

This screenshot shows the Dataflow expression builder interface. It features a sidebar with a tree view showing a node named 'filter1'. The main area is titled 'Expression' and contains the code 'Color != 'Red''. Below the expression input field are operators: '+', '−', '∗', '÷', and '||'. On the far left, there's a vertical toolbar with icons for different data types and operations.

Dataflow expression builder

filter1

Expression

Expression elements

All Functions Input schema Parameters Cached lookup

Expression values

Filter by keyword

Create new

abc ProductNumber abc Color

Data preview Refresh

Output: ✓ Color abc

✓	Red
✓	Black
✓	White
✓	White
X	Blue
✓	Multi

Save and finish Cancel Clear contents

for dc.services.visualstudio.com...

The screenshot shows the Dataflow expression builder interface. At the top, there's a toolbar with various operators like +, -, *, /, ||, &&, !, ^, ==, ===, <=>, !=. Below the toolbar is a sidebar titled 'Expression elements' containing 'All', 'Functions', 'Input schema', 'Parameters', and 'Cached lookup'. To the right of the sidebar is a panel titled 'Expression values' with a search bar 'Filter by keyword' and a 'Create new' button. Under 'Expression values' are two entries: 'abc ProductNumber' and 'abc Color'. Below these panels is a 'Data preview' section with a 'Refresh' button. The main area shows a table with one column labeled 'Output: ✓' and another column labeled 'Color abc'. The table contains six rows: the first four have checkmarks in the 'Output' column and are listed under 'Color' as Red, Black, White, and White respectively; the fifth row has an 'X' in the 'Output' column and is listed under 'Color' as Blue; the sixth row has a checkmark in the 'Output' column and is listed under 'Color' as Multi. At the bottom of the interface are buttons for 'Save and finish', 'Cancel', and 'Clear contents'.

Question 2:

2. Create a pipeline to join the two files (Customer, Customer Address) based on customer id and save the result as a JSON file.

Solution:

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analy

Data Factory Validate all Publish all 2

Factory Resources Assignment_8_Que... Ass_8_Q2_DataFlow Ass_8_Q

Filter resources by name +

Assignment_6_part6_threshold_file Pipeline

Assignment_8_Ques_1 Change Data Capture (preview)

Copy_Customer_JSON_To_Folder Dataset

Copy_Customer_Table_To_CSV Data flow

Copy_Customer_To_CSV_Pipe Power Query

Copy_Customer_To_JSON Flowlet

Copy_ProductTable_To_CSV Copy Data tool

Ingestion_Customer_SQLDB_ADLS

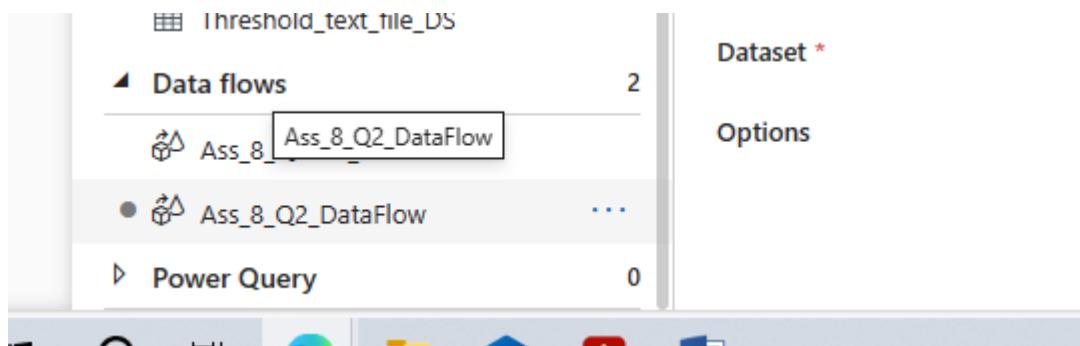
Ingestion_Customer_SQLDB_ADLS_F...

Properties

General Related

Name * Ass_8_Q2_DataFlow

Description Create a pipeline to join the two files (Customer, Customer Address) based on customer id and save the result as a



This screenshot shows the configuration for the "Customer" dataset. At the top, there are tabs for "Assignment_8_Que..." (disabled), "Ass_8_Q2_DataFlow" (selected), and "Ass_8_Q1_Customer_CSV". Below that are buttons for "Validate" (checked), "Data flow debug" (checked), and "Debug Settings".

The main area displays a preview of the dataset, showing a single row with 17 columns. The first column is labeled "Customer".

Below the preview, the "Source settings" tab is active. The configuration includes:

- Output stream name ***: Customer
- Description**: Import data from Ass_8_Q1_Customer_CSV
- Source type ***: Dataset (selected)
- Dataset ***: Ass_8_Q1_Customer_CSV (selected)
- Options**:
 - Allow schema drift
 - Infer drifted column types
 - Validate schema

Screenshot of the Azure Data Factory interface showing the 'Ass_8_Q1_Customer...' pipeline configuration.

The pipeline consists of three stages:

- Assignment_8_Queue...**: Represented by a Queue icon.
- Ass_8_Q2_DataFlow...**: Represented by a Data Flow icon.
- Ass_8_Q1_Customer...**: Represented by a Table icon.

The 'Ass_8_Q1_Customer...' stage is currently selected. It is configured as a DelimitedText source named 'Ass_8_Q1_Customer_CSV'. The connection is set to 'ADLS_Azure_DataLakeStorage1_Link...', which has a successful connection test result.

Configuration details for the 'Ass_8_Q1_Customer...' stage:

- Linked service**: ADLS_Azure_DataLakeStorage1_Link... (selected)
- File path**: landing / Ass_8_input_files / SalesLT.Customer-xRyuEu2...
- Compression type**: Select...
- Column delimiter**: Comma (,)
- Row delimiter**: Default (\r\n, or \n)
- Encoding**: Default(UTF-8)
- Quote character**: Double quote ("")
- Escape character**: Backslash (\)
- First row as header**:
- Null value**: (empty field)

Other buttons visible include Test connection, Edit, New, Learn more, Browse, Preview data, and Detect format.

Preview data

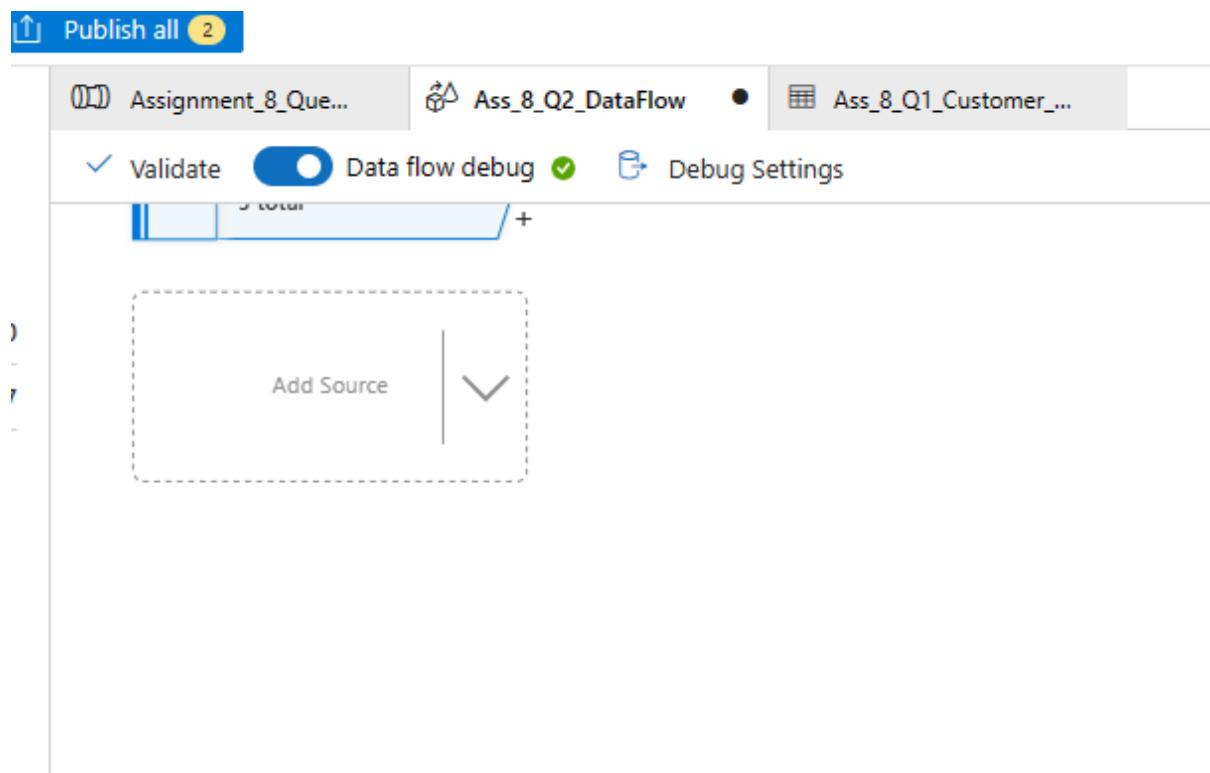
Linked service: ADLS_Azure_DataLakeStorage1_Linked_service

Object: SalesLT.Customer-xRyuaEu2p7-WdCbzuWllo-80V7BfCQtP.txt

#	CustomerID	NameStyle	Title	FirstName	MiddleName	LastName	Suffix	CompanyName	SalesPerson
1	1	False	Mr.	Orlando	N.	Gee		A Bike Store	adven works
2	2	False	Mr.	Keith		Harris		Progressive Sports	adven works
3	3	False	Ms.	Donna	F.	Carreras		Advanced Bike Components	adven works
4	4	False	Ms.	Janet	M.	Gates		Modular Cycle Systems	adven works
								Metropolitan	adven works

◀ ▶ ⏪ ⏩

Preview



CustomerAddress
Columns:
5 total

Source settings Source options Projection Optimize Inspect Data preview ●

Output stream name * CustomerAddress Learn more ↗

Description Import data from Ass_8_Q2_CustomerAddress_CSVClear Reset

Source type * Dataset Inline

Dataset * Ass_8_Q2_CustomerAddress_CSVClear Test connection Open + New

Options Allow schema drift ⓘ Infer drifted column types ⓘ Validate schema ⓘ

Skip line count

Sampling * ⓘ Enable Disable

CustomerAddress Learn more ↗

Import data from Ass_8_Q2_CustomerAddress_CSVClear Reset

Source type * Dataset Inline

Dataset * Ass_8_Q2_CustomerAddress_CSVClear Test connection Open + New New

Options Allow schema drift ⓘ Infer drifted column types ⓘ Validate schema ⓘ

Skip line count

Sampling * ⓘ Enable Disable

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

 Search

All Azure Database File Generic protocol NoSQL Services and apps



Amazon S3



Azure Blob Storage



Azure Cosmos DB for
NoSQL



Azure Data Explorer
(Kusto)



Azure Data Lake Storage
Gen1



Azure Data Lake Storage
Gen2



Azure Database for
MySQL



Azure Database for
PostgreSQL



Azure SQL Database



Continue

Cancel

Select format

Choose the format type of your data



Avro



DelimitedText



Excel



JSON



ORC



Parquet



XML



Binary

Continue

Back

Cancel

Set properties

Name

DelimitedText1

Linked service *

ADLS_Azure_DataLakeStorage1_Linked_service

**File path**

File system

/ Directory

/ File name

**First row as header****Import schema** From connection/store From sample file None

> Advanced

OK**Back****Cancel**

Browse

Select a file or folder.

Root folder > landing > Ass_8_input_files

- SalesLT.Customer-xRyuaEu2p7-WdCbzuWllo-80V7BfCQtP.txt
- SalesLT.CustomerAddress-nnCI751651 (1)-0A4ckSV5gM-jBPnfa8HWQ.txt
- SalesLT.Product-760RKVH4kL-x8c0uQ9qBG-puI1LFOEyO.txt

Set properties

Name

DelimitedText1

Linked service *

ADLS_Azure_DataLakeStorage1_Linked_service



File path

landing

/ Ass_8_input_files

/ SalesLT.CustomerAddre...



First row as header



Import schema

From connection/store

From sample file

None

> Advanced

Publish all (3)

Assignment_8_Que... **Ass_8_Q2_DataFlow** • **Ass_8_Q1_Customer...**

Validate Data flow debug Debug Settings

0 18

Source settings **Source options** **Projection** **Optimize** **Inspect** **Data preview**

Output stream name * CustomerAddress [Learn more](#)

Description Import data from Ass_8_Q2_CustomerAddress_CS...

Source type * Dataset Inline

Dataset * Ass_8_Q2_CustomerAddress_CS... [Test connection](#) [Open](#) [New](#)

Options Allow schema drift [①](#) Infer drifted column types [①](#) Validate schema [①](#)

Skip line count

Sampling * Enable Disable

Assignment_8_Que... **Ass_8_Q2_DataFlow** • **Ass_8_Q1_Customer...**

Validate Data flow debug Debug Settings

3

Source settings **Source options** **Projection** **Optimize** **Inspect** **Data preview**

Number of rows **INSERT** 100 **UPDATE** 0 **DELETE** 0 **UPSERT** 0 **LOOKUP**

CustomerID **AddressID** **AddressType** **rowguid** **ModifiedDate**

CustomerID	AddressID	AddressType	rowguid	ModifiedDate
29485	1086	Main Office	16765338-dbe4-442...	2007-09-01 00:00:00...
29486	621	Main Office	22b3e910-14af-4ed...	2005-09-01 00:00:00...
29489	1069	Main Office	a095c88b-d7e6-417...	2005-07-01 00:00:00...
29490	887	Main Office	f12e1702-d897-403...	2006-09-01 00:00:00...
29492	618	Main Office	5b3b3eb2-3f43-47e...	2006-12-01 00:00:00...
29494	537	Main Office	492d92b6-31af-47e...	2005-09-01 00:00:00...

Assignment_8_Queue

Ass_8_Q2_DataFlow

Ass_8_Q1_Customer...

Validate Data flow debug Debug Settings

CustomerAddress
Columns: 5 total

Source settings Source options Projection Optimize Inspect Data preview

Number of columns Total 5

Order ↑↓	Column ↑↓	Type ↑↓
1	CustomerID	abc string
2	AddressID	abc string
3	AddressType	abc string
4	rowguid	abc string
5	ModifiedDate	abc string

Publish

Assignment_8_Queue

Ass_8_Q2_DataFlow

Ass_8_Q1_Customer...

Validate Data flow debug Debug Settings

CustomerAddress
Columns: 5 total

Source settings Source options **Projection** Optimize Inspect Data preview

Define default format Detect data type Import projection Reset schema

Column name	Type	Format
CustomerID	abc string	Specify format
AddressID	abc string	Specify format
AddressType	abc string	Specify format
rowguid	abc string	Specify format
ModifiedDate	abc string	Specify format

New to Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Assignment_8_Que... **Ass_8_Q2_DataFlow** **Ass_8_Q1_Customer_...**

✓ Validate **Data flow debug** ✓ [Debug Settings](#)

0 18

Customer
Columns: 15 total

CustomerAddress
Import data from Ass_8_Q1_Customer_CS...

Source settings Source options Projection Optimize Inspect Data preview

Output stream name * Customer Learn more [Customer](#)

Description Import data from Ass_8_Q1_Customer_CS... [Reset](#)

Source type * Dataset [Dataset](#) [Inline](#)

Dataset * Ass_8_Q1_Customer_CS... [Test connection](#) [Open](#) [New](#)

Options Allow schema drift [①](#) Infer drifted column types [①](#) Validate schema [①](#)

2

Customer
Columns: 15 total

Source settings Source options Projection Optimize Inspect Data preview

Number of rows **100** **INSERT 100** **UPDATE 0** **DELETE 0** **UPSERT**

Refresh Typecast Modify Map drifted Statistics Remove Export to CSV

Custom...	NameSt...	Title	FirstName	Middle...	LastName	Suffix	Compan...
1	False	Mr.	Orlando	N.	Gee	NULL	A Bike St..
2	False	Mr.	Keith	NULL	Harris	NULL	Progress...
3	False	Ms.	Donna	F.	Carreras	NULL	Advance...
4	False	Ms.	Janet	M.	Gates	NULL	Modular...
5	False	Mr.	Lucy	NULL	Harrington	NULL	Metropo...
6	False	Ms.	Rosmarie	J.	Carroll	NULL	Aerobic ...
7	False	Mr.	Dominic	P.	Gash	NIJII	Associat...

2. Create a pipeline to join the two files (Customer, Customer Address) based on customer id and save the result as a JSON file.

MICROSOFT FABRIC, A DYNAMIC AND EXCITING WAY TO BUILD CLOUD-FIRST DATA ANALYTICS. [CLICK HERE](#) TO GET STARTED WITH FABRIC

Publish all 3

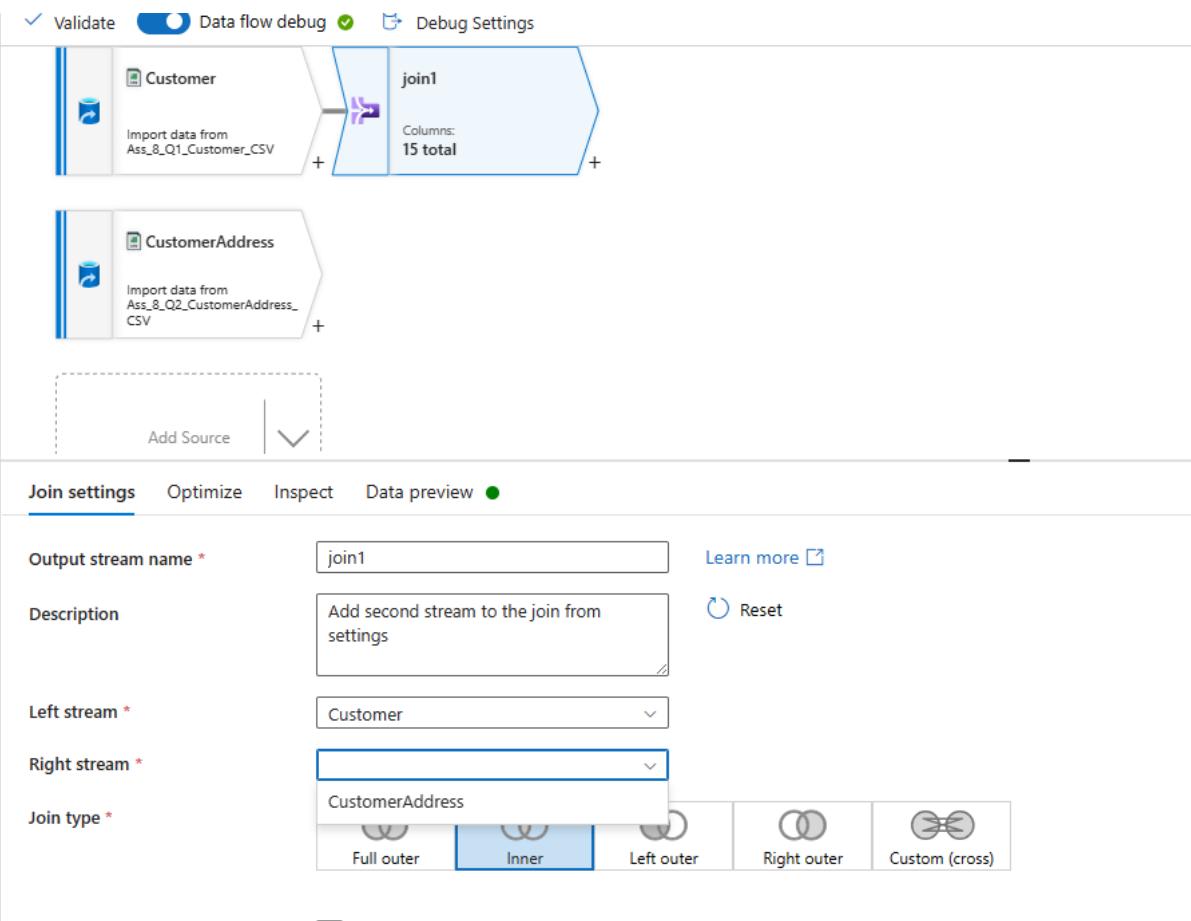
Assignment_8_Que... Ass_8_Q2_DataFlow Ass_8_Q1_Customer_...

✓ Validate Data flow debug Debug Settings

The screenshot shows the Microsoft Fabric Data Flow interface. At the top, there are three tabs: "Assignment_8_Que...", "Ass_8_Q2_DataFlow" (which is selected), and "Ass_8_Q1_Customer...". Below the tabs are buttons for "Validate", "Data flow debug" (which is turned on), and "Debug Settings".
The main area displays two data sources: "Customer" and "CustomerAddress". The "Customer" source has 15 columns. The "CustomerAddress" source is imported from a CSV file named "Ass_8_Q2_CustomerAddress_CSV". A "Join" operation is currently selected, indicated by a plus sign (+) between the two sources.
A sidebar on the right contains a search bar and a list of operations:

- Multiple inputs/outputs
- Join
- Conditional Split
- Exists
- Union
- Lookup

Below this, under "Source settings", there are fields for "Output stream name *" and "Description". Under "Schema modifier", there are options for "Derived Column", "Select", "Aggregate", and "Surrogate Key".
At the bottom right of the sidebar, there are buttons for "Data preview", "Learn more", and "Reset".



Validate Data flow debug Debug Settings

join1
Columns: 20 total

Customer
Import data from Ass_8_Q1_Customer_CSV

CustomerAddress
Import data from Ass_8_Q2_CustomerAddress_CSV

join1
Columns: 20 total

Add Source

Join settings **Optimize** **Inspect** **Data preview**

Output stream name * join1 [Learn more](#)

Description Inner join on 'Customer' and 'CustomerAddress'

Left stream * Customer

Right stream * CustomerAddress

Join type * Full outer Inner Left outer Right outer Custom (cross)

Use fuzzy matching

Join conditions *

Left: Customer's column	Right: CustomerAddress's column		
Select column...	=	Select column...	+ -

Add Source

Join settings **Optimize** **Inspect** **Data preview**

Join type * Full outer Inner Left outer Right outer Custom (cross)

Use fuzzy matching

Join conditions *

Left: Customer's column	Right: CustomerAddress's column
12s CustomerID Filter...	= Select column... + -
12s CustomerID NameStyle Title	Type mismatch

2

Left stream *

Customer

Right stream *

CustomerAddress

Join type *

Full outer Inner Left outer Right outer Custom (cross)

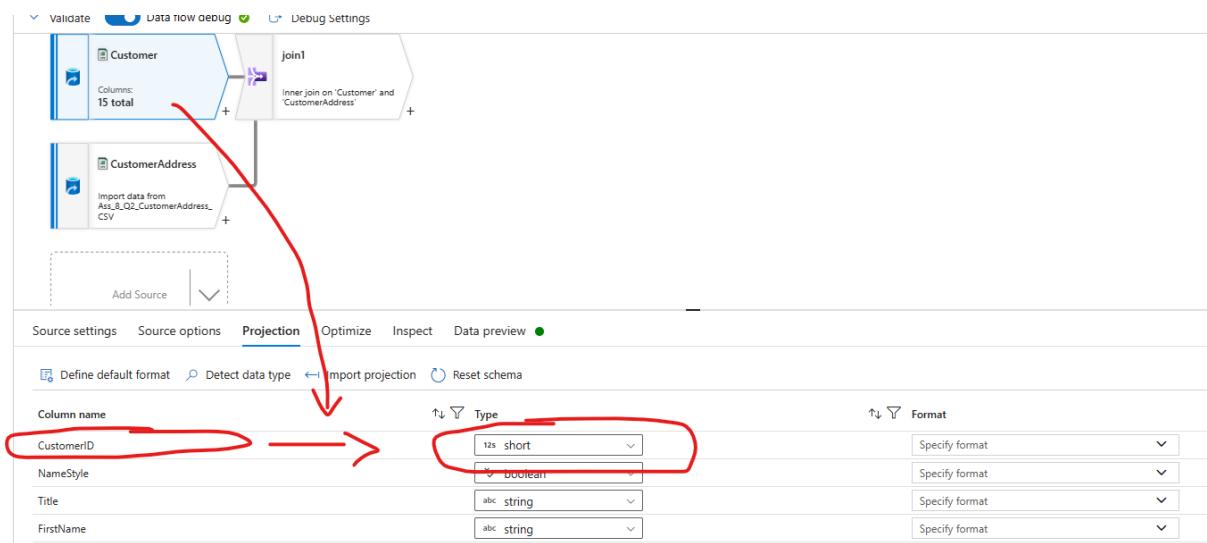
Use fuzzy matching ⓘ

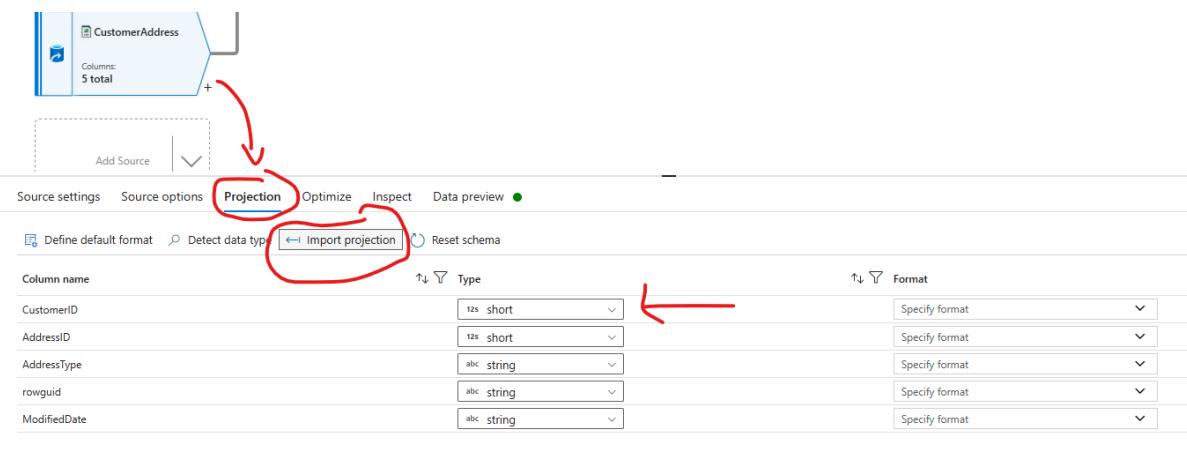
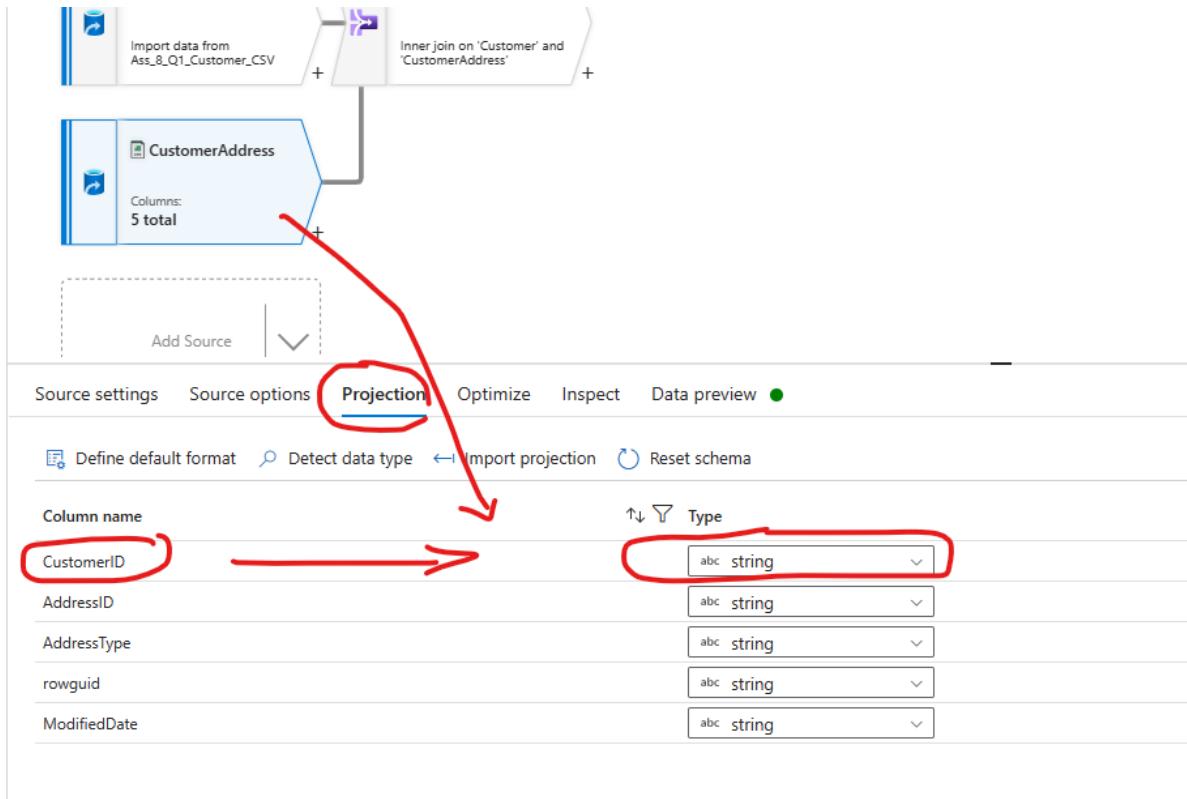
Join conditions *

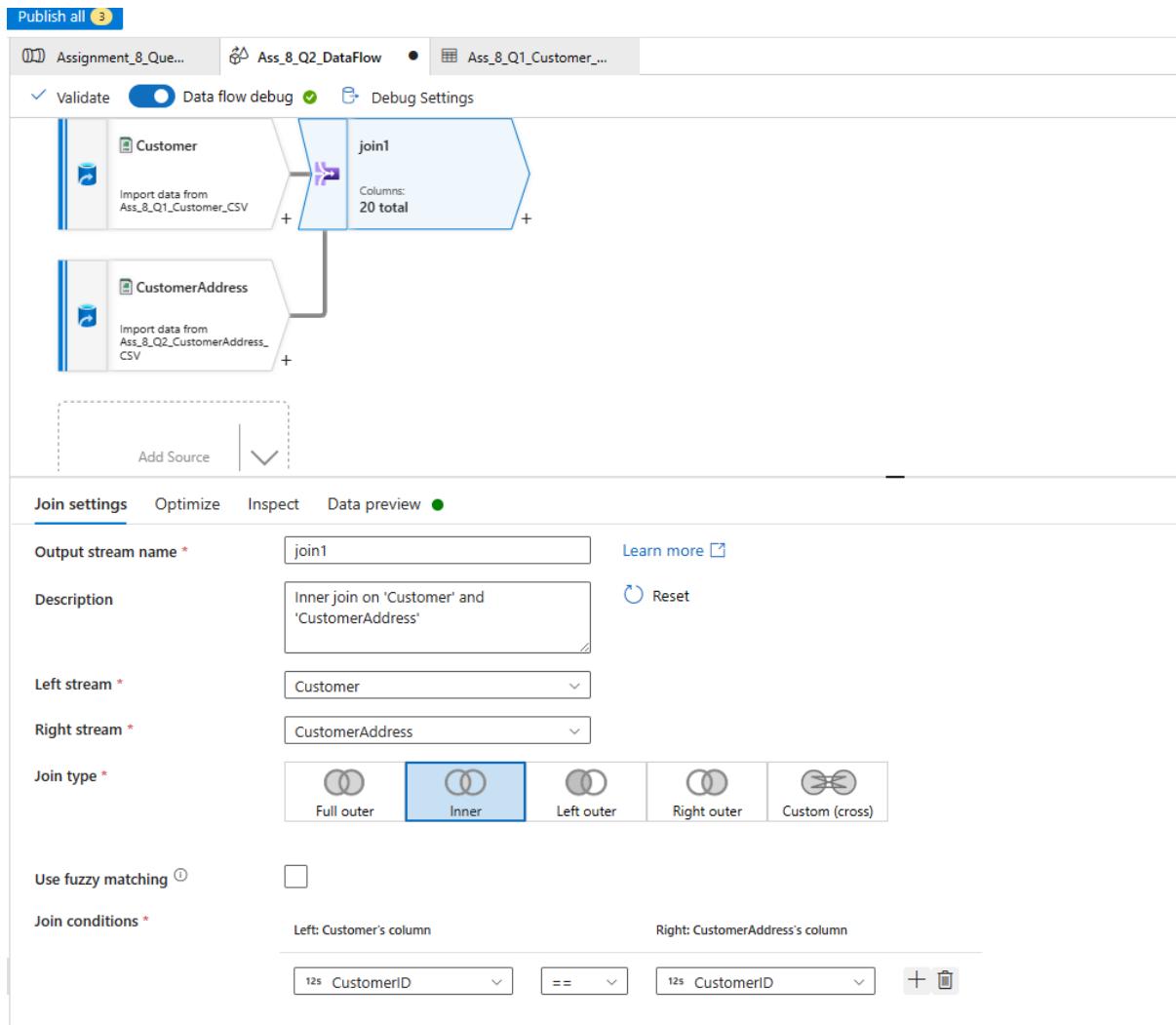
Left: Customer's column Right: CustomerAddress's column

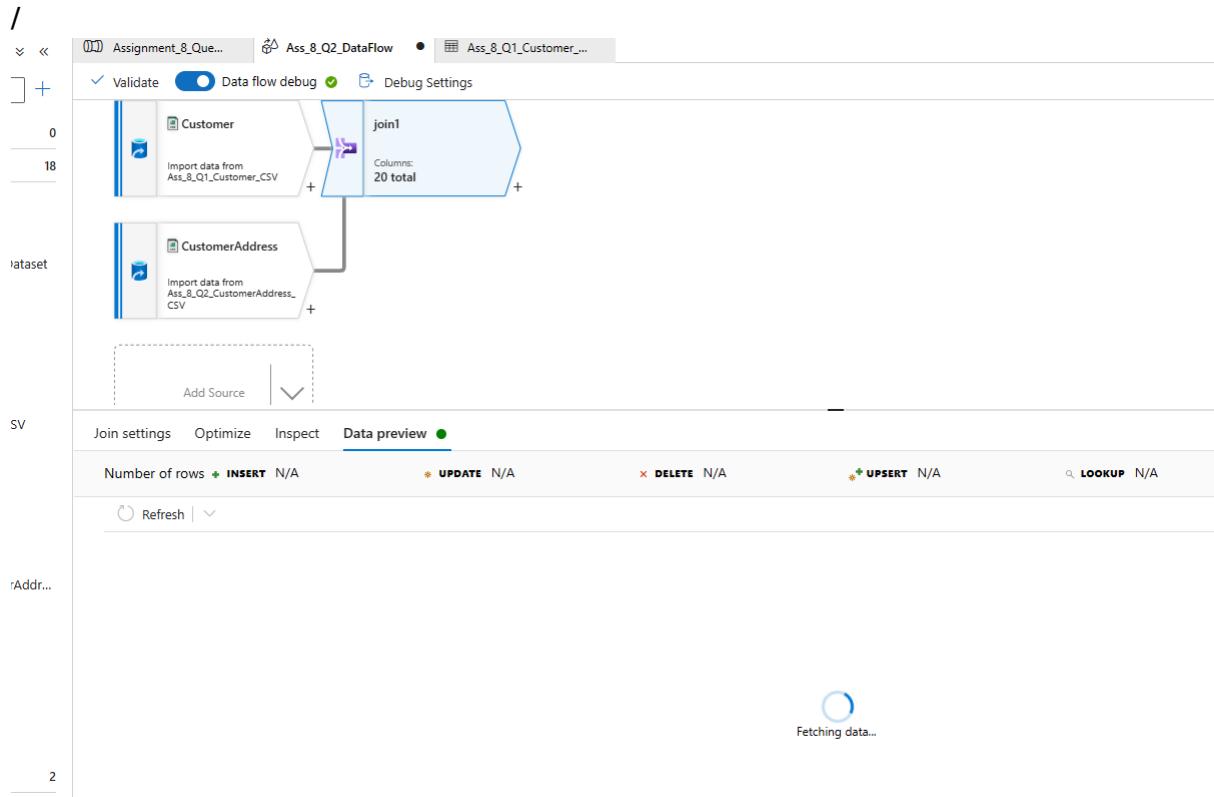
12s CustomerID == abc CustomerID

Type mismatch









Validate Data flow debug Debug Settings

Customer
Import data from Ass_8_Q1_Customer_CSV

CustomerAddress
Import data from Ass_8_Q2_CustomerAddress_CSV

join1
Columns: 20 total

Add Source

Join settings Optimize Inspect Data preview

Number of columns Left: Customer 15 Right: CustomerAddress 5 Total 2

Order ↑↓	Column ↑↓	Type ↑↓	Fed by ↑↓	Used in ↑↓
1	CustomerID	12s short	Customer	✓
2	NameStyle	✗ boolean	Customer	
3	Title	abc string	Customer	
4	FirstName	abc string	Customer	
5	MiddleName	abc string	Customer	
6	LastName	abc string	Customer	
7	Suffix	abc string	Customer	
8	CompanyName	abc string	Customer	
9	SalesPerson	abc string	Customer	
10	EmailAddress	abc string	Customer	

- Create a pipeline to join the two files (Customer and customer id) and save the result as a JSON file.

View of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Assignment_8_Que... **Ass_8_Q2_DataFlow** **Ass_8_Q1_Customer...**

Validate **Data flow debug** **Debug Settings**

The screenshot shows a Data Flow activity named 'join1' with the following configuration:

- Customer** dataset: Import data from 'Ass_8_Q1_Customer_CS1V'.
- CustomerAddress** dataset: Import data from 'Ass_8_Q2_CustomerAddress_CS1V'.
- join1**: Inner join on 'Customer' and 'CustomerAddress'.
- Destination**: Sink.
- Columns: 20 total**

Join settings tab is selected, showing the mapping between columns:

Order ↑	Column ↑	Type ↑	Fed by ↑
1	CustomerID	12s short	Customer
2	NameStyle	x boolean	Customer
3	Title	abc string	Customer
4	FirstName	abc string	Customer
5	MiddleName	abc string	Customer

View of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Assignment_8_Que... **Ass_8_Q2_DataFlow** **Ass_8_Q1_Customer...**

Data flow debug **Debug Settings**

The screenshot shows a Data Flow activity named 'sink1' with the following configuration:

- join1**: Inner join on 'Customer' and 'CustomerAddress'.
- sink1**: Columns: 17 total.
- Save the result in JSON Dataset**

Format dropdown is set to 'Dataset'.

New dataset pane is open, showing a grid of Azure data stores:

All	Azure	Database	File	Generic protocol	NoSQL	Services and apps
Amazon S3	Azure Blob Storage	Azure Cosmos DB for NoSQL				
Azure Data Explorer (Kusto)	Azure Data Lake Storage Gen1	Azure Data Lake Storage Gen2				
Azure Database for MySQL	Azure Database for PostgreSQL	Azure SQL Database				

DEFAULT DIRECTORY

Select format

Choose the format type of your data



Avro



DelimitedText



JSON



ORC



Parquet



Binary

Set properties

Name

Linked service *

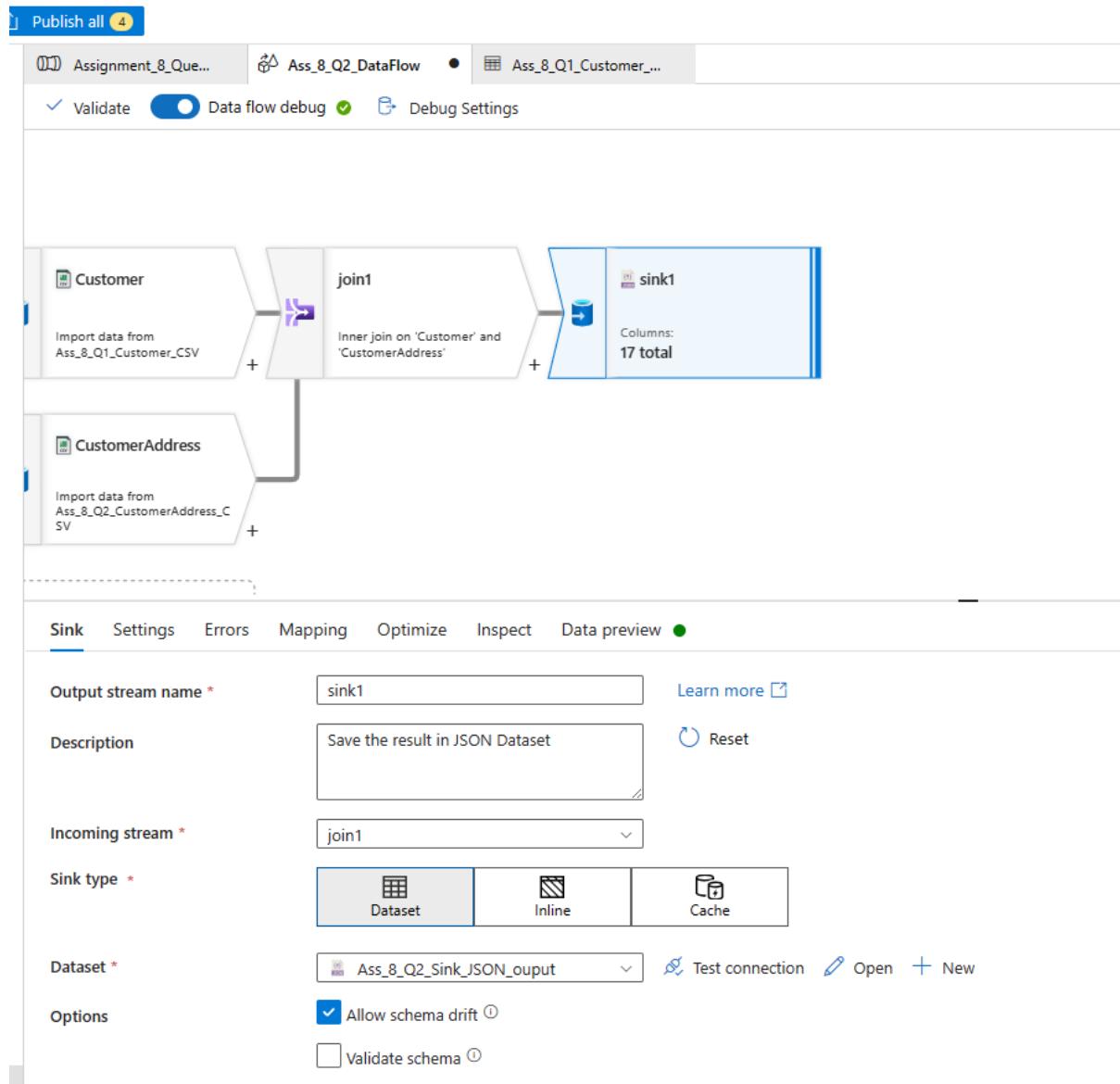
File path

Import schema

- From connection/store From sample file None

> Advanced



Publish all 4

Assignment_8_Que... Ass_8_Q2_DataFlow Ass_8_Q1_Customer_...

Validate Data flow debug Debug Settings

0 19

```

graph LR
    Customer[Customer] --> join1((join1))
    CustomerAddress[CustomerAddress] --> join1
    join1 --> sink1[sink1]
    sink1 --> Output
    subgraph Output [ ]
        direction TB
        C1[Import data from Ass_8_Q1_Customer_CS] --- C2[Customer]
        C3[Import data from Ass_8_Q2_CustomerAddress_CS] --- C4[CustomerAddress]
        C5[Inner join on 'Customer' and 'CustomerAddress'] --- C6[join1]
        C7[Columns: 17 total] --- C8[sink1]
    end

```

Sink Settings Errors Mapping Optimize Inspect Data preview

Sink

Output stream name * sink1 Learn more

Description Save the result in JSON Dataset Reset

Incoming stream * join1

Sink type * Dataset Inline Cache

Dataset * Ass_8_Q2_Sink_JSON_ouput Testing connection... Open New

Options Allow schema drift ⓘ Validate schema ⓘ

2

19

```

graph LR
    Customer[Customer] --> join1((join1))
    CustomerAddress[CustomerAddress] --> join1
    join1 --> sink1[sink1]
    sink1 --> Output
    subgraph Output [ ]
        direction TB
        C1[Import data from Ass_8_Q1_Customer_CS] --- C2[Customer]
        C3[Import data from Ass_8_Q2_CustomerAddress_CS] --- C4[CustomerAddress]
        C5[Inner join on 'Customer' and 'CustomerAddress'] --- C6[join1]
        C7[Columns: 17 total] --- C8[sink1]
    end

```

Sink Settings Errors Mapping Optimize Inspect Data preview

Number of rows + INSERT N/A * UPDATE N/A × DELETE N/A * UPD N/A

Custom...	NameSt...	Title	FirstName	MiddleN...	LastName	Suffix	Compan...	SalesPer...	EmailAd...	Phone
29485	X	Ms.	Catherine	R.	Abel	NULL	Professio...	adventur...	catherin...	747-555...
29486	X	Ms.	Kim	NULL	Abercro...	NULL	Riders C...	adventur...	kim2@a...	334-555...
29489	X	Ms.	Frances	B.	Adams	NULL	Area Bik...	adventur...	frances0...	991-555...
29490	X	Ms.	Margaret	J.	Smith	NULL	Bicycle A...	adventur...	margaret...	959-555...
29492	X	Mr.	Jay	NULL	Adams	NULL	Valley Bi...	adventur...	iav1@ad...	158-555...
29494	X	Mr.	Samuel	N.	Adcaoil	NULL	Vinyl an...	adventur...	samuel0...	554-555...
29496	X	Mr.	Robert	E.	Ahlering	NULL	Fun Toys...	adventur...	robert1...	678-555...

2

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Preview experience Off

Factory Resources Activities Properties

Pipelines

- Ass_5_3_JOIN_Ingestion_Customer
- Ass_6_3_Foreach_Example_2
- Ass_6_4_Foreach_Example_2
- Ass_6_5_Cust_YYYY_MM_DD
- Ass_7_1_Get_MetaData_Example
- Ass_7_4_Table_HighWaterMarkFile
- Ass_7_ques_3
- Ass_7_ques_5_Foreach_Example2

Activities

- Pipeline
- Change Data Capture (preview)
- Dataset
- Data flow
- Power Query
- Copy Data tool

Properties

General Related

Name * pipeline1

Description

Annotations + New

Import from pipeline template

Preview experience Off

Trigger Data flow debug

Ass_8_Q2_Sink_JSON... Ass_8_Q2_Pipeline...

Properties

General Related

Name * Ass_8_Q2_Pipeline_Dataflow

Description Ass_8_Q2_Pipeline_Dataflow
Calls Dataflow to join Cust & Custom

Annotations + New

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory Validate all Publish all Preview experience

Factory Resources

- Assignment_6_part6_threshold_file
- Assignment_8_Ques_1
- Copy_Customer_JSON_To_Folder
- Copy_Customer_Table_To_CSV
- Copy_Customer_To_CSV_Pipe
- Copy_Customer_To_JSON
- Copy_ProductTable_To_CSV
- Ingestion_Customer_SQLDB_ADLS
- Ingestion_Customer_SQLDB_ADLS_F...
- Ingestion_Product_To_JSON
- Ass_8_Q2_Pipeline_Dataflow
- Without_Foreach_Example_Pipeline
- Change Data Capture (preview) 0
- Datasets 19

Activities

- Move and transform
- Copy data
- Data flow
- Azure Data Explorer
- Azure Data Explorer C...
- Databricks
- Notebook
- Jar
- Python
- Data Lake Analytics
- U-SQL

Data flow

Data flow1

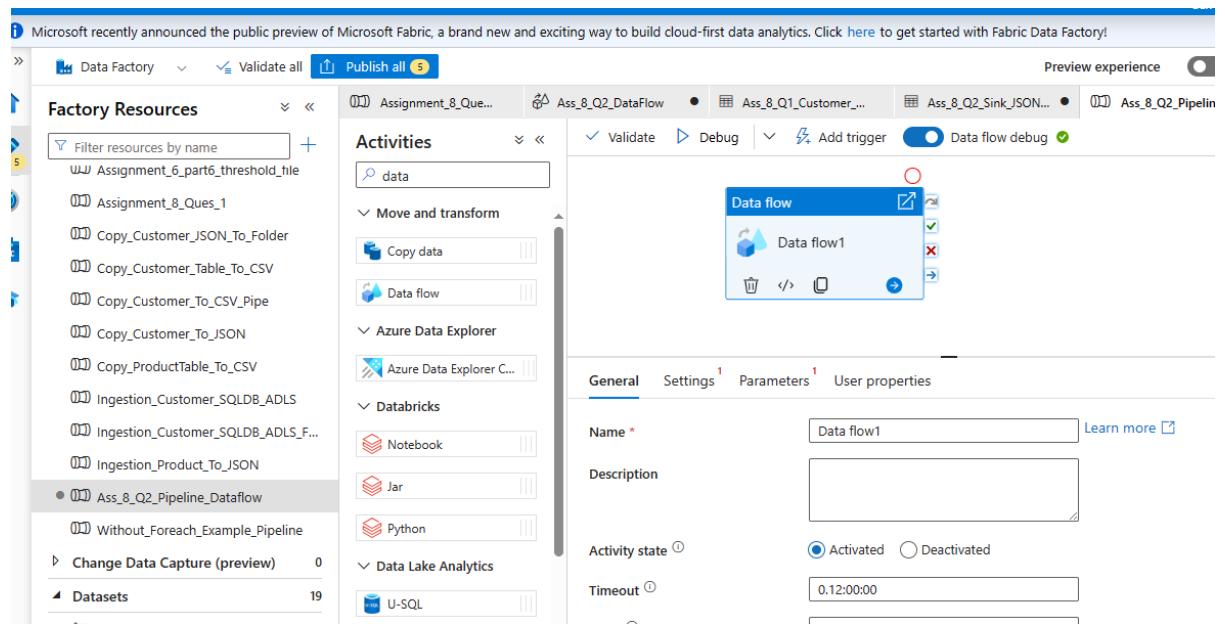
General Settings Parameters User properties

Name * Data flow1

Description

Activity state Activated Deactivated

Timeout 0:12:00:00



Ass_8_Q2_DataFlow Ass_8_Q1_Customer... Ass_8_Q2_Sink_JS

Validate Debug Add trigger Data flow d

Data flow

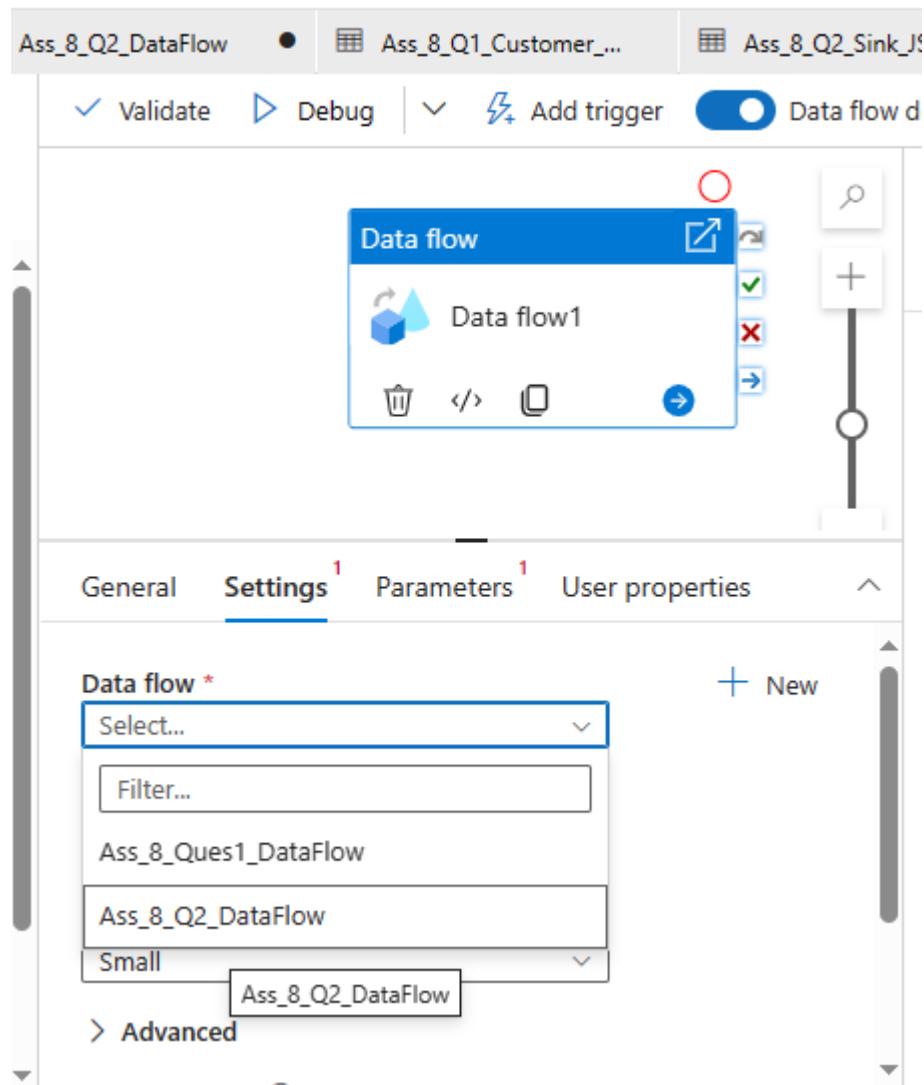
Data flow1

General Settings Parameters User properties

Data flow *

Select... Filter... Ass_8_Ques1_DataFlow Ass_8_Q2_DataFlow Small Ass_8_Q2_DataFlow

New > Advanced



Preview experience

Ass_8_Q2_DataFlow • Ass_8_Q1_Customer_... • Ass_8_Q2_Sink_JSON... • Ass_8_Q2_Pipeline...

✓ Validate ▶ Debug ▾ ⚡ Add trigger Data flow debug ✓ {

Data flow

Data flow1

Data flow * Ass_8_Q2_DataFlow

Run on (Azure IR) * AutoResolveIntegrationRuntime

Compute size * Small

> Advanced

Logging level * Verbose Basic None

> Sink properties

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory Preview experience

Factory Resources +

- Assignment_6_part6_threshold_file
- Assignment_8_Ques_1
- Copy_Customer_JSON_To_Folder
- Copy_Customer_Table_To_CSV
- Copy_Customer_To_CSV_Pipe
- Copy_Customer_To_JSON
- Copy_ProductTable_To_CSV
- Ingestion_Customer_SQLDB_ADLS
- Ingestion_Customer_SQLDB_ADLS_F...
- Ingestion_Product_To_JSON
- Ass_8_Q2_Pipeline_Dataflow
- Without_Foreach_Example_Pipeline

Activities

Ass_8_Q2_DataFlow • Ass_8_Q1_Customer_... • Ass_8_Q2_Sink_JSON... • Ass_8_Q2_Pipeline...

✓ Validate ▶ Debug ▾ ⚡ Add trigger Data flow debug ✓ {

Data flow

Dataflow Ass 8 Q2

General Settings Parameters User properties

Name * Dataflow Ass_8_Q2_...

Description Calls Dataflow Ass_8_Q2_Dataflow to join Cust & Custom Address

Activity state Activated Deactivated

Timeout 0:12:00:00

To run Dataflow,

- 1) we need to create a Pipeline and call DataFlow from Pipeline
- 2) This will generate the output for us running JOINS

The screenshot shows the Azure Data Factory pipeline editor interface. On the left, the 'Factory Resources' sidebar lists various pipelines and datasets. The 'Ass_8_Q2_Pipeline_Dataflow' pipeline is currently selected. In the center, the 'Activities' tab is open, displaying a list of available activities: Move and transform (Copy data, Data flow), Azure Data Explorer, Databricks (Notebook, Jar, Python), Data Lake Analytics (U-SQL), and General. A specific 'Data flow' activity is highlighted with a red circle around its icon. Above the activities list, a toolbar includes buttons for Validate, Debug (which is circled in red), Add trigger, and Data flow debug. On the right, the 'General' tab of the pipeline configuration is visible, showing the pipeline's name ('Dataflow Ass_8_Q2_'), description ('Calls Dataflow Ass_8_Q2_Dataflow to join Cust & Custom Address'), activity state (Activated), timeout (0.12:00:00), and retry settings.

d new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Preview experience

The screenshot shows the Azure Data Factory preview experience. At the top, there are tabs for 'Validate', 'Debug' (which is selected), and 'Add trigger'. A red arrow points from the 'Validate' tab to the 'Debug' tab. Below the tabs is a 'Data flow' card with a blue header containing the name 'Ass_8_Q2_DataFlow'. The card has icons for trash, copy, and run. A red circle highlights the entire card. To the right of the card is a vertical toolbar with various icons. The main area is divided into sections: 'General', 'Settings' (which is selected), 'Parameters', and 'User properties'. Under 'Settings', the 'Data flow' dropdown is set to 'Ass_8_Q2_DataFlow' (highlighted by a red circle). The 'Run on (Azure IR)' dropdown is set to 'AutoResc Ass_8_Q2_DataFlow'. The 'Compute size' dropdown is set to 'Small'. There are also sections for 'Advanced', 'Logging level' (set to 'Verbose'), and 'Sink properties'.

OUTPUT JSON generated in ADLS container

The screenshot shows the Azure Storage Explorer interface. At the top, there are buttons for 'Upload', 'Add Directory', 'Refresh', 'Rename', 'Delete', 'Change tier', 'Acquire lease', 'Break lease', and 'Give feedback'. Below the buttons, it says 'Authentication method: Access key ([Switch to Microsoft Entra user account](#))' and 'Location: landing / Ass_8_output_files'. A red arrow points from the 'Access key' link to the 'Location' text. There is a search bar for 'Search blobs by prefix (case-sensitive)'. A red circle highlights the 'Search blobs by prefix' input field. Below the search bar is a table with columns: 'Name', 'Modified', 'Access tier', and 'Archive status'. The table contains two rows: '_SUCCESS' (modified 3/2/2024, 12:57:44 AM, Hot (Inferred)) and 'part-00000-a90c8241-fe0e-4558-91b6-062da00cca46-c000.json' (modified 3/2/2024, 12:57:43 AM, Hot (Inferred)). The second row is highlighted with a red circle.

s > gopi01012024adls1 | Containers >

blobs

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: [landing](#) / Ass_8_output_files

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> _SUCCESS	3/2/2024, 12:57:44 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> part-00000-a90c8241-fe0e-4558-97b6-062da00cca46-c...	3/2/2024, 12:57:43 AM	Hot (Inferred)		Block blob

Ass_8_output_files/part-00000-a90c8241-fe0e-4558-97...

Blob

 Save Discard Download Refresh | Delete

Overview Versions Edit Generate SAS

```
1 [{"CustomerID":29485,"NameStyle":false,"Title":"Ms.", "FirstName":"Catherine", "LastName":"Fleming"},  
2 {"CustomerID":29486,"NameStyle":false,"Title":"Ms.", "FirstName":"Kim", "LastName":"Lynch"},  
3 {"CustomerID":29489,"NameStyle":false,"Title":"Ms.", "FirstName":"Frances", "MiddleName":"Mae"},  
4 {"CustomerID":29490,"NameStyle":false,"Title":"Ms.", "FirstName":"Margaret", "MiddleName":"Mae"},  
5 {"CustomerID":29492,"NameStyle":false,"Title":"Mr.", "FirstName":"Jay", "LastName":"Wong"},  
6 {"CustomerID":29494,"NameStyle":false,"Title":"Mr.", "FirstName":"Samuel", "MiddleName":"Sam"},  
7 {"CustomerID":29496,"NameStyle":false,"Title":"Mr.", "FirstName":"Robert", "MiddleName":"Robert"},  
8 {"CustomerID":29497,"NameStyle":false,"Title":"Mr.", "FirstName":"François", "LastName":"Lavigne"},  
9 {"CustomerID":29499,"NameStyle":false,"Title":"Ms.", "FirstName":"Amy", "MiddleName":"Amy"},  
10 {"CustomerID":29502,"NameStyle":false,"Title":"Mr.", "FirstName":"Paul", "MiddleName":"Paul"},  
11 {"CustomerID":29503,"NameStyle":false,"Title":"Mr.", "FirstName":"Gregory", "MiddleName":"Gregory"},  
12 {"CustomerID":29503,"NameStyle":false,"Title":"Mr.", "FirstName":"Gregory", "MiddleName":"Gregory"},  
13 {"CustomerID":29505,"NameStyle":false,"Title":"Ms.", "FirstName":"Michelle", "MiddleName":"Michelle"},  
14 {"CustomerID":29506,"NameStyle":false,"Title":"Mr.", "FirstName":"Sean", "MiddleName":"Sean"},  
15 {"CustomerID":29508,"NameStyle":false,"Title":"Mr.", "FirstName":"Marvin", "MiddleName":"Marvin"},  
16 {"CustomerID":29510,"NameStyle":false,"Title":"Mr.", "FirstName":"Cecil", "MiddleName":"Cecil"}]
```

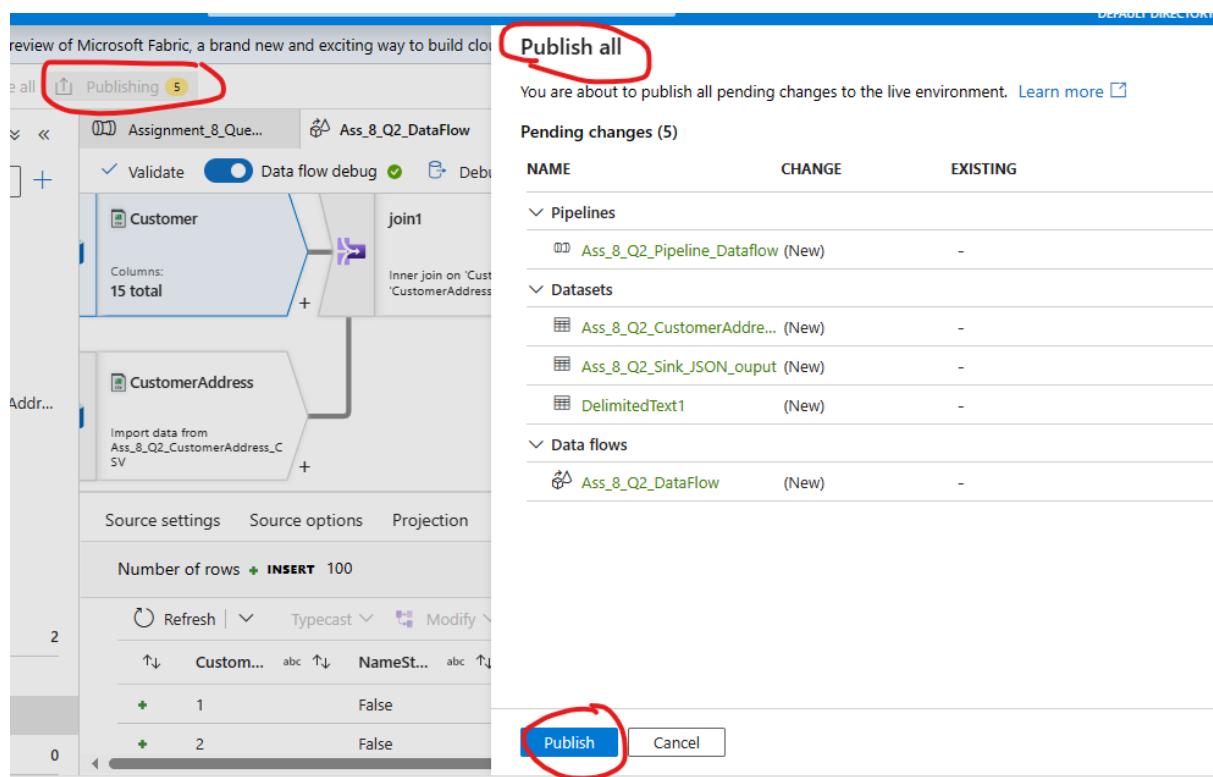
Json

▽

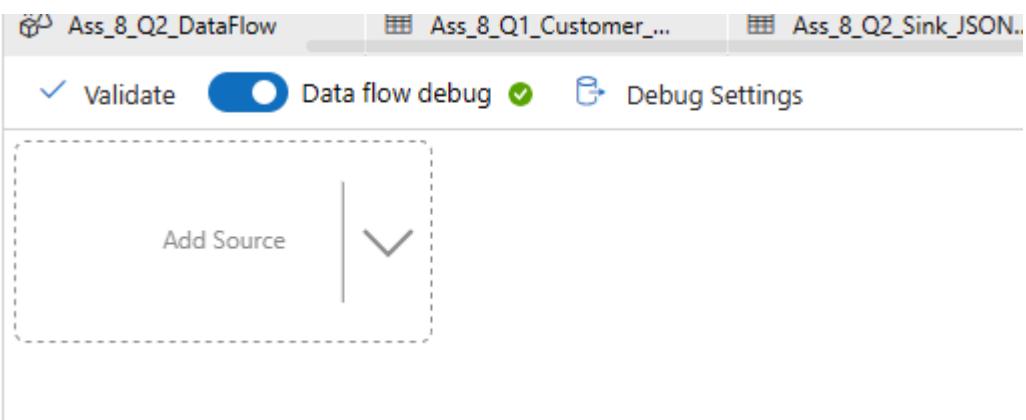
 Preview

For Saving the Pipeline – Do not forget to Publish it

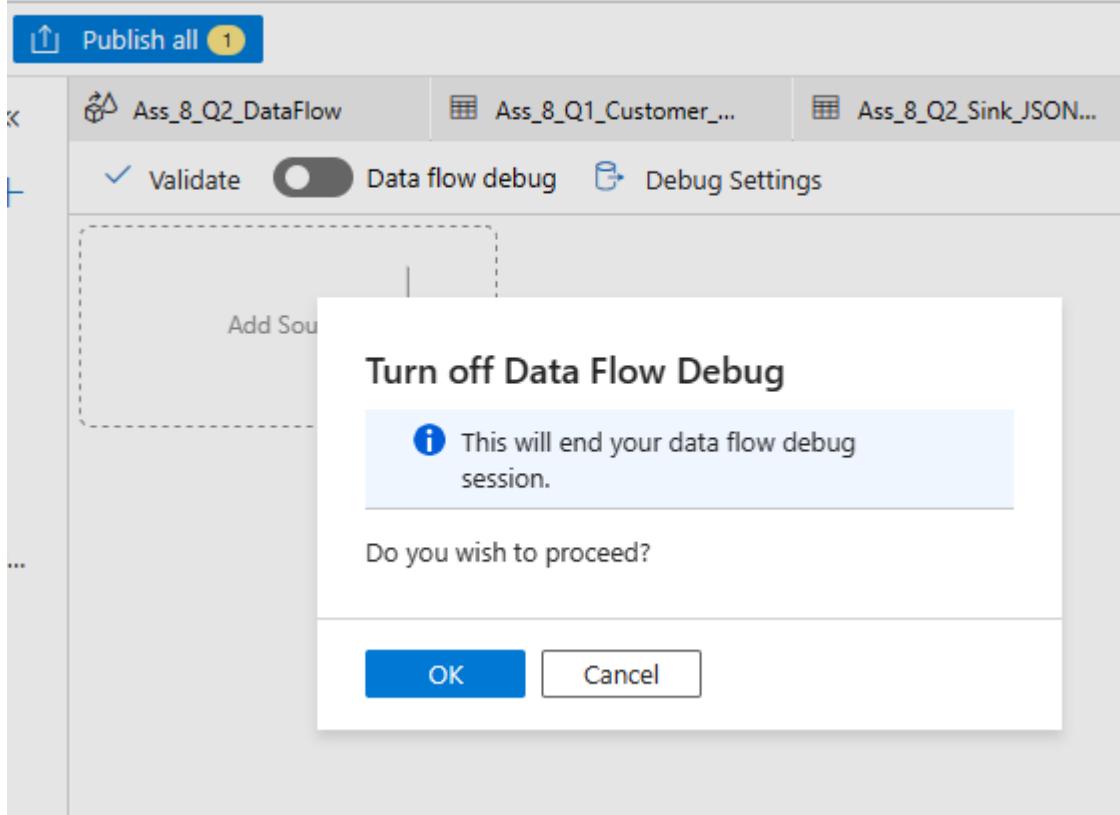
Publish All – will save all the output



Finally to save costs turn off your “Data flow Debug” mode



of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#)



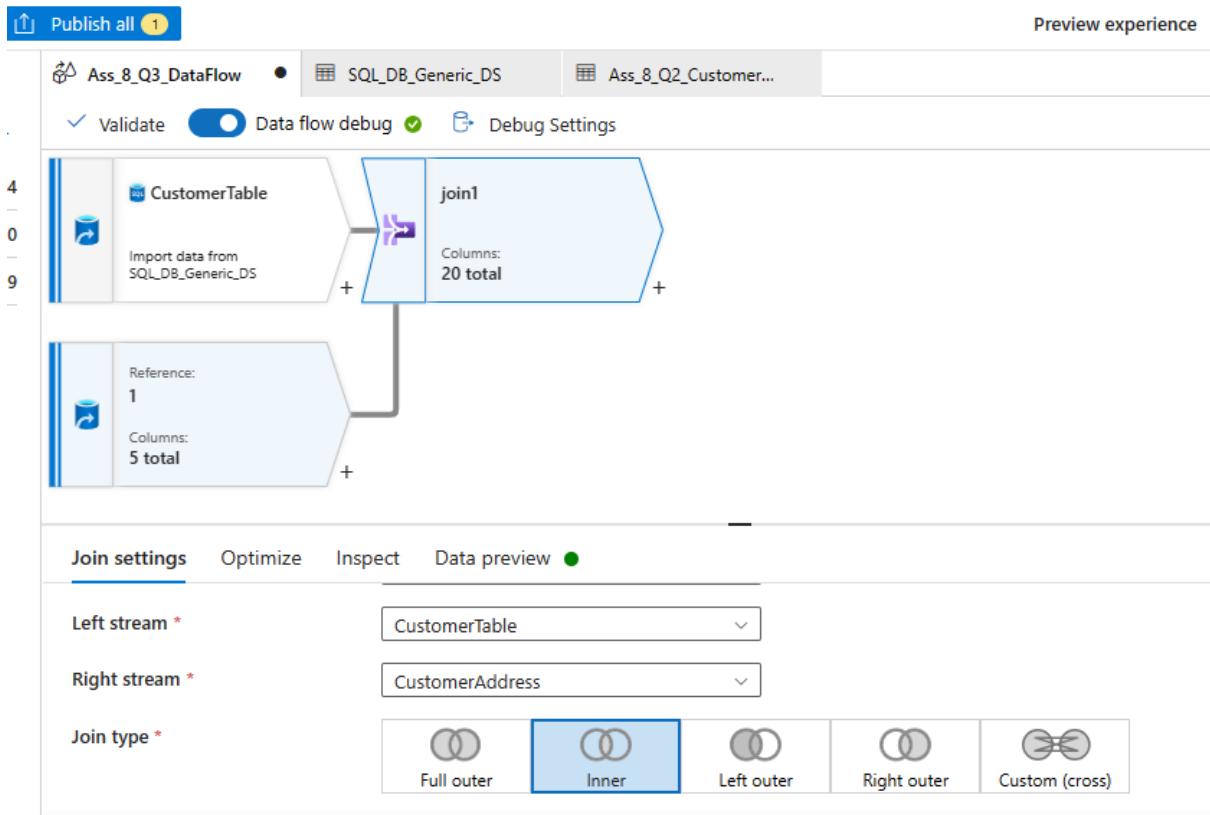
Question 3:

3. Create a pipeline to read the Customer table data from SQL and CustomerAddress data from CSV, join both of them, and then save the result where customer id > 1000 & Customer id < 2000 in ascending order as a Parquet file.

Solution 3:

Same as Q1 & Q2 , extra thing is after join, use “Filter”

customer id > 1000 & Customer id < 2000 in ascending order as a Parquet file.



Publish all 1

Ass_8_Q3_DataFlow • SQL_DB_Generic_DS Ass_8_Q2_Customer...

Validate Data flow debug Debug Settings

CustomerTable
Columns: 15 total

join1
Inner join on 'CustomerTable' and 'CustomerAddress'

CustomerAddress

Source settings Source options Projection Optimize Inspect Data preview

Description Import data from SQL_DB_Generic_DS Reset

Source type * Dataset Inline

Dataset * SQL_DB_Generic_DS Test connection Open New

Options Allow schema drift ① Infer drifted column types ①

Publish all 1

Ass_8_Q3_DataFlow • SQL_DB_Generic_DS Ass_8_Q2_Customer...

Validate Data flow debug Debug Settings

CustomerTable
Import data from SQL_DB_Generic_DS

join1
Inner join on 'CustomerTable' and 'CustomerAddress'

CustomerAddress

Source settings Source options Projection Optimize Inspect Data preview

Output stream name * CustomerAddress Learn more

Description Import data from Ass_8_Q2_CustomerAddress_CSV Reset

Source type * Dataset Inline

Dataset * Ass_8_Q2_CustomerAddress_CSV Test connection Open New

✓ Connection successful

all Publish all 1 Preview experience

Ass_8_Q3_DataFlow • SQL_DB_Generic_DS Ass_8_Q2_Customer...

✓ Validate Data flow debug Debug Settings

CustomerTable
Import data from SQL_DB_Generic_DS

join1
Columns: 20 total

Join settings Optimize Inspect Data preview

Output stream name *

Description

Left stream *

Right stream *

Join type *

Inner

all Publish all 1 Preview experience

Ass_8_Q3_DataFlow • SQL_DB_Generic_DS Ass_8_Q2_Customer...

✓ Validate Data flow debug Debug Settings

CustomerTable
Import data from SQL_DB_Generic_DS

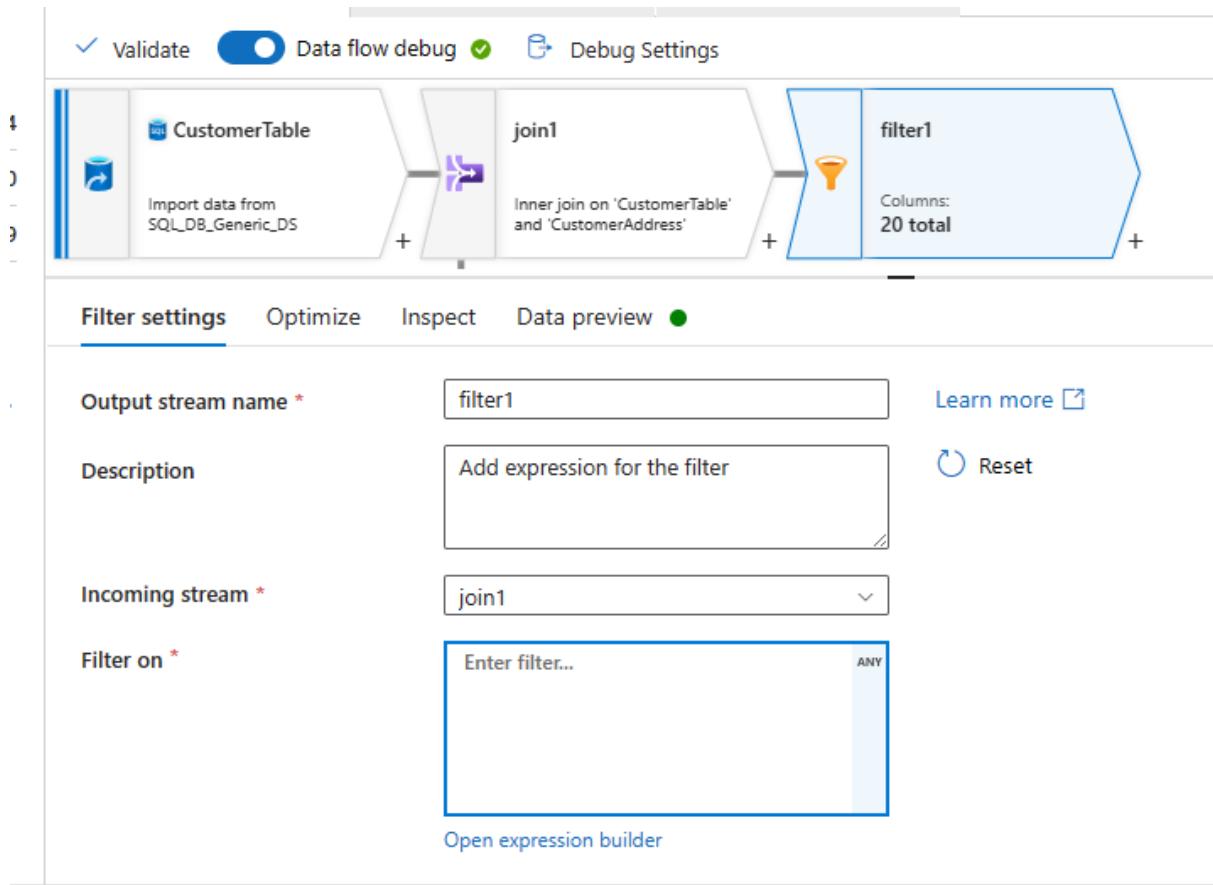
join1
Columns: 20 total

Join settings Optimize Inspect Data preview

Output stream name *

Description

Row modifier



Dataflow expression builder

filter1

Expression

`CustomerTable@CustomerID > 1000 && CustomerTable@CustomerID < 2000`

+ - * / || && ! ^ == === <=> !=

Expression elements

- All
- Functions
- Input schema
- Parameters
- Cached lookup
- Data flow library functions

Expression values

-
-
- 123 CustomerTable@CustomerID
- x NameStyle
- abc Title

Data preview

Input schema
Parameters
Cached lookup
Data flow library functions

123 CustomerTable@CustomerID
x✓ NameStyle
abc Title

Data preview Refresh

Save and finish **Cancel** **Clear contents**

CustomerTable@CustomerID

Data preview Refresh

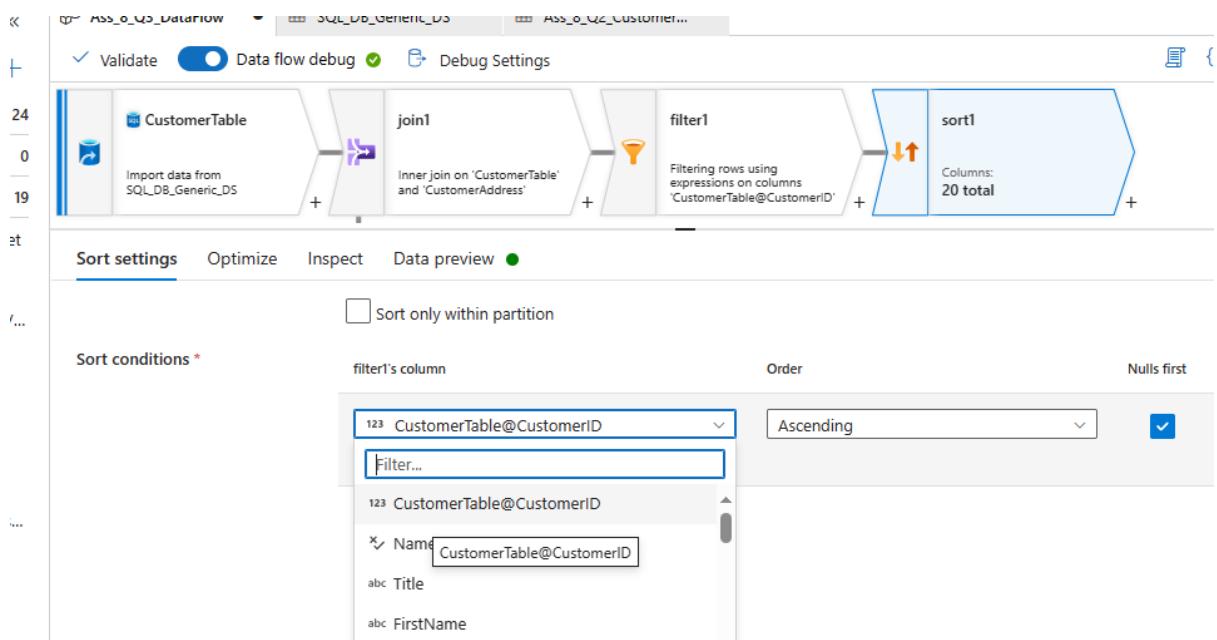
Output: x✓ CustomerTable@CustomerID 123

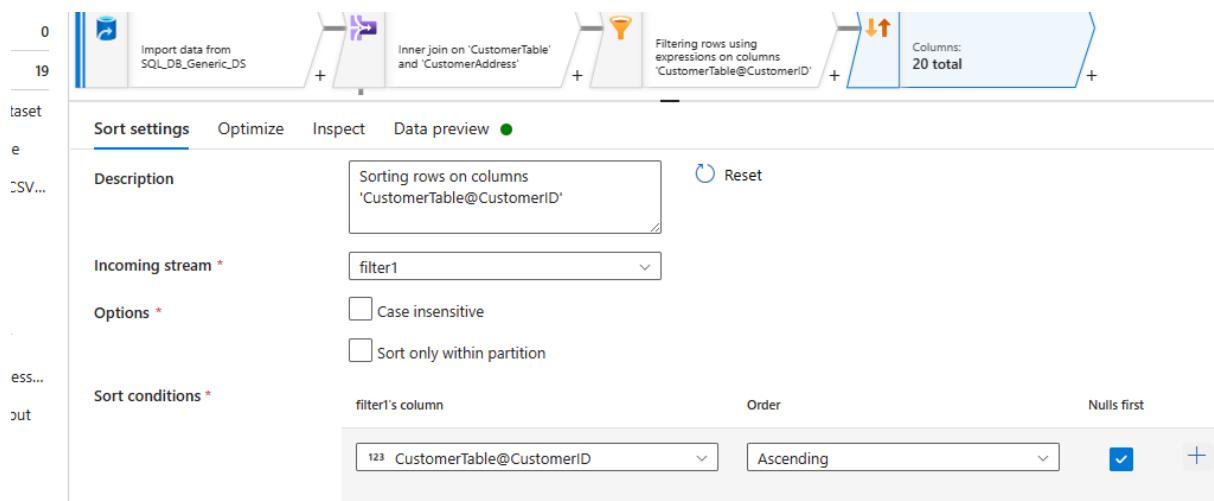
x	29485
x	29486
x	29489
x	29490

Save and finish **Cancel** **Clear contents**

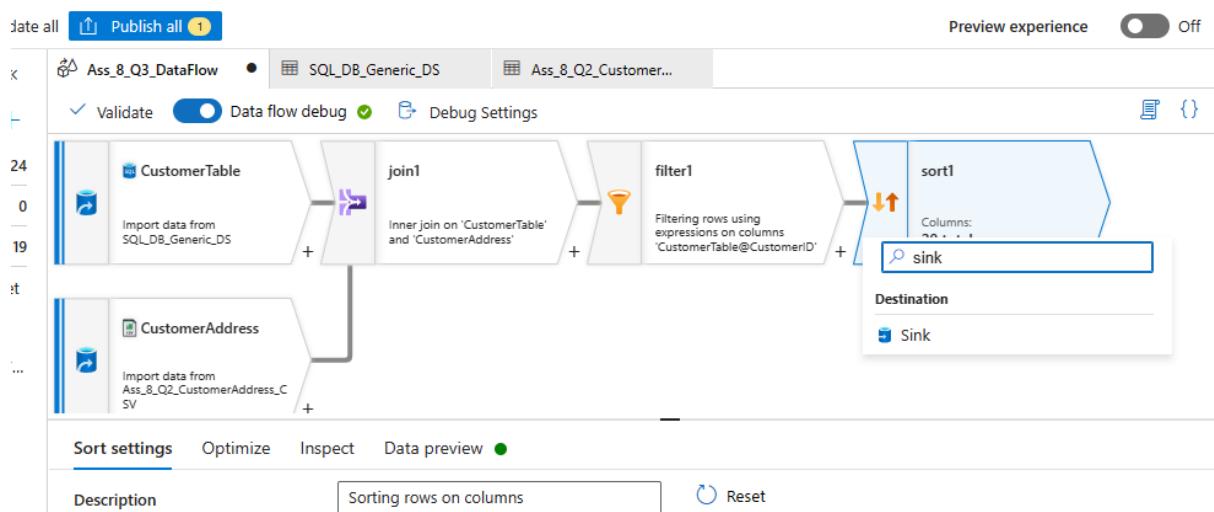
Next Step is ORDER BY → so use “Sort”

data from CSV, join l and then save the result where customer id> 1000 & Customer id <2000 in ascending order as a Parquet file.





Save it as Parquet file: So use “Sink”



New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

All Azure Database File Generic protocol NoSQL Services and apps



Amazon S3



Azure Blob Storage



Azure Cosmos DB for
NoSQL



Azure Data Explorer
(Kusto)



Azure Data Lake Storage
Gen1



Azure Data Lake Storage
Gen2

Continue

Cancel

Select format

Choose the format type of your data



Avro



DelimitedText



JSON



ORC



Parquet



Binary

[Continue](#)[Back](#)[Cancel](#)

DEFAULT DIRECTORY

Set properties

Name
Ass_8_Q3_Customer_Parquet_file

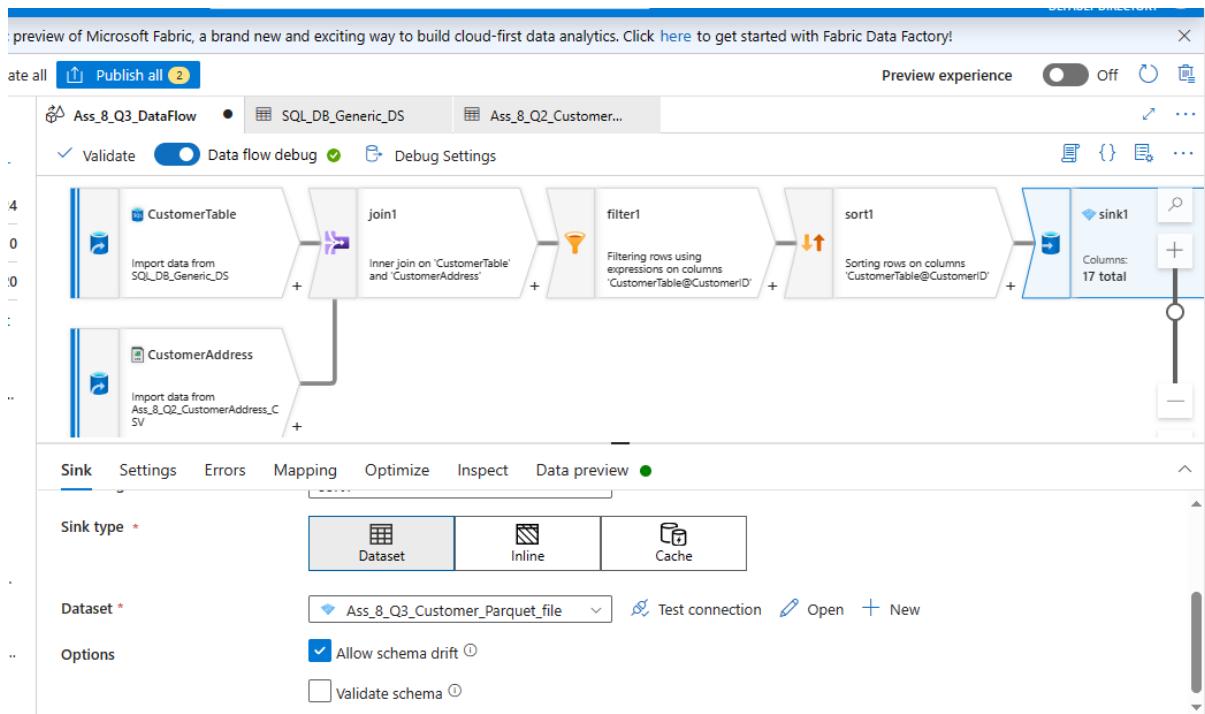
Linked service *
ADLS_Azure_DataLakeStorage1_Linked_service

File path
landing / Ass_8_output_files / File name [edit]

Import schema
 From connection/store From sample file None

> Advanced

OK Back Cancel



Lastly I create one pipeline which will call this “Ass_8_Q3_Dataflow”

Find and exciting way to build cloud first data analysis click [here](#) to get started with Azure Data Factory.

Preview experience Off

SQL_DB_Generic_DS Ass_8_Q2_Customer... Ass_8_Q3_Pipeline... ●

« Validate Debug Add trigger Data flow debug

Validate the current resource

Data flow Ass_8_Q3_Dataflow

General Settings Parameters User properties

Data flow * Ass_8_Q3_DataFlow

CustomerTable parameters

Name	Value	Type
schemaName	Value	string
tableName	Value	string

Run on (Azure IR) *

Compute size *

17:1

The screenshot shows the Azure Data Factory interface for managing pipelines. At the top, there are tabs for 'General', 'Settings' (which is currently selected), 'Parameters' (with a red exclamation mark), and 'User properties'. Under the 'Settings' tab, the 'Data flow' dropdown is set to 'Ass_8_Q3_DataFlow'. Below this, under 'CustomerTable parameters', there are two entries: 'schemaName' and 'tableName', both set to 'Value'. The 'Run on' dropdown is set to 'AutoResolveIntegrationRuntime' and the 'Compute size' dropdown is set to 'Small'. A red circle is drawn around the 'Data flow' tab in the top navigation bar.

Search factory and documentation

nd new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

DEFAULT DIRECTORY

Preview experience Off

SQL_DB_Generic_DS Ass_8_Q2_Customer... Ass_8_Q3_Pipeline... ●

Validate Cancel options Add trigger Data flow debug

Data flow Ass_8_Q3_Dataflow

Parameters Variables Settings Output

Pipeline run ID: b7e11b9e-24a5-4dfc-9a5d-ce98eeb3be91 Pipeline status In progress

All status Monitor in Azure Metrics Export to CSV

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration
Ass_8_Q3_Dataflow	In progress	Data flow	3/2/2024, 5:12:13 PM	2s

and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Preview experience Off

SQL_DB_Generic_DS Ass_8_Q2_Customer... Ass_8_Q3_Pipeline... ●

Validate Debug Add trigger Data flow debug

Validate the current resource

Data flow Ass_8_Q3_Dataflow

Parameters Variables Settings Output

Pipeline run ID: b7e11b9e-24a5-4dfc-9a5d-ce98eeb3be91 Pipeline status Succeeded View debug run consumption

All status Monitor in Azure Metrics Export to CSV

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration
Ass_8_Q3_Dataflow	Succeeded	Data flow	3/2/2024, 5:12:13 PM	50s

landers > landing >

Ass_8_output_files/part-00000-c9a42b9a-dbf7-4ae8-a...

blob

Upload Add Directory ...

Authentication method: Access key ([Switch to Microsoft Entra user account](#))

Location: landing / Ass_8_output_files

Search blobs by prefix (case-sensitive)

Show deleted objects

Name

	Name
<input type="checkbox"/>	_SUCCESS
<input type="checkbox"/>	part-00000-a90c8241-fe0e-4558-97b6-062da00cca46-c000.json
<input checked="" type="checkbox"/>	part-00000-c9a42b9a-dbf7-4ae8-a6ff-6e0b65874b5d-c000.snappy.parquet

Save Discard Download Refresh Delete

Overview Versions Edit Generate SAS

⚠ The file 'Ass_8_output_files/part-00000-c9a42b9a-dbf7-4ae8-a6ff-6e0b65874b5d-c000.snappy.parquet' may not render correctly as it contains an unrecognized extension.

1 PAR1%20%20%20H%20spark_schema%20%20%20%20
2 CustomerID%20%20%20NameStyle%20%20%20Title%20%20%20FirstName%20%20%20
3 MiddleName%20%20%20LastName%20%20%20Suffix%20%20%20CompanyName%20%20%20

Upload Add Directory Refresh Rename Delete Change tier Acquire lease

Authentication method: Access key ([Switch to Microsoft Entra user account](#))

Location: landing / Ass_8_output_files

Search blobs by prefix (case-sensitive)

Name	Modified	Access
_SUCCESS	3/2/2024, 5:12:23 PM	Hot
part-00000-a90c8241-fe0e-4558-97b6-062da00cca46-c000.json	3/2/2024, 12:57:43 AM	Hot
<input checked="" type="checkbox"/> part-00000-c9a42b9a-dbf7-4ae8-a6ff-6e0b65874b5d-c000.snappy.parquet	3/2/2024, 5:12:22 PM	Hot

ublic preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Validate all Publish all Preview experience Off

Ass_8_Q3_DataFlow SQL_DB_Generic_DS Ass_8_Q2_Customer... Ass_8_Q3_Peipelin...

Activities

25

Customer_J...
e_2
e_2
.DD
example
rMarkFile
example2
low
n_Product....
eline_2
n_Product....
example
shold_file

Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query

Ass_8_Q3_Dataflow

Data flow

Parameters Variables Settings Output

Pipeline run ID: b7e11b9e-24a5-4dfc-9a5d-ce98eeb3be91
Pipeline status Succeeded

All status Monitor in Azure Metrics Export to CSV

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration
Ass_8_Q3_Dataflow	Succeeded	Data flow	3/2/2024, 5:12:13 PM	50s

Question 4:

create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.

Ensure that product shouldn't be of blue in color and save the result as CSV file inside ProductResult folder.

Solution 4:

Since question has csv file, so we need to use “Data flow”

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists various pipelines and datasets. In the center, the 'Ass_8_Q4_DataFlow' pipeline is selected. A 'Properties' panel on the right displays the pipeline's name ('Ass_8_Q4_DataFlow') and a detailed description: 'create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.' The 'Description' field also contains a link to the question. The 'Data flows' section at the bottom shows other existing data flows like 'Ass_8_Q2_DataFlow' and 'Ass_8_Q3_DataFlow', along with options to 'New data flow', 'New flowlet', and 'New folder'.

review of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factor

Ass_8_Q3_Pipeline... **Ass_8_Q4_DataFlow**

Validate Data flow debug

Source settings **Source options** **Projection** **Optimize** **Inspect** **Data preview**

Source type * Dataset Inline

Dataset * Ass_8_Q4_ProductCSV [Test connection](#) [Open](#) [+](#)

Options

- Allow schema drift ①
- Infer drifted column types ①
- Validate schema ①

recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud

Copy Resources

- Copy_Customer_ID_Csv_Pipe
- Copy_Customer_To_JSON
- Copy_ProductTable_To_CSV
- Ingestion_Customer_SQLDB_ADLS...
- Ingestion_Customer_SQLDB_ADLS_F...
- Ingestion_Product_To_JSON
- Without_Foreach_Example_Pipeline

Change Data Capture (preview) 0

Datasets 20

Data flows 4

- Ass_8_Q2_DataFlow
- Ass_8_Q3_DataFlow
- Ass_8_Ques1_DataFlow
- Ass_8_Q4_DataFlow

Power Query 0

Ass_8_Q3_Pipeline... **Ass_8_Q4_DataFlow**

Validate Data flow debug

Source settings **Source options** **Projection**

Source type * Dataset Inline

Dataset * Select... [Select...](#)

Options

- Allow schema drift ①
- Infer drifted column types ①
- Validate schema ①

Browse

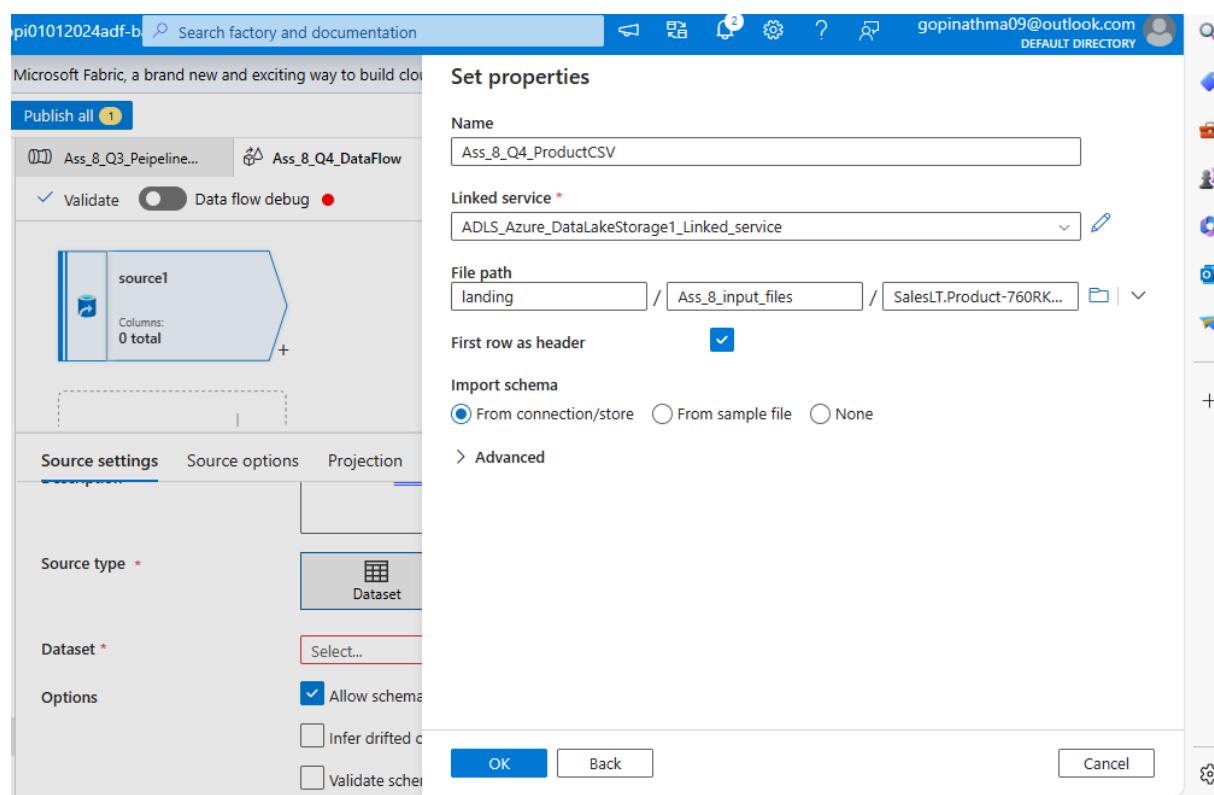
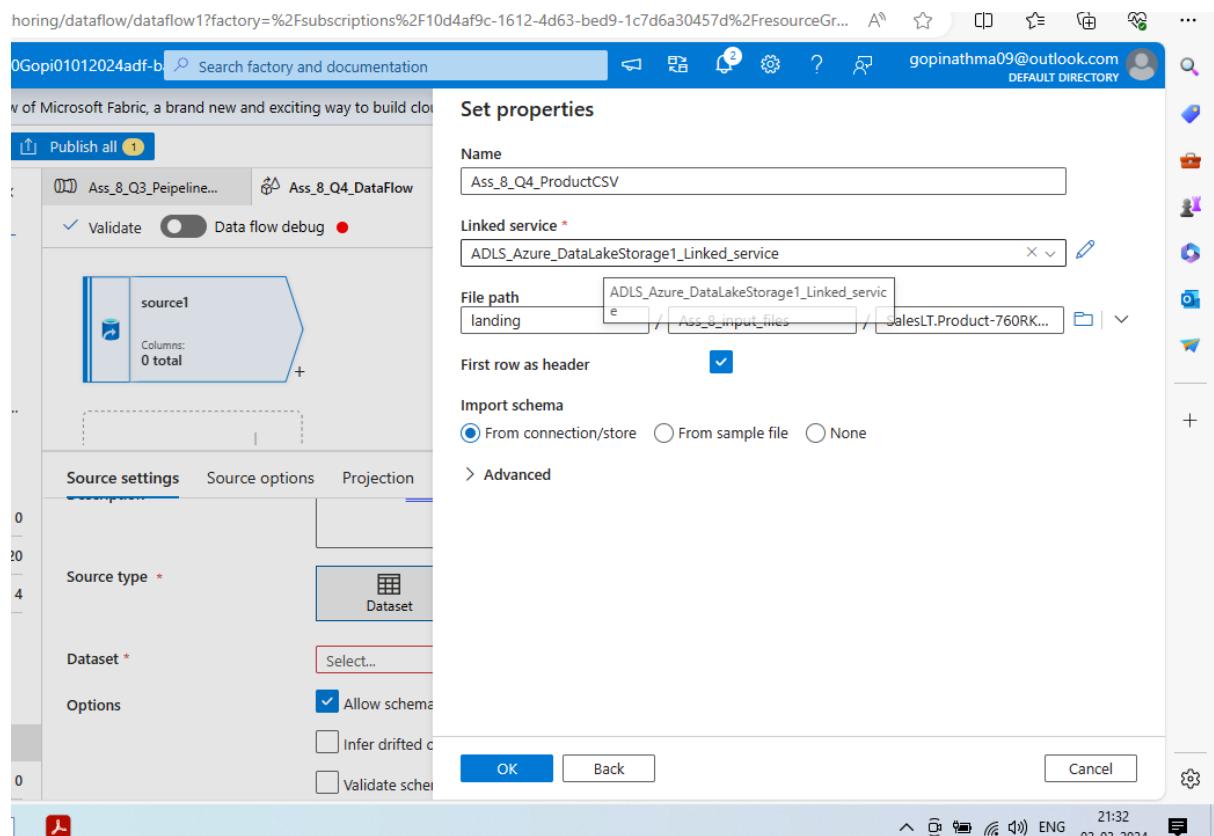
Select a file or folder.

Root folder > landing > Ass_8_input_files

- SalesLT.Customer-xRyuaEu2p7-WdCbzuWllo-80V78fCQtP.txt
- SalesLT.CustomerAddress-nnCI751651 (1)-0A4ckSV5gM-jBPrnf8HWQ.txt
- SalesLT.Product-760RKVH4kL-x8c0uQ9qBG-pui1LFOEyO.txt

Showing 1 - 3 of 3 items

OK **Cancel**



Publish all 2

Ass_8_Q3_Pipeline... Ass_8_Q4_DataFlow

Validate Data flow debug Debug Settings

source1

Columns: 17 total

Source settings Source options Projection Optimize Inspect Data preview

Number of rows + INSERT 100 * UPDATE 0

Refresh Typecast Modify Map drifted Statistics Remove Export to CSV

	ProductID	Name	ProductN...	Color	Standard...	ListPrice
+	680	HL Road F...	FR-R92B-58	Black	1059.3100	1431.5000
+	706	HL Road F...	FR-R92R-58	Red	1059.3100	1431.5000
+	707	Sport-100 ...	HL-U509-R	Red	13.0863	34.9900
+	708	Sport-100 ...	HL-U509	Black	13.0863	34.9900
+	709	Mountain ...	SO-B909-M	White	3.3963	9.5000

of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Publish all 2

Ass_8_Q3_Pipeline... Ass_8_Q4_DataFlow

Validate Data flow debug Debug Settings

source1

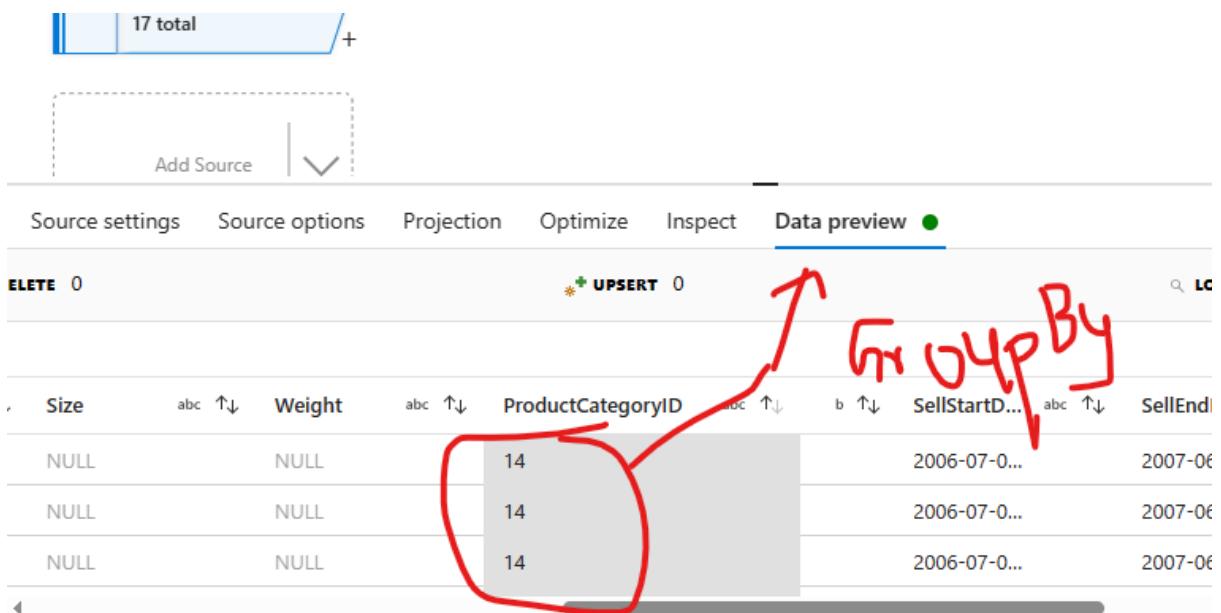
Columns: 17 total

Source settings Source options **Projection** Optimize Inspect Data preview

Define default format Detect data type Import projection Reset schema

Column name	Type	Format
ProductID	12s short	Specify format
Name	abc string	Specify format
ProductNumber	abc string	Specify format
Color	abc string	Specify format
StandardCost	1.2 double	Specify format
ListPrice	1.2 double	Specify format
Size	abc string	Specify format
Weight	1.2 double	Specify format

4. create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.
 Ensure that product shouldn't be of blue in color and save the result as CSV file inside ProductResult folder.



17 total

Add Source ✓

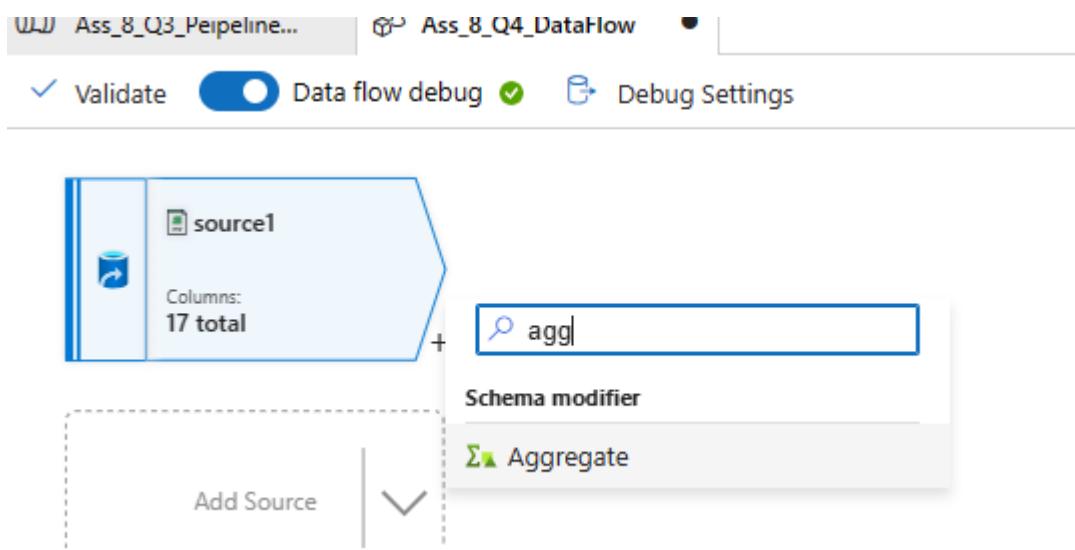
Source settings Source options Projection Optimize Inspect Data preview

~~DELETE 0~~ UPSERT 0

Size abc ↑↓ Weight abc ↑↓ ProductCategoryID abc ↑↓ b ↑↓ SellStartD... abc ↑↓ SellEndI

Size	Weight	ProductCategoryID	b	SellStartD...	SellEndI
NULL	NULL	14		2006-07-0...	2007-06
NULL	NULL	14		2006-07-0...	2007-06
NULL	NULL	14		2006-07-0...	2007-06

Do Aggregate function to get groupby by category id



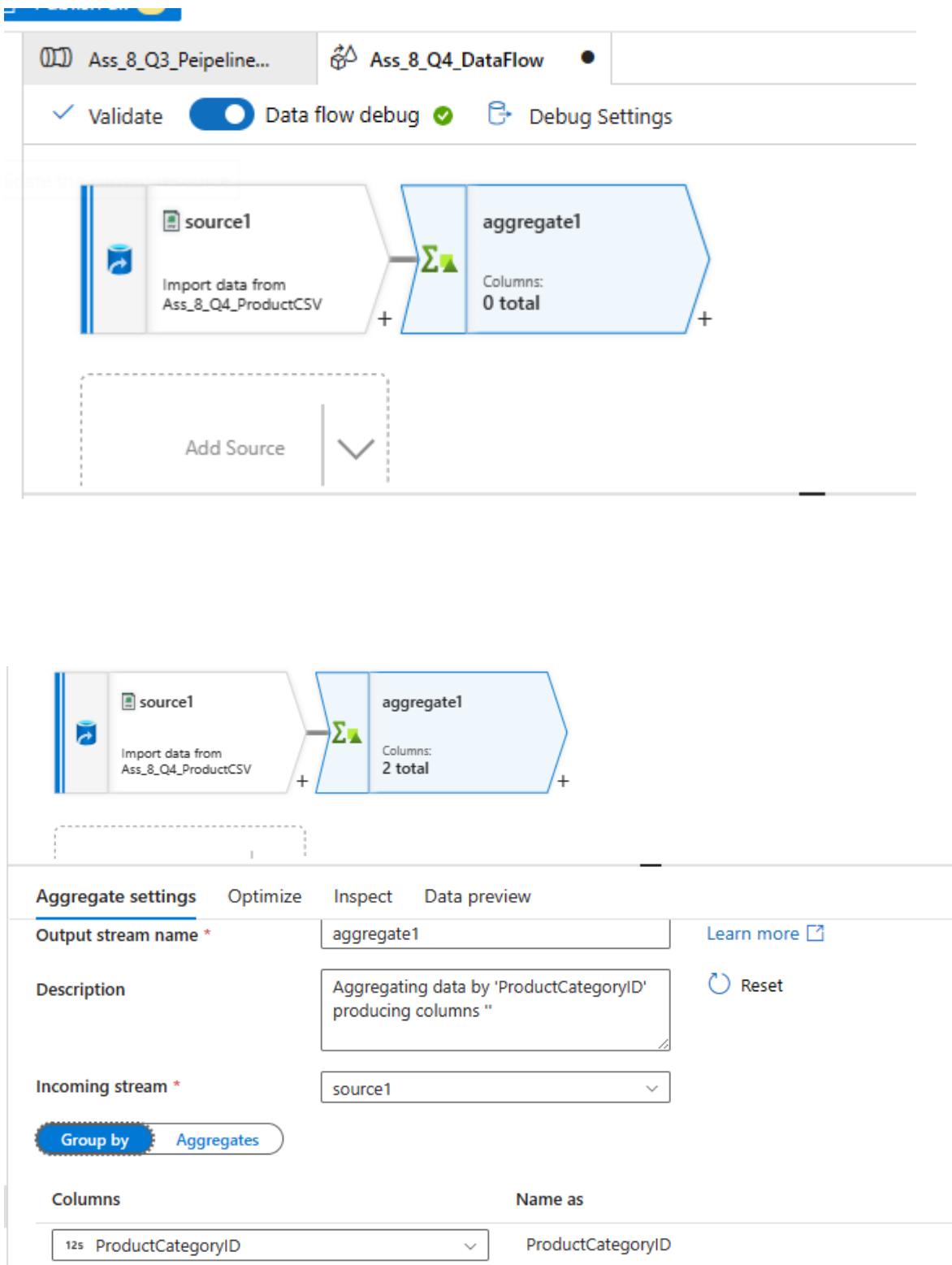


Diagram showing a data flow from 'source1' to 'aggregate1'. 'source1' imports data from 'Ass_8_Q4_ProductCSV'. 'aggregate1' has 2 total columns.

Aggregate settings

Incoming stream *: source1

Aggregates

Grouped by: ProductCategoryID

Add Clone Delete Open expression builder

Column	Expression
HighestListPrice	max(ProductCategoryID)

Column name *: HighestListPrice

Expression

max()

Expression elements

- All
- Functions
- Input schema
- Parameters
- Cached lookup
- Data flow library functions

Expression values

- Filter by keyword
- Create new
- StandardCost
- ListPrice
- Size
- Weight

Column name *

HighestListPrice

Expression

`max(ListPrice)`

+ - * / || && ! ^ == === <=>

Expression elements

All

Functions

Input schema

Parameters

Cached lookup

Data flow library functions

Expression values

Filter by keyword

+ Create new ▾

123 StandardCost

123 ListPrice

abc Size

123 Weight

Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory.

Publish all (2) **Preview experience**

Ass_8_Q3_Pipeline... **Ass_8_Q4_DataFlow**

Validate **Data flow debug** **Debug Settings**

Aggregate settings **Optimize** **Inspect** **Data preview**

Incoming stream *: source1

Group by **Aggregates**

Grouped by: ProductCategoryID

Add **Clone** **Delete** **Open expression builder**

Column	Expression
HighestListPrice	max(ListPrice)

Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Publish all **Preview experience** **Off**

Ass_8_Q3_Pipeline... **Ass_8_Q4_DataFlow**

Validate **Data flow debug** **Debug Settings**

Aggregate settings **Optimize** **Inspect** **Data preview**

Schema **Input** **Output**

Number of columns	New * 1	Dropped 16	Unchanged 1	Total 2
Order ↑↓	Column ↑↓	Type ↑↓	Aggregated as ↑↓	Based on ↑↓
1	ProductCategoryID	12s short	Group by	ProductCategoryID
2	HighestListPrice	1.2 double	Aggregate	ListPrice

Publish all 2

Preview experience

Ass_8_Q3_Pipeline... Ass_8_Q4_DataFlow

Validate Data flow debug Debug Settings

Reference: 1 Columns: 17 total

aggregate1 Columns: 2 total

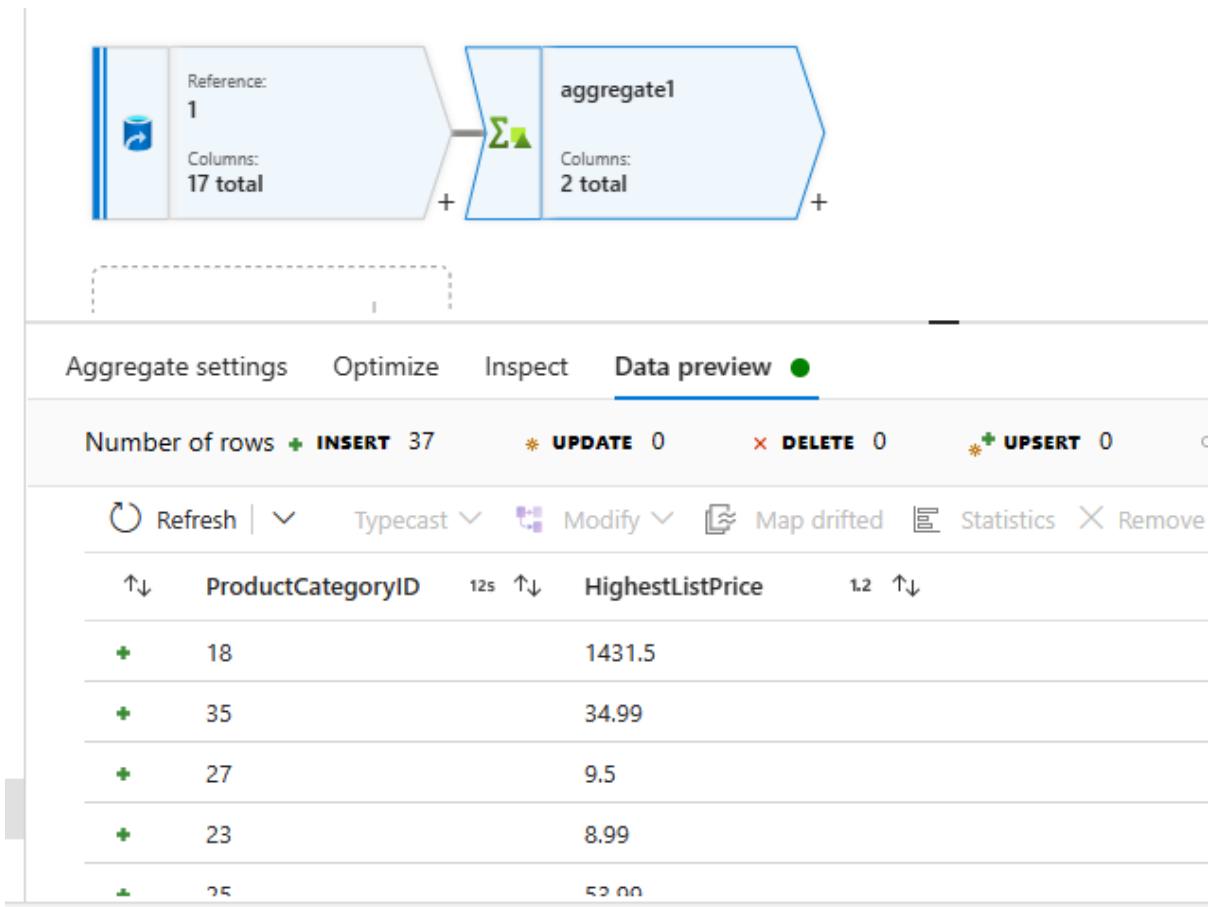
Aggregate settings Optimize Inspect Data preview

Number of rows: 0 INSERT N/A UPDATE N/A DELETE N/A UPSERT N/A LOOKUP N/A ERROR N/A

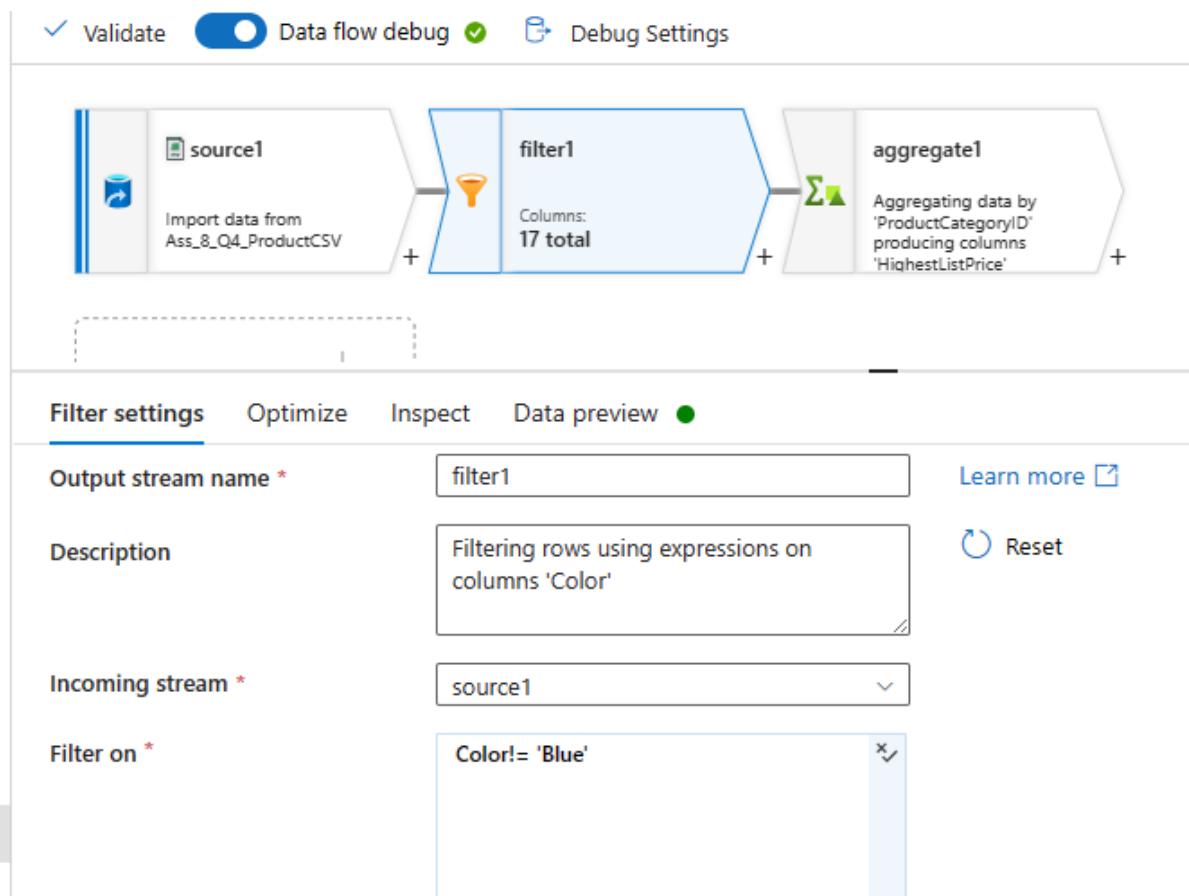
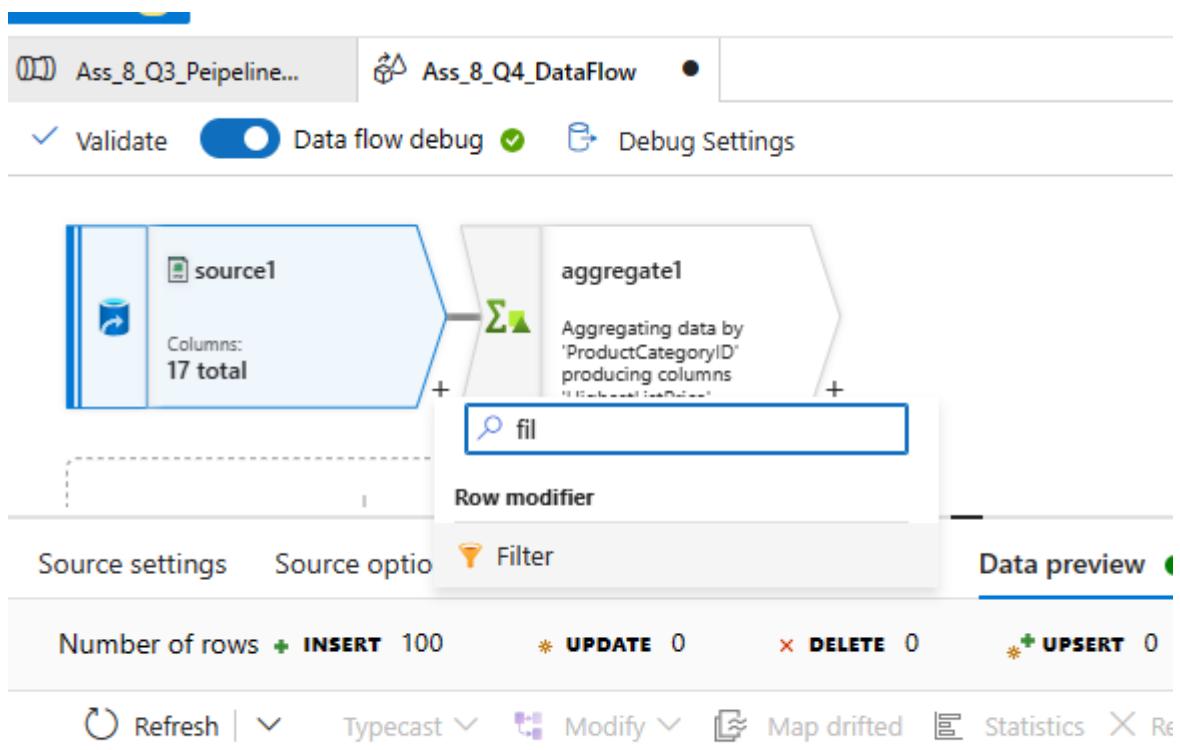
Refresh |

Fetching data...

The screenshot shows the 'Data preview' tab of the Azure Data Flow interface. At the top, there are tabs for 'Aggregate settings', 'Optimize', 'Inspect', and 'Data preview'. The 'Data preview' tab is selected, indicated by a green dot. Below the tabs, there are sections for 'Number of rows' (0), 'INSERT' (N/A), 'UPDATE' (N/A), 'DELETE' (N/A), 'UPSERT' (N/A), 'LOOKUP' (N/A), and 'ERROR' (N/A). A 'Refresh' button is also present. The main area displays a data flow diagram with two components: a 'Reference' component (labeled '1') and an 'aggregate1' component. The 'aggregate1' component has a summation symbol (Σ) and is described as having '2 total' columns. The status bar at the bottom indicates 'Fetching data...'.



Ensure that product shouldn't be of blue in color and save the result as CSV file inside ProductResult folder.



Dataflow expression builder

filter1

Expression

```
Color != 'Blue'
```

+ - * / || && ! ^ == ===

Expression elements

- All
- Functions
- Input schema
- Parameters
- Cached lookup
- Data flow library functions

Expression values

Filter by keyword

Create new

- Name
- ProductNumber
- Color

Data preview Refresh

Save and finish Cancel Clear contents

Dataflow expression builder

filter1

Expression

```
color != 'Blue'
```

Expression elements

- All
- Functions
- Input schema

Expression values

Filter by keyword

Create new

123 PRODUDS

Data preview

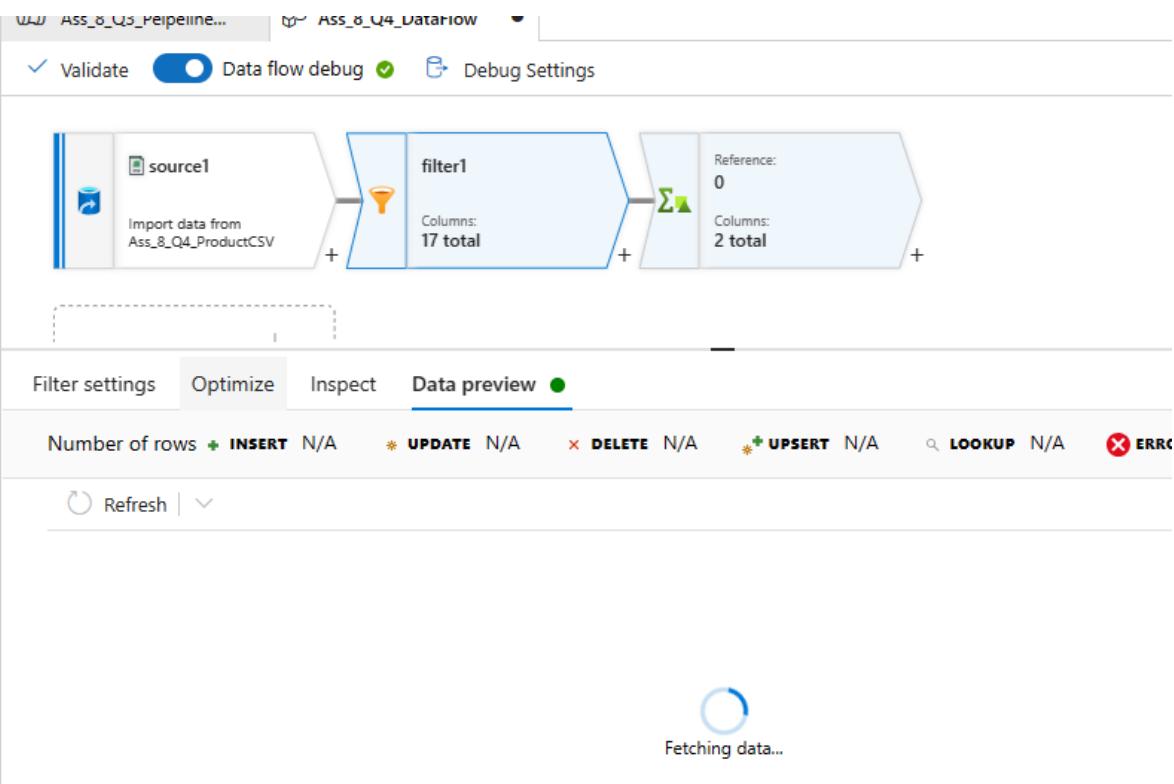
Refresh

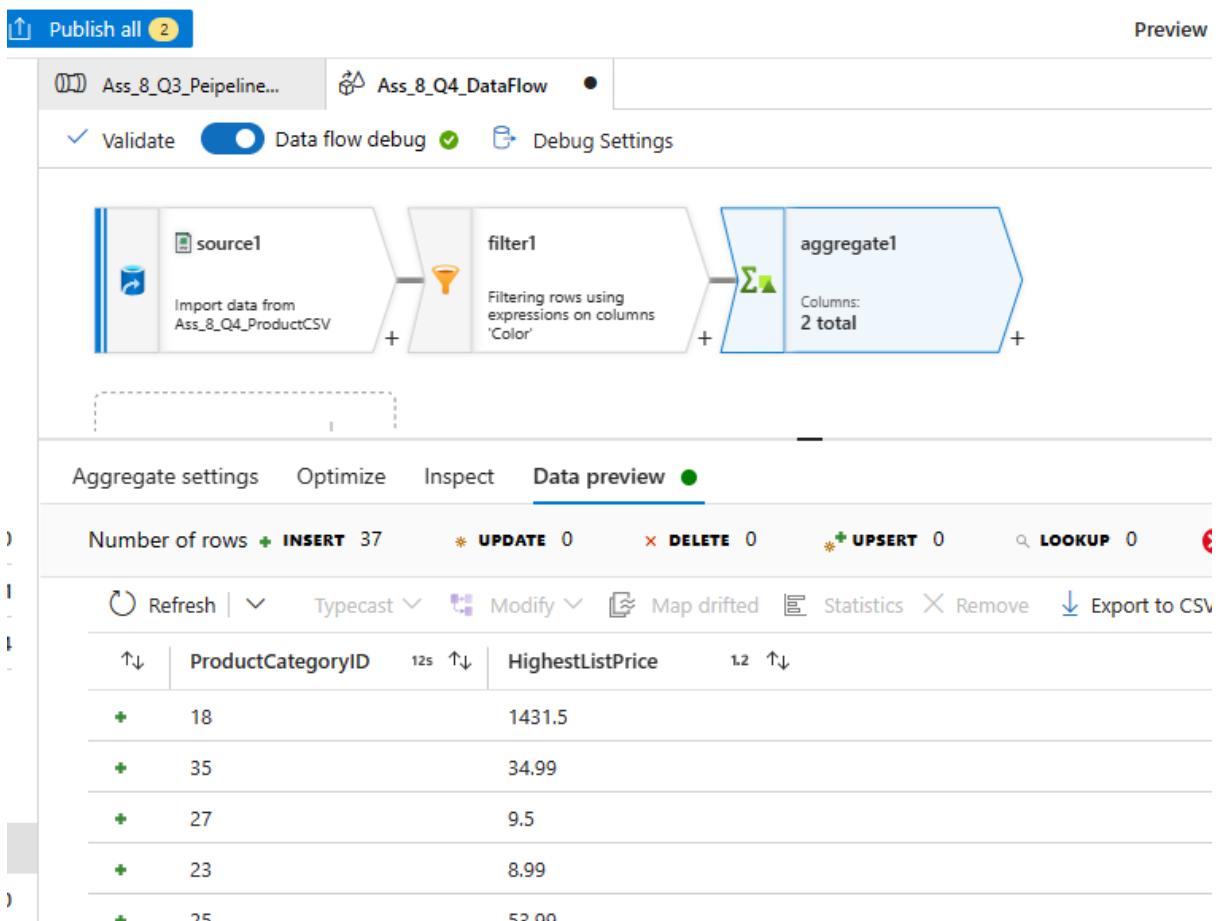
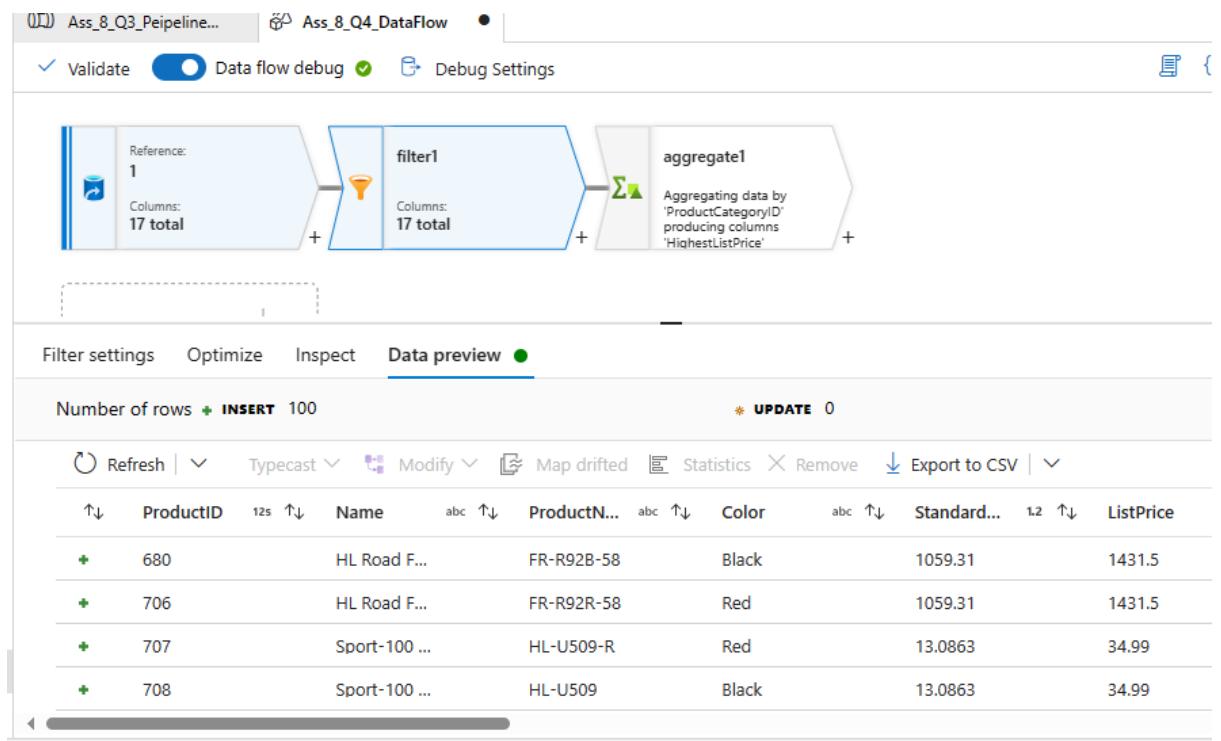
✓		White
✓		White
✗	Blue	Blue
✓		Multi
✓		Multi

Save and finish

Cancel

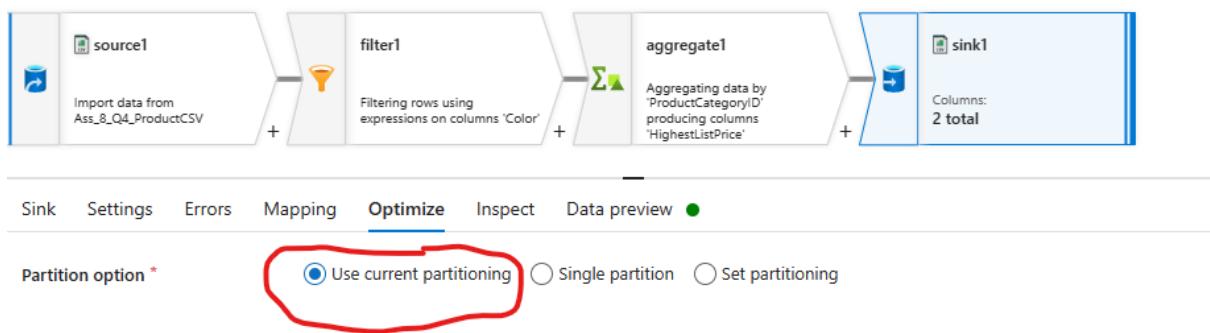
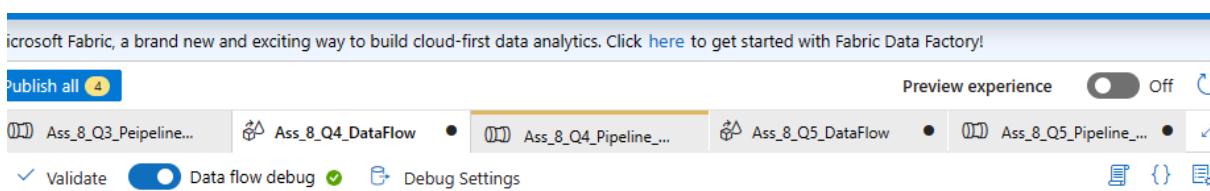
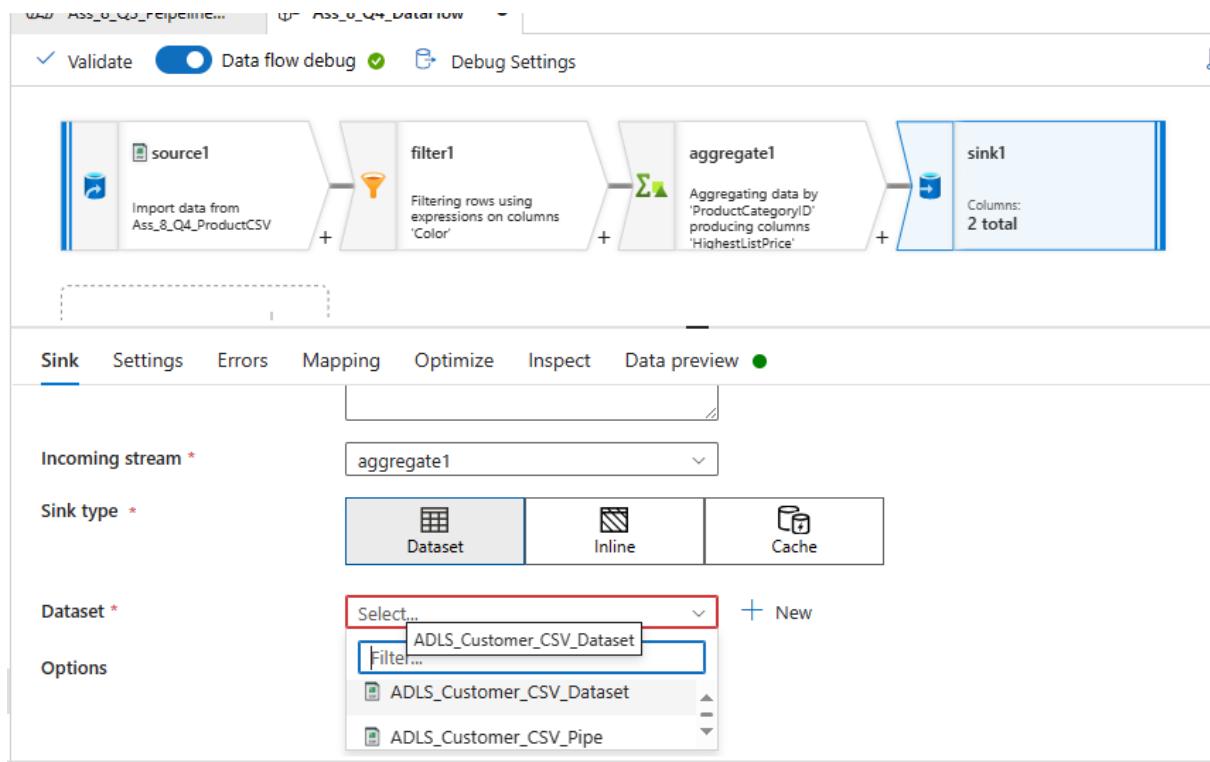
Clear contents





save the result as CSV file inside ProductResult folder.

→ So save in “Sink” file



Set properties

Name

Linked service *

File path

 / /  

First row as header



Import schema

From connection/store From sample file None

> Advanced

OK

Back

Cancel



Preview experience Off

ADLS_Generic_CSV_DS Ass_8_Q4_ProductCS... ● Ass_8_Q2_Pipeline... Ass_8_Q4_Pipeline... ●

Validate Debug Add trigger Data flow debug

Data flow Ass_8_Q4_Pipeline_Dataflow

General Settings Parameters 1 User properties

Data flow * Ass_8_Q4_DataFlow

Run on (Azure IR) * AutoResolveIntegrationRuntime

Compute size * Small

Advanced

Logging level * Verbose Basic None

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Preview experience Off

Data Factory Validate all Publish all

Factory Resources ADLS_Customer_CSV... ADLS_Generic_CSV_DS Ass_8_Q4_ProductCS... ● Ass_8_Q2_Pipeline... Ass_8_Q4_Pipeline... ●

Activities

Move and transform

Azure Data Explorer

Databricks

Data Lake Analytics

General Settings Parameters 1 User properties

Data flow * Ass_8_Q4_DataFlow

Run on (Azure IR) * AutoResolveIntegrationRuntime

Compute size * Small

Properties

Name * Ass_8_Q4_Pipeline_Dataflow

Description Call Ass_8_Q4_Dataflow create a pipeline to read the Product

Annotations

Ass_8_Q4_Pipeline_Dataflow

Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

all 4

Preview experience Off

[_S_Customer_CSV...](#) [ADLS_Generic_CSV_DS](#) [Ass_8_Q4_ProductCS...](#) [Ass_8_Q2_Pipeline...](#) [Ass_8_Q4_Pipeline...](#)

Activities

Copy and transform

Use Data Explorer

Notebooks

Workbooks

Python

Data Lake Analytics

J-SQL

Data flow Ass_8_Q4_Pipeline_Dataflow

Output

Pipeline run ID: 1fe8c2ab-32b5-4887-9ddc-e398a50f43b2 Pipeline status In progress

All status Monitor in Azure Metrics Export to CSV

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration
Ass_8_Q4_Pipeline_Dataflow	In progress	Data flow	3/2/2024, 11:18:00 PM	27s

https://portal.azure.com/#view/Microsoft_Azure_Storage/BlobPropertiesBladeV2/storageAccountId/%2Fsubscriptions%2F10d4af9c-1612-4...

Microsoft Azure Search resources, services, and docs (G/)

Storage accounts > gopi01012024adls1 | Containers > landing

landing Container

Overview Authentication method: Access key (Switch to Microsoft Entra user account) Location: landing / Ass_8_output_files / Ass_8_Q4_Product_Result

Search blobs by prefix (cas... Show deleted objects

Name

- _SUCCESS
- part-00000-3207fc50-ace4-455f-b4
- part-00014-3207fc50-ace4-455f-b4
- part-00019-3207fc50-ace4-455f-b4
- part-00024-3207fc50-ace4-455f-b4
- part-00030-3207fc50-ace4-455f-b4
- part-00048-3207fc50-ace4-455f-b4
- part-00049-3207fc50-ace4-455f-b4

Ass_8_output_files/part-00000-a90c8241-fe0e-4558-97...

Save Discard Download Refresh Delete

Overview Versions Edit Generate SAS

```

1  [{"CustomerID":29485,"NameStyle":false,"Title":"Ms.","FirstName":"Catherine","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29486,"NameStyle":false,"Title":"Ms.","FirstName":"Kim","LastName":"Hill"}, {"CustomerID":29488,"NameStyle":false,"Title":"Ms.","FirstName":"Frances","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29490,"NameStyle":false,"Title":"Ms.","FirstName":"Margaret","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29492,"NameStyle":false,"Title":"Mr.","FirstName":"Jay","LastName":null}, {"CustomerID":29494,"NameStyle":false,"Title":"Mr.","FirstName":"Samuel","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29496,"NameStyle":false,"Title":"Mr.","FirstName":"Robert","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29497,"NameStyle":false,"Title":"Mr.","FirstName":"Francois","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29499,"NameStyle":false,"Title":"Ms.","FirstName":"Amy","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29502,"NameStyle":false,"Title":"Mr.","FirstName":"Paul","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29503,"NameStyle":false,"Title":"Mr.","FirstName":"Gregory","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29503,"NameStyle":false,"Title":"Mr.","FirstName":"Gregory","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29505,"NameStyle":false,"Title":"Ms.","FirstName":"Michelle","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29506,"NameStyle":false,"Title":"Mr.","FirstName":"Sean","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29508,"NameStyle":false,"Title":"Mr.","FirstName":"Marvin","MiddleName":null,"LastName":"Hill"}, {"CustomerID":29510,"NameStyle":false,"Title":"Mr.","FirstName":"Cecil","MiddleName":null,"LastName":"Hill"}]

```

Json Preview

« [Upload](#) [Add Directory](#) [Refresh](#) | [Rename](#) [Delete](#) [Change tier](#) [Acquire lease](#) [Break lease](#) [Give feedback](#)

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: landing / Ass_8_output_files / Ass_8_Q4_Product_Result

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/>  [.]				
<input type="checkbox"/>  _SUCCESS	3/2/2024, 11:18:16 PM	Hot (Inferred)		Block blob
<input type="checkbox"/>  part-00000-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv	3/2/2024, 11:18:13 PM	Hot (Inferred)		Block blob
<input type="checkbox"/>  part-00014-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv	3/2/2024, 11:18:11 PM	Hot (Inferred)		Block blob
<input type="checkbox"/>  part-00019-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv	3/2/2024, 11:18:11 PM	Hot (Inferred)		Block blob
<input type="checkbox"/>  part-00024-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv	3/2/2024, 11:18:11 PM	Hot (Inferred)		Block blob
<input type="checkbox"/>  part-00030-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv	3/2/2024, 11:18:11 PM	Hot (Inferred)		Block blob
<input type="checkbox"/>  part-00048-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv	3/2/2024, 11:18:12 PM	Hot (Inferred)		Block blob
<input type="checkbox"/>  part-00049-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv	3/2/2024, 11:18:12 PM	Hot (Inferred)		Block blob
<input type="checkbox"/>  part-00053-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv	3/2/2024, 11:18:12 PM	Hot (Inferred)		Block blob
<input type="checkbox"/>  part-00066-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv	3/2/2024, 11:18:12 PM	Hot (Inferred)		Block blob

Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Publish all 4 Preview experience

ADLS_Customer_CSV... ADLS_Generic_CSV_DS Ass_8_Q4_ProductCS... Ass_8_Q2_Pipeline... Ass_8_Q4_Pipeline...

✓ Validate ▶ Debug ✎ Add trigger Data flow debug ✓

Data flow Ass_8_Q4_Pipeline_Dataflow

General Settings Parameters User properties

Name * Ass_8_Q4_Pipeline_Dataflow Learn more

Description Ass_8_Q4_Pipeline_Dataflow

Activity state Activated Deactivated

Timeout 0:12:00:00

Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Publish all 4 Preview experience Off

ADLS_Customer_CSV... ADLS_Generic_CSV_DS Ass_8_Q4_ProductCS... Ass_8_Q2_Pipeline... Ass_8_Q4_Pipeline...

✓ Validate ▶ Debug ✎ Add trigger Data flow debug ✓

Data flow Ass_8_Q4_Pipeline_Dataflow

Parameters Variables Settings Output

Pipeline run ID: f4ad7167-37b5-4ae7-9d44-064194d521f6 Pipeline status ✓ Succeeded View debug run consumption

All status ▾ Monitor in Azure Metrics Export to CSV

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime
Ass_8_Q4_Pipeline_Dataflow	✓ Succeeded	Data flow	3/2/2024, 11:24:26 PM	1m 21s	debugpool-8Cores-G

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: landing / Ass_8_output_files / Ass_8_Q4_Product_Result

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> _SUCCESS	3/3/2024, 2:05:13 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> part-00000-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	3/3/2024, 2:05:10 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> part-00014-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	3/3/2024, 2:05:08 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> part-00019-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	3/3/2024, 2:05:08 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> part-00024-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	3/3/2024, 2:05:08 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> part-00030-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	3/3/2024, 2:05:08 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> part-00048-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	3/3/2024, 2:05:09 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> part-00049-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	3/3/2024, 2:05:09 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> part-00053-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	3/3/2024, 2:05:09 AM	Hot (Inferred)		Block blob
<input type="checkbox"/> part-00066-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	3/3/2024, 2:05:09 AM	Hot (Inferred)		Block blob

Search resources, services, and docs (G+)

12024adls1 | Containers > landing >

Ass_8_output_files/Ass_8_Q4_Product_Result/part-0001.

Upload Add Directory ...

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: landing / Ass_8_output_files / Ass_8_Q4_Product_Result

Search blobs by prefix (case-sensitive) Show deleted objects

Name

<input type="checkbox"/> _SUCCESS	***
<input type="checkbox"/> part-00000-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	***
<input type="checkbox"/> part-00014-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	***
<input type="checkbox"/> part-00019-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	***
<input type="checkbox"/> part-00024-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	***
<input type="checkbox"/> part-00030-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	***
<input type="checkbox"/> part-00048-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	***
<input type="checkbox"/> part-00049-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	***
<input type="checkbox"/> part-00053-4e78fe72-ead9-4c66-adc7-114cc795de82-c...	***

Save Discard Download Refresh Delete

Overview Versions Edit Generate SAS

1 ProductCategoryID,HighestListPrice
2 28,74.99
3

Csv

Search resources, services, and docs (G+/)

1012024adls1 | Containers > landing >

Ass_8_output_files/Ass_8_Q4_Product_Result/part-merged.csv

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: landing / Ass_8_output_files / Ass_8_Q4_Product_Result

Search blobs by prefix (case-insensitive)

Show deleted objects

Name

part-merged.csv

	ProductCategoryID, HighestListPrice
1	28,74.99
2	22,89.99
3	17,80.99
4	35,34.99
5	10,106.5
6	13,121.46
7	5,3399.99
8	7,2384.07
9	21,357.06
10	18,1431.5
11	16,1364.5
12	6,3578.27
13	23,8.99
14	39,125.0
15	27,9.5
16	

Csv Preview

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud data solutions.

Data Factory Publishing

Factory Resources

- Ingestion_Customer_SQLDB_DLDS
- Ingestion_Customer_SQLDB_DLDS_F...
- Ingestion_Product_To_JSON
- Ass_8_Q4_Pipeline_Dataflow
- Without_Foreach_Example_Pipeline
- Change Data Capture (preview) 0
- Datasets 22
 - ADLS_Customer_CSV_Dataset
 - ADLS_Customer_CSV_Pipe
 - ADLS_Datalake_storage_CSV_Dataset
 - ADLS_Generic_CSV_DS
 - ADLS_Generic_JSON_DS
 - ADLS_JSON_Dataset
 - Ass_8_Q1_Customer_CSV

ADLS_Customer_CSV... ADLS_Generic_CSV_DS

Validate Debug Add trigger

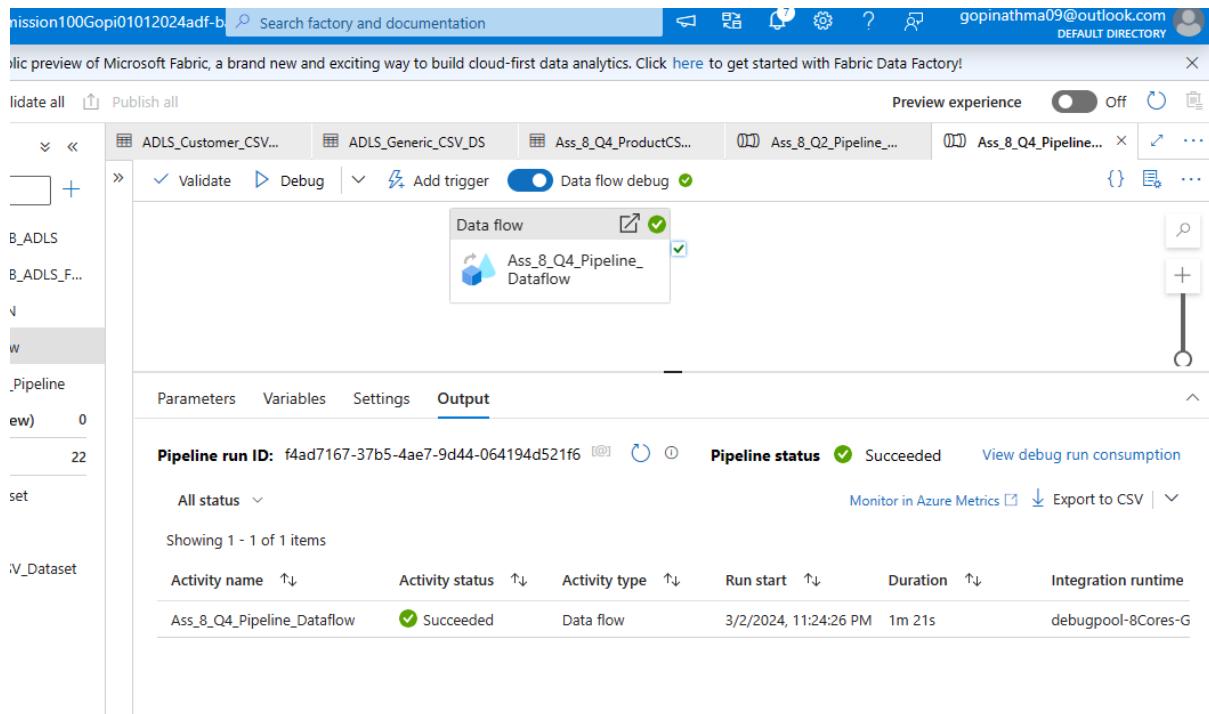
Parameters Variables Settings Output

Pipeline run ID: f4ad7167-37b5-4ae7-9d44-0

All status Showing 1 - 1 of 1 items

Activity name Ass_8_Q4_Pipeline_Dataflow Activity status Succeeded

Publish Cancel



Question 5:

create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.

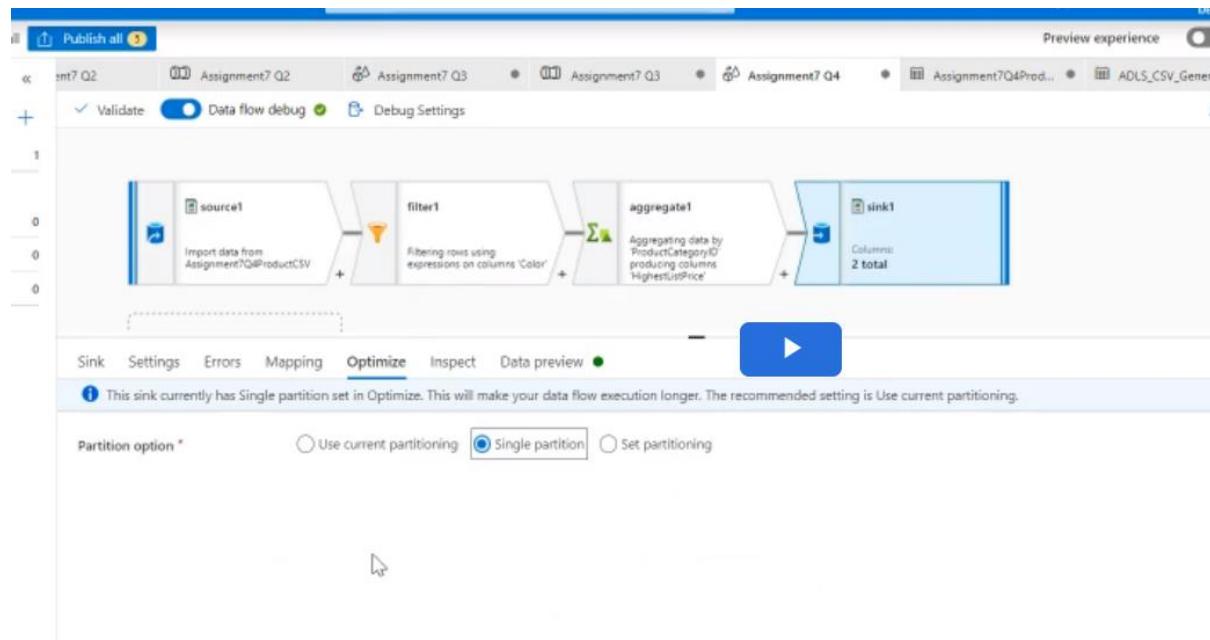
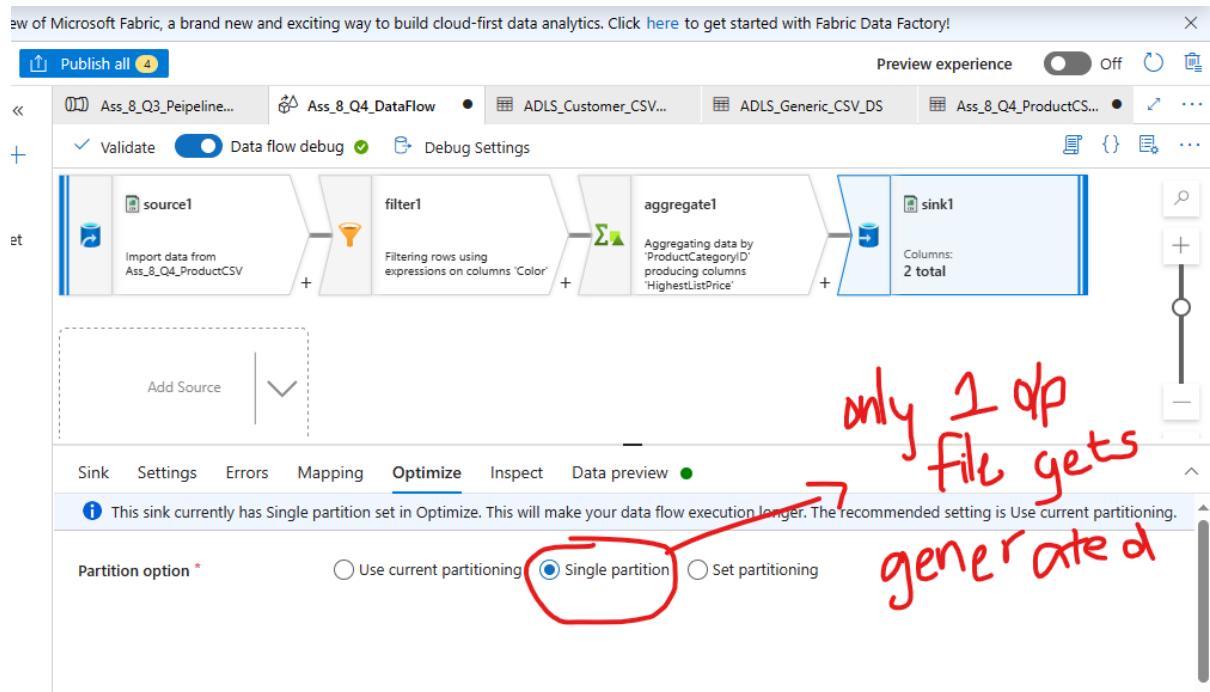
Ensure that product shouldn't be of blue in color and save the result as a SINGLE CSV file inside ProductSingleResult folder.

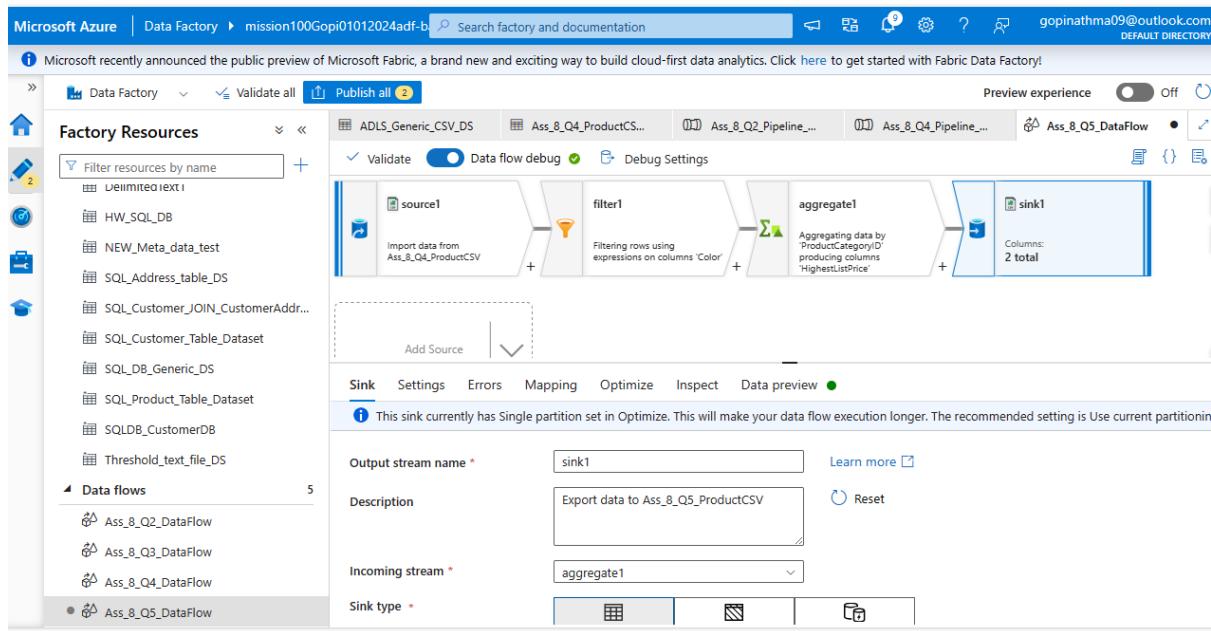
Same Solution as Question 4:

Only difference is in the Sink “Optimize ” tab instead of “use current partitioning”

Select “Single partition” that’s it == All output will be saved in SINGLE CSV file

Earlier it was multi partitioning (so multiple processing and multipleoutput csv files) – Now only single CSV file





Azure | Data Factory > mission100Gopi01012024adf-b | Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory | Validate all | Publish all | Preview experience | Off | [Fabric Data Factory](#)

Factory Resources

- Filter resources by name
- HW_SQL_DB
- NEW_Meta_data_test
- SQL_Address_table_DS
- SQL_Customer_JOIN_CustomerAddr...
- SQL_Customer_Table_Dataset
- SQL_DB_Generic_DS
- SQL_Product_Table_Dataset
- SQLDB_CustomerDB
- Threshold_text_file_DS

Data flows 5

- Ass_8_Q2_DataFlow
- Ass_8_Q3_DataFlow
- Ass_8_Q4_DataFlow
- Ass_8_Q5_DataFlow
- Ass_8_Ques1_DataFlow

Set properties

Name Ass_8_Q5_ProductCSV

Linked service ADLS_Azure_DataLakeStorage1_Linked_service

File path landing / Ass_8_output_files/Ass_... / File name

First row as header

Import schema
 From connection/store
 From sample file
 None

> Advanced

Incoming stream * aggregate1

Sink type * Dataset

Dataset * Ass_8_Q4_P

Skip line count

Options
 Allow schema
 Validate schema

OK | Back | Cancel

New Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Publish all 4

Preview experience Off

Validate Data flow debug Debug Settings

Ass_8_Q4_DataFlow

source1 Import data from Ass_8_Q4_ProductCSV

filter1 Filtering rows using expressions on columns 'Color'

aggregate1 Aggregating data by 'ProductCategoryID' producing columns 'HighestListPrice'

sink1 Columns: 2 total

Add Source

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

Optimize This sink currently has Single partition set in Optimize. This will make your data flow execution longer. The recommended setting is Use current partitioning.

Partition option * Use current partitioning **Single partition** Set partitioning

only 1 dp file gets generated

New Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Publish all 4

Preview experience Off

Validate Add trigger Data flow debug

Ass_8_Q4_Pipeline...

Data flow Ass_8_Q5_Pipeline_Dataflow

Activities

- > Move and transform
- > Synapse
- > Azure Data Explorer
- > Azure Function
- > Batch Service
- > Databricks
- > Data Lake Analytics
- > General
- > HDInsight
- > Iteration & conditionals
- > Machine Learning
- > Power Query

Parameters Variables Settings **Output**

Pipeline run ID: 8ff84448-f51f-49fe-9508-ba67e758ce57

Pipeline status Succeeded View debug run consumption

All status Monitor in Azure Metrics Export to CSV

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration
Ass_8_Q5_Pipeline_Dataflow	Succeeded	Data flow	3/3/2024, 2:09:15 AM	52s

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: landing / Ass_8_output_files / Ass_8_Q4_Product_Result

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> part-merged.csv	3/2/2024, 11:24:56 PM	Hot (Inferred)		Block blob

Single partition file

Upgrade Search resources, services, and docs (G+/)

gopi@1012024adls1 | Containers > landing > Ass_8_output_files/Ass_8_Q5_ProductSingleResult/part...

Upload Add Directory ...

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: landing / Ass_8_output_files / Ass_8_Q5_ProductSingleResult

Search blobs by prefix (case-sensitive) Show deleted objects

Name

<input type="checkbox"/> part-merged.csv	...
--	-----

ProductCategoryID,HighestListPrice

- 1 28,74.99
- 2 22,89.99
- 3 17,80.99
- 4 35,34.99
- 5 10,106.5
- 6 13,121.46
- 7 5,3399.99
- 8 7,2384.07
- 9 21,357.06
- 10 18,1431.5
- 11 16,1364.5
- 12 6,3578.27
- 13 23,8.99
- 14 39,125.0
- 15 27,9.5

Csv Preview

Ass_8_output_files/Ass_8_Q5_ProductSingleResult /part...

Authentication method: Access key ([Switch to Microsoft Entra user account](#))

Location: landing / Ass_8_output_files / Ass_8_Q5_ProductSingleResult

Name

	ProductCategoryID, HighestListPrice
1	28, 74.99
2	22, 89.99
3	17, 80.99
4	35, 34.99
5	10, 106.5
6	13, 121.46
7	5, 3399.99
8	7, 2384.07
9	21, 357.06
10	18, 1431.5
11	16, 1364.5
12	6, 3578.27
13	23, 8.99
14	39, 125.0
15	27, 9.5
16	

Finally Publish all

Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (4)

NAME	CHANGE	EXISTING
Ass_8_Q5_Pipeline_Dataflow (New)	-	
Ass_8_Q5_ProductCSV (New)	-	
Ass_8_Q4_DataFlow (Edited)	Ass_8_Q4_DataFlow	
Ass_8_Q5_DataFlow (New)	-	

Pipeline run ID: Ass_8_Q5_Pipeline_Dataflow

Pipeline status: All status

Activity name: Ass_8_Q5_Pipeline_Dataflow

Buttons: Publish, Cancel

