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

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.

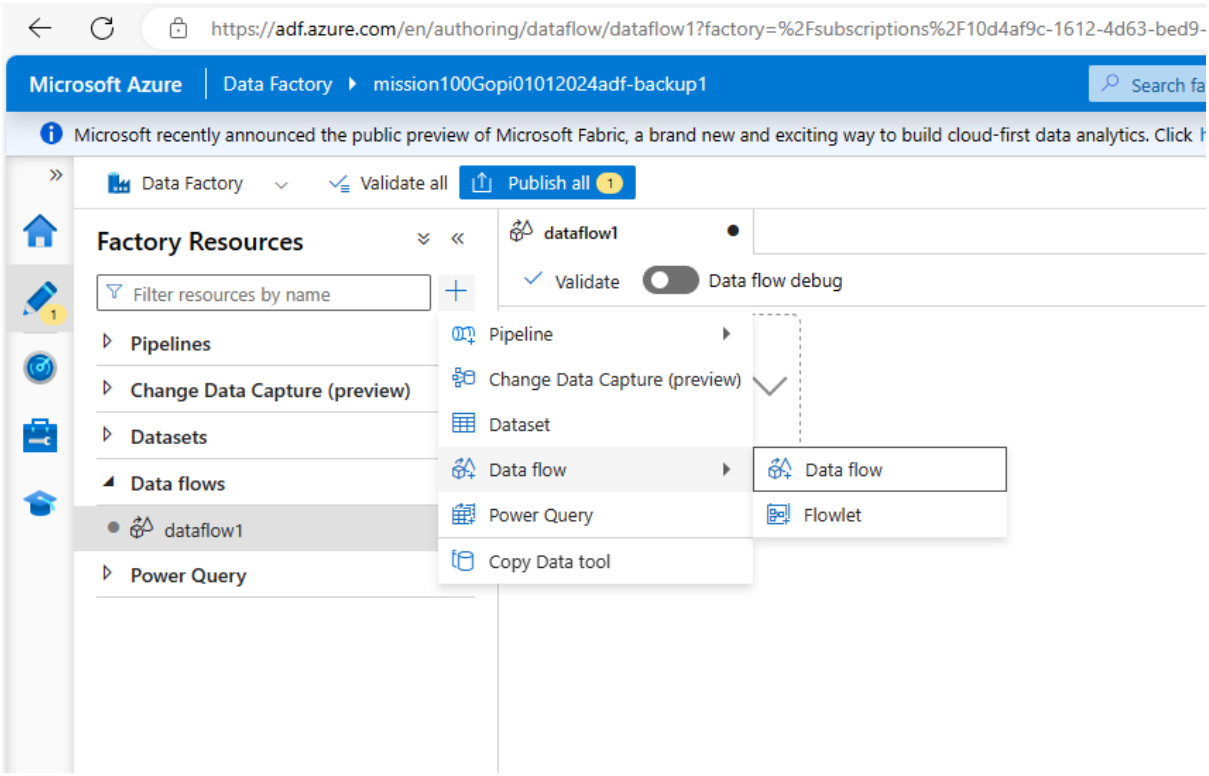
Solutions:

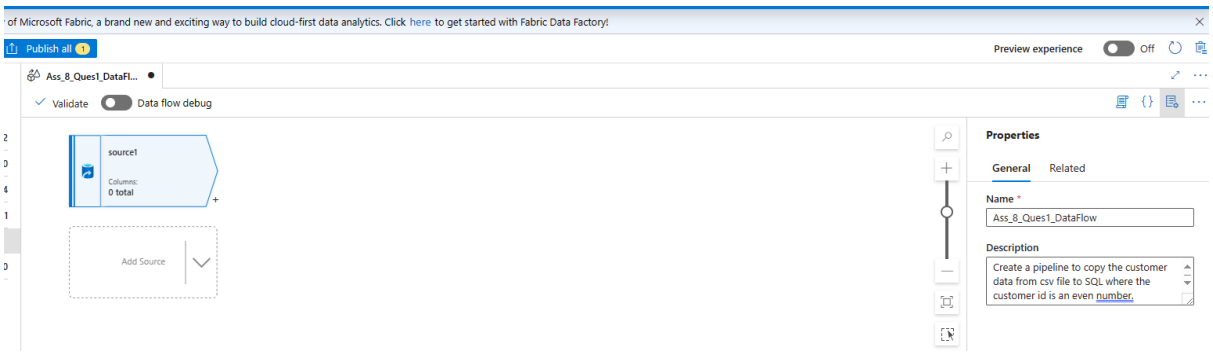
Solutions:
Assignment 8: ADF Data Flow 5 Question

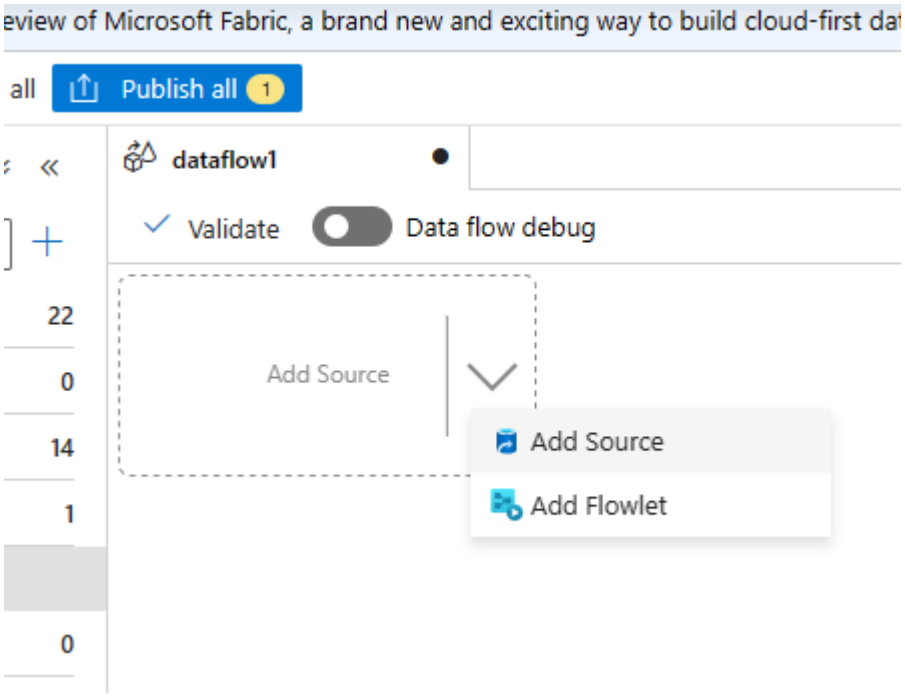
1. Create a pipeline to copy the customer data from csv file to SQL where the customer id is an even number.(Attaching the Customer File)
   File: SalesLT.Customer-xRyuaEu2p7-WdCbzuWlIo-80V7BfCQtP

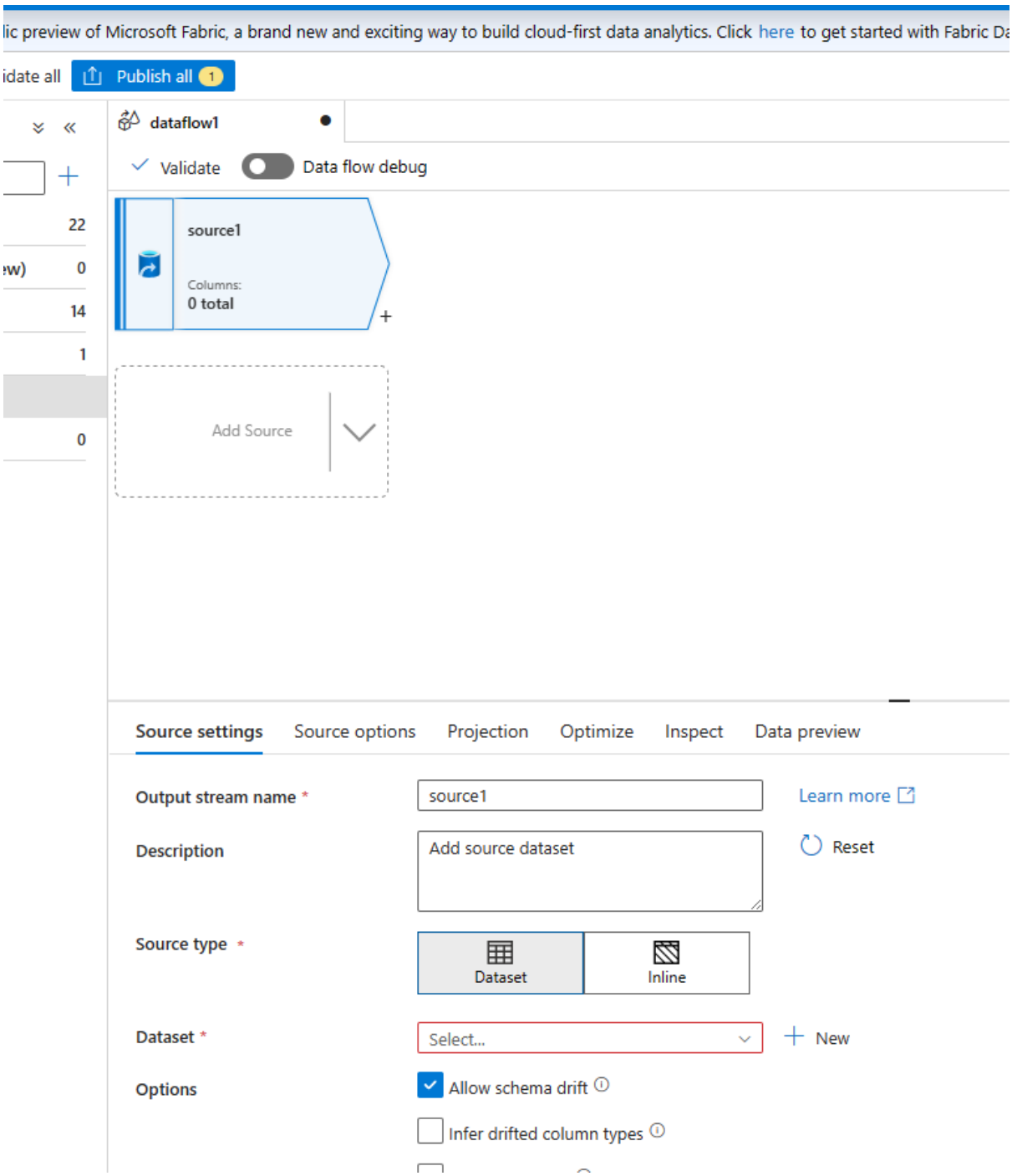
Solution :

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

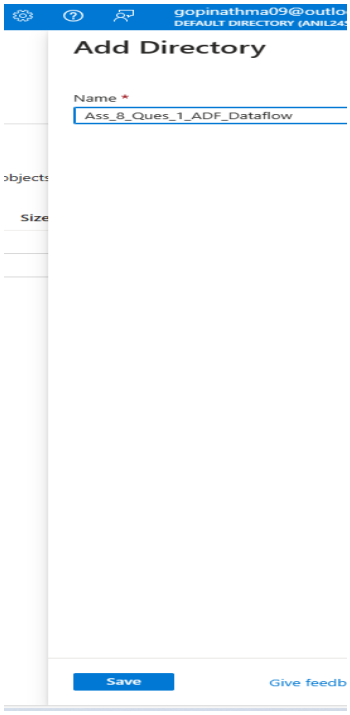


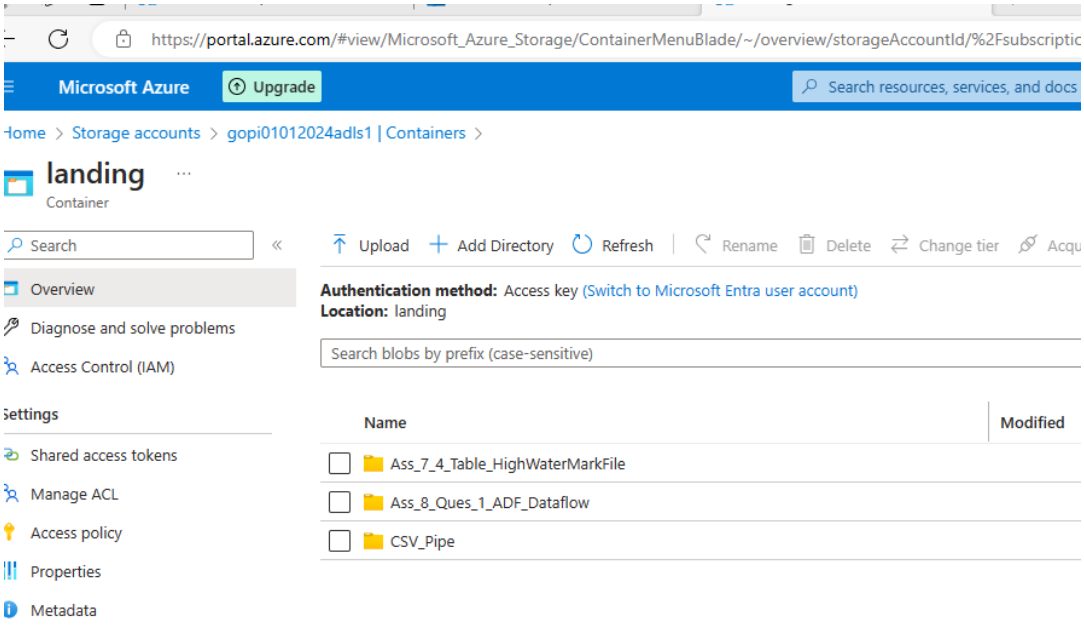


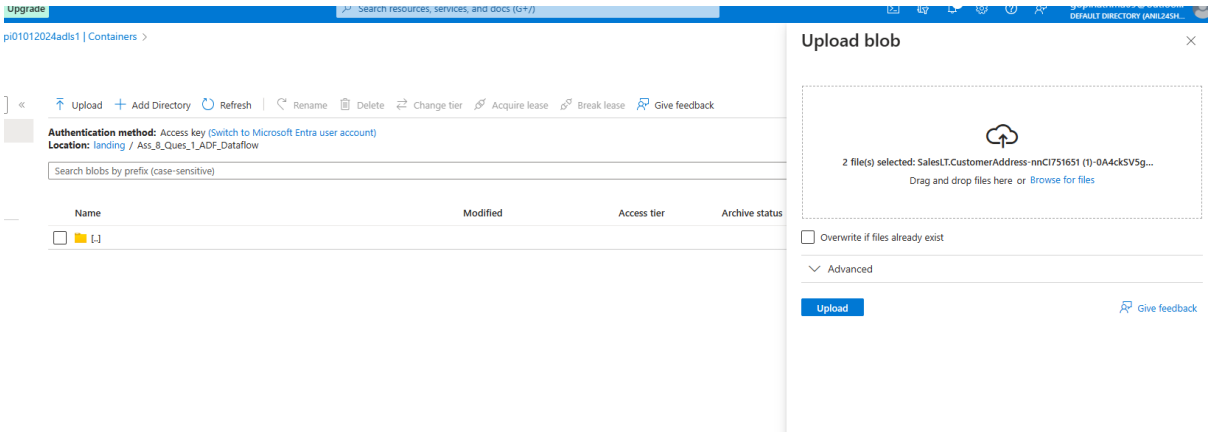

eview of Microsoft Fabric, a brand new and exciting way to build cloud-first dat
all
Publish all 1
dataflow1
Validate
Data flow debug
22
0
14
1
0
Add Source
Add Source
Add Flowlet

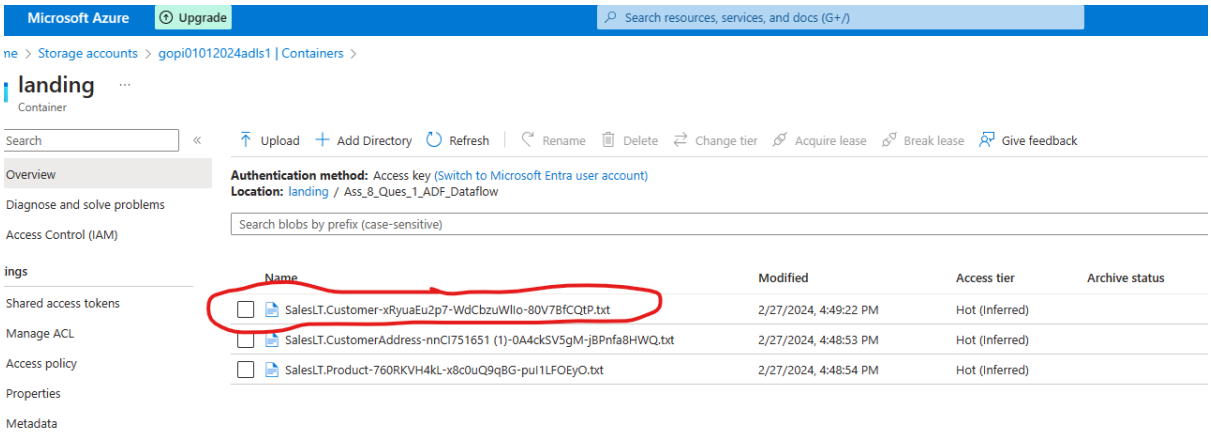


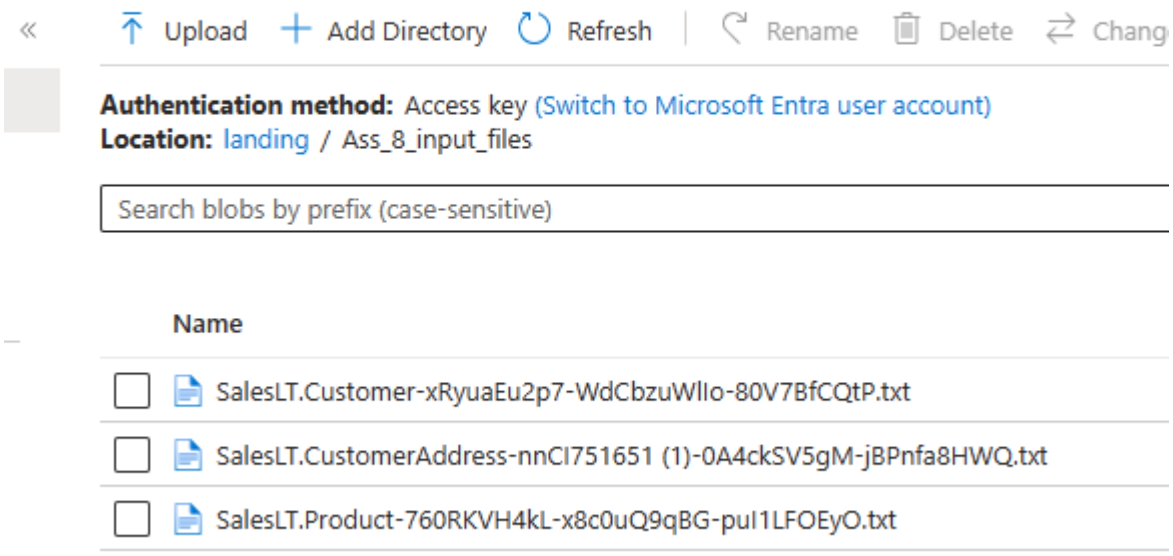
Since CSV is input file, It is not avaialble , we need to upload to ADLS storage container account first .


Add Directory
Name *
Ass_8_Ques_1_ADF_Dataflow
Save
Give feedb

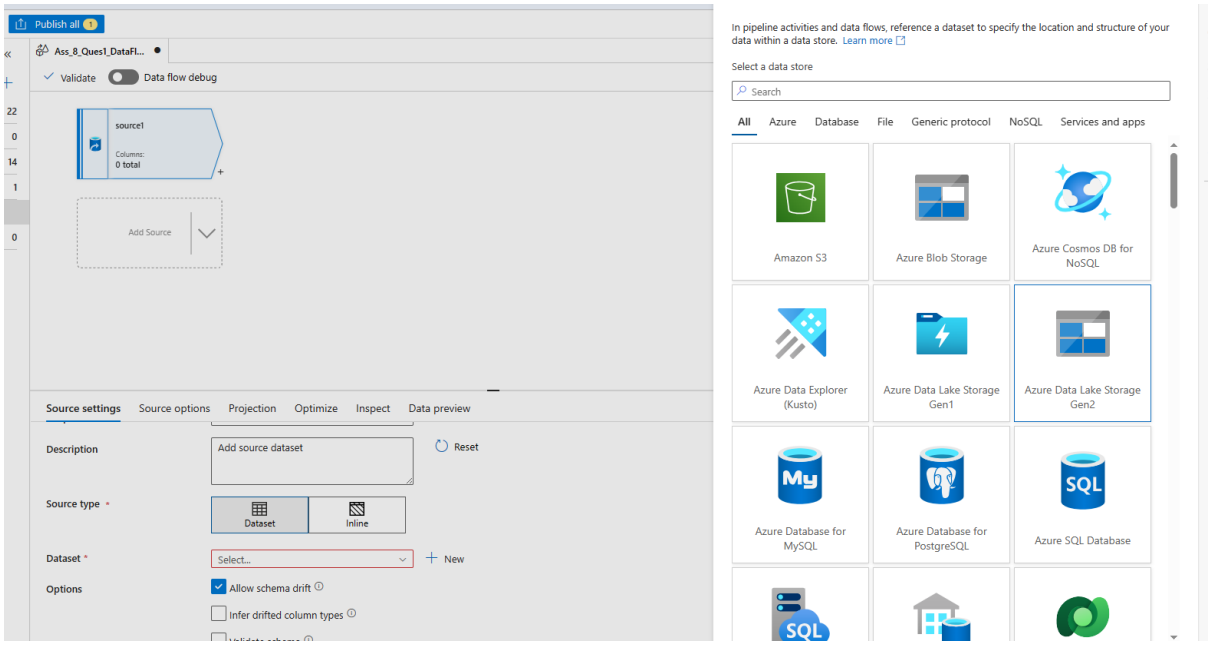

https://portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptic
Microsoft Azure
Upgrade
Search resources, services, and docs
Home > Storage accounts > gopi01012024adls1 | Containers >
landing
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: landing
Search blobs by prefix (case-sensitive)
Name
Modified
Ass_7_4_Table_HighWaterMarkFile
Ass_8_Ques_1_ADF_Dataflow
CSV_Pipe

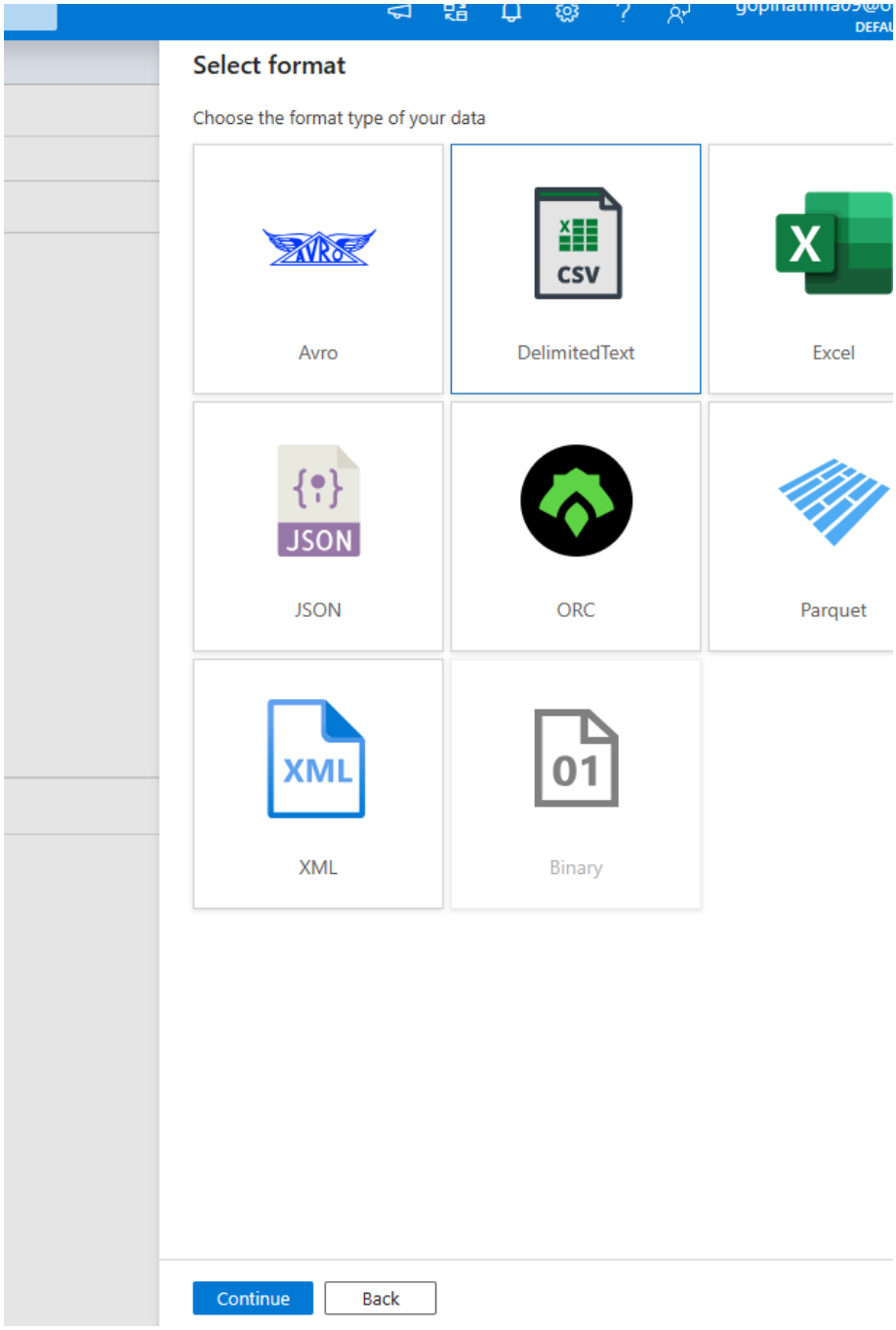

Upgrade
Search resources, services, and docs (G+/)
pi01012024adls1 | Containers
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: landing / Ass_8_Ques_1_ADF_Dataflow
Search blobs by prefix (case-sensitive)
Name
Modified
Access tier
Archive status
[..]
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

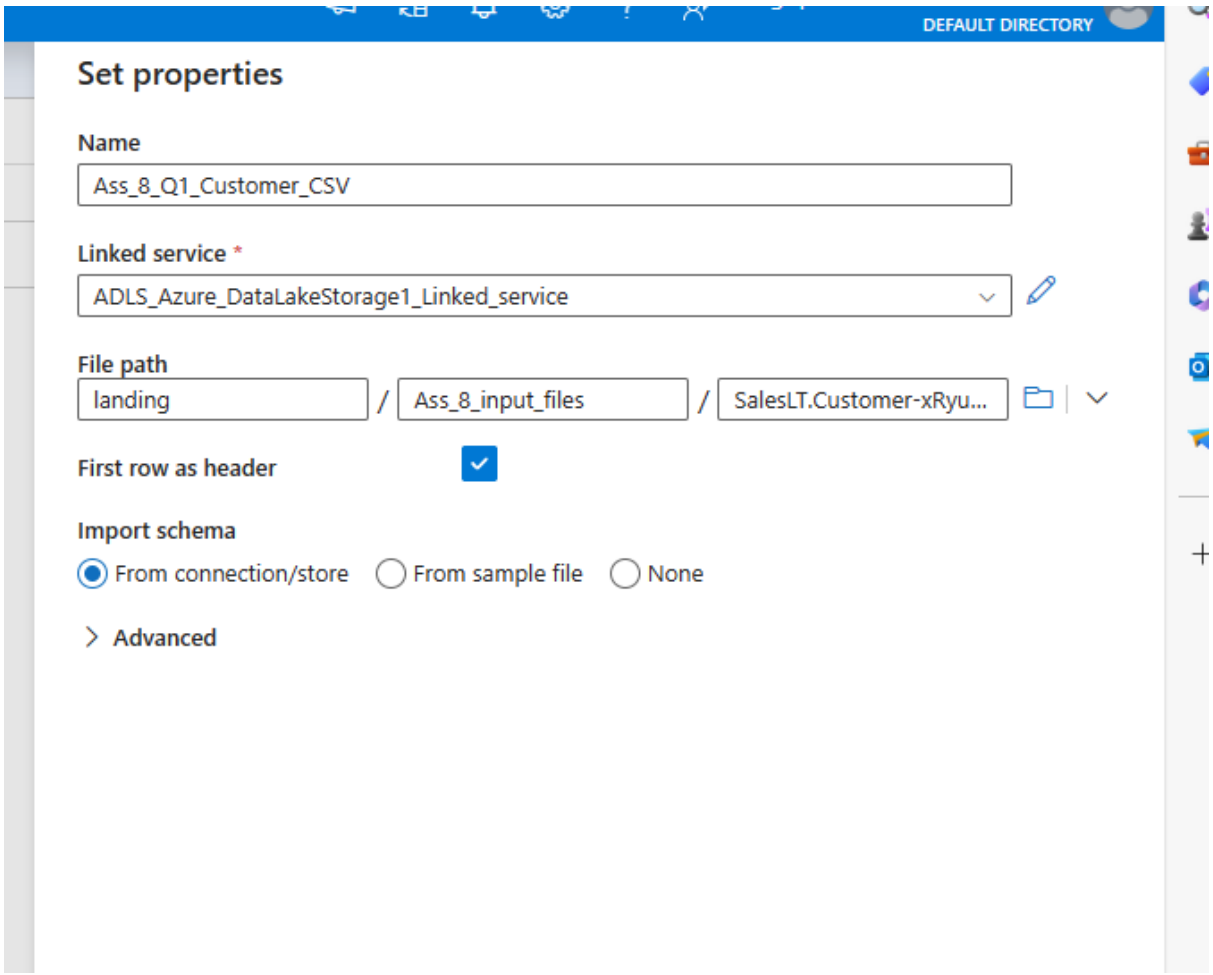

Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
ne > Storage accounts > gopi01012024adls1 | Containers
landing
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
ings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Upload
Add Directory
Refresh
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
Search blobs by prefix (case-sensitive)
Name
Modified
Access tier
Archive status
SalesLT.Customer-xRyuaEu2p7-WdCbzuWIlo-80V7BfCQtP.txt
2/27/2024, 4:49:22 PM
Hot (Inferred)
SalesLT.CustomerAddress-nnCI751651 (1)-0A4ckSV5gM-jBPnfa8HWQ.txt
2/27/2024, 4:48:53 PM
Hot (Inferred)
SalesLT.Product-760RKVH4kL-x8c0uQ9qBG-pul1LFOEyO.txt
2/27/2024, 4:48:54 PM
Hot (Inferred)



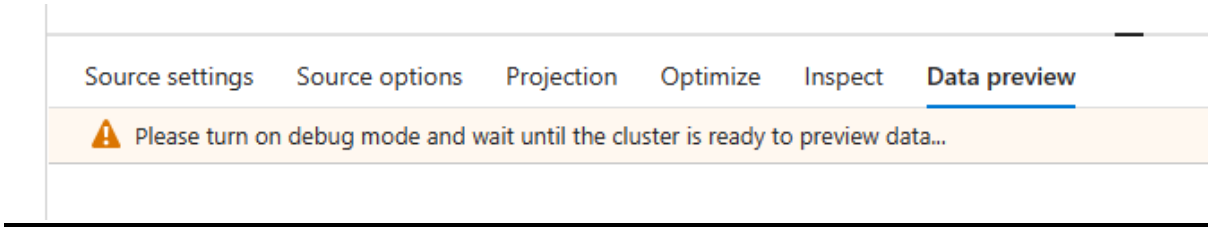
Now from DataFlow – laod this Dataset of CSV file from ADLS containter

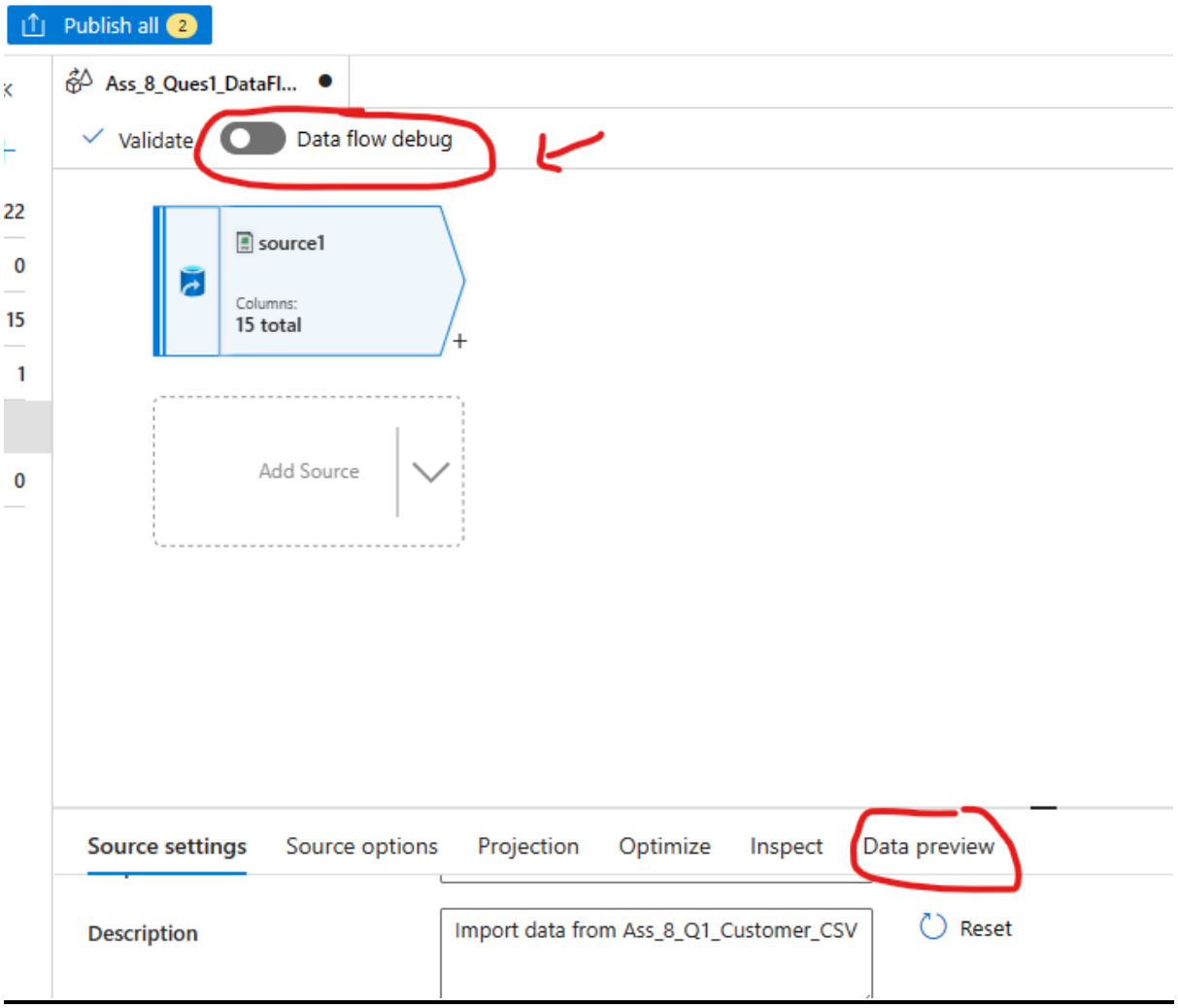


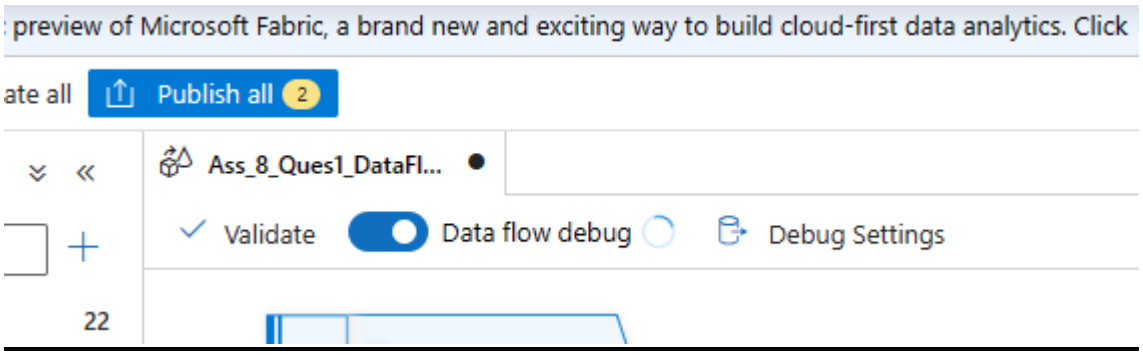

Select format
Choose the format type of your data
Avro
DelimitedText
Excel
JSON
ORC
Parquet
XML
Binary
Continue
Back

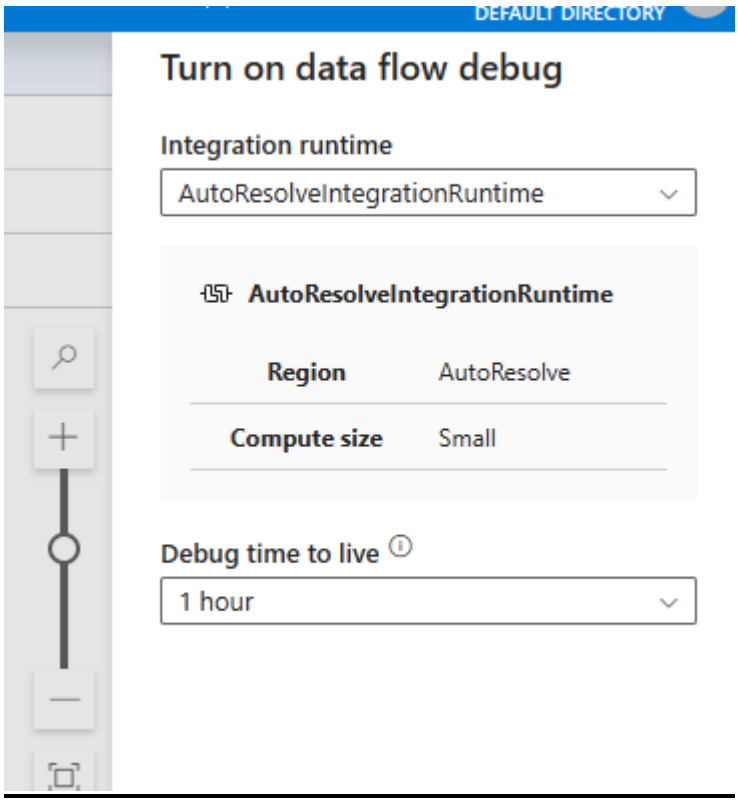


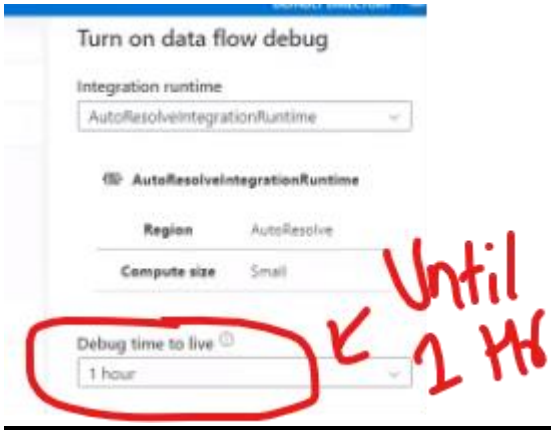
I cannot "Data Preview" my data , until I enable the "Data flow debug" button

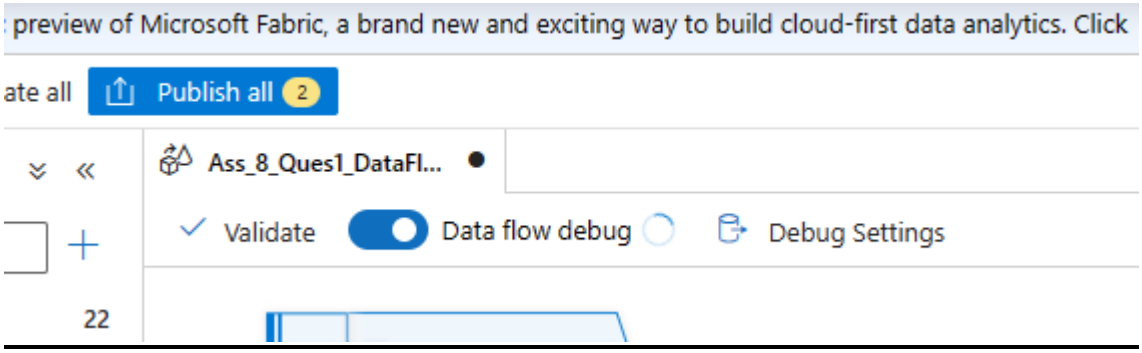


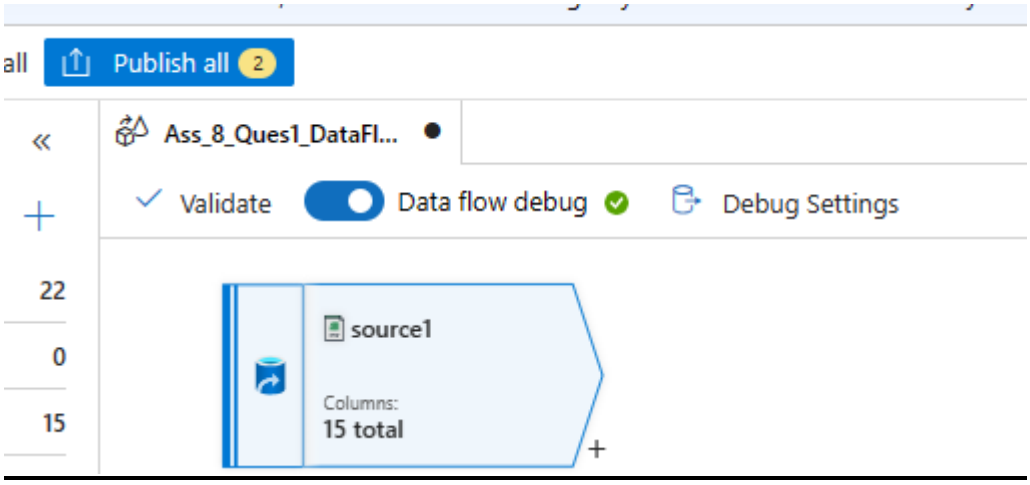

Publish all 2
Ass_8_Ques1_DataFl...
Validate
Data flow debug
source1
Columns:
15 total
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Description
Import data from Ass_8_Q1_Customer_CSV
Reset

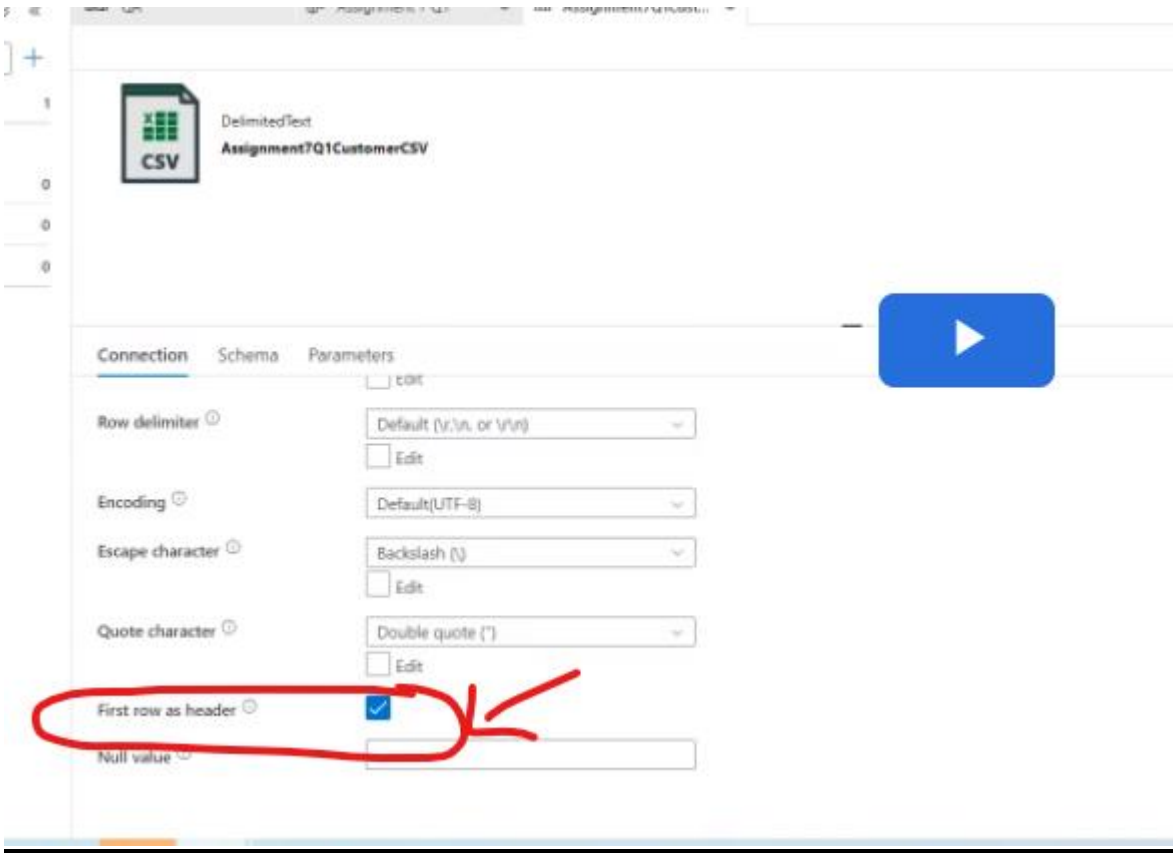

preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click
ate all
Publish all 2
Ass_8_Ques1_DataFl...
Validate
Data flow debug
Debug Settings
22

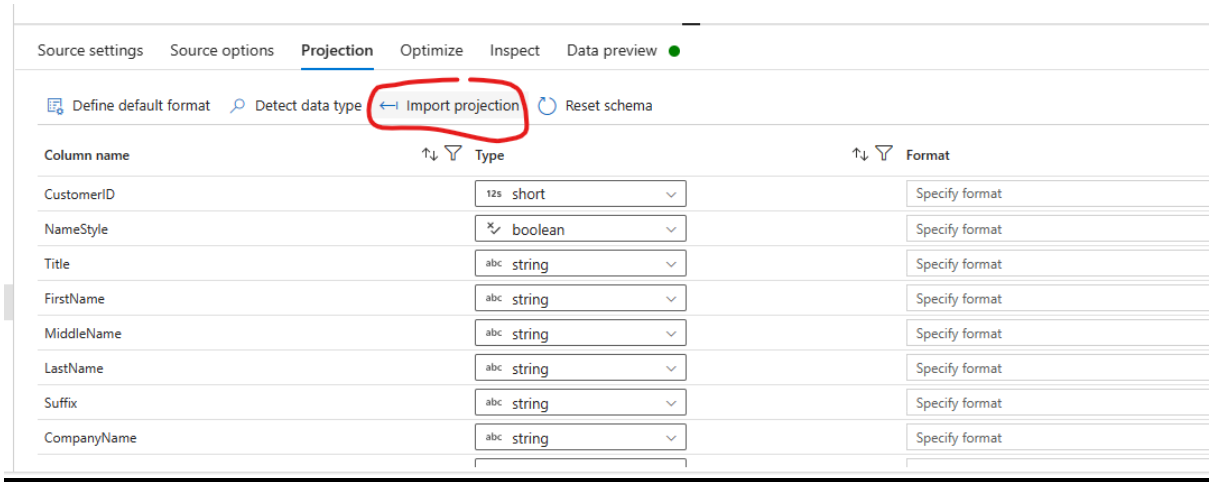

DEFAULT DIRECTORY
Turn on data flow debug
Integration runtime
AutoResolveIntegrationRuntime
AutoResolveIntegrationRuntime
Region
AutoResolve
Compute size
Small
Debug time to live
1 hour

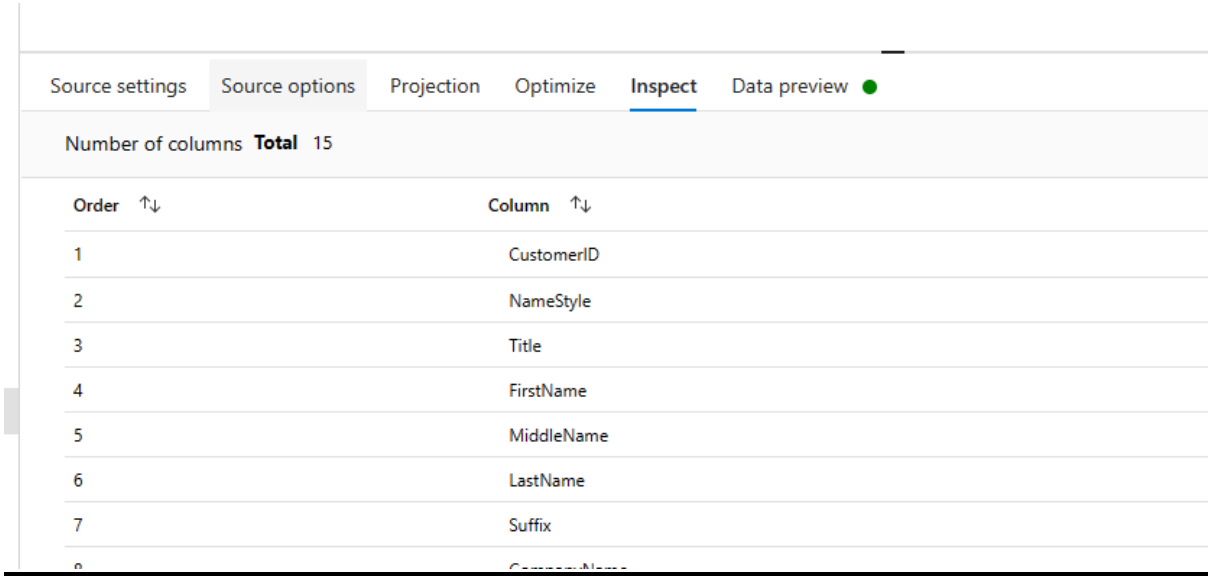

Turn on data flow debug
Integration runtime
AutoResolveIntegrationRuntime
AutoResolveIntegrationRuntime
Region
AutoResolve
Compute size
Small
Debug time to live
1 hour
Until
2 Hr

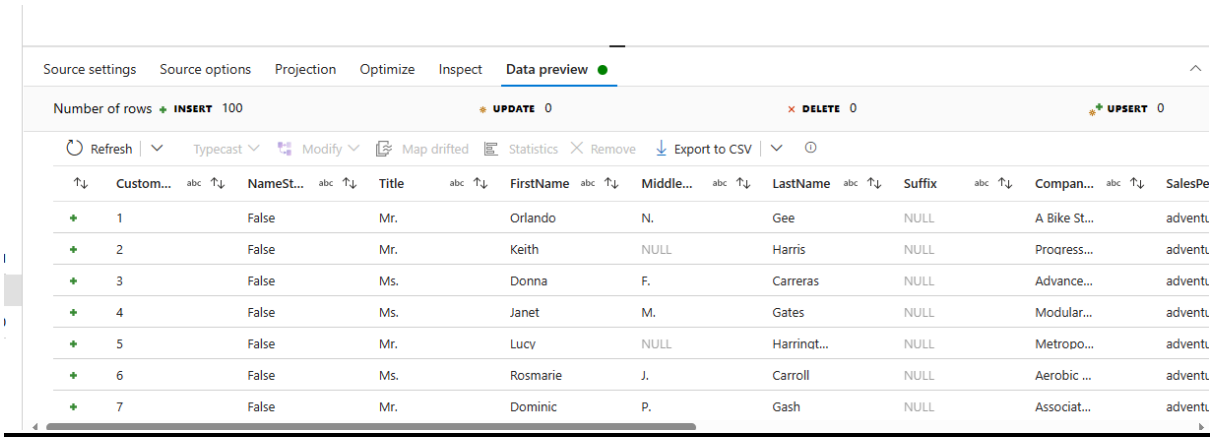

preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click
ate all
Publish all 2
Ass_8_Ques1_DataFl...
Validate
Data flow debug
Debug Settings
22


all
Publish all 2
Ass_8_Ques1_DataFl...
Validate
Data flow debug
Debug Settings
22
0
15
source1
Columns:
15 total

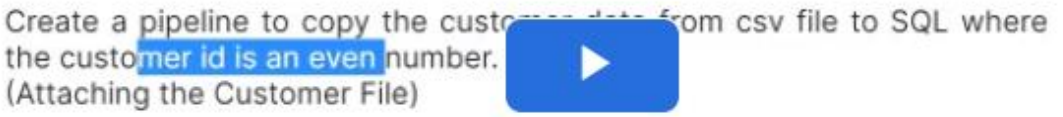

CSV
DelimitedText
Assignment7Q1CustomerCSV
Connection
Schema
Parameters
Row delimiter
Default (\r,\n, or \r\n)
Edit
Encoding
Default(UTF-8)
Escape character
Backslash (\)
Edit
Quote character
Double quote (")
Edit
First row as header
Null value

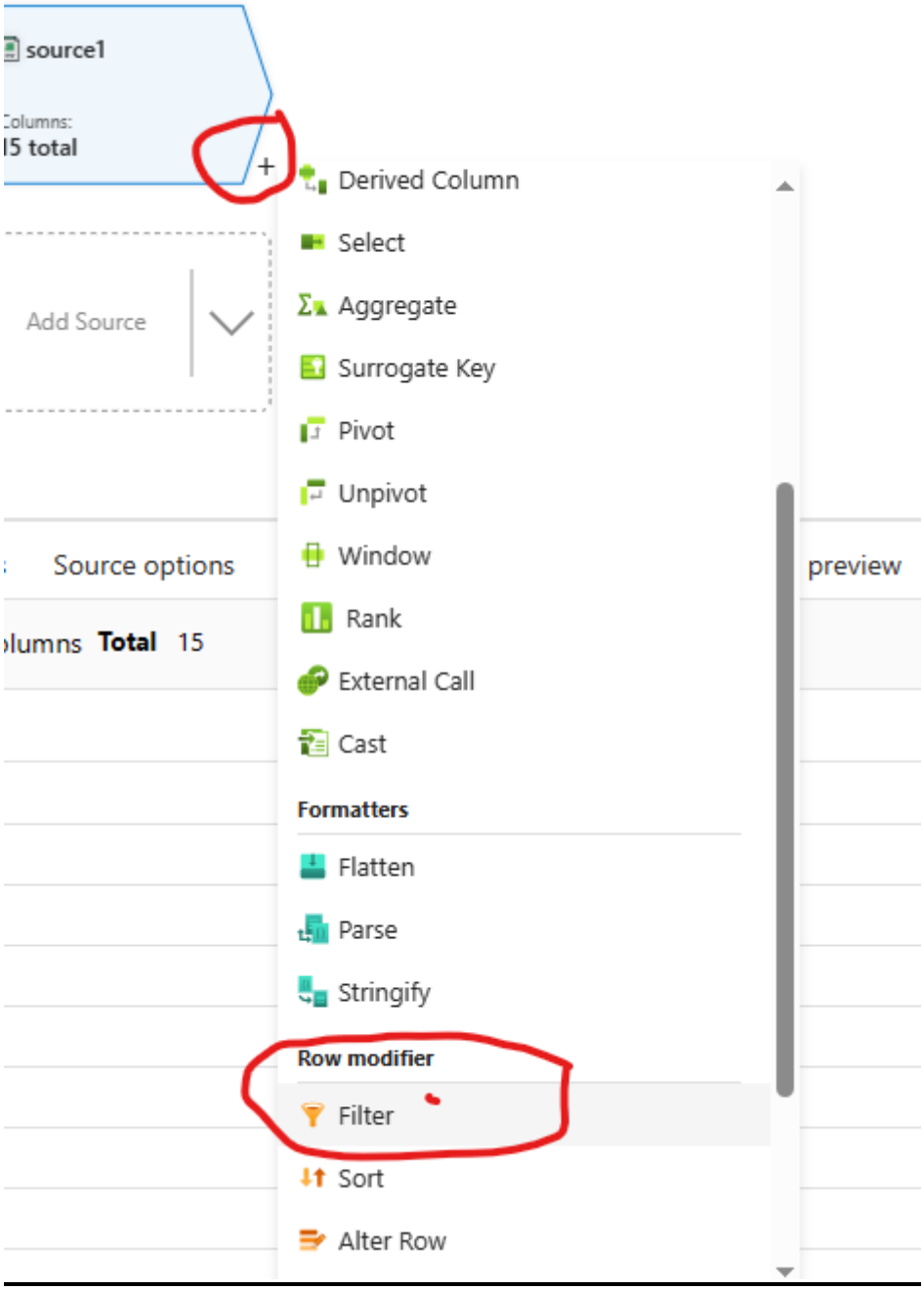

Source settings
Source options
Projection
Optimize
Inspect
Data preview
Define default format
Detect data type
Import projection
Reset schema
Column name
Type
Format
CustomerID
short
NameStyle
boolean
Title
string
FirstName
string
MiddleName
string
LastName
string
Suffix
string
CompanyName
string
Specify format

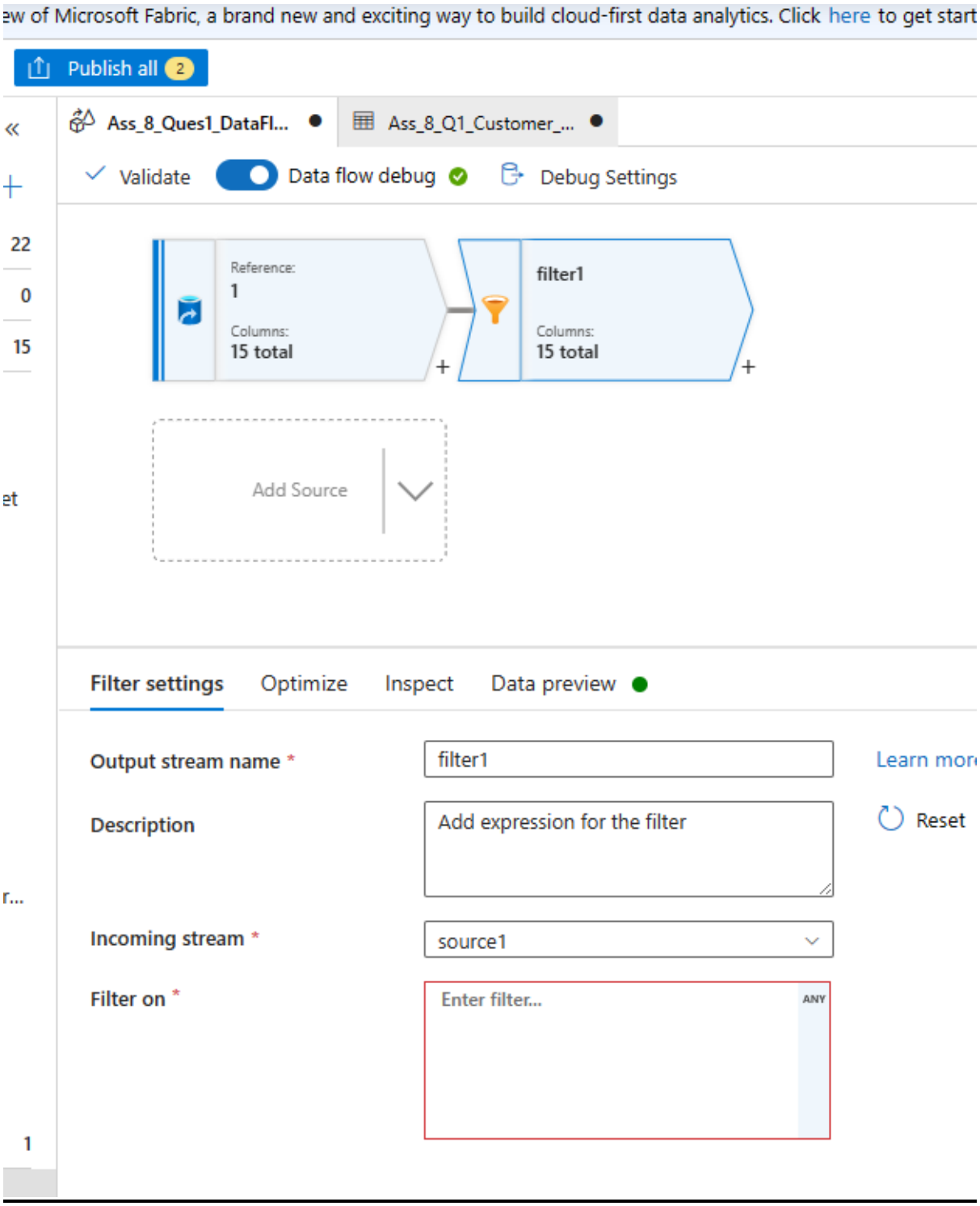

Source settings
Source options
Projection
Optimize
Inspect
Data preview
Number of columns Total 15
Order
Column
1 CustomerID
2 NameStyle
3 Title
4 FirstName
5 MiddleName
6 LastName
7 Suffix

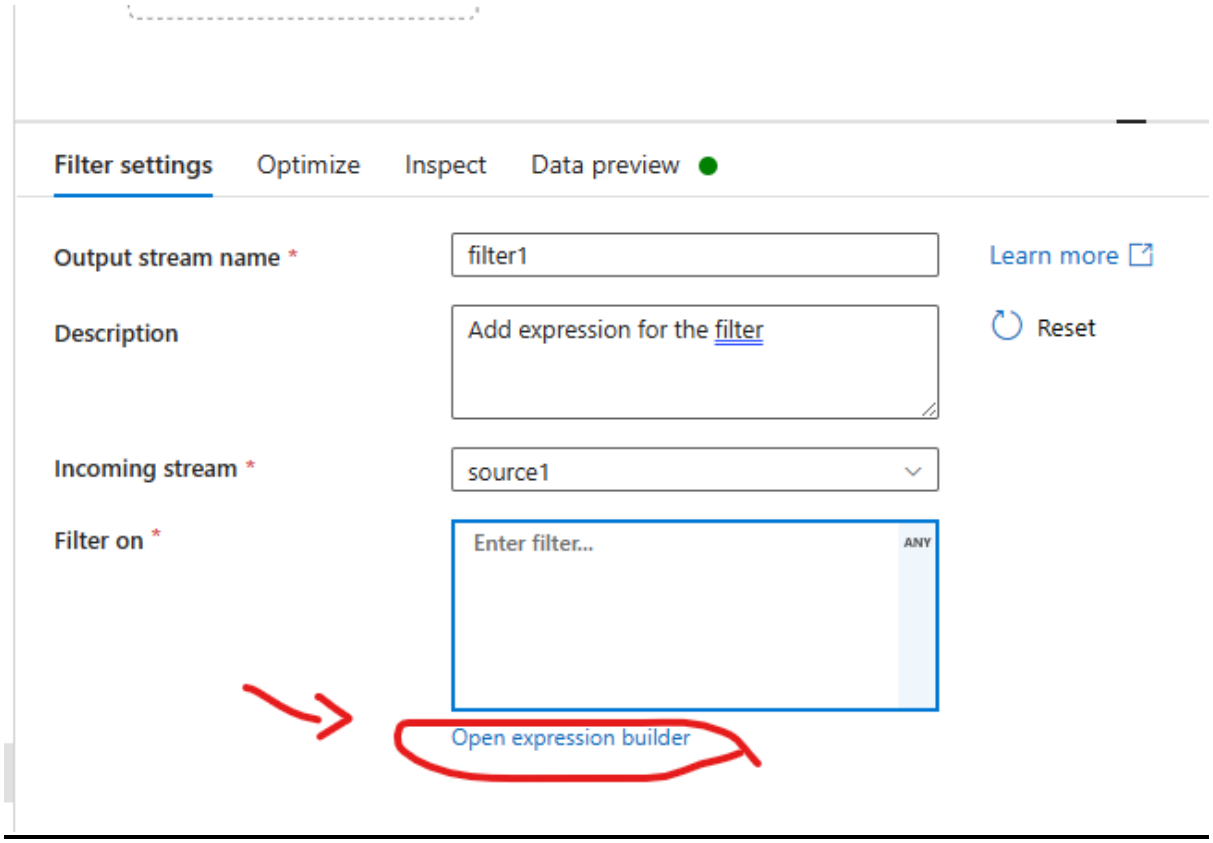

Source settings
Source options
Projection
Optimize
Inspect
Data preview
Number of rows INSERT 100
UPDATE 0
DELETE 0
UPSERT 0
Refresh
Typecast
Modify
Map drifted
Statistics
Remove
Export to CSV
Custom...
NameSt...
Title
FirstName
Middle...
LastName
Suffix
Compan...
SalesPe
1 False Mr. Orlando N. Gee NULL 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

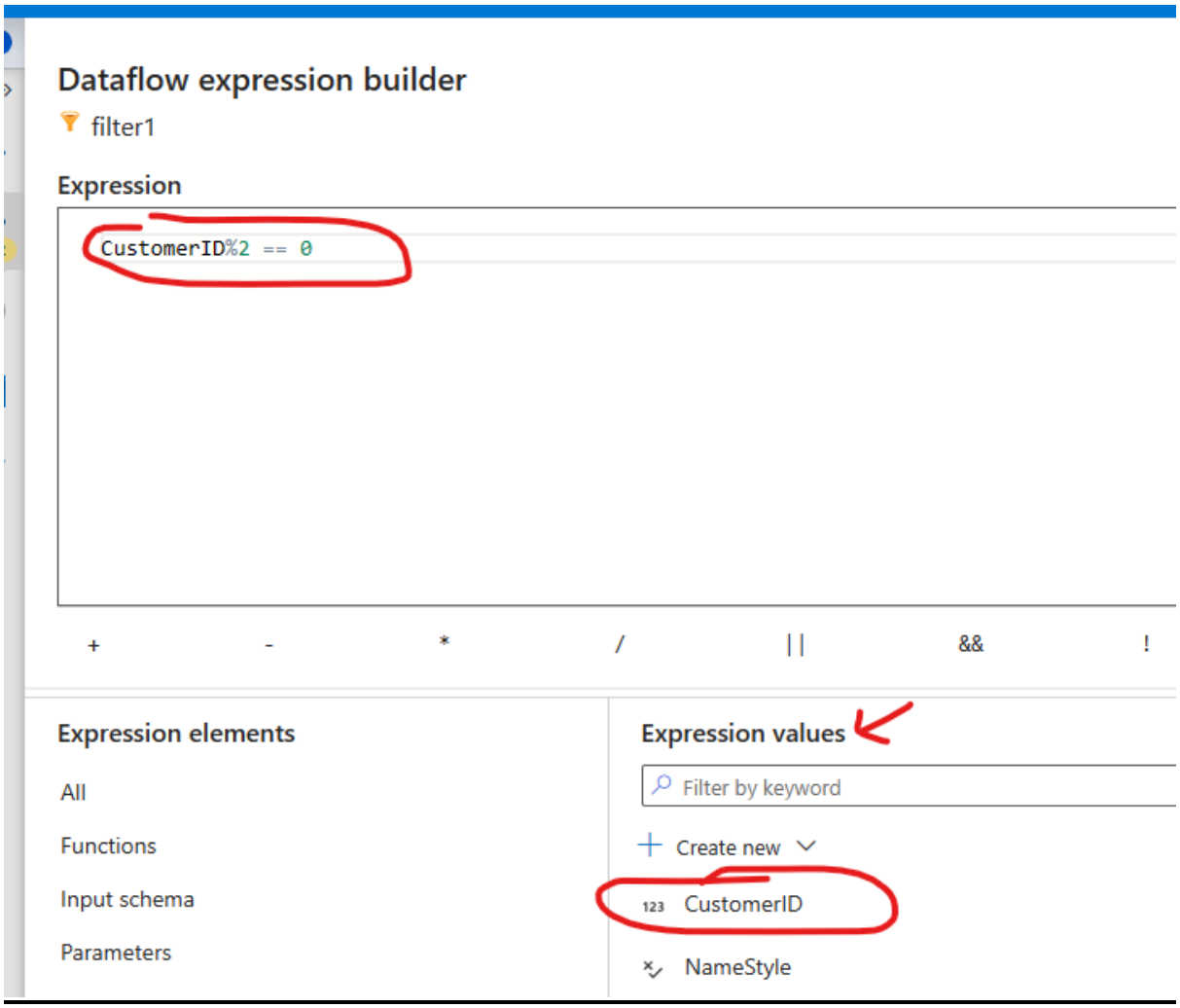
Next stept is to pick customer ID == Even number? SO I need to "Filter" it for even number--

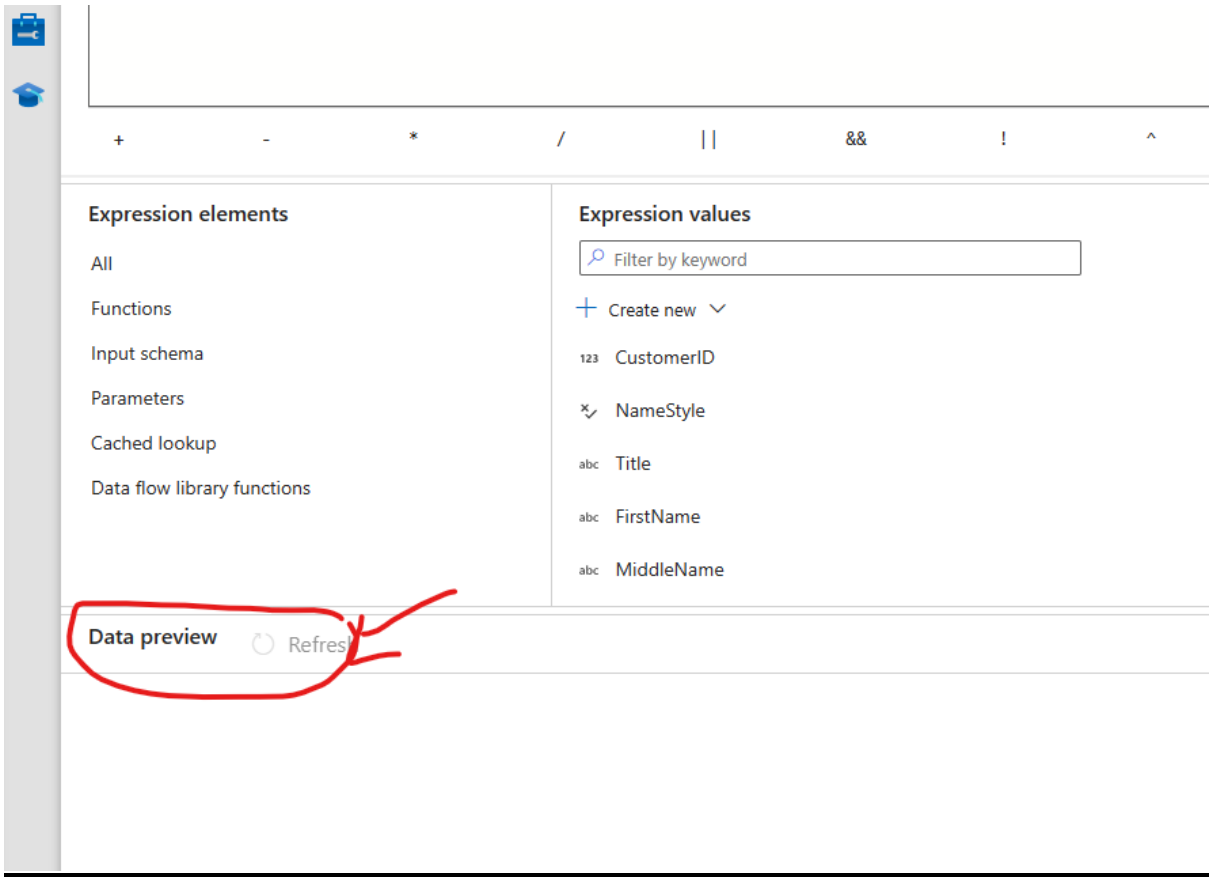


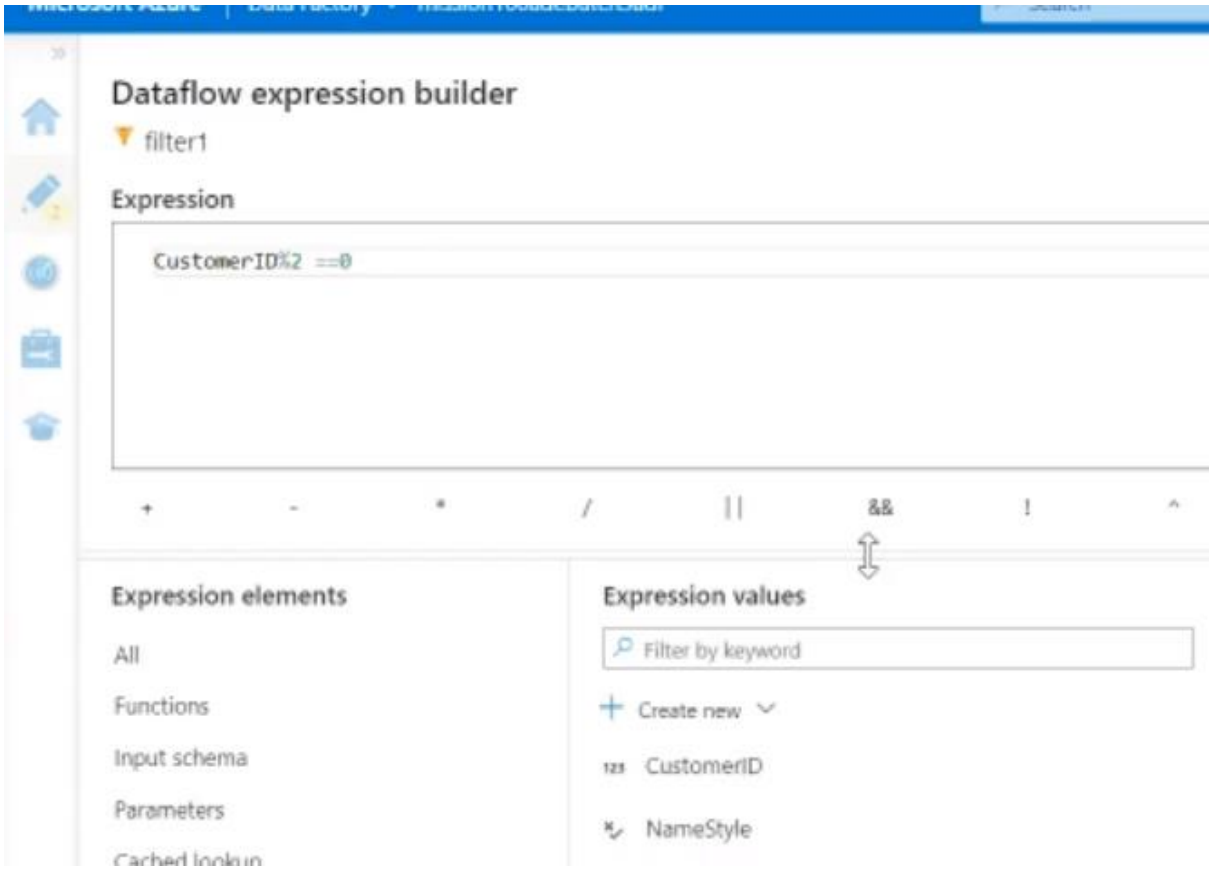


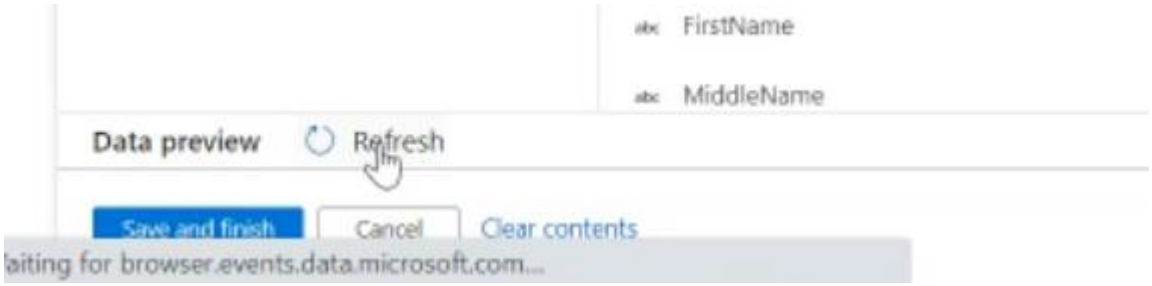

ew of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get start
Publish all 2
Ass_8_Ques1_DataFl...
Ass_8_Q1_Customer_...
Validate
Data flow debug
Debug Settings
Reference:
1
Columns:
15 total
filter1
Columns:
15 total
Add Source
Filter settings
Optimize
Inspect
Data preview
Output stream name *
filter1
Learn mor
Description
Add expression for the filter
Reset
Incoming stream *
source1
Filter on *
Enter filter...
ANY


Filter settings
Optimize
Inspect
Data preview
Output stream name *
filter1
Learn more
Description
Add expression for the filter
Reset
Incoming stream *
source1
Filter on *
Enter filter...
ANY
Open expression builder

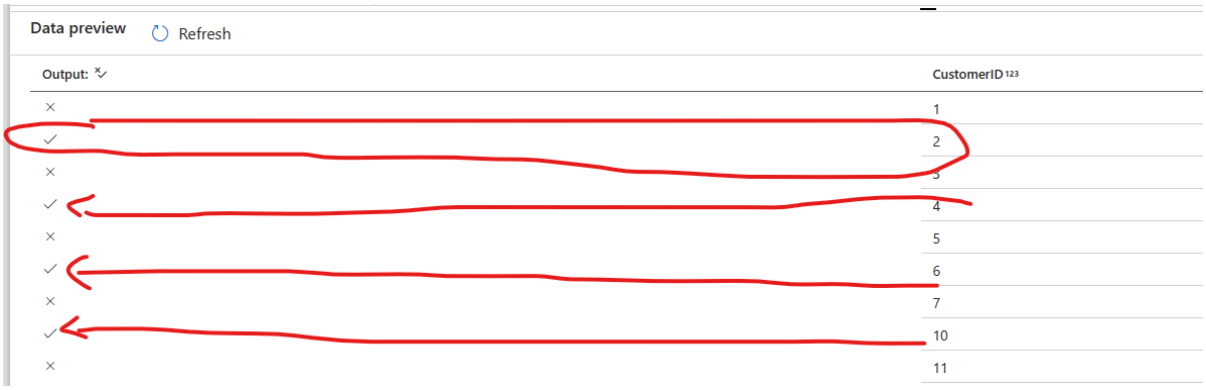

Dataflow expression builder
filter1
Expression
CustomerID%2 == 0
+
-
*
/
||
&&
!
Expression elements
All
Functions
Input schema
Parameters
Expression values
Filter by keyword
Create new
CustomerID
NameStyle


+ - * / || && ! ^
Expression elements
All
Functions
Input schema
Parameters
Cached lookup
Data flow library functions
Expression values
Filter by keyword
Create new
CustomerID
NameStyle
Title
FirstName
MiddleName
Data preview
Refresh

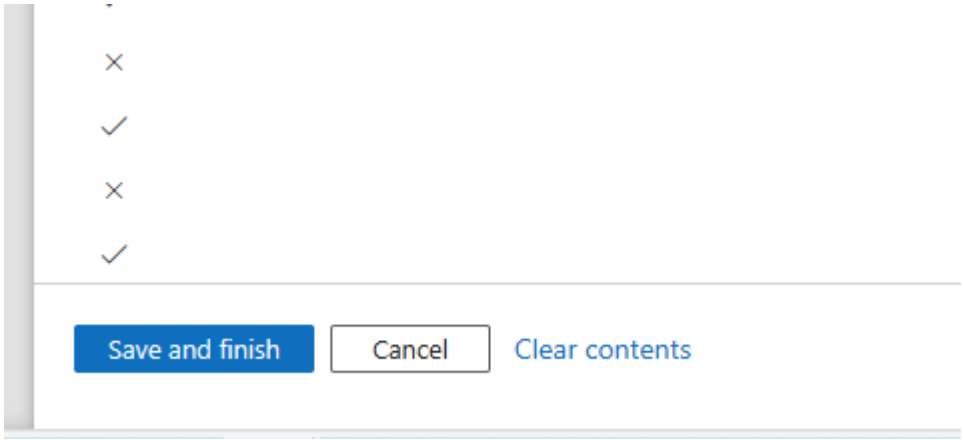

Dataflow expression builder
filter1
Expression
CustomerID%2 ==0
+ - * / || && ! ^
Expression elements
All
Functions
Input schema
Parameters
Expression values
Filter by keyword
Create new
CustomerID
NameStyle

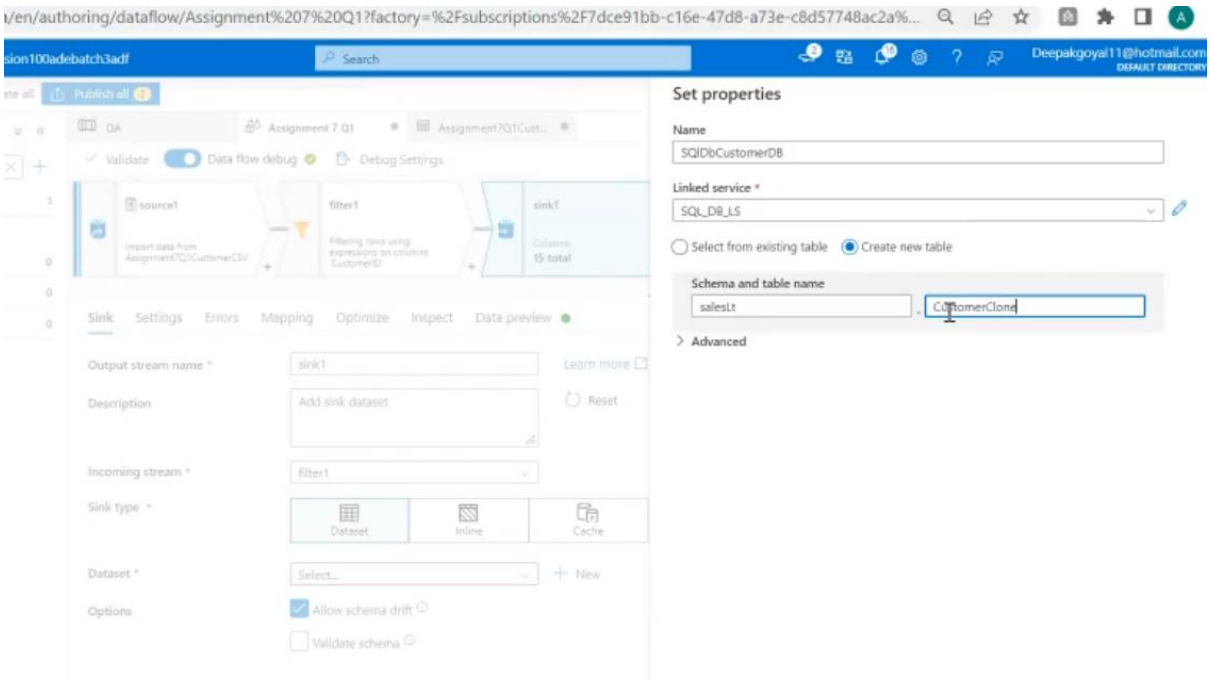


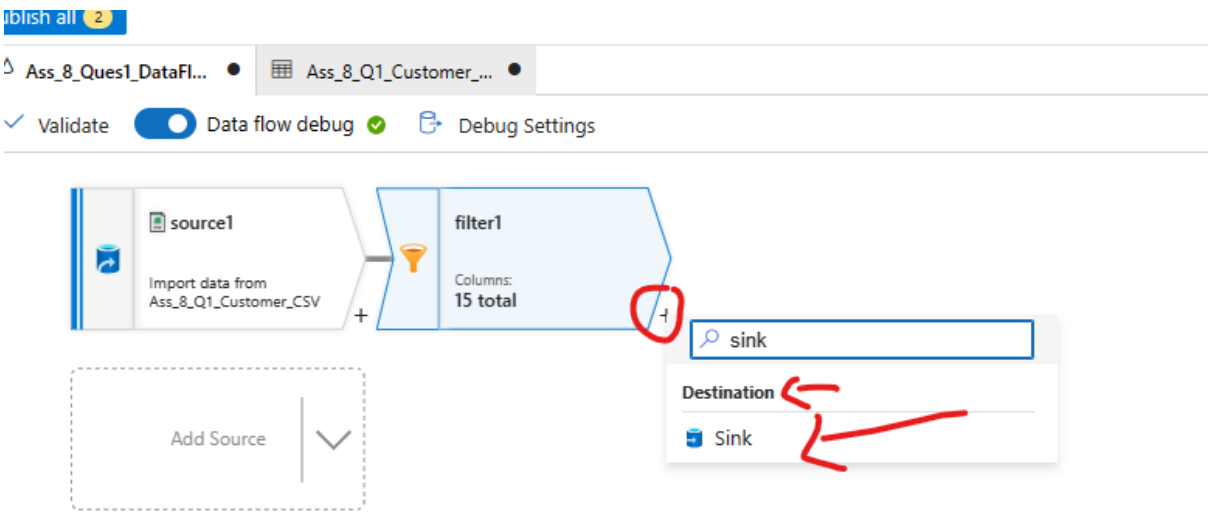
Refresh the Data Preview shows the preview data:

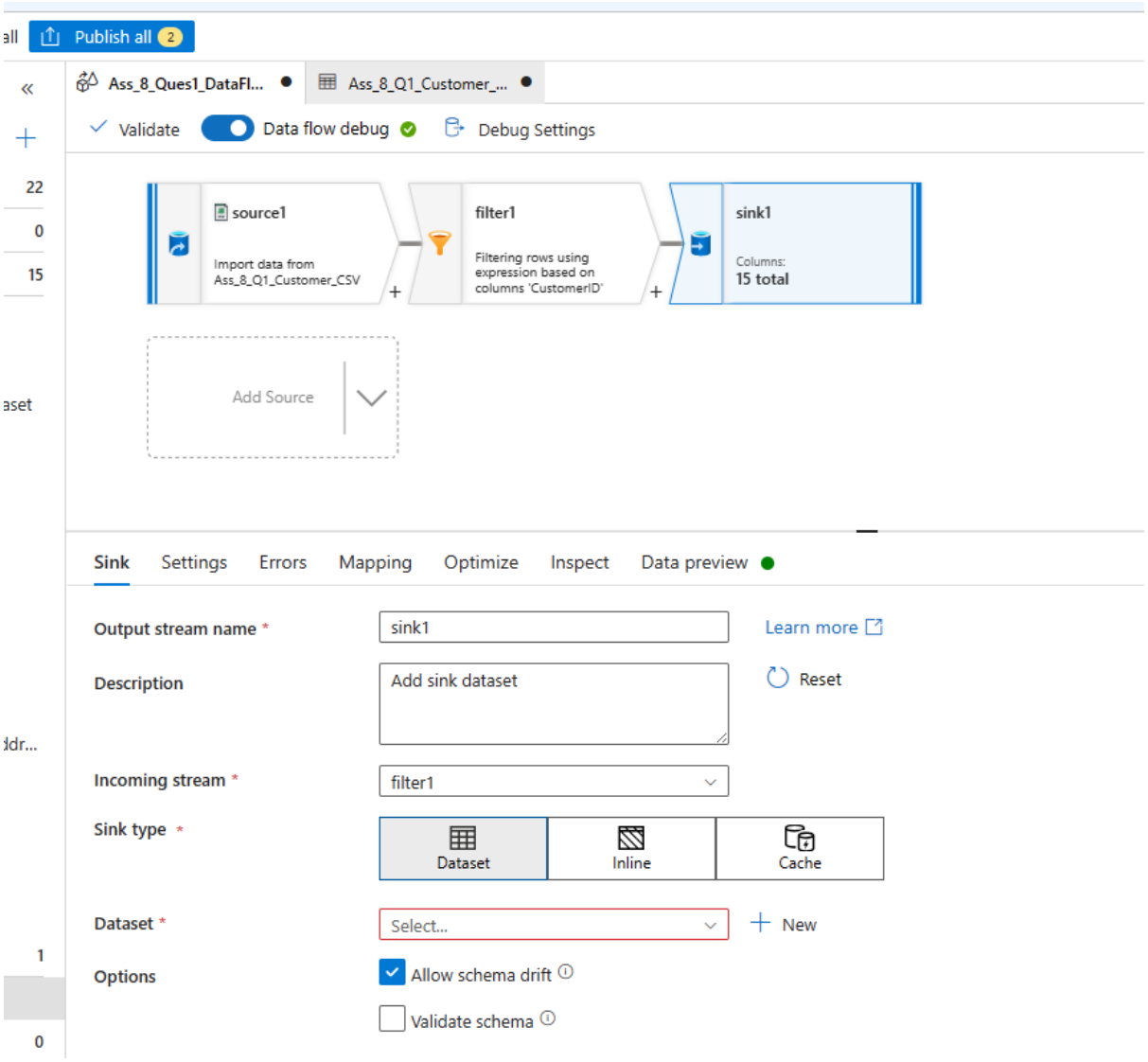

Data preview
Refresh
Output:
CustomerID 123
1
2
3
4
5
6
7
10
11

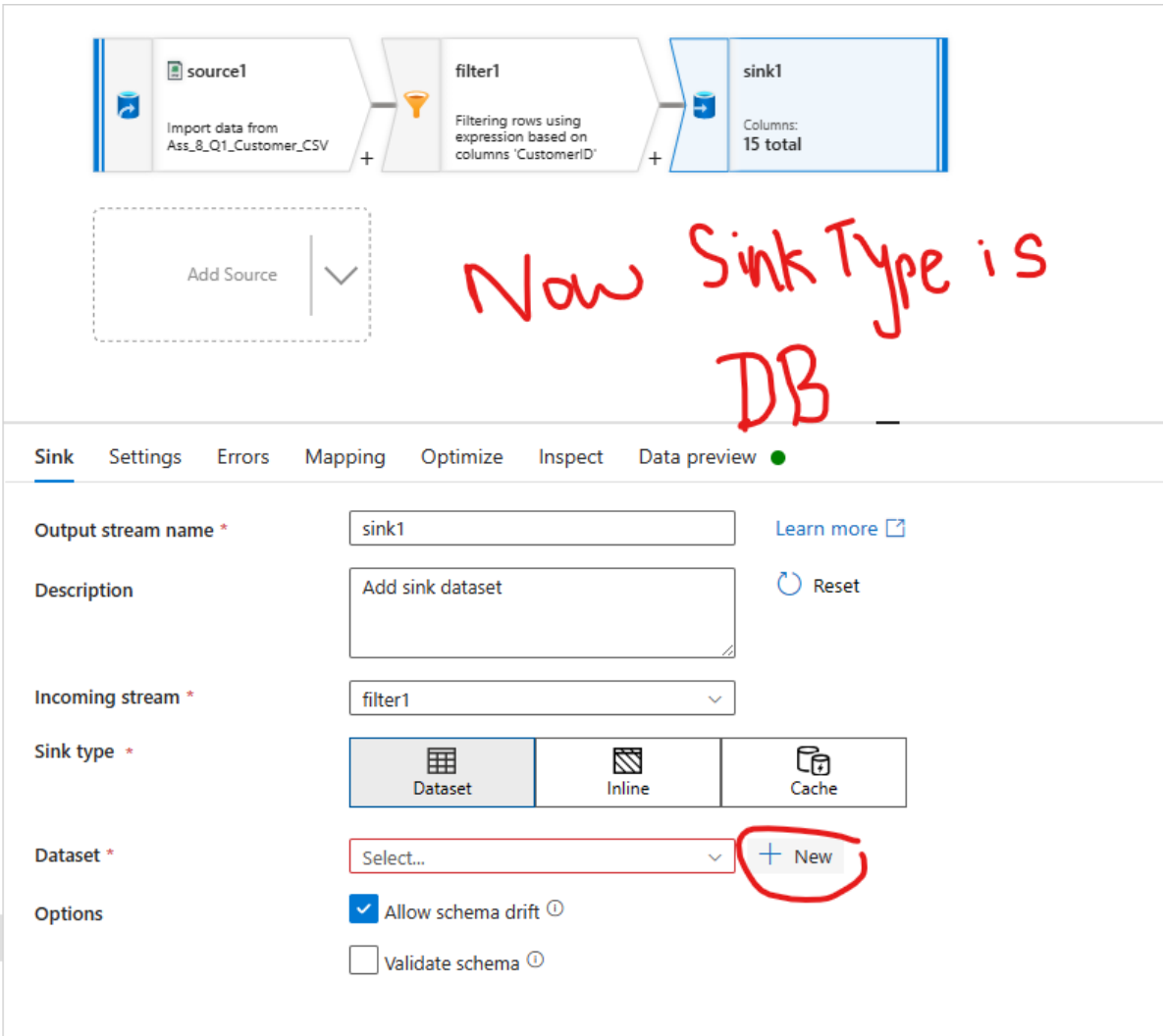

Data preview
Refresh
Output:
CustomerID 123
1
2
3


Save and finish
Cancel
Clear contents


Set properties
Name
SQlDbCustomerDB
Linked service *
SQL_DB_LS
Select from existing table
Create new table
Schema and table name
salesLt
CustomerClone
Advanced
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Output stream name *
sink1
Description
Add sink dataset
Incoming stream *
filter1
Sink type *
Dataset
Inline
Cache
Dataset *
Select...
New
Options
Allow schema drift
Validate schema
Reset


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


Publish all 2
Ass_8_Ques1_DataFl...
Ass_8_Q1_Customer_...
Validate
Data flow debug
Debug Settings
source1
Import data from Ass_8_Q1_Customer_CSV
filter1
Filtering rows using expression based on columns 'CustomerID'
sink1
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
Inline
Cache
Dataset *
Select...
New
Options
Allow schema drift
Validate schema



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

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structur
data within a data store. Learn more

Select a data store

sql

All | Azure | Database | File | Generic protocol | NoSQL | Services and a

Azure Cosmos DB for NoSQL

Azure Database for MySQL

Azure Database for PostgreSQL

Azure SQL Database

Azure SQL Database Managed Instance

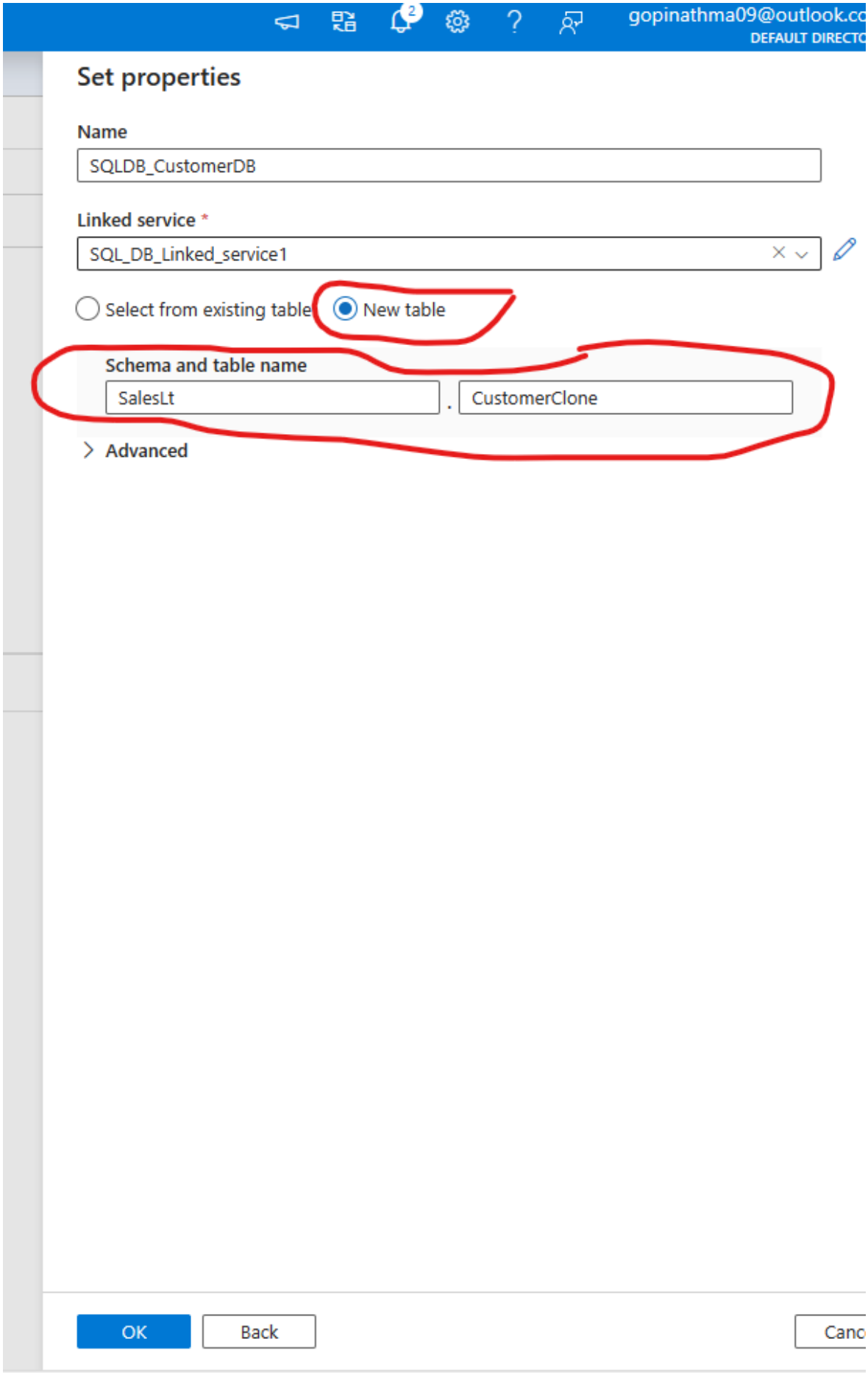
SQL server

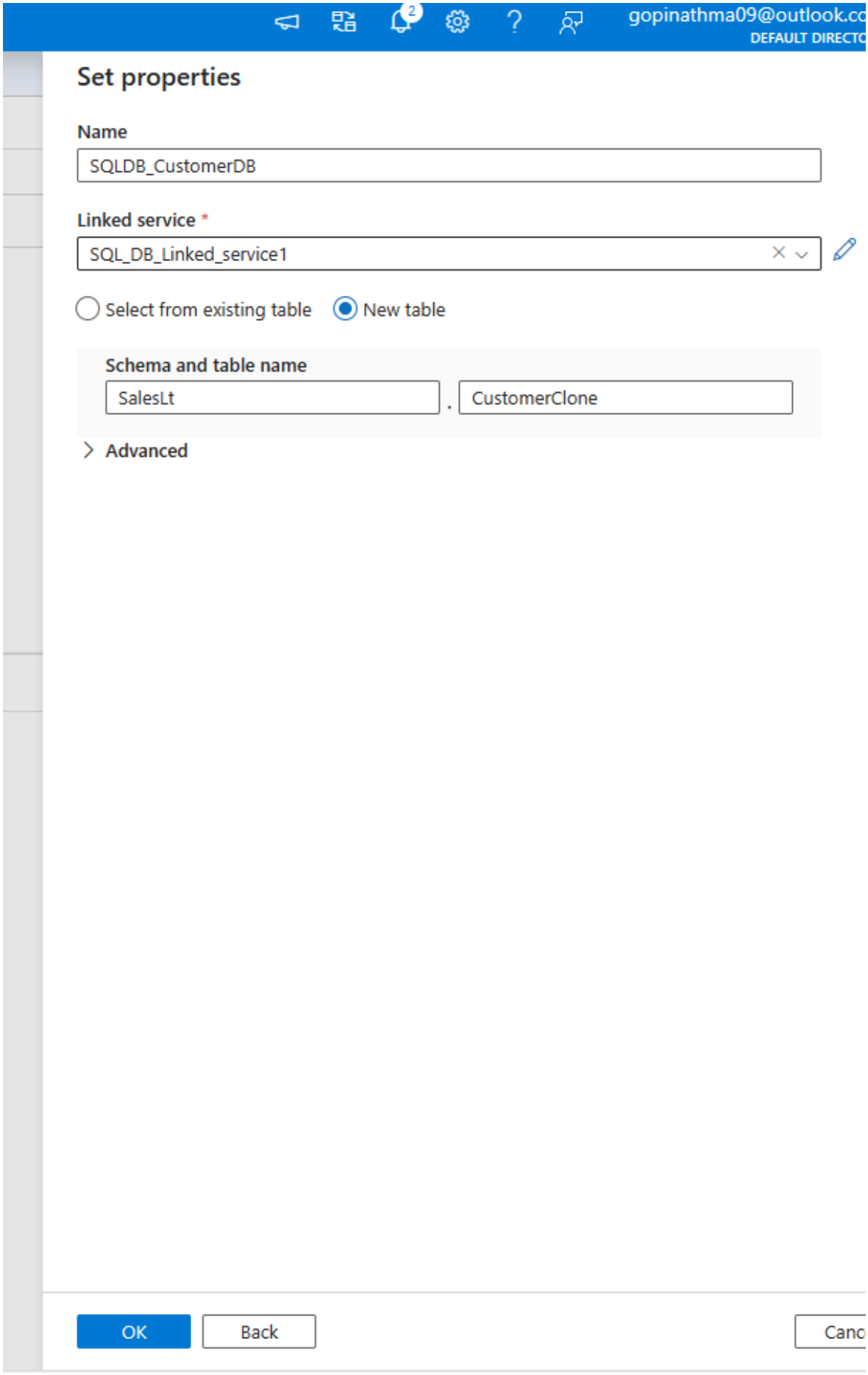
Amazon RDS for SQL Server

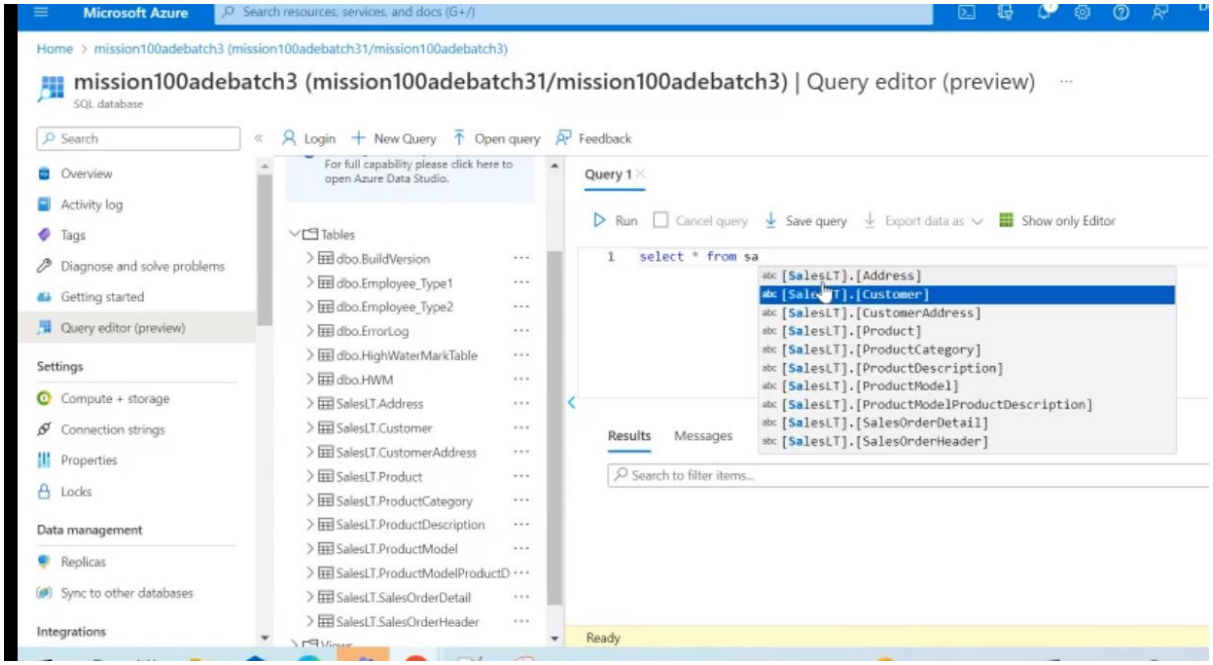
MySQL

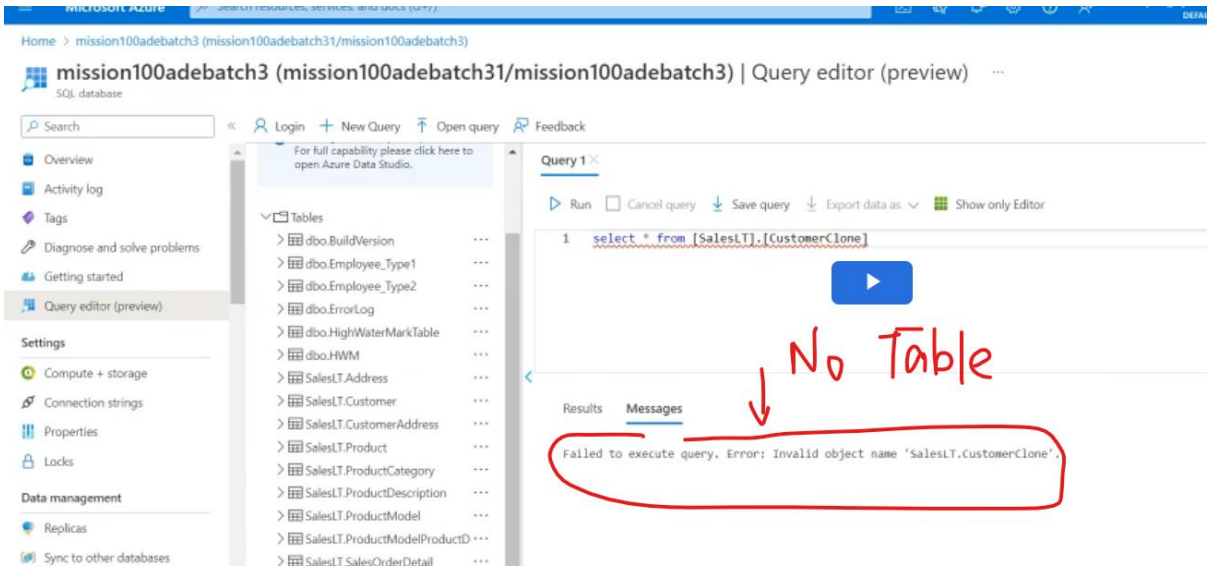
PostgreSQL

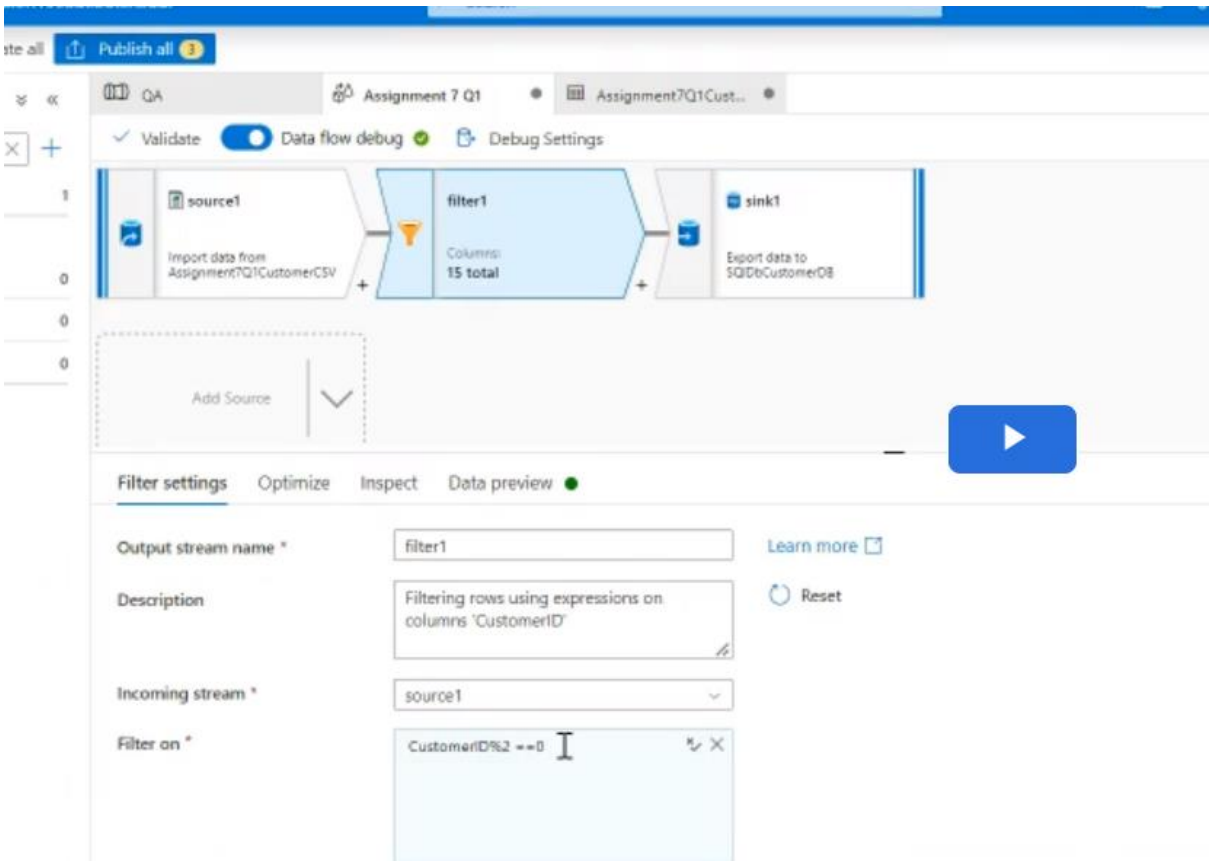
Continue

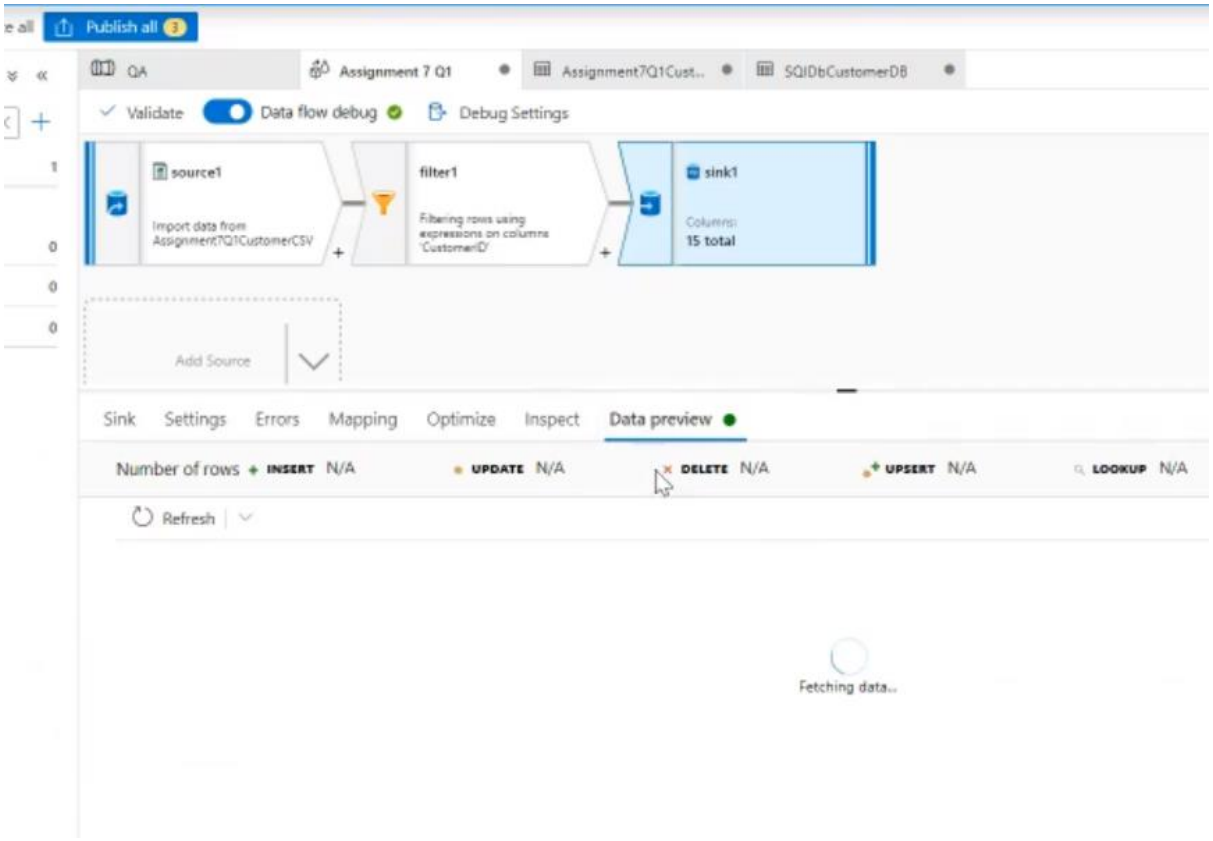

gopinathma09@outlook.co
DEFAULT DIRECTO
Set properties
Name
SQLDB_CustomerDB
Linked service *
SQL_DB_Linked_service1
Select from existing table
New table
Schema and table name
SalesLt
CustomerClone
Advanced
OK
Back
Canc

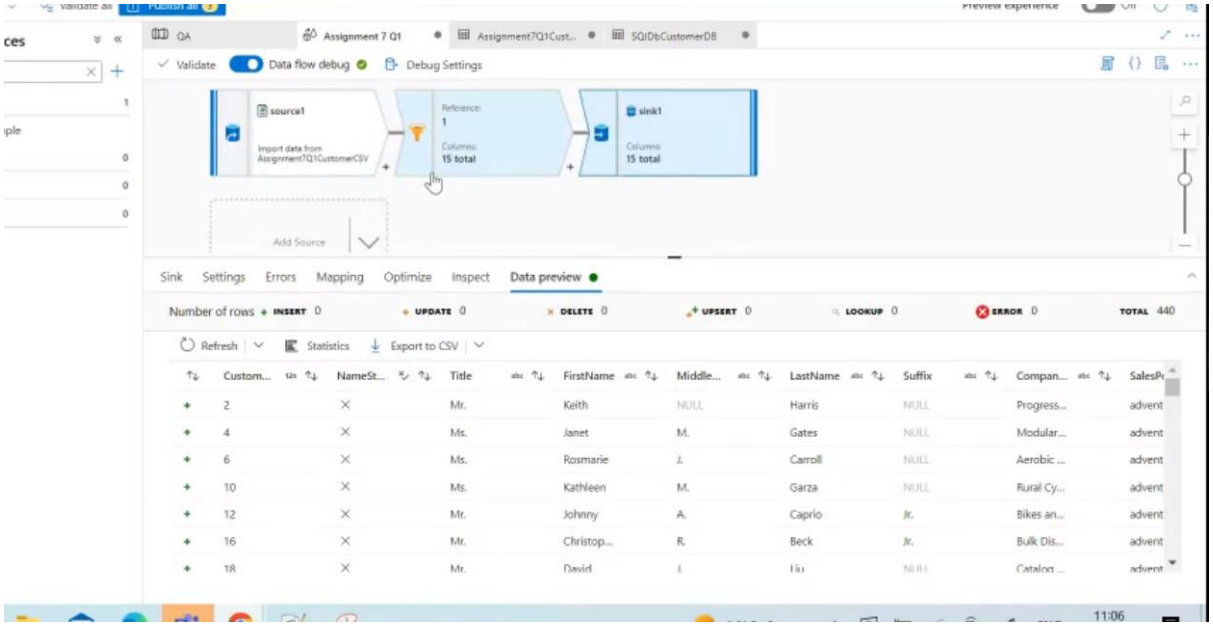

gopinathma09@outlook.co
DEFAULT DIRECTO
Set properties
Name
SQLDB_CustomerDB
Linked service *
SQL_DB_Linked_service1
Select from existing table
New table
Schema and table name
SalesLt
CustomerClone
Advanced
OK
Back
Canc

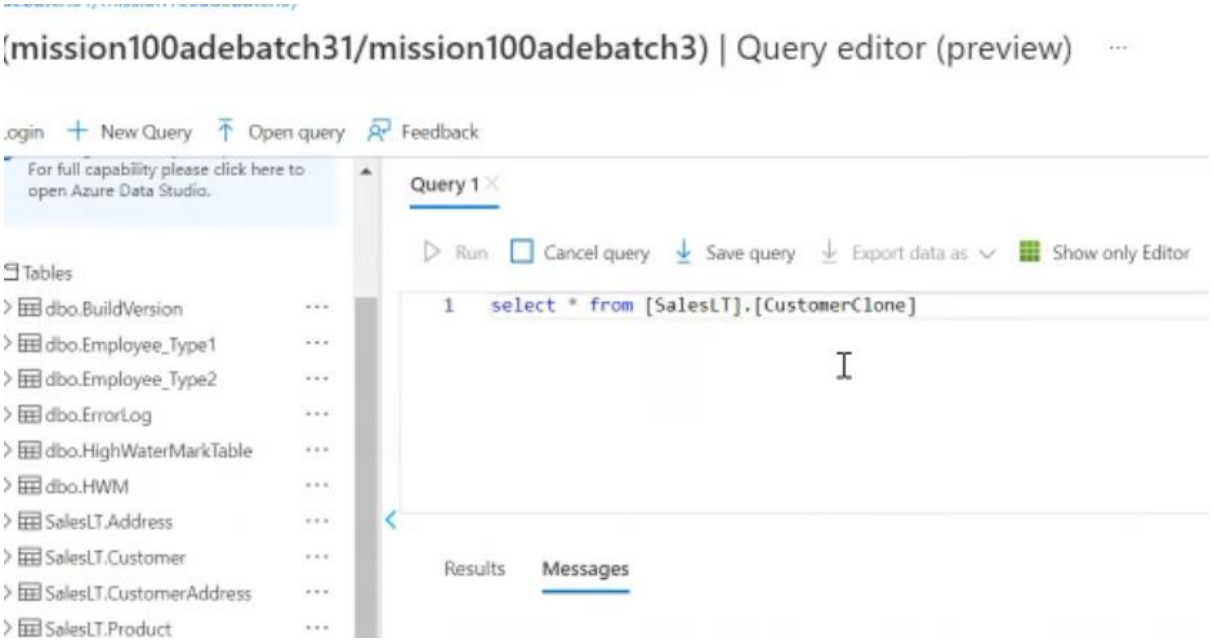

Microsoft Azure
Search resources, services, and docs (G+/)
Home > mission100adebatch3 (mission100adebatch31/mission100adebatch3)
mission100adebatch3 (mission100adebatch31/mission100adebatch3) | Query editor (preview)
SQL database
Search
Login
New Query
Open query
Feedback
Overview
Activity log
Tags
Diagnose 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
[SalesLT].[Address]
[SalesLT].[Customer]
[SalesLT].[CustomerAddress]
[SalesLT].[Product]
[SalesLT].[ProductCategory]
[SalesLT].[ProductDescription]
[SalesLT].[ProductModel]
[SalesLT].[ProductModelProductDescription]
[SalesLT].[SalesOrderDetail]
[SalesLT].[SalesOrderHeader]
Results
Messages
Search to filter items...
Ready

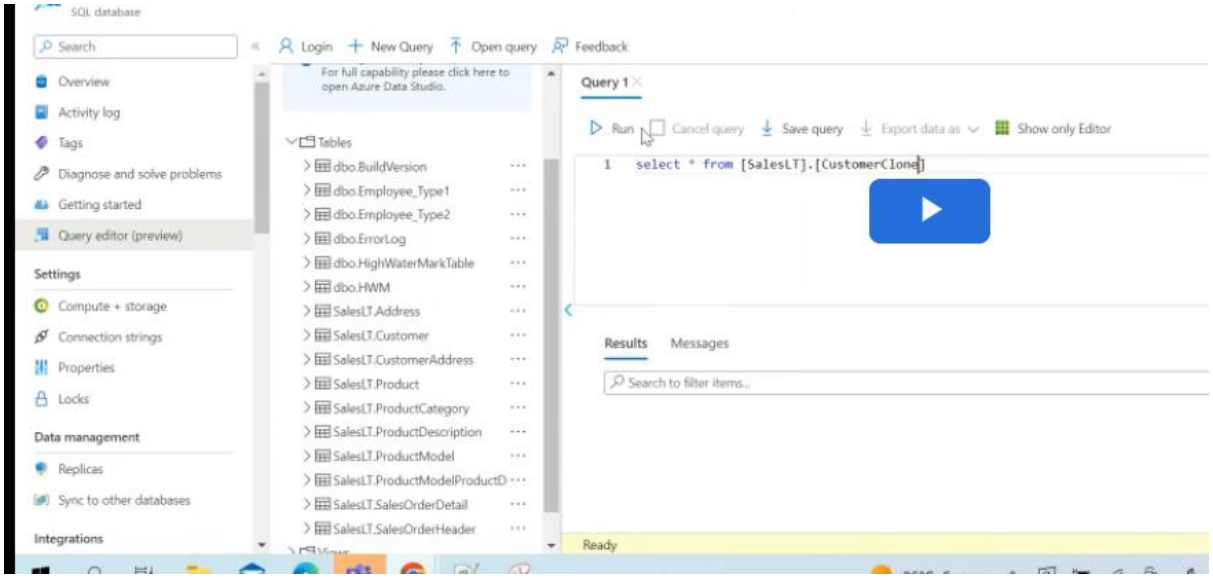

Home > mission100adebatch3 (mission100adebatch31/mission100adebatch3)
mission100adebatch3 (mission100adebatch31/mission100adebatch3) | Query editor (preview)
SQL database
Search
Login
New Query
Open query
Feedback
Overview
Activity log
Tags
Diagnose and solve problems
Getting started
Query editor (preview)
Settings
Compute + storage
Connection strings
Properties
Locks
Data management
Replicas
Sync to other databases
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'.

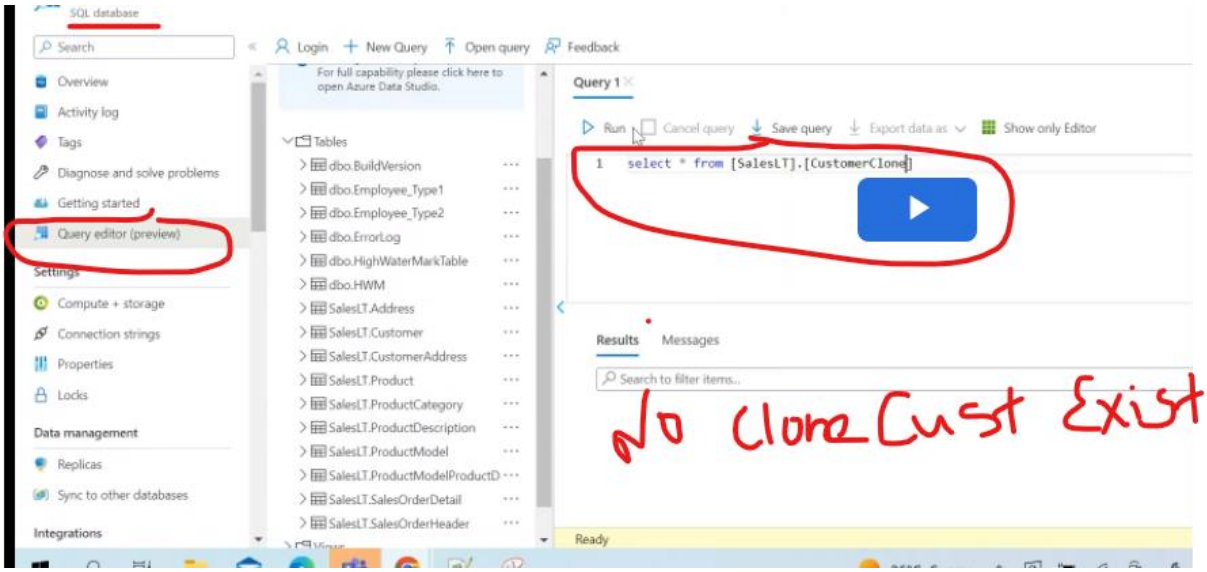

Publish all
QA
Assignment 7 Q1
Assignment7Q1Cust...
Validate
Data flow debug
Debug Settings
source1
Import data from Assignment7Q1CustomerCSV
filter1
Columns:
15 total
sink1
Export data to SQLDbCustomerDB
Add Source
Filter settings
Optimize
Inspect
Data preview
Output stream name *
filter1
Learn more
Description
Filtering rows using expressions on columns 'CustomerID'
Reset
Incoming stream *
source1
Filter on *
CustomerID%2 ==0

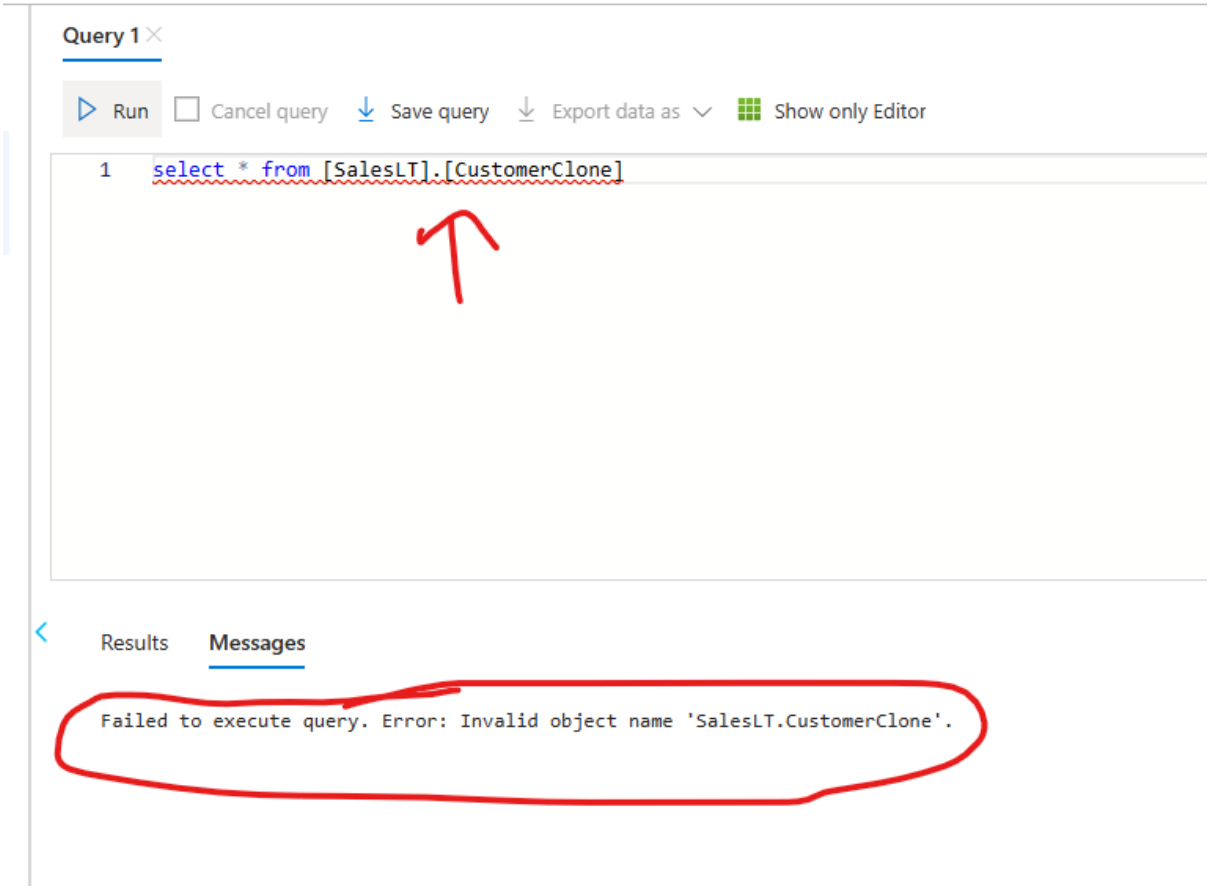

Publish all
QA
Assignment 7 Q1
Assignment7Q1Cust...
SQLDbCustomerDB
Validate
Data flow debug
Debug Settings
source1
Import data from Assignment7Q1CustomerCSV
filter1
Filtering rows using expressions on columns 'CustomerID'
sink1
Columns:
15 total
Add Source
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Number of rows
INSERT N/A
UPDATE N/A
DELETE N/A
UPSERT N/A
LOOKUP N/A
Refresh
Fetching data...

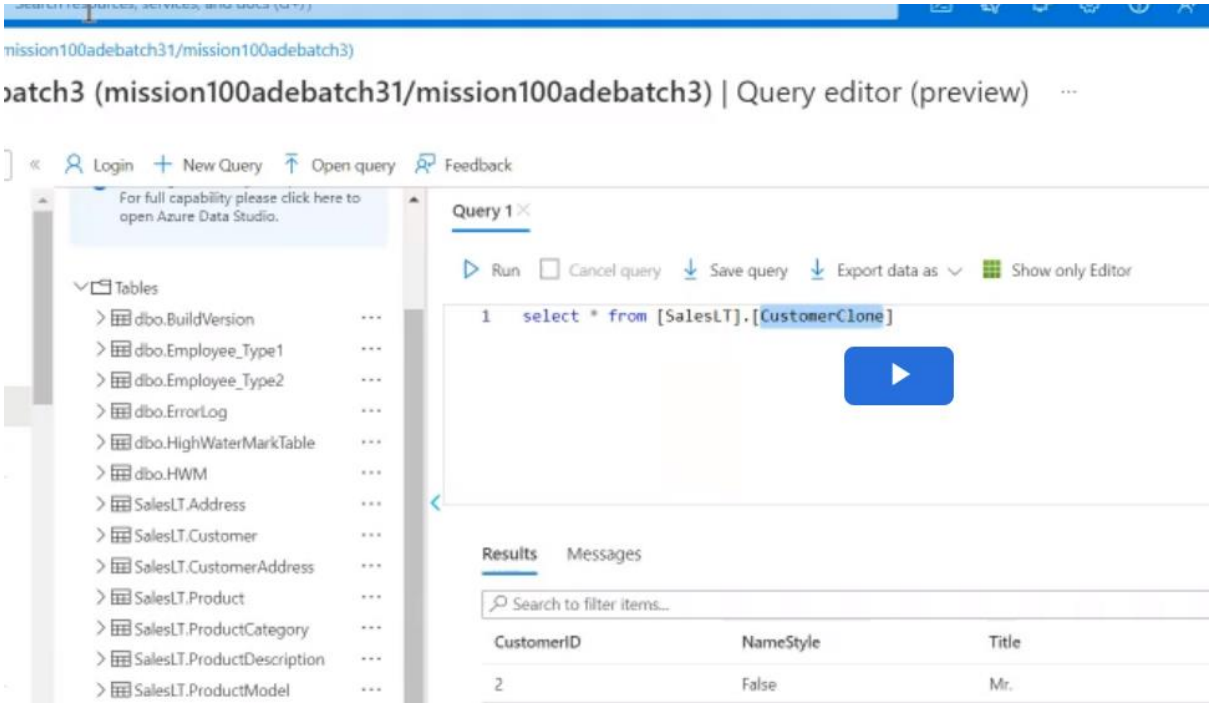

QA
Assignment 7 Q1
Validate
Data flow debug
Debug Settings
source1
Import data from Assignment7Q1CustomerCSV
Reference:
1
Columns:
15 total
sink1
Columns:
15 total
Add Source
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Number of rows
INSERT 0
UPDATE 0
DELETE 0
UPSERT 0
LOOKUP 0
ERROR 0
TOTAL 440
Refresh
Statistics
Export to CSV
Custom...
NameSt...
Title
FirstName
Middle...
LastName
Suffix
Compan...
2 Mr. Keith NULL Harris NULL Progress... advent
4 Ms. Janet M. Gates NULL Modular... advent
6 Ms. Rosmarie J. Carroll NULL Aerobic ... advent
10 Ms. Kathleen M. Garza NULL Rural Cy... advent
12 Mr. Johnny A. Caprio Jr. Bikes an... advent
16 Mr. Christop... R. Beck Jr. Bulk Dis... advent
18 Mr. David J. Liu NULL Catalog ... advent

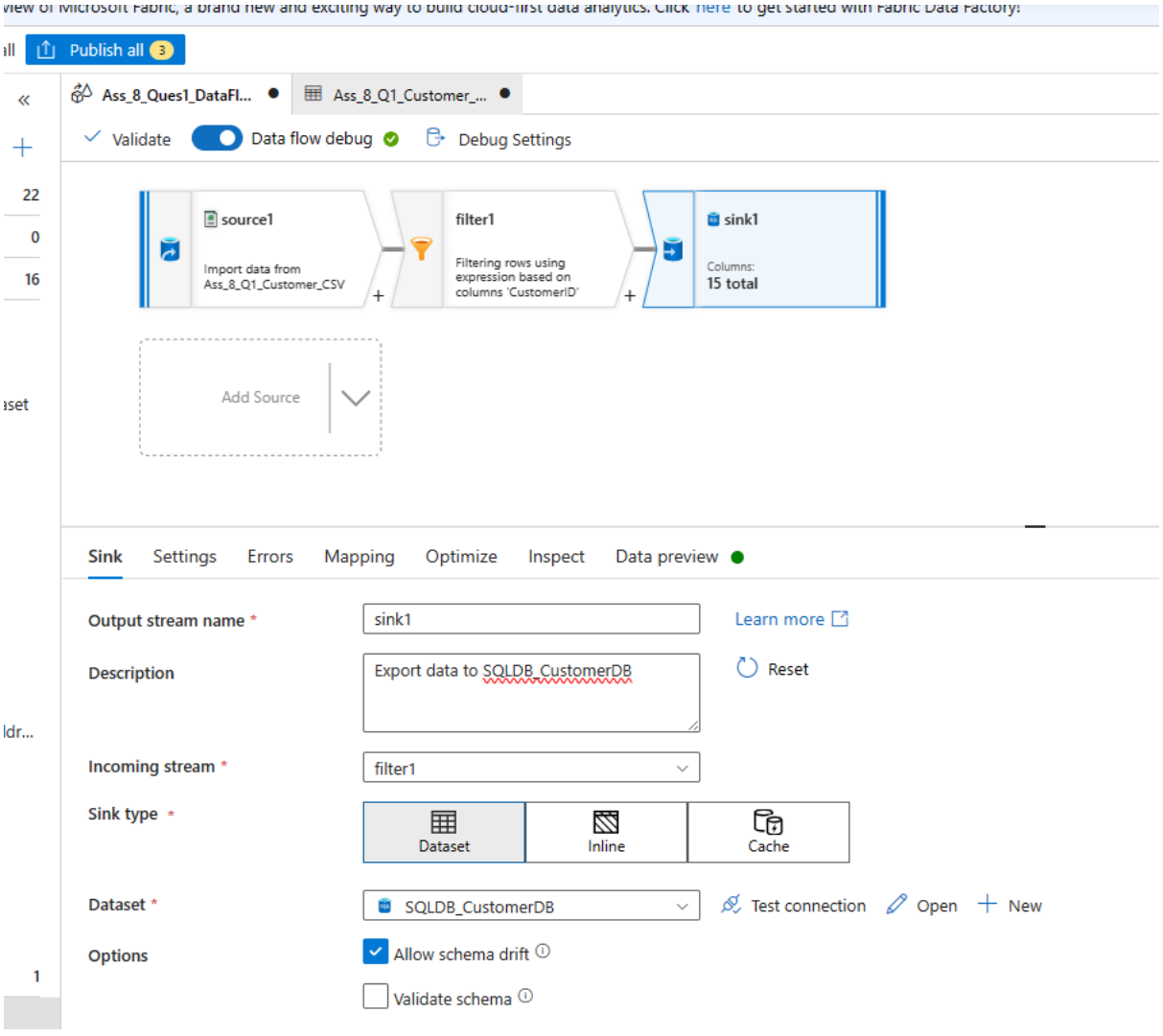

(mission100adebatch31/mission100adebatch3) | Query editor (preview)
Login
New Query
Open query
Feedback
For full capability please click here to open Azure Data Studio.
Tables
dbo.BuildVersion
dbo.Employee_Type1
dbo.Employee_Type2
dbo.ErrorLog
dbo.HighWaterMarkTable
dbo.HWM
SalesLT.Address
SalesLT.Customer
SalesLT.CustomerAddress
SalesLT.Product
Query 1
Run
Cancel query
Save query
Export data as
Show only Editor
1 select * from [SalesLT].[CustomerClone]
Results
Messages

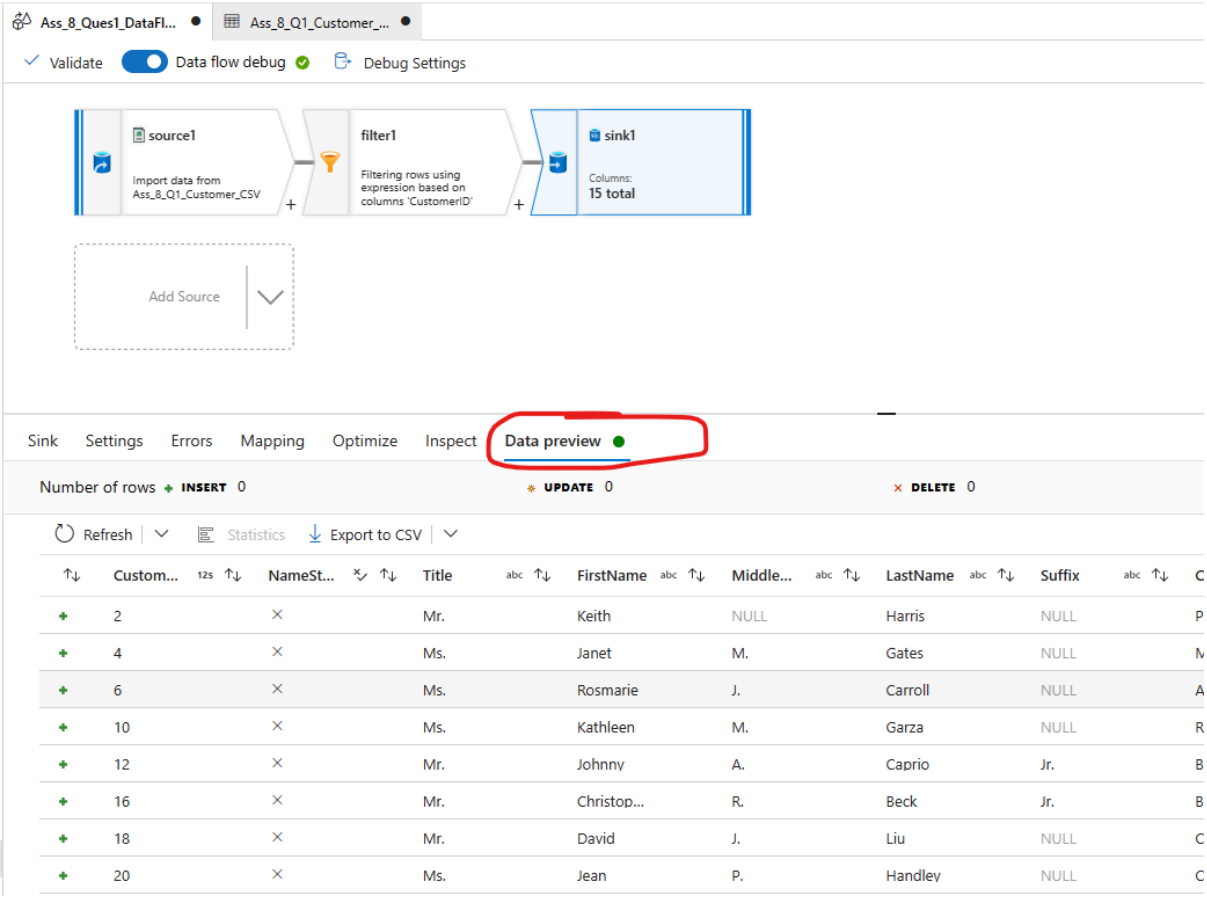

SQL database
Search
Overview
Activity log
Tags
Diagnose and solve problems
Getting started
Query editor (preview)
Settings
Compute + storage
Connection strings
Properties
Locks
Data management
Replicas
Sync to other databases
Integrations
Login
New Query
Open query
Feedback
For full capability please 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
Search to filter items...
Ready


SQL database
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 [SalesLT].[CustomerClone]
Results
Messages
Search to filter items...
No CloneCust Exist
Ready

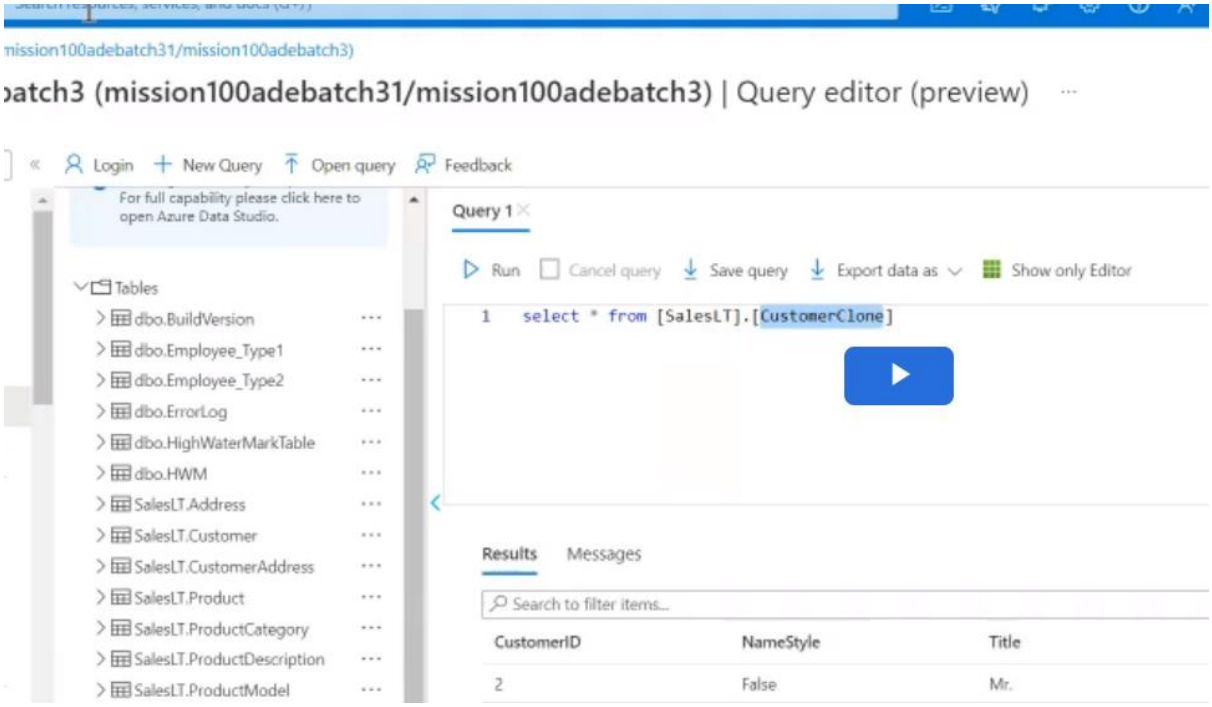

Query 1
Run
Cancel query
Save query
Export data as
Show only Editor
1 select * from [SalesLT].[CustomerClone]
Results
Messages
Failed to execute query. Error: Invalid object name 'SalesLT.CustomerClone'.


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.

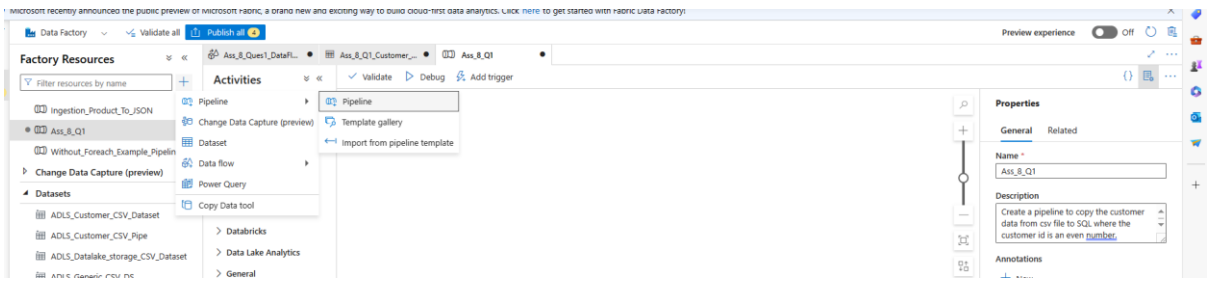

Publish all 3
Ass_8_Ques1_DataFl...
Ass_8_Q1_Customer_...
Validate
Data flow debug
Debug Settings
source1
Import data from Ass_8_Q1_Customer_CSV
filter1
Filtering rows using expression based on columns 'CustomerID'
sink1
Columns:
15 total
Add Source
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data 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

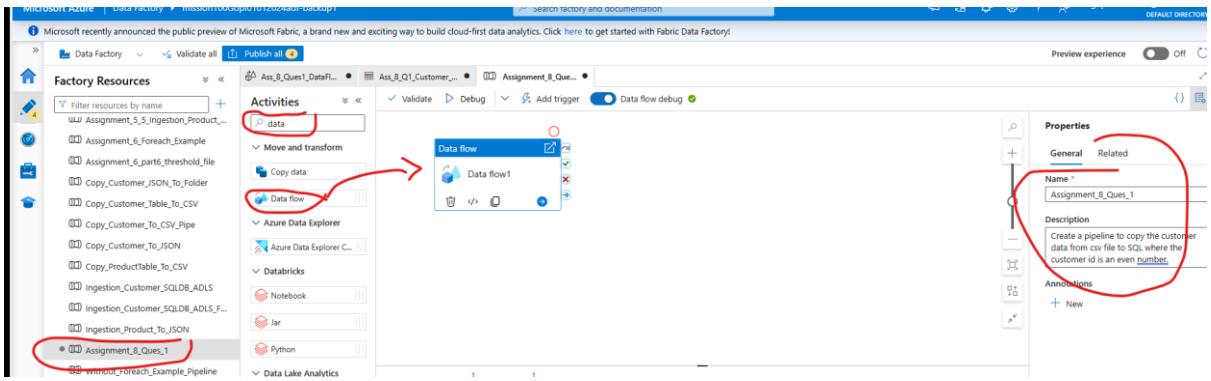


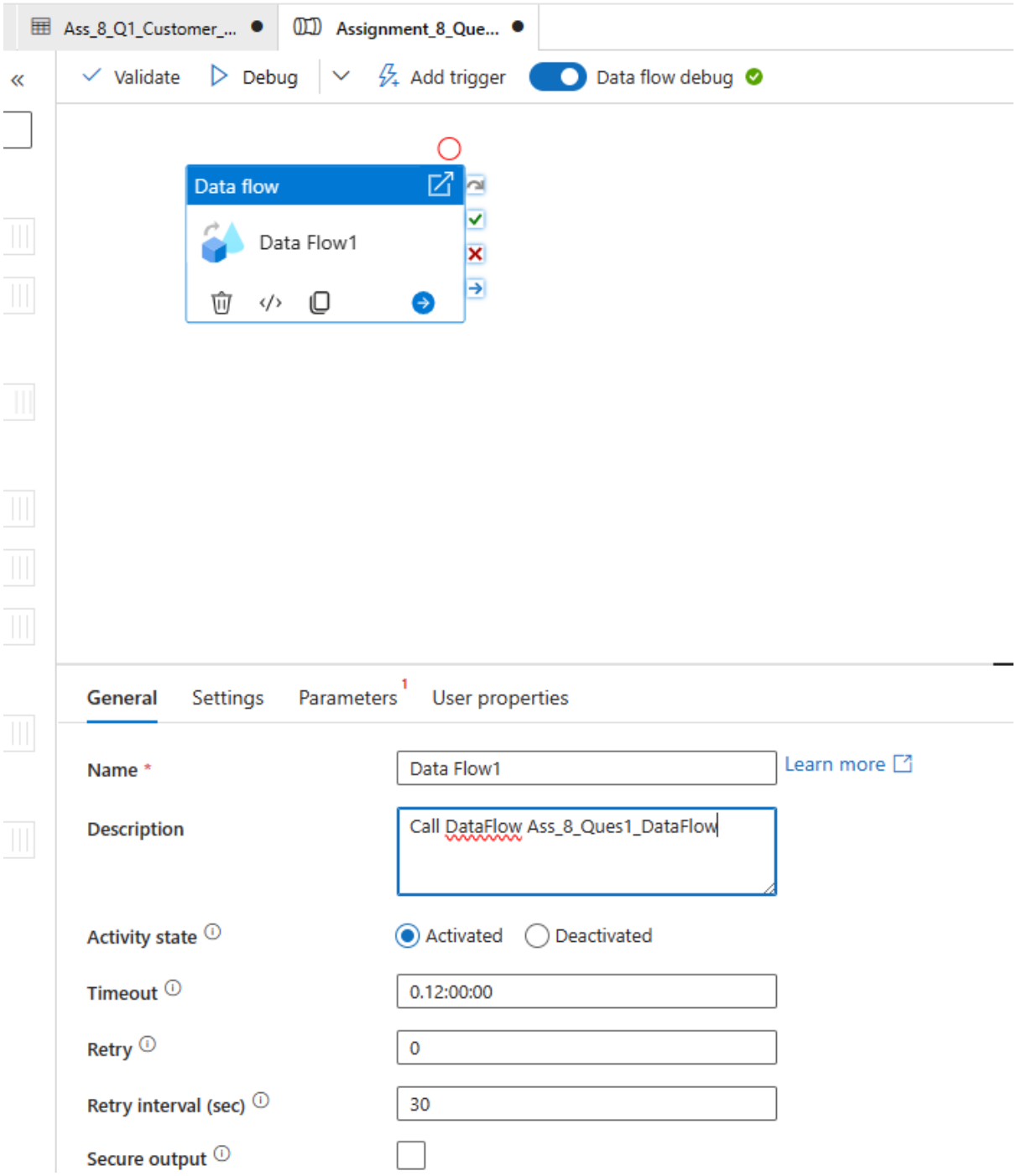
Custom...	NameSt...	Title	FirstName	Middle...	LastName	Suffix	C
2	×	Mr.	Keith	NULL	Harris	NULL	P
4	×	Ms.	Janet	M.	Gates	NULL	M
6	×	Ms.	Rosmarie	J.	Carroll	NULL	A
10	×	Ms.	Kathleen	M.	Garza	NULL	R
12	×	Mr.	Johnny	A.	Caprio	Jr.	B
16	×	Mr.	Christop...	R.	Beck	Jr.	B
18	×	Mr.	David	J.	Liu	NULL	C
20	×	Ms.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

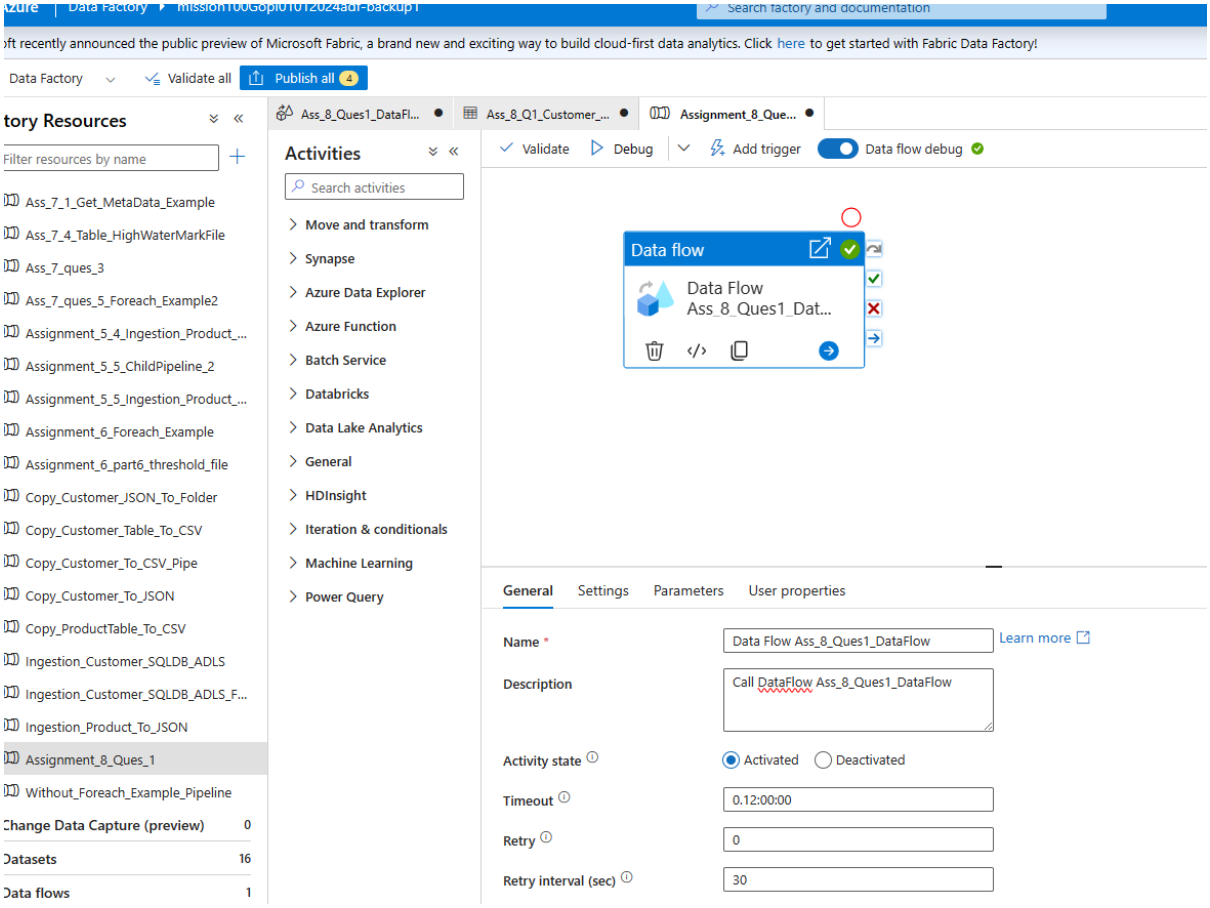


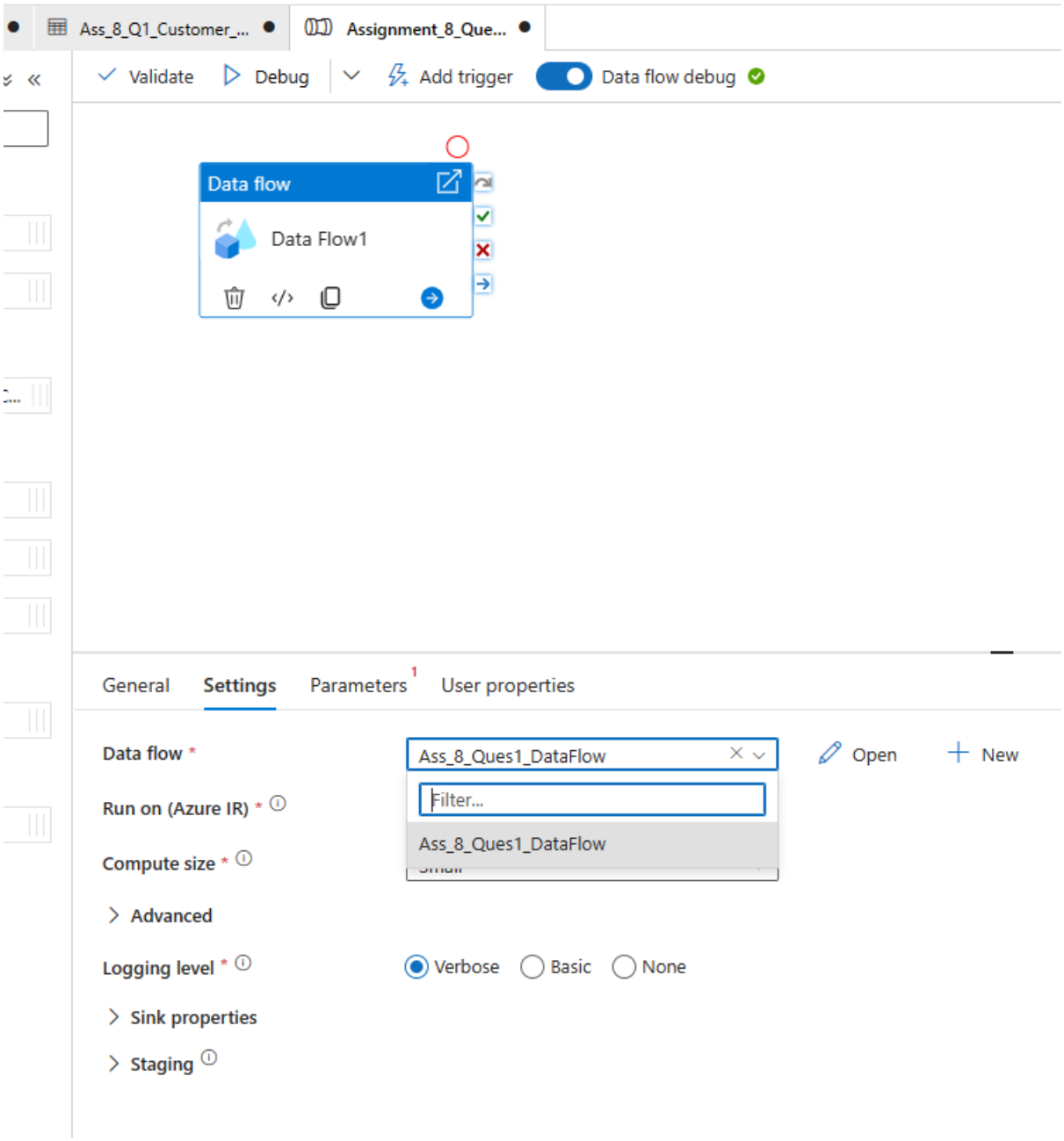
Data flow does not run its by own –so u need to create a pipeline

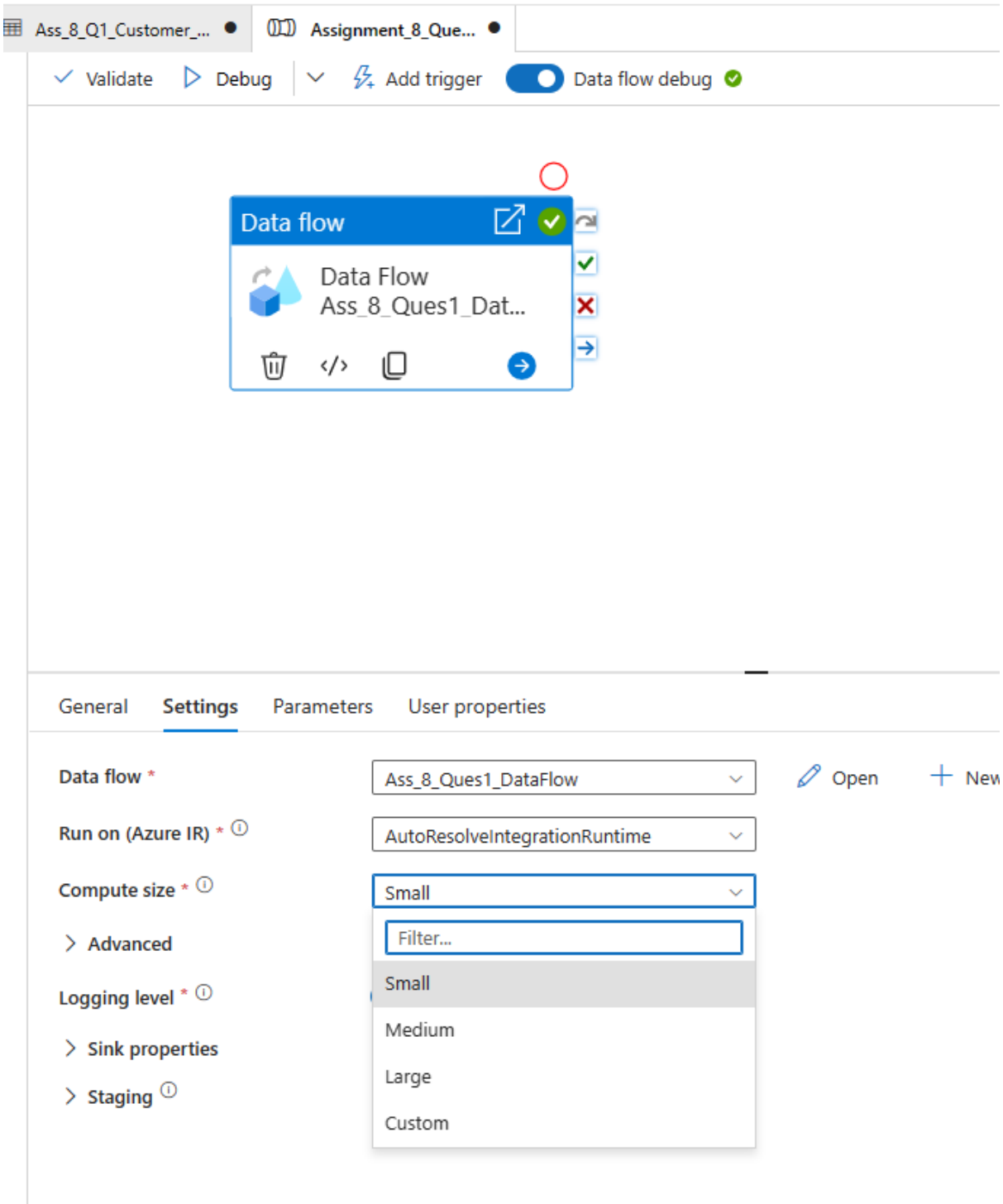


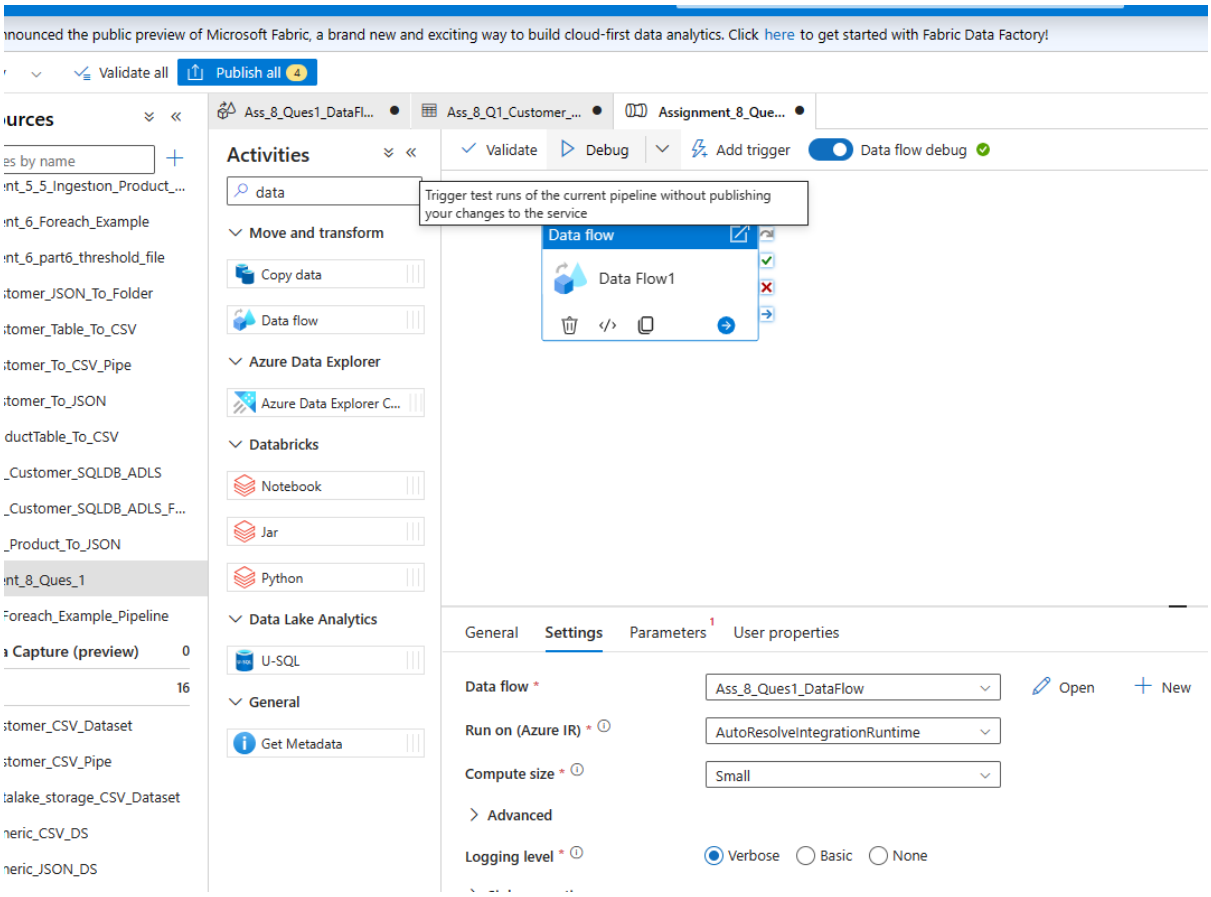


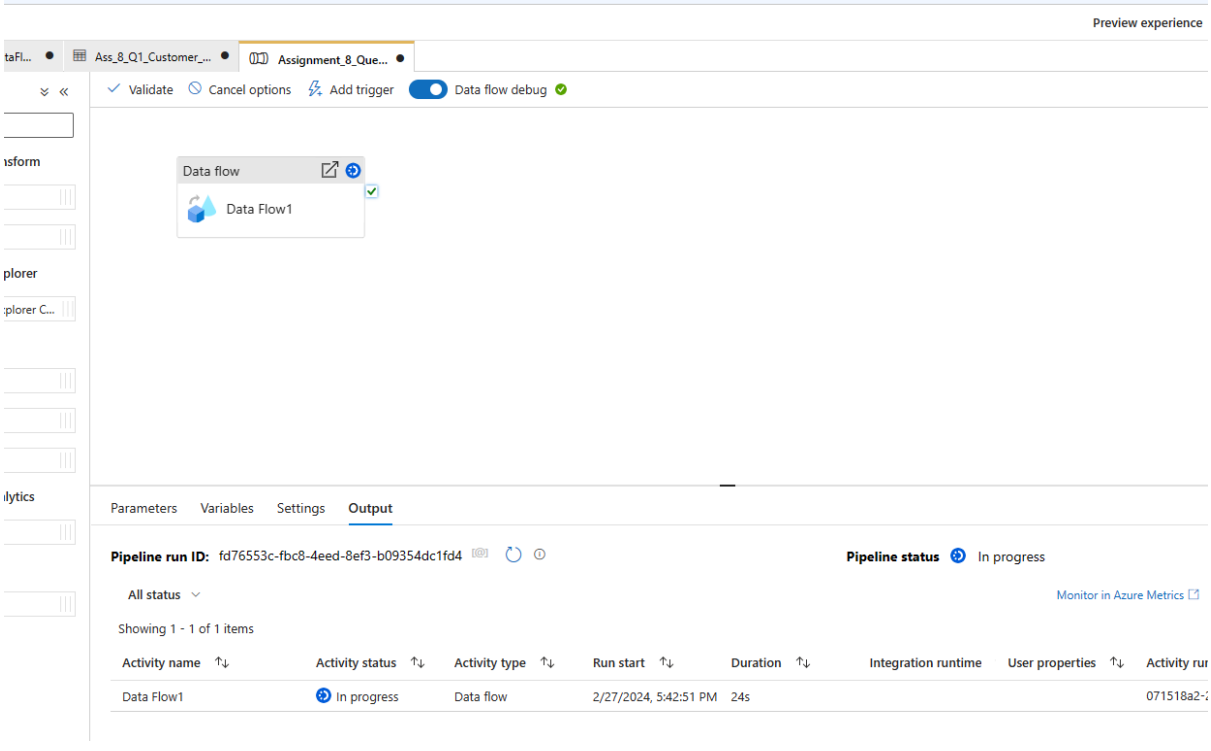

Ass_8_Q1_Customer_...
Assignment_8_Que...
Validate
Debug
Add trigger
Data flow debug
Data flow
Data Flow1
General
Settings
Parameters 1
User properties
Name *
Data Flow1
Learn more
Description
Call DataFlow Ass_8_Ques1_DataFlow
Activity state
Activated
Deactivated
Timeout
0.12:00:00
Retry
0
Retry interval (sec)
30
Secure output


Data Factory
Search factory and documentation
ft 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
tory Resources
Filter resources by name
Ass_7_1_Get_MetaData_Example
Ass_7_4_Table_HighWaterMarkFile
Ass_7_ques_3
Ass_7_ques_5_Foreach_Example2
Assignment_5_4_Ingestion_Product_...
Assignment_5_5_ChildPipeline_2
Assignment_5_5_Ingestion_Product_...
Assignment_6_Foreach_Example
Assignment_6_part6_threshold_file
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
Assignment_8_Ques_1
Without_Foreach_Example_Pipeline
Change Data Capture (preview) 0
Datasets 16
Data flows 1
Ass_8_Ques1_DataFl...
Ass_8_Q1_Customer_...
Assignment_8_Que...
Activities
Search activities
Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query
Validate
Debug
Add trigger
Data flow debug
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


Ass_8_Q1_Customer_...
Assignment_8_Que...
Validate
Debug
Add trigger
Data flow debug
Data flow
Data Flow1
General
Settings
Parameters
1
User properties
Data flow *
Ass_8_Ques1_DataFlow
Open
New
Filter...
Ass_8_Ques1_DataFlow
Run on (Azure IR) *
Compute size *
Advanced
Logging level *
Verbose
Basic
None
Sink properties
Staging


Ass_8_Q1_Customer_...
Assignment_8_Que...
Validate
Debug
Add trigger
Data flow debug
Data flow
Data Flow
Ass_8_Ques1_Dat...
General
Settings
Parameters
User properties
Data flow *
Ass_8_Ques1_DataFlow
Open
New
Run on (Azure IR) *
AutoResolveIntegrationRuntime
Compute size *
Small
Advanced
Filter...
Small
Medium
Large
Custom
Logging level *
Sink properties
Staging


nounced 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
Ass_8_Ques1_DataFl...
Ass_8_Q1_Customer_...
Assignment_8_Que...
Activities
data
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
Validate
Debug
Add trigger
Data flow debug
Data flow
Data Flow1
General
Settings
Parameters 1
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
Ass_8_Q1_Customer_...
Assignment_8_Que...
Validate
Cancel options
Add trigger
Data flow debug
Data flow
Data Flow1
Parameters
Variables
Settings
Output
Pipeline run ID: fd76553c-fbc8-4eed-8ef3-b09354dc1fd4
Pipeline status In progress
All status
Monitor in Azure Metrics
Showing 1 - 1 of 1 items
Activity name
Activity status
Activity type
Run start
Duration
Integration runtime
User properties
Activity ru
Data Flow1
In progress
Data flow
2/27/2024, 5:42:51 PM
24s

Parameters Variables Settings Output

Pipeline run ID: fd76553c-fbc8-4eed-8ef3-b09354dc1fd4 **Pipeline status** Succeeded View d

All status Monitor in Azure Metrics

Showing 1 - 1 of 1 items

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

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

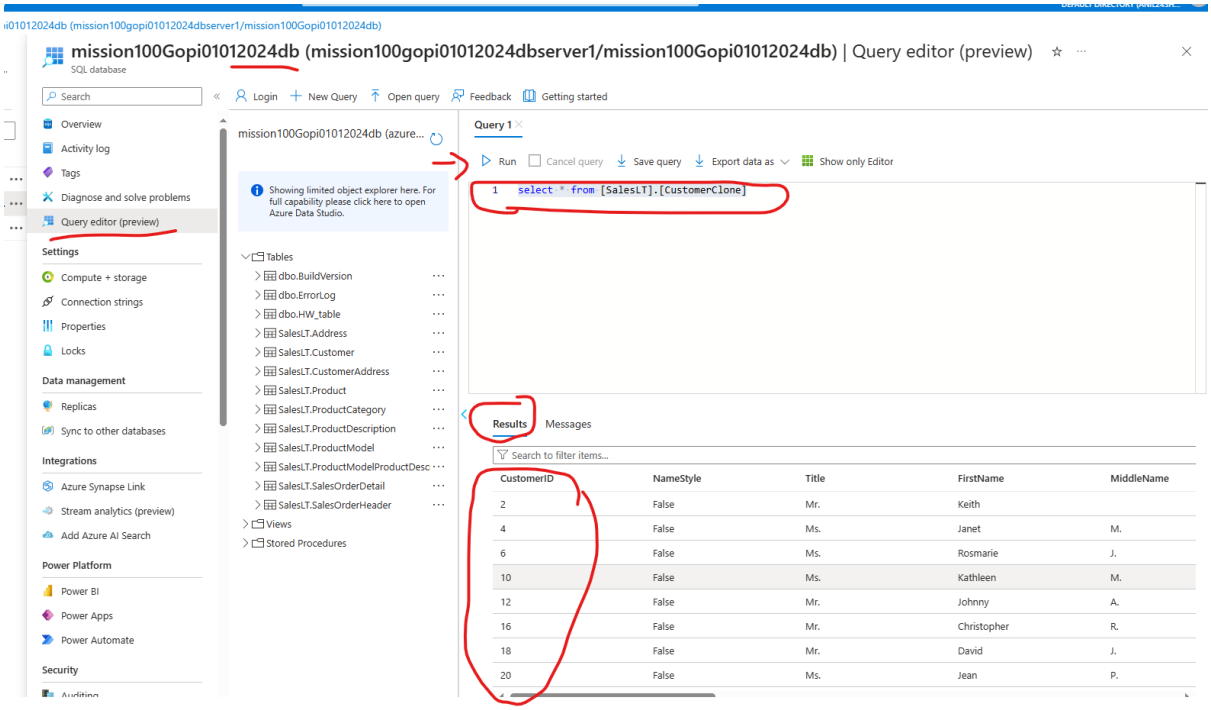
Refresh **Auto refresh** On Edit dataflow

re (previ...

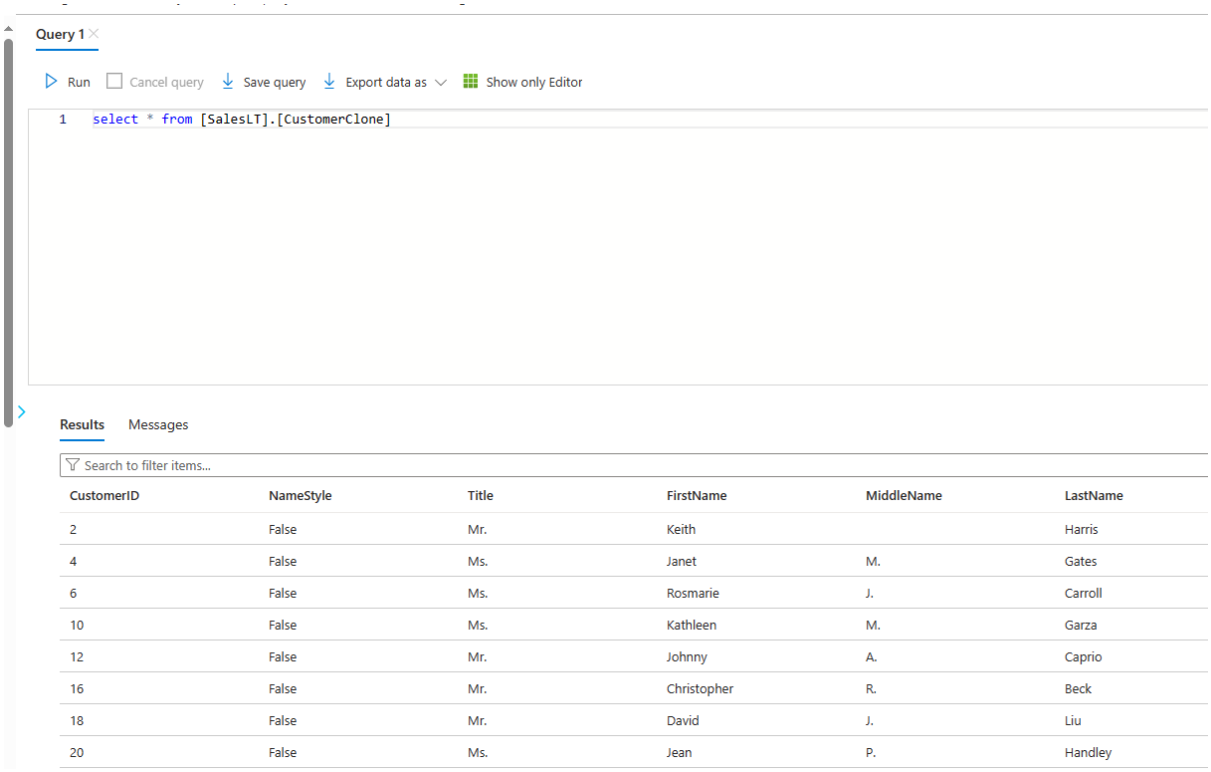
s

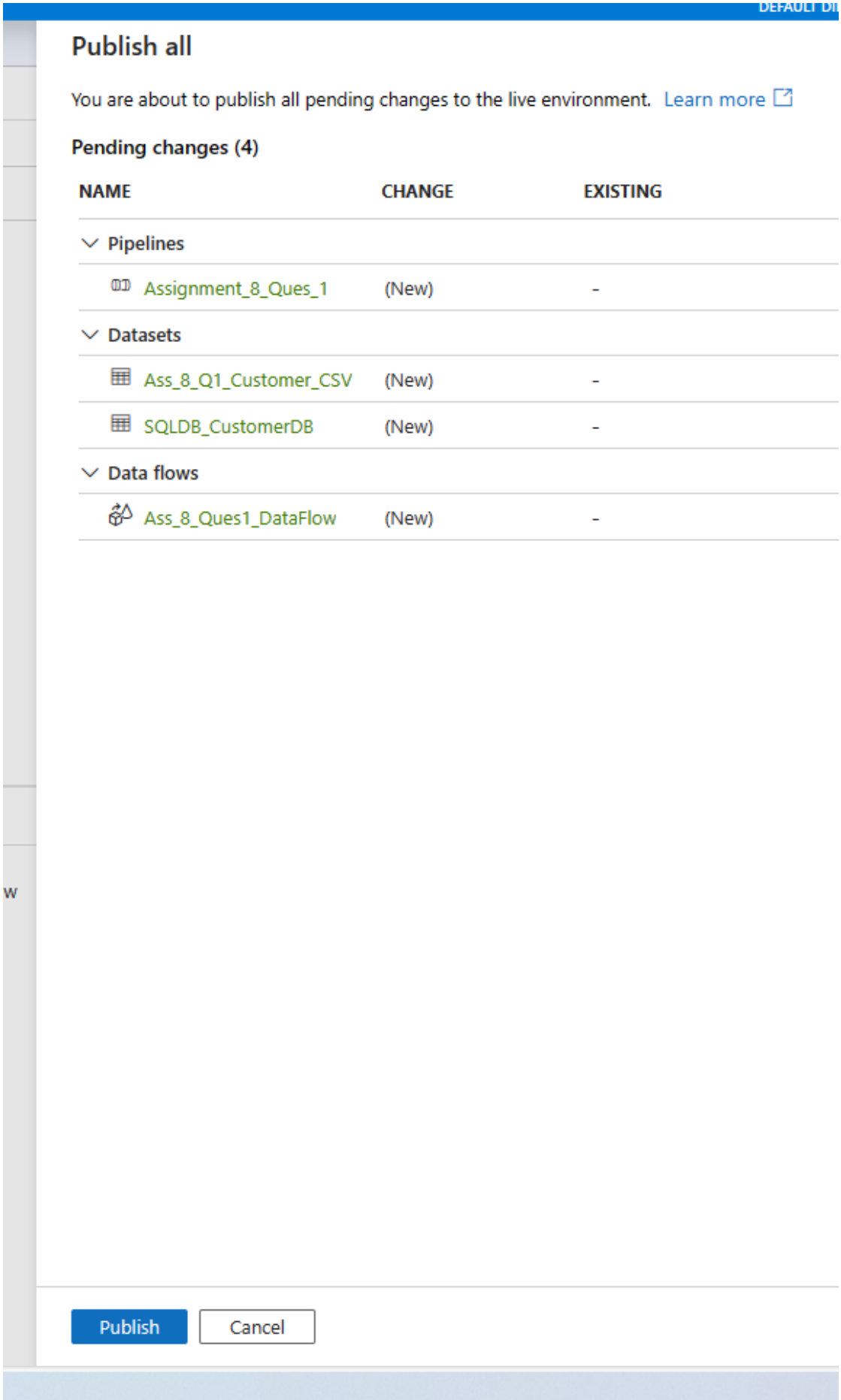
Sinks All streams

Sink	Status	Processing time
sink1	Succeeded	1s 954ms



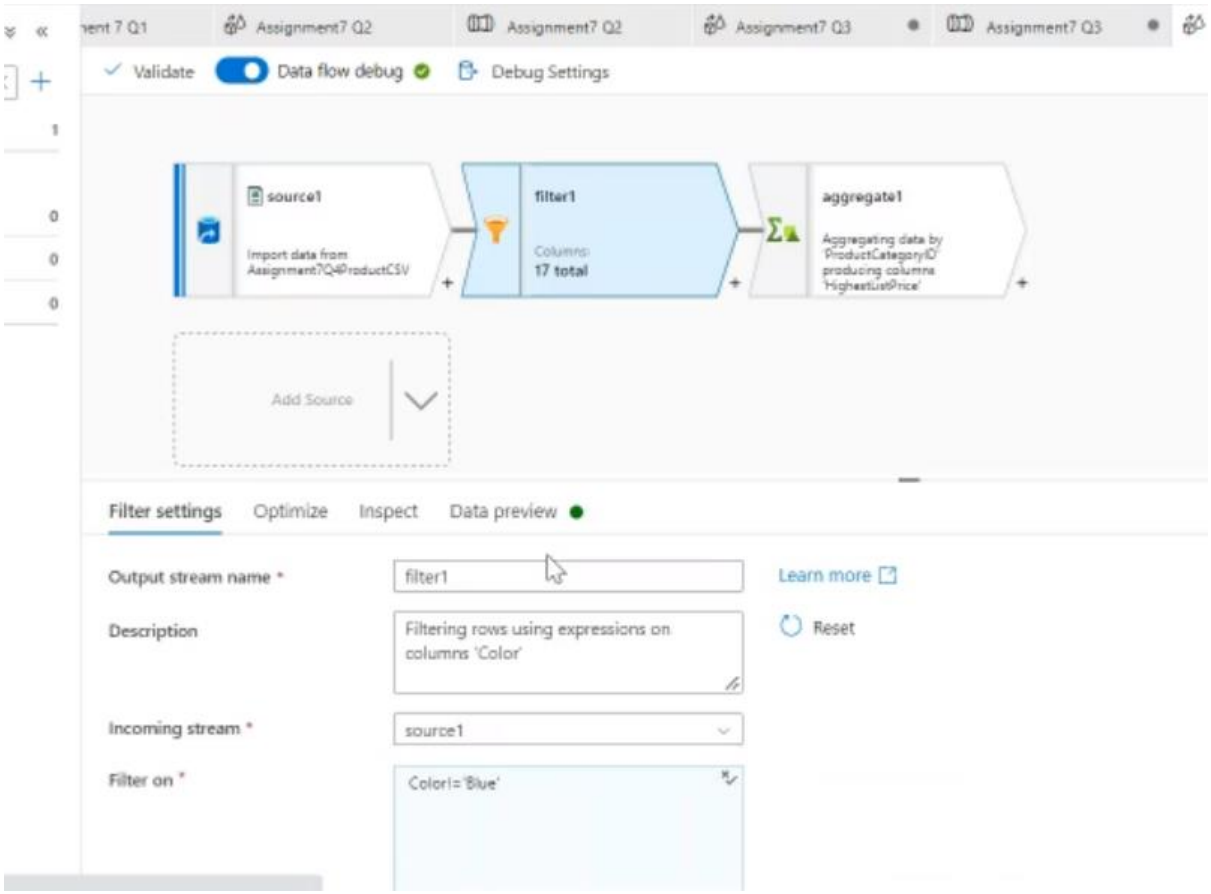
Only Even Customer ID are displayed, hence our pipeline is correct

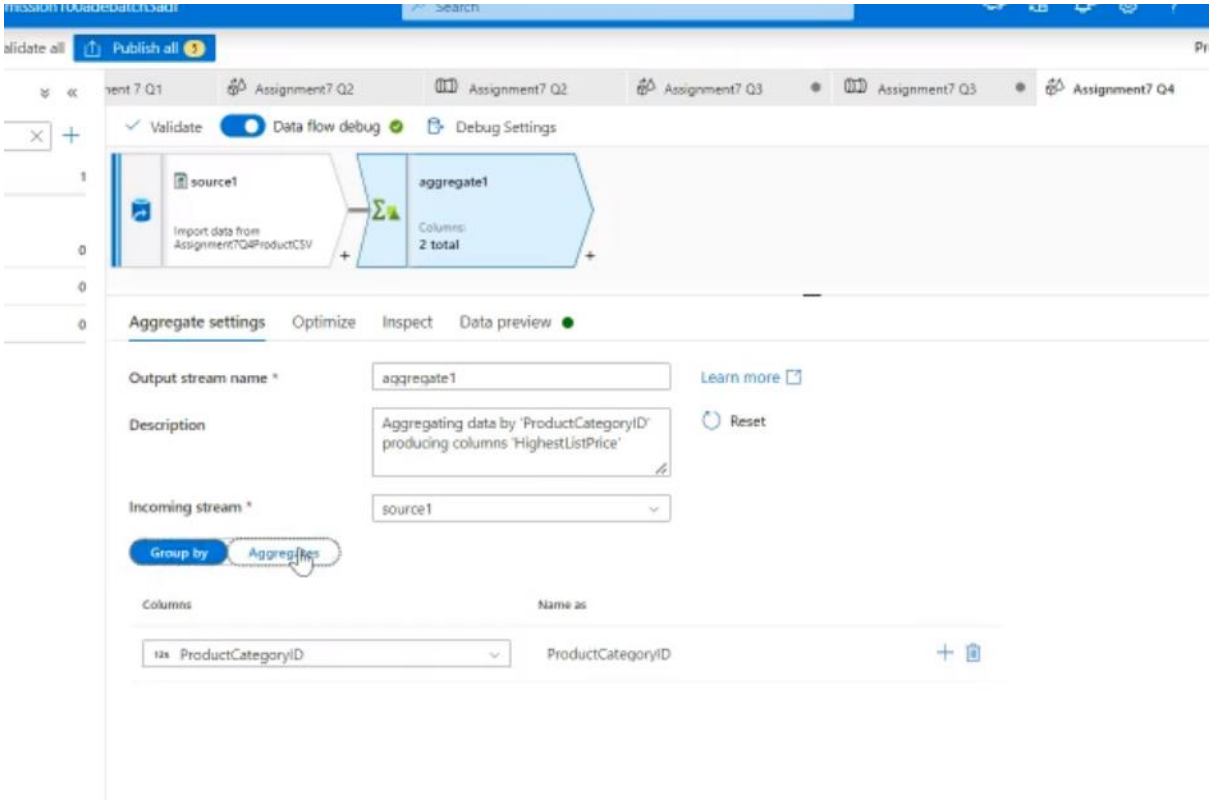


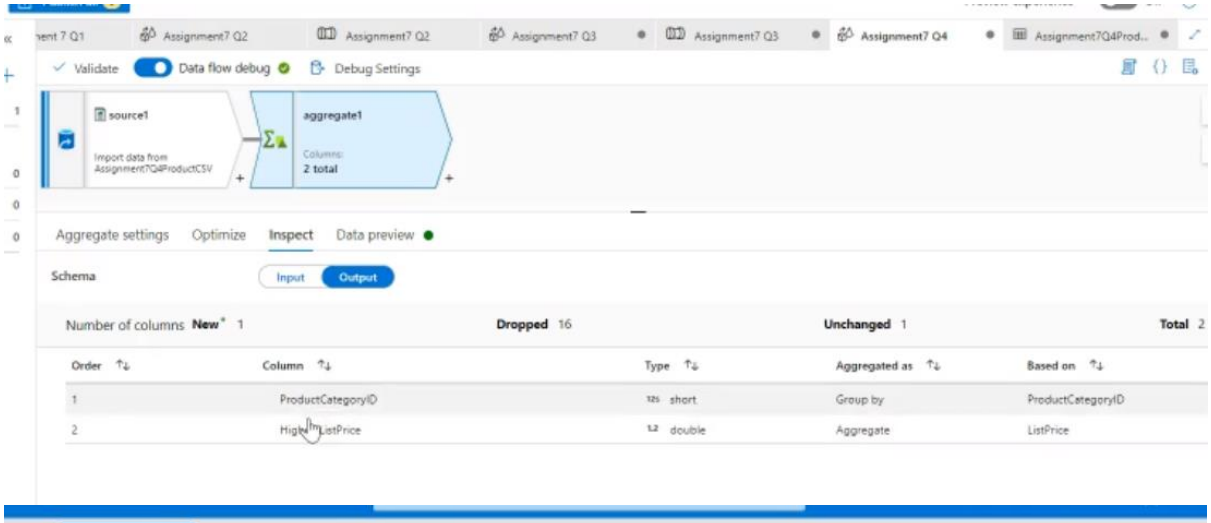

Publish all
You are about to publish all pending changes to the live environment. Learn more
Pending changes (4)
NAME
CHANGE
EXISTING
Pipelines
Assignment_8_Ques_1
(New)
-
Datasets
Ass_8_Q1_Customer_CSV
(New)
-
SQLDB_CustomerDB
(New)
-
Data flows
Ass_8_Ques1_DataFlow
(New)
-
Publish
Cancel

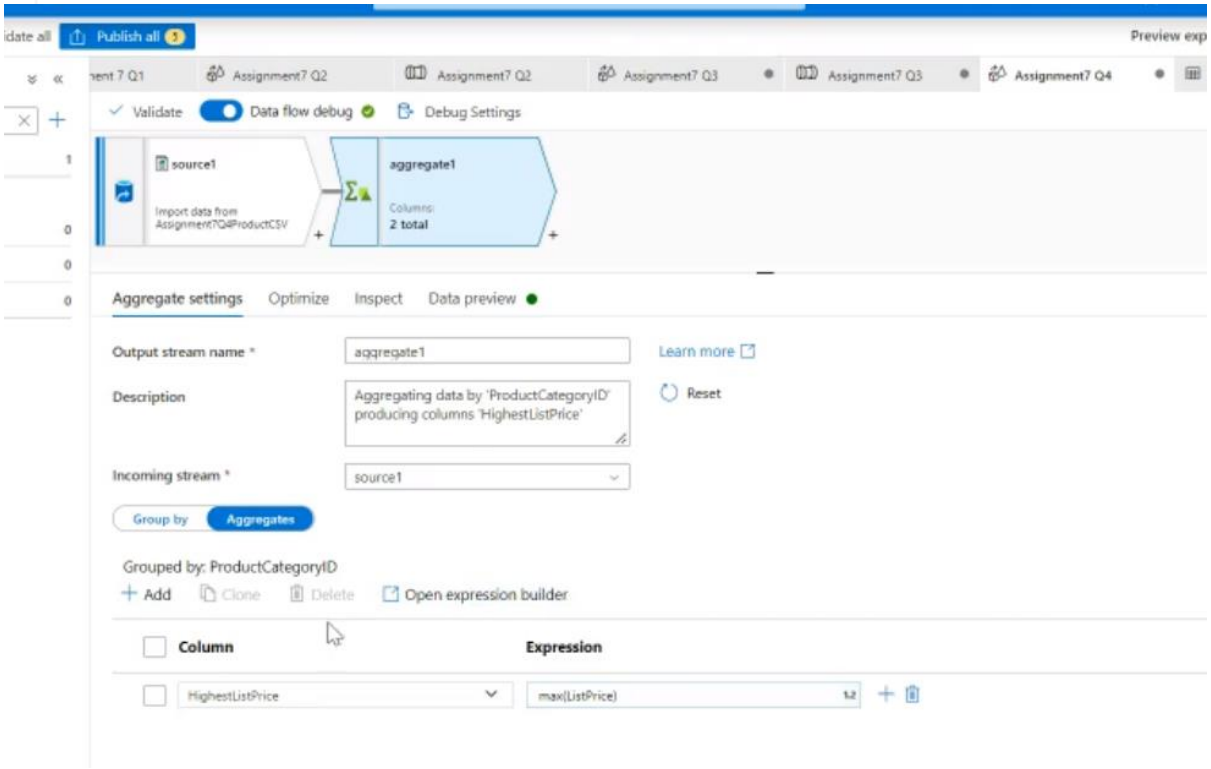
Question 4:
create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.
Ensure that product shouldn't be of blue in color and save the result as CSV file inside ProductResult folder.

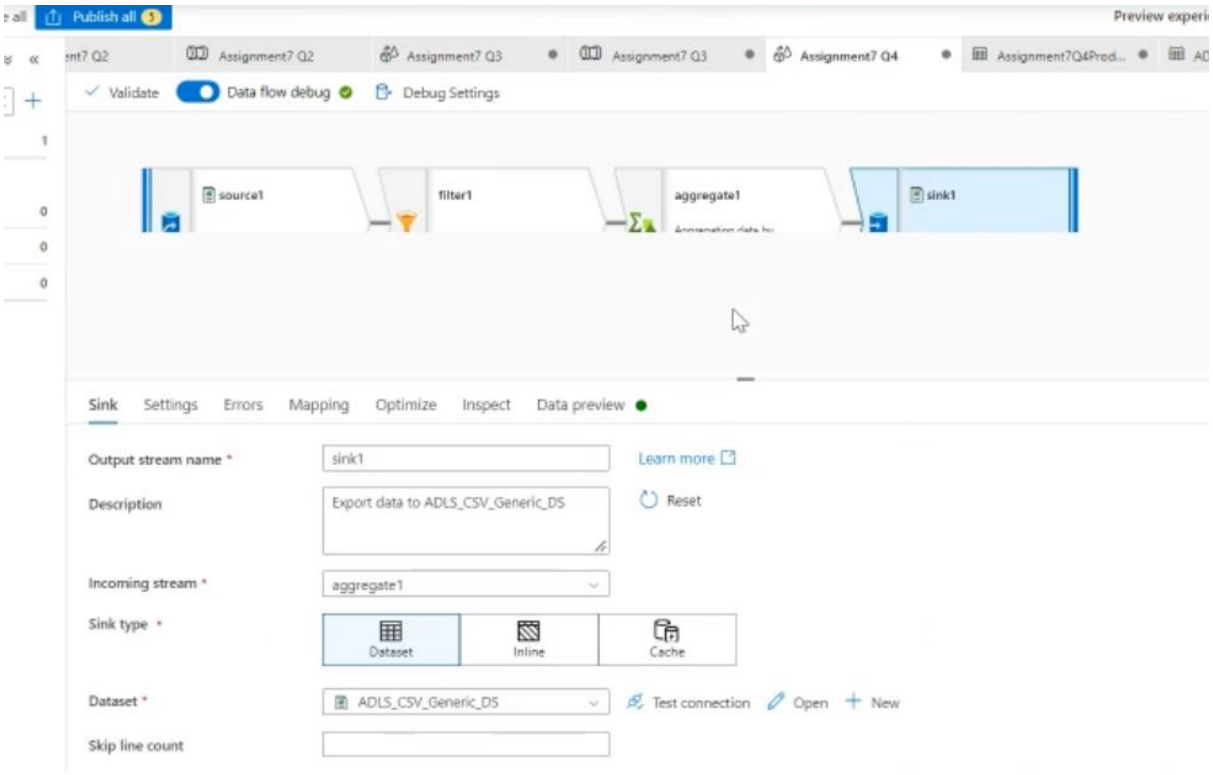
Solution:




Publish all
Assignment7 Q2
Assignment7 Q2
Assignment7 Q3
Assignment7 Q3
Assignment7 Q4
Validate
Data flow debug
Debug Settings
source1
Import data from Assignment7CsvProductCSV
aggregate1
Columns:
2 total
Aggregate settings
Optimize
Inspect
Data preview
Output stream name *
aggregate1
Learn more
Description
Aggregating data by 'ProductCategoryID' producing columns 'HighestListPrice'
Reset
Incoming stream *
source1
Group by
Aggregates
Columns
Name as
ProductCategoryID
ProductCategoryID


Assignment7 Q2
Assignment7 Q2
Assignment7 Q3
Assignment7 Q3
Assignment7 Q4
Validate
Data flow debug
Debug Settings
source1
Import data from Assignment7Q4ProductCSV
aggregate1
Columns:
2 total
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
short
Group by
ProductCategoryID
2
HighestListPrice
double
Aggregate
ListPrice


Publish all
Preview exp
Assignment7 Q2
Assignment7 Q2
Assignment7 Q3
Assignment7 Q3
Assignment7 Q4
Validate
Data flow debug
Debug Settings
source1
Import data from Assignment7Q4ProductCSV
aggregate1
Columns:
2 total
Aggregate settings
Optimize
Inspect
Data preview
Output stream name *
aggregate1
Learn more
Description
Aggregating data by 'ProductCategoryID' producing columns 'HighestListPrice'
Reset
Incoming stream *
source1
Group by
Aggregates
Grouped by: ProductCategoryID
Add
Clone
Delete
Open expression builder
Column
Expression
HighestListPrice
max(ListPrice)



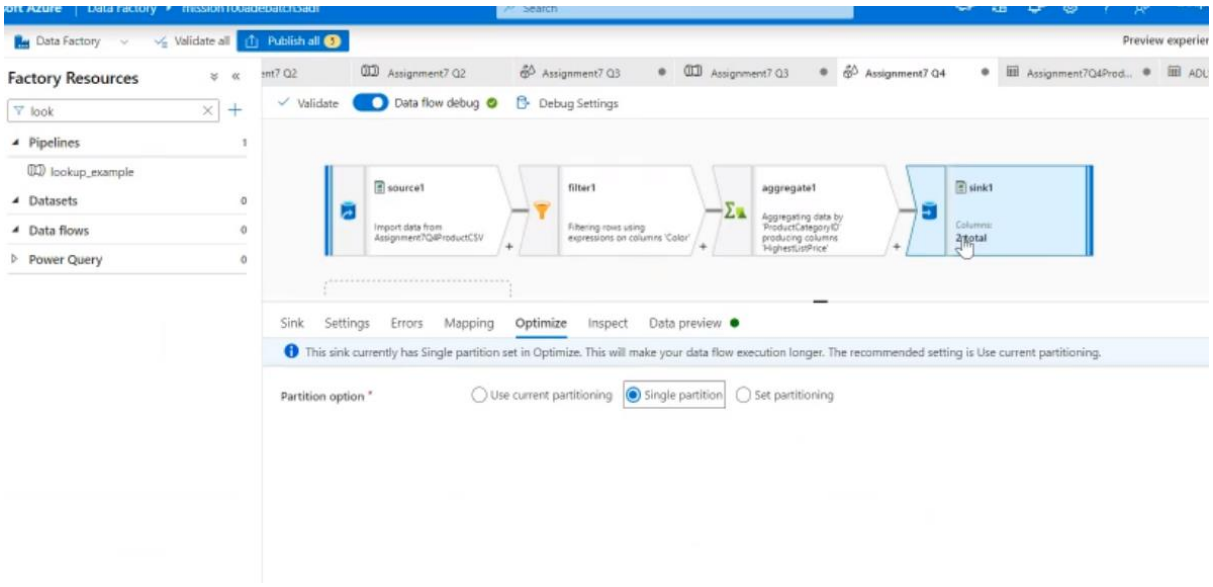
Question 5**:**

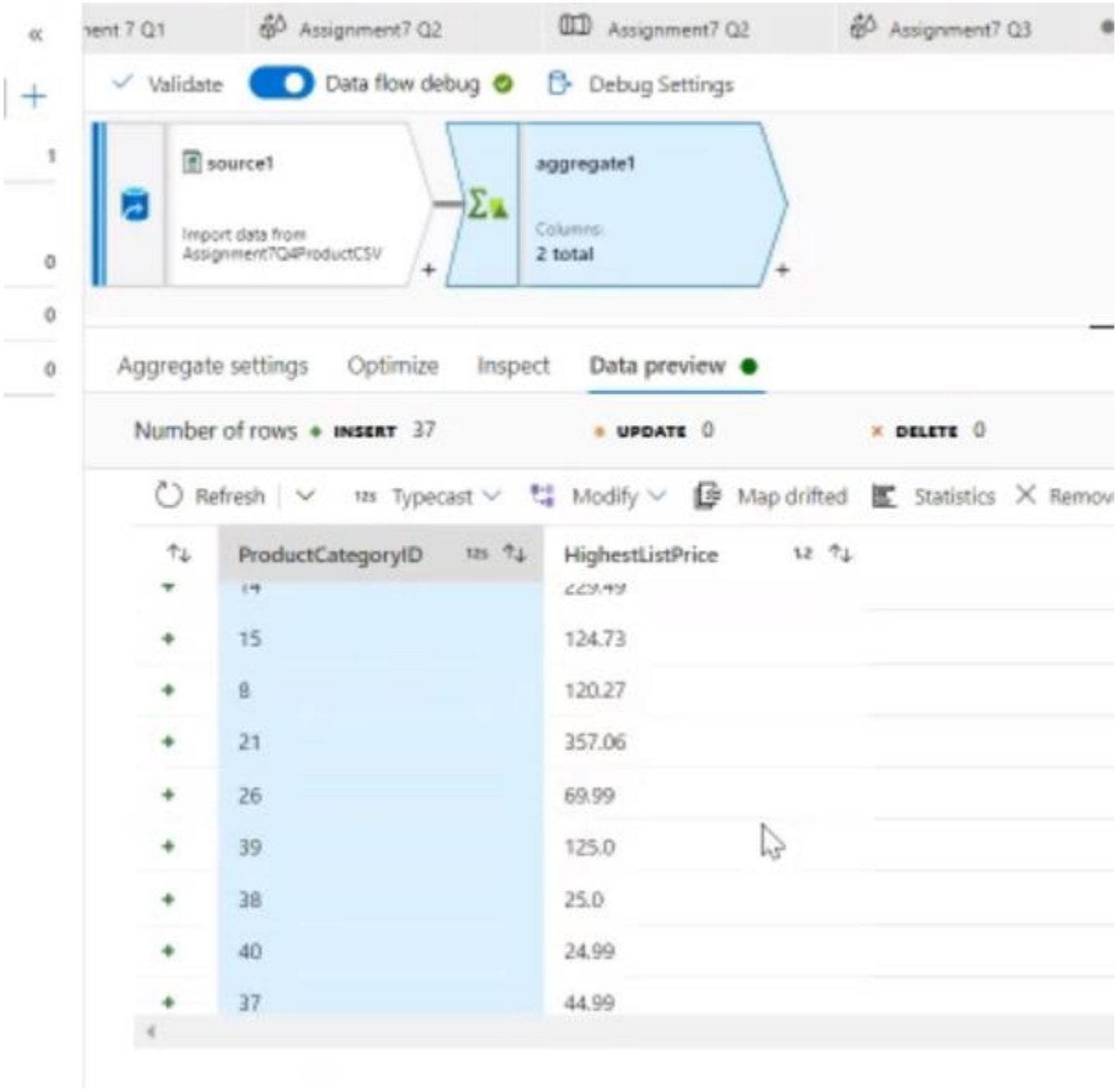
5. create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.
Ensure that product shouldn't be of blue in color and save the result as a SINGLE CSV file inside ProductSingleResult folder

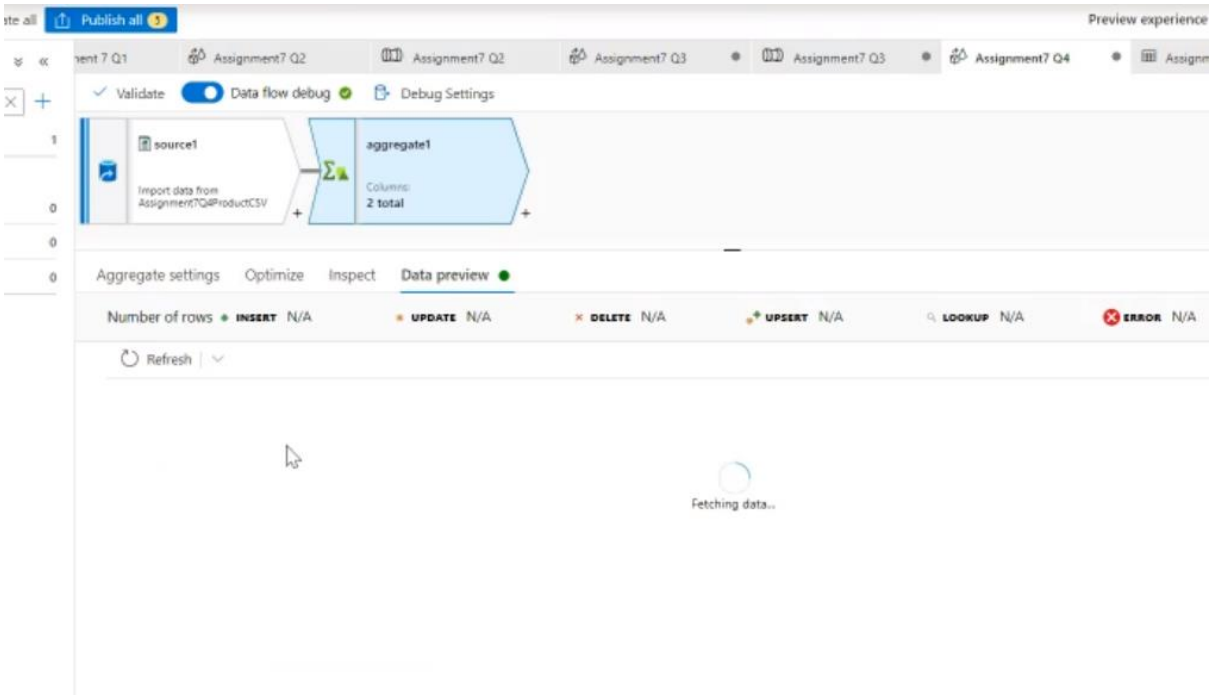
Solution Q5:

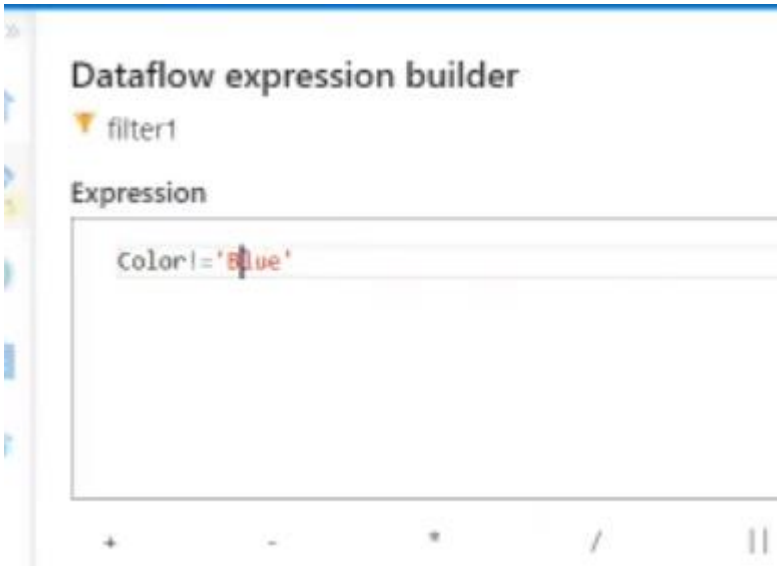
Choose "**Single Partitio**n" in optimize tab (so there wont be multiple partition)

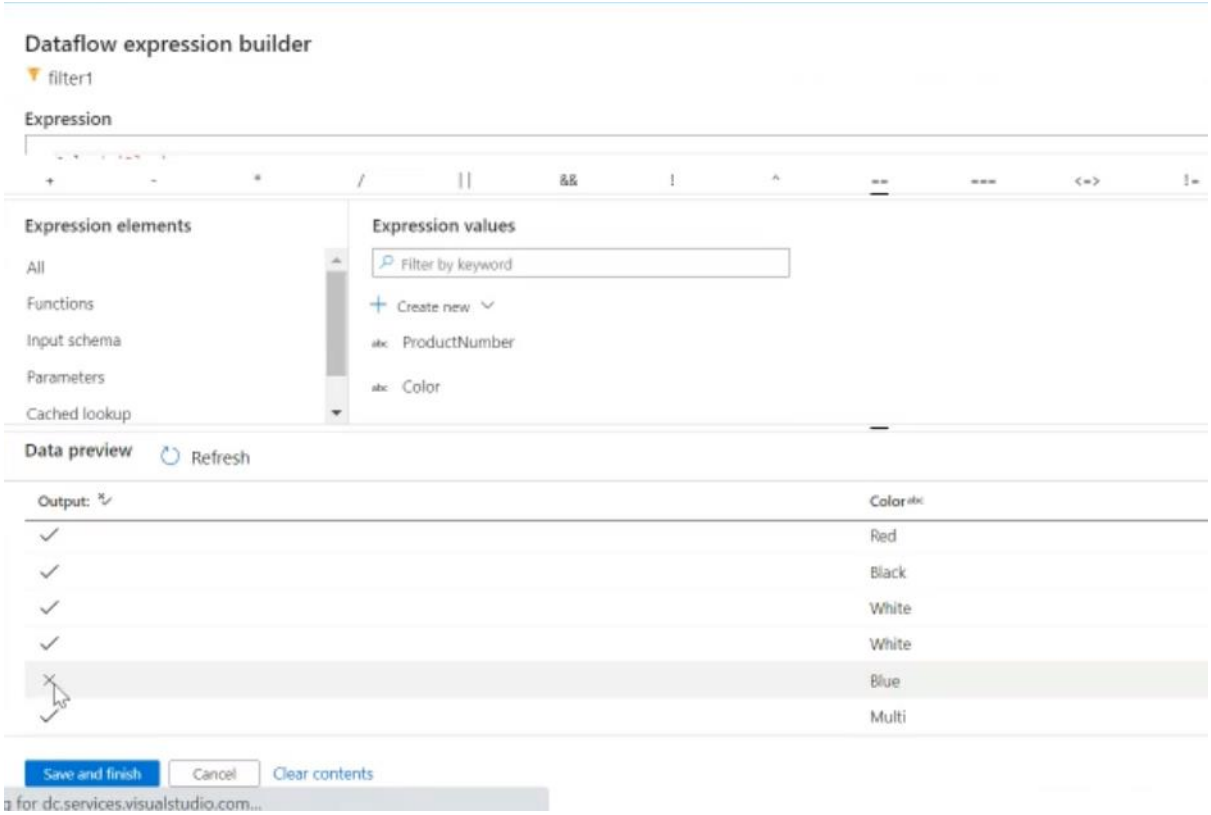
Rather than using parallel processng/ paralleising multiple files, it willconvert/ merge into single file/one file only instead of multiple fils


Data Factory
Validate all
Publish all 5
Preview experier
Factory Resources
look
Pipelines 1
lookup_example
Datasets 0
Data flows 0
Power Query 0
Assignment7 Q2
Assignment7 Q3
Assignment7 Q3
Assignment7 Q4
Assignment7Q4Prod...
Validate
Data flow debug
Debug Settings
source1
Import data from Assignment7Q4ProductCSV
filter1
Filtering rows using expressions on columns 'Color'
aggregate1
Aggregating data by 'ProductCategoryID' producing columns 'HighestListPrice'
sink1
Columns:
2 total
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
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


ent 7 Q1
Assignment7 Q2
Assignment7 Q2
Assignment7 Q3
Validate
Data flow debug
Debug Settings
source1
Import data from Assignment7Q4ProductCSV
aggregate1
Columns:
2 total
Aggregate settings
Optimize
Inspect
Data preview
Number of rows
INSERT 37
UPDATE 0
DELETE 0
Refresh
Typecast
Modify
Map drifted
Statistics
ProductCategoryID
HighestListPrice
15 124.73
8 120.27
21 357.06
26 69.99
39 125.0
38 25.0
40 24.99
37 44.99


Publish all
Preview experience
Assignment7 Q1
Assignment7 Q2
Assignment7 Q2
Assignment7 Q3
Assignment7 Q3
Assignment7 Q4
Validate
Data flow debug
Debug Settings
source1
Import data from Assignment7CsvProductCSV
aggregate1
Columns:
2 total
Aggregate settings
Optimize
Inspect
Data preview
Number of rows
INSERT N/A
UPDATE N/A
DELETE N/A
UPSERT N/A
LOOKUP N/A
ERROR N/A
Refresh
Fetching data...

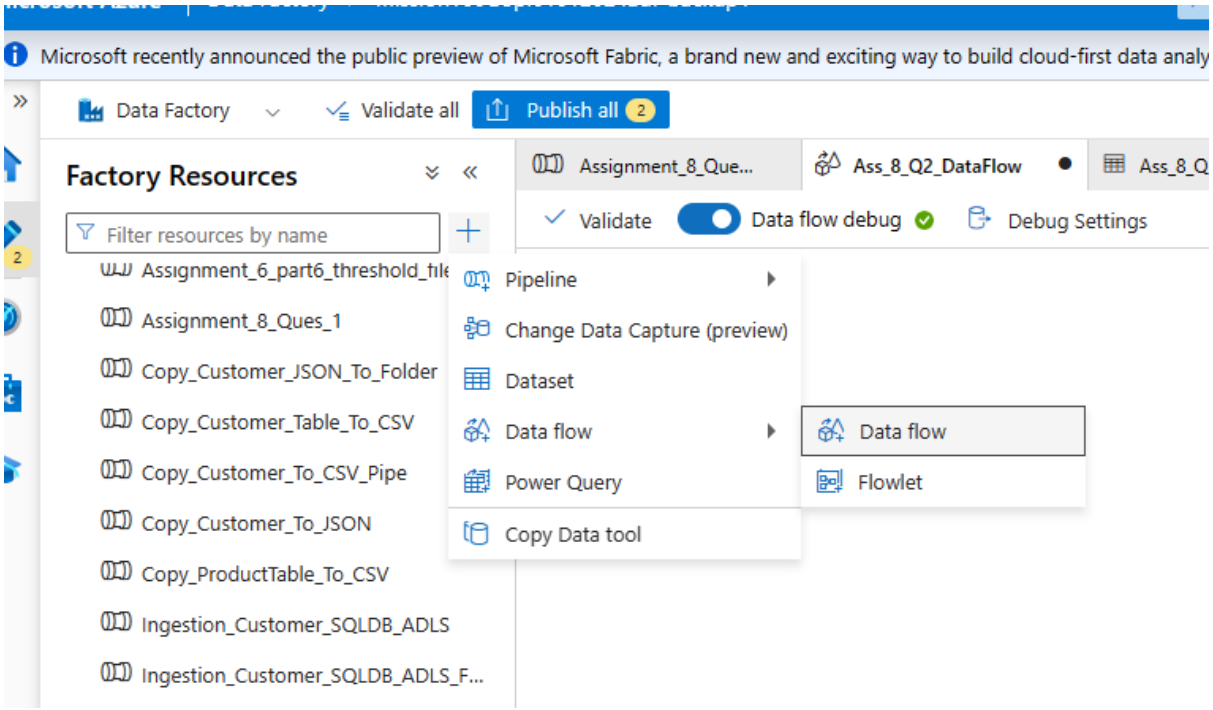

Dataflow expression builder
filter1
Expression
Color!='Blue'

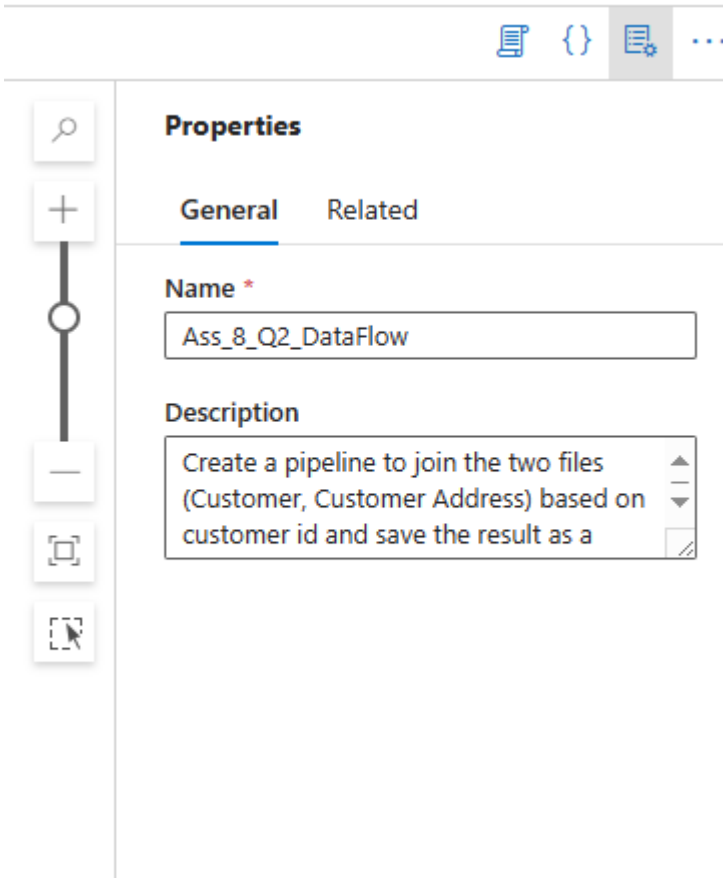


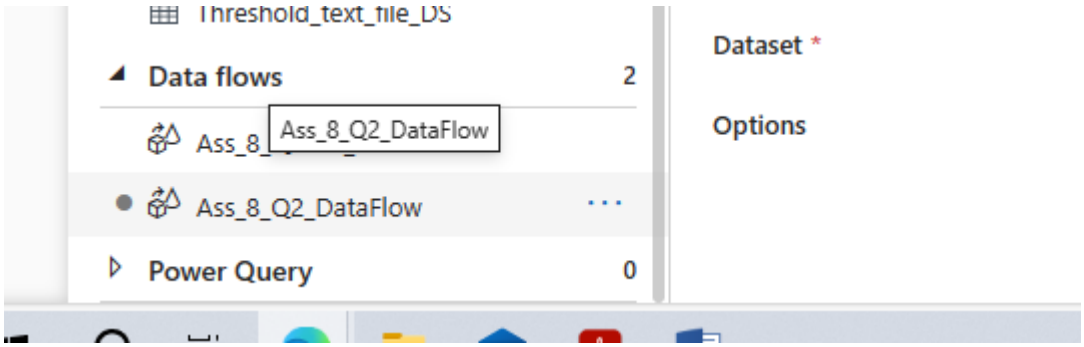
Question 2:

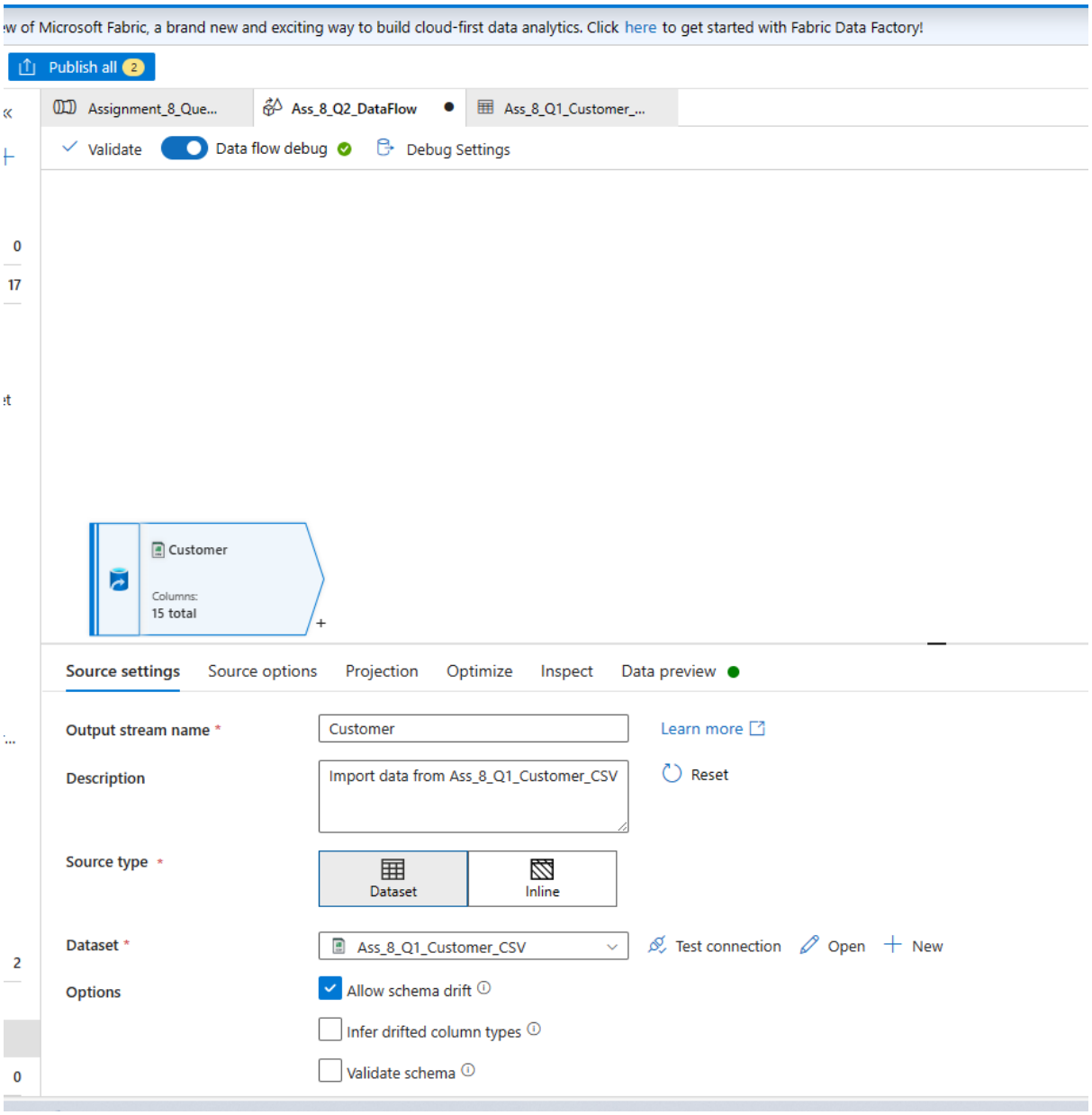
2. Create a pipeline to join the two files (Customer, Customer Address) based on customer id and save the result as a JSON file.

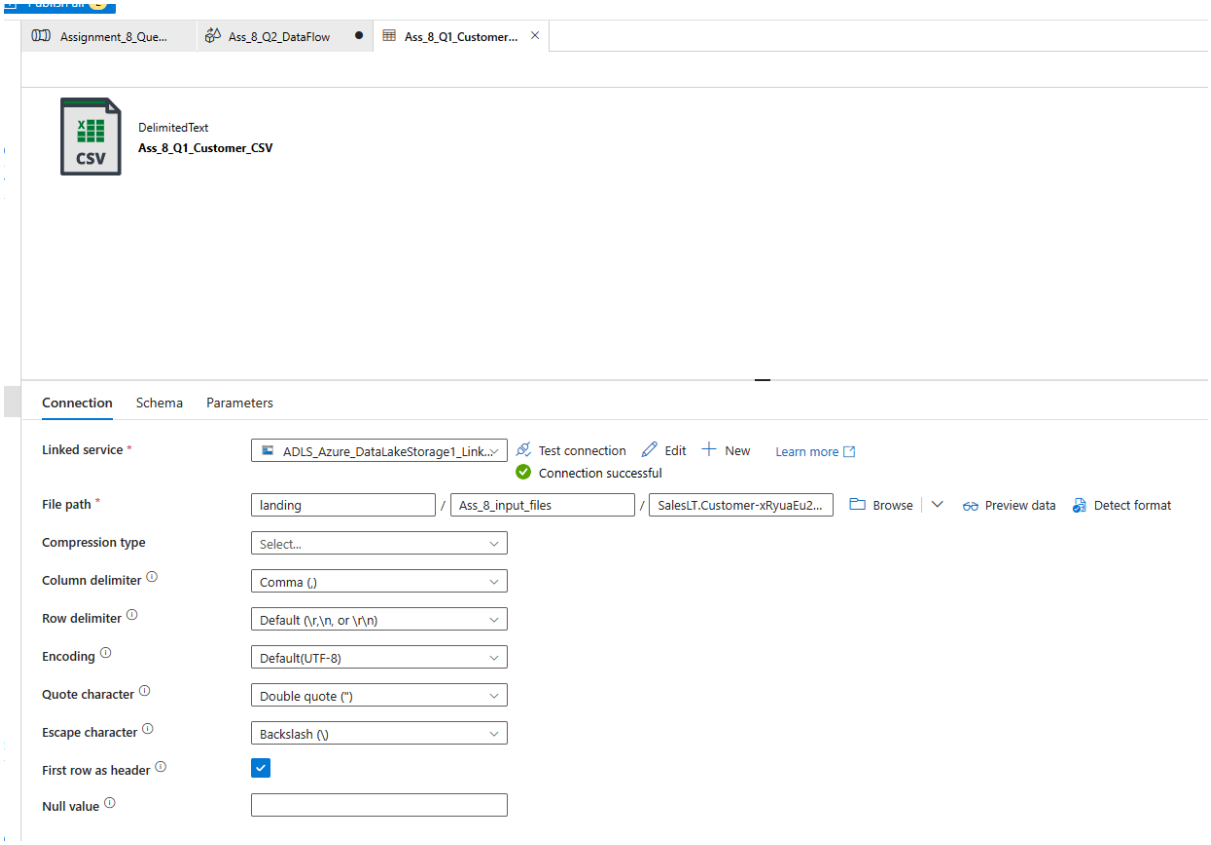
Solution:


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
Filter resources by name
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...
Assignment_8_Que...
Ass_8_Q2_DataFlow
Ass_8_Q
Validate
Data flow debug
Debug Settings
Pipeline
Change Data Capture (preview)
Dataset
Data flow
Power Query
Copy Data tool
Data flow
Flowlet


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


Threshold_text_file_DS
Data flows 2
Ass_8_Q2_DataFlow
Ass_8_Q2_DataFlow
Power Query 0
Dataset *
Options


w 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
Assignment_8_Que...
Ass_8_Q2_DataFlow
Ass_8_Q1_Customer_...
Validate
Data flow debug
Debug Settings
Customer
Columns:
15 total
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Output stream name *
Customer
Learn more
Description
Import data from Ass_8_Q1_Customer_CSV
Reset
Source type *
Dataset
Inline
Dataset *
Ass_8_Q1_Customer_CSV
Test connection
Open
New
Options
Allow schema drift
Infer drifted column types
Validate schema

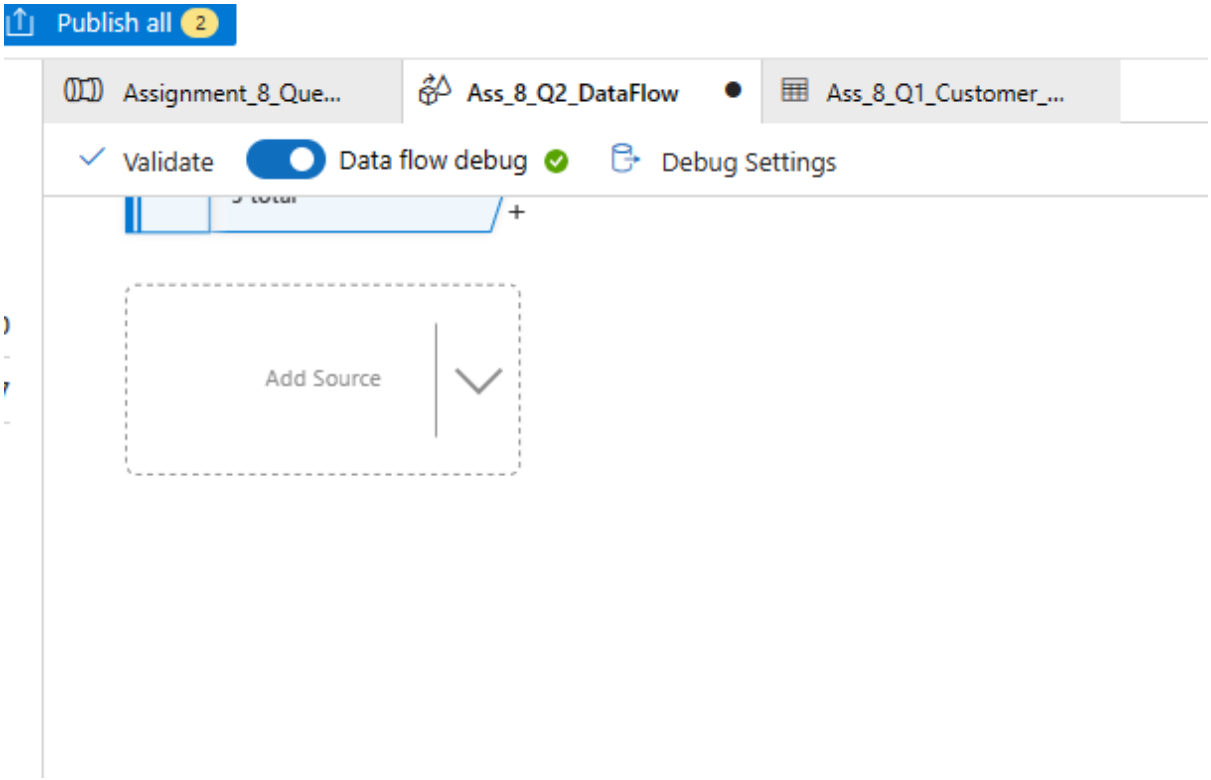

Assignment_8_Que...
Ass_8_Q2_DataFlow
Ass_8_Q1_Customer...
DelimitedText
Ass_8_Q1_Customer_CSV
Connection
Schema
Parameters
Linked service *
ADLS_Azure_DataLakeStorage1_Link...
Test connection
Edit
New
Learn more
Connection successful
File path *
landing
Ass_8_input_files
SalesLT.Customer-xRyuaEu2...
Browse
Preview data
Detect format
Compression type
Select...
Column delimiter
Comma (,)
Row delimiter
Default (\r,\n, or \r\n)
Encoding
Default(UTF-8)
Quote character
Double quote (")
Escape character
Backslash (\)
First row as header
Null value

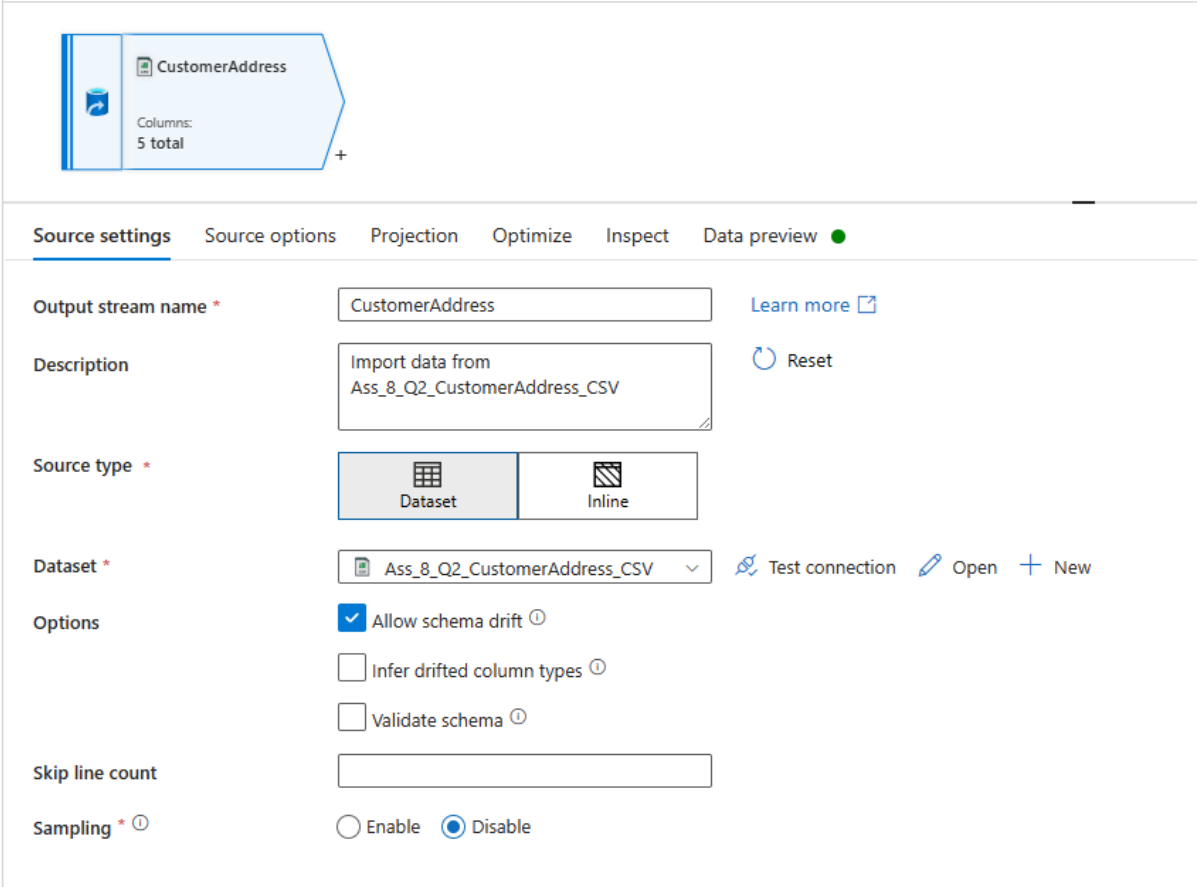
Preview data

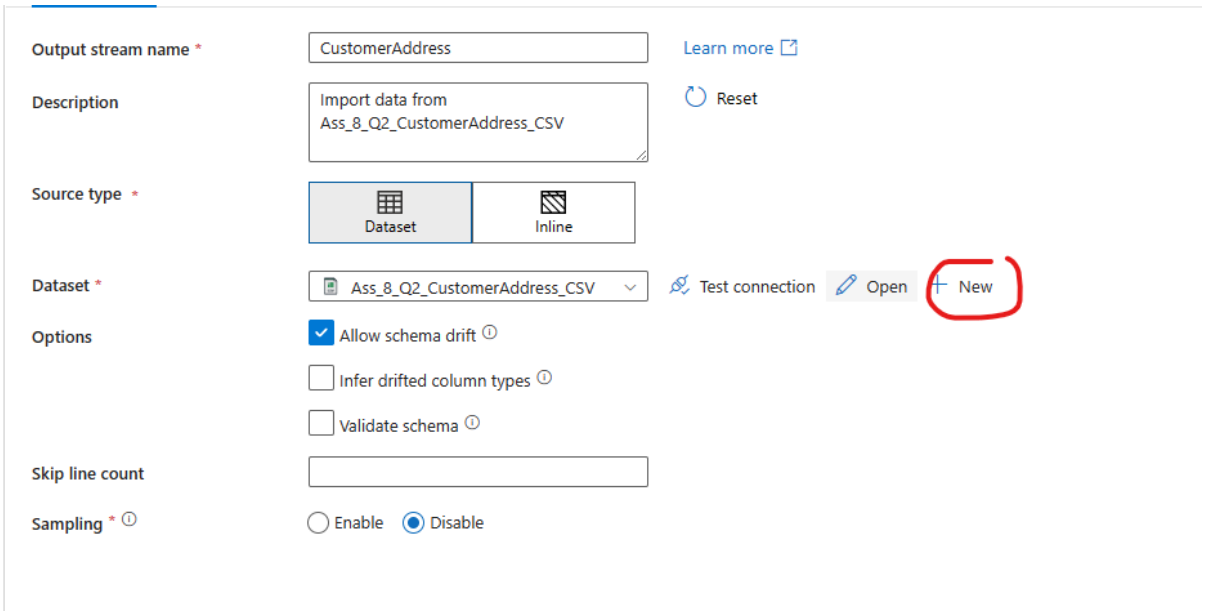
Linked service: ADLS_Azure_DataLakeStorage1_Linked_service

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

	CustomerID	NameStyle	Title	FirstName	MiddleName	LastName	Suffix	CompanyName	SalesF
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




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_CSV
Reset
Source type *
Dataset
Inline
Dataset *
Ass_8_Q2_CustomerAddress_CSV
Test connection
Open
New
Options
Allow schema drift
Infer drifted column types
Validate schema
Skip line count
Sampling *
Enable
Disable


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
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

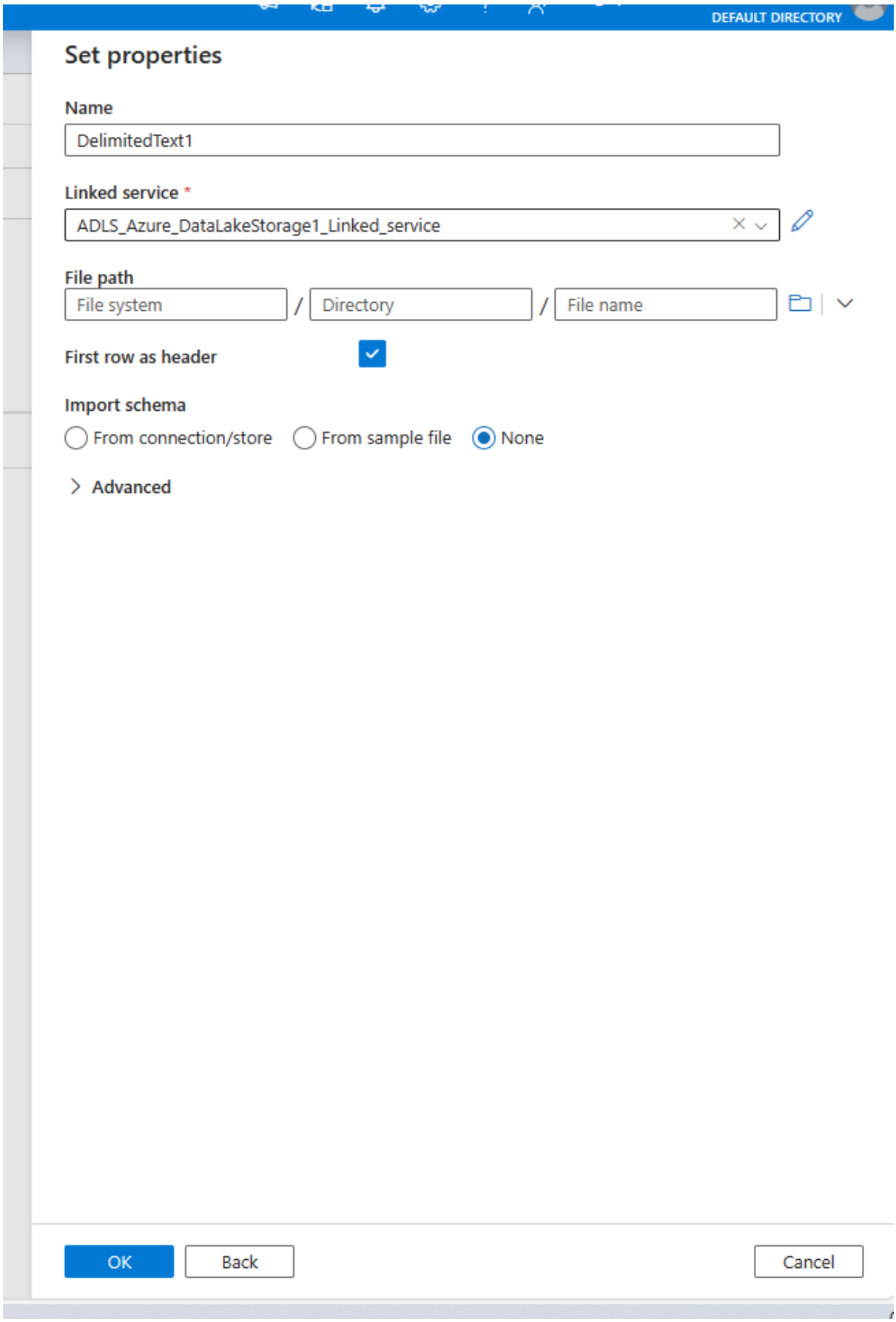
ORC

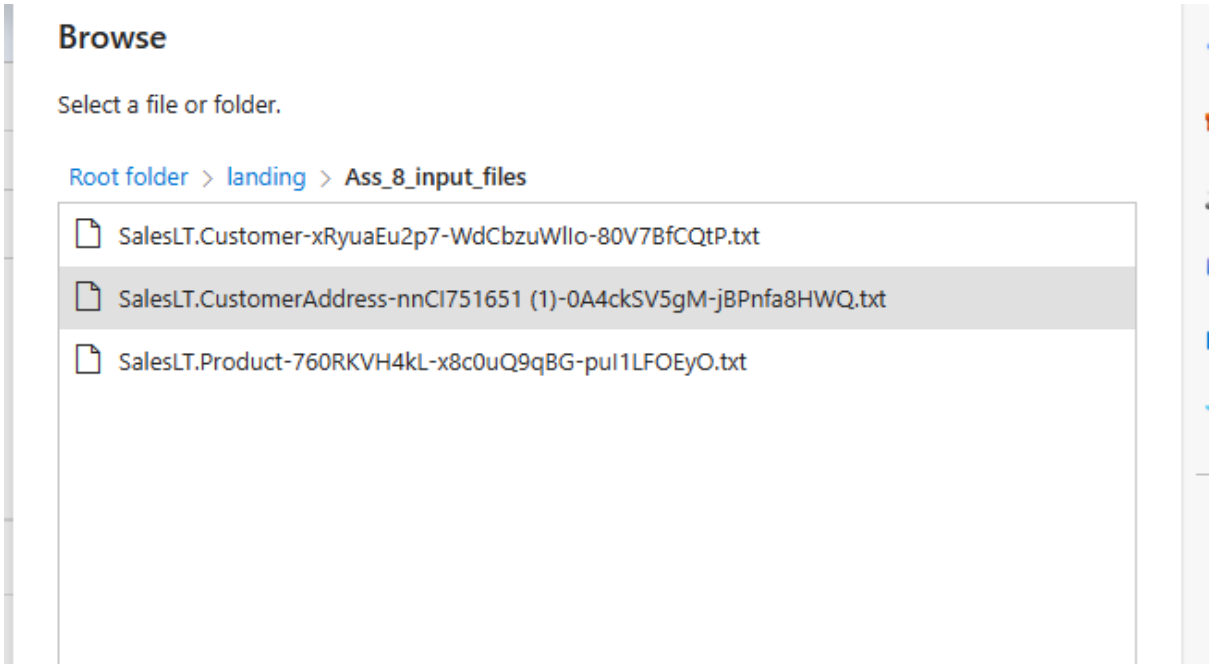
Parquet

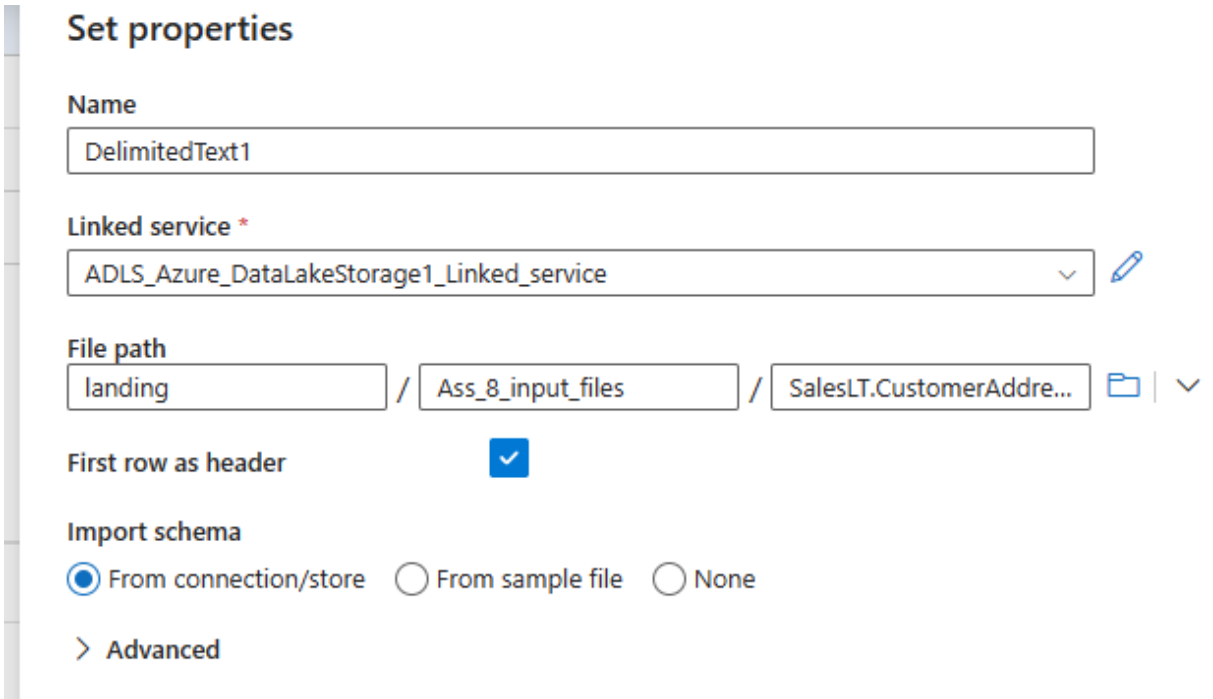
XML

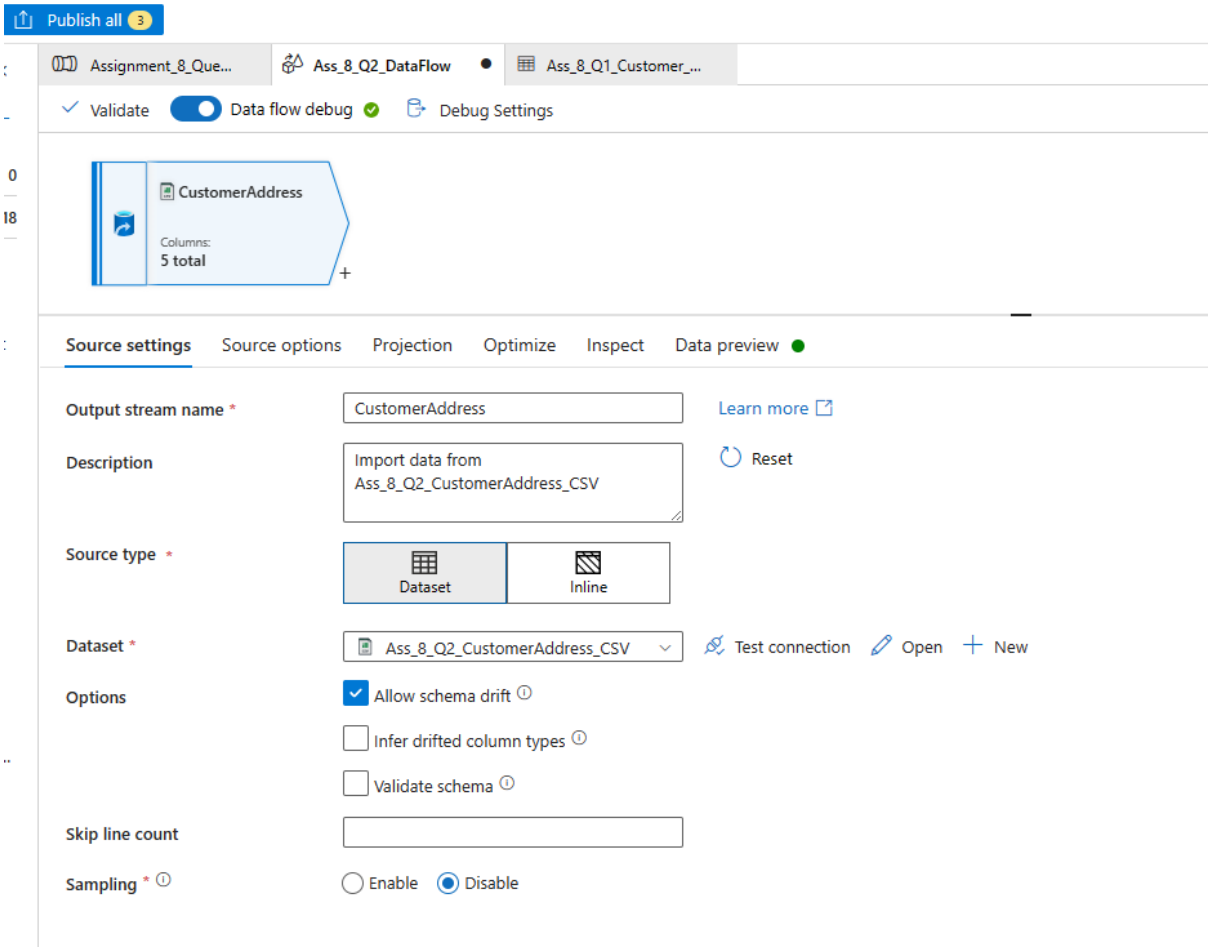
Binary

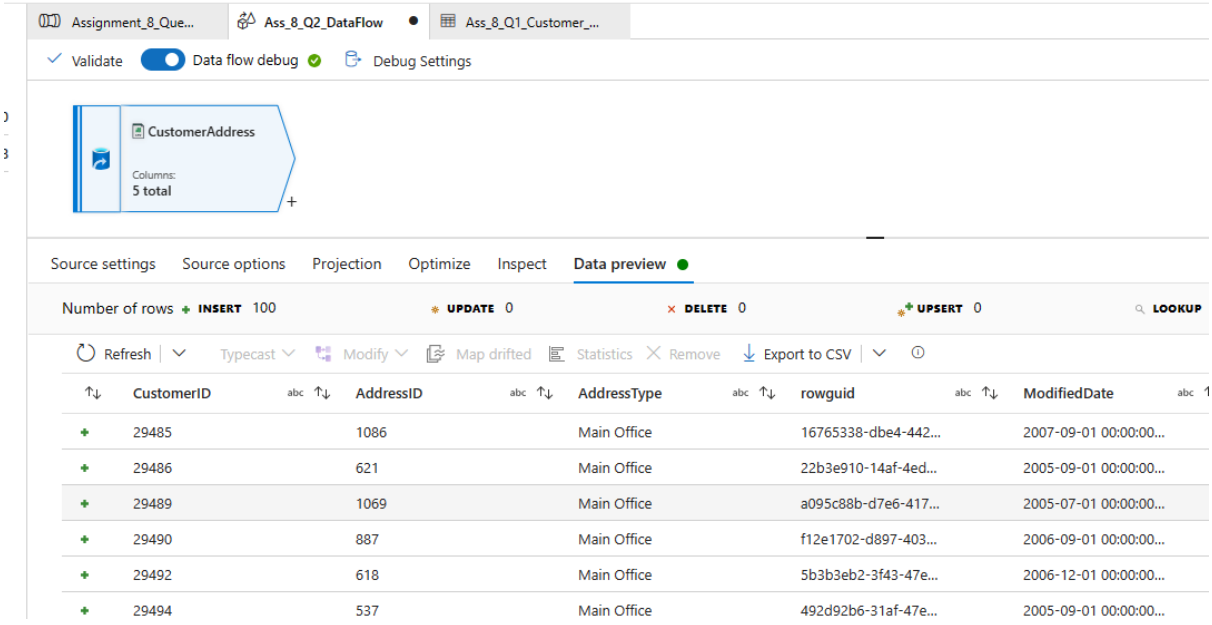
Continue | Back | Cancel


DEFAULT DIRECTORY
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

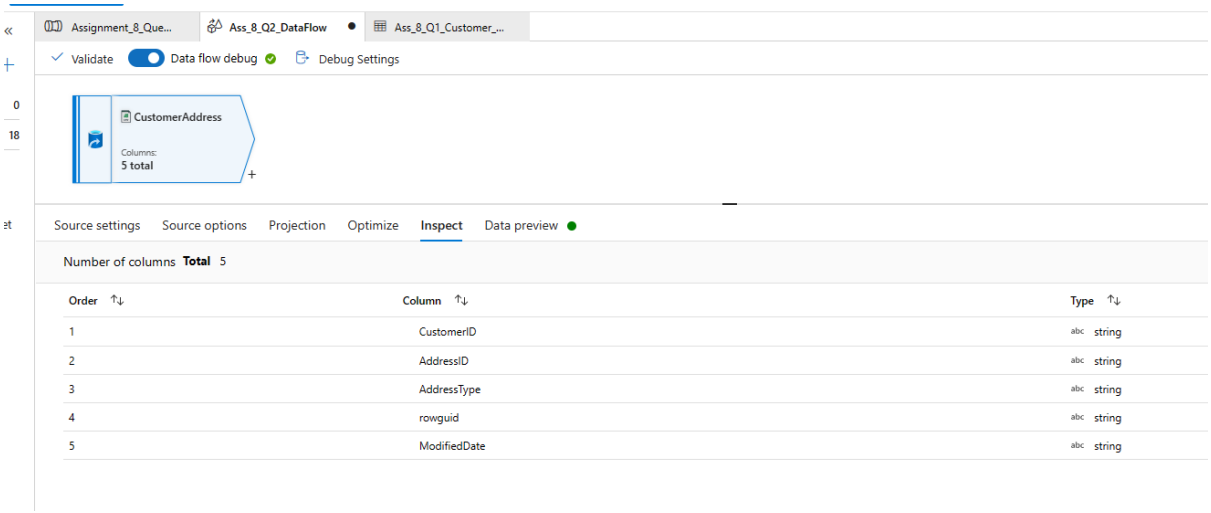

Browse
Select a file or folder.
Root folder > landing > Ass_8_input_files
SalesLT.Customer-xRyuaEu2p7-WdCbzuWIIo-80V7BfCQtP.txt
SalesLT.CustomerAddress-nnCI751651 (1)-0A4ckSV5gM-jBPnfa8HWQ.txt
SalesLT.Product-760RKVH4kL-x8c0uQ9qBG-puI1LFOEyO.txt

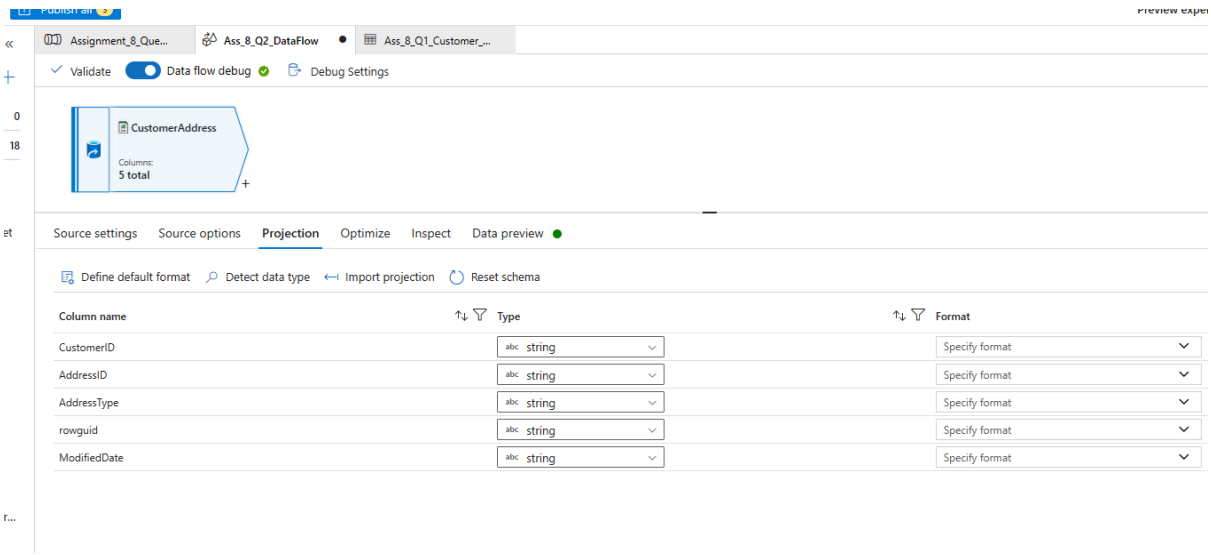

Set properties
Name
DelimitedText1
Linked service *
ADLS_Azure_DataLakeStorage1_Linked_service
File path
landing
Ass_8_input_files
SalesLT.CustomerAddre...
First row as header
Import schema
From connection/store
From sample file
None
Advanced

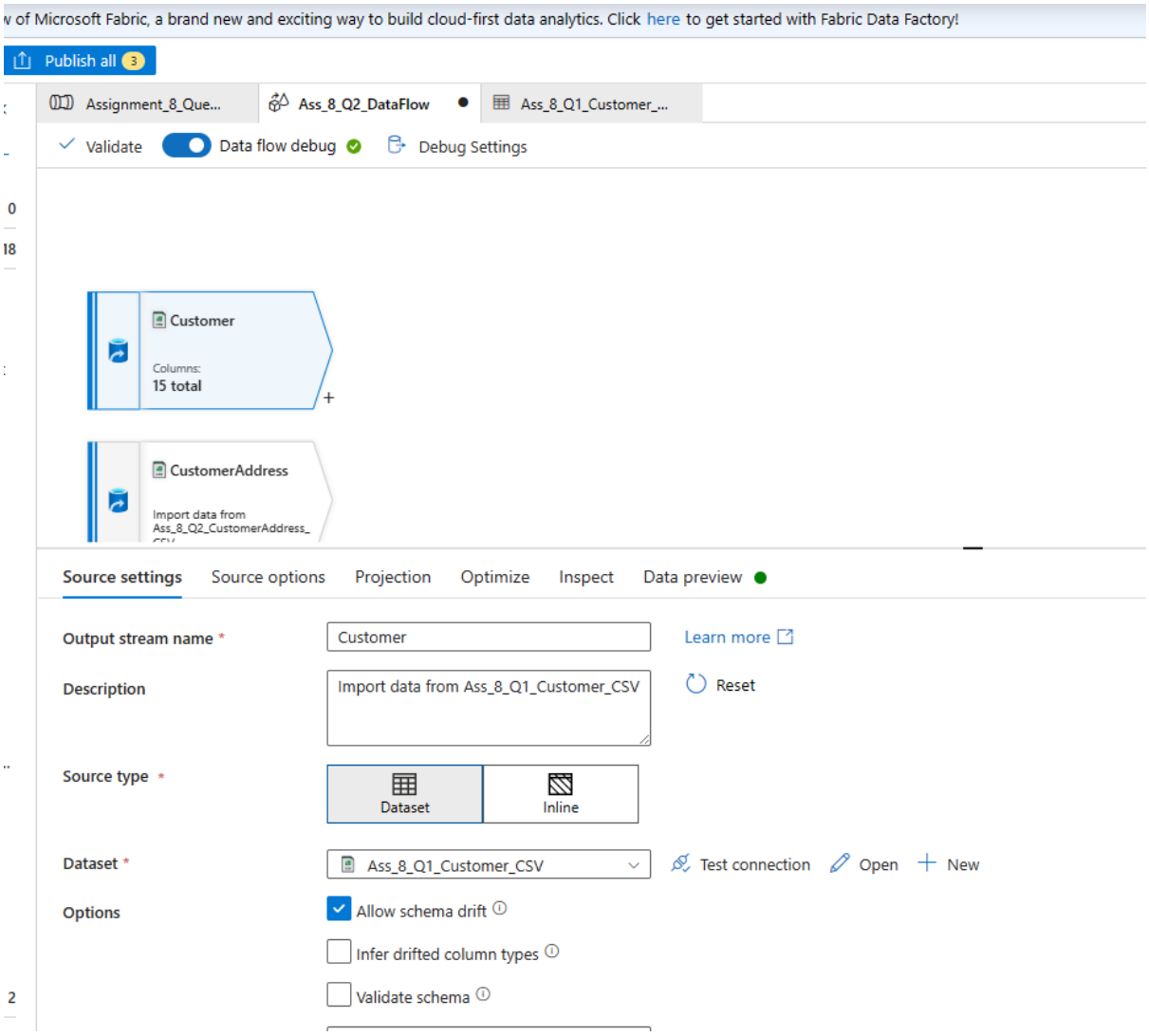


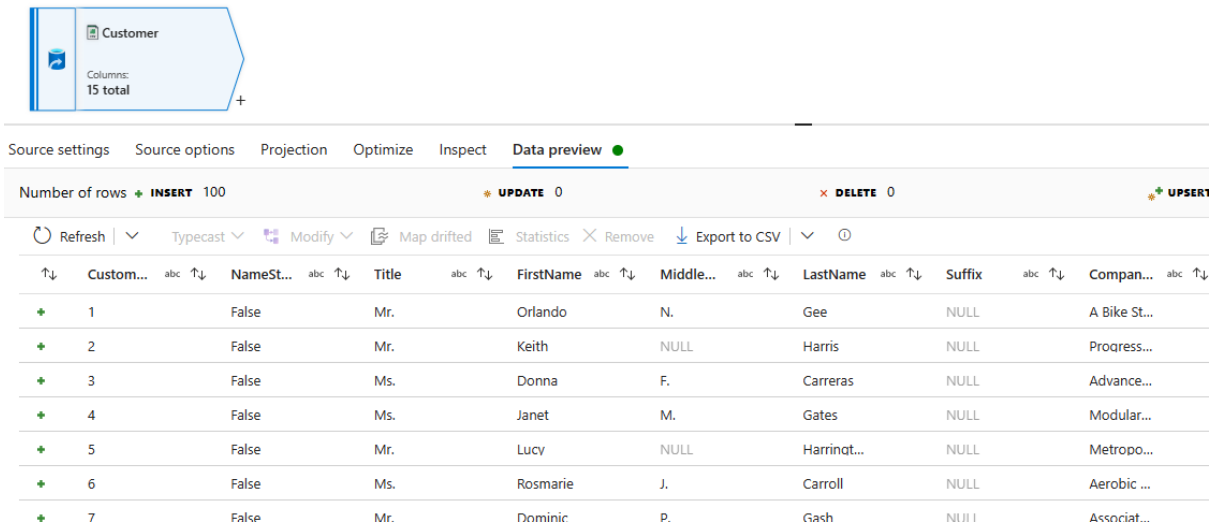


CustomerID	AddressID	AddressType	rowguid	ModifiedDate
29485	1086	Main Office	16765338-dbe4-442...	2007-09-01 00:00:00...
29486	621	Main Office	22b3e910-14af-4ed...	2005-09-01 00:00:00...
29489	1069	Main Office	a095c88b-d7e6-417...	2005-07-01 00:00:00...
29490	887	Main Office	f12e1702-d897-403...	2006-09-01 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_Que...
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 string
2 AddressID string
3 AddressType string
4 rowguid string
5 ModifiedDate string

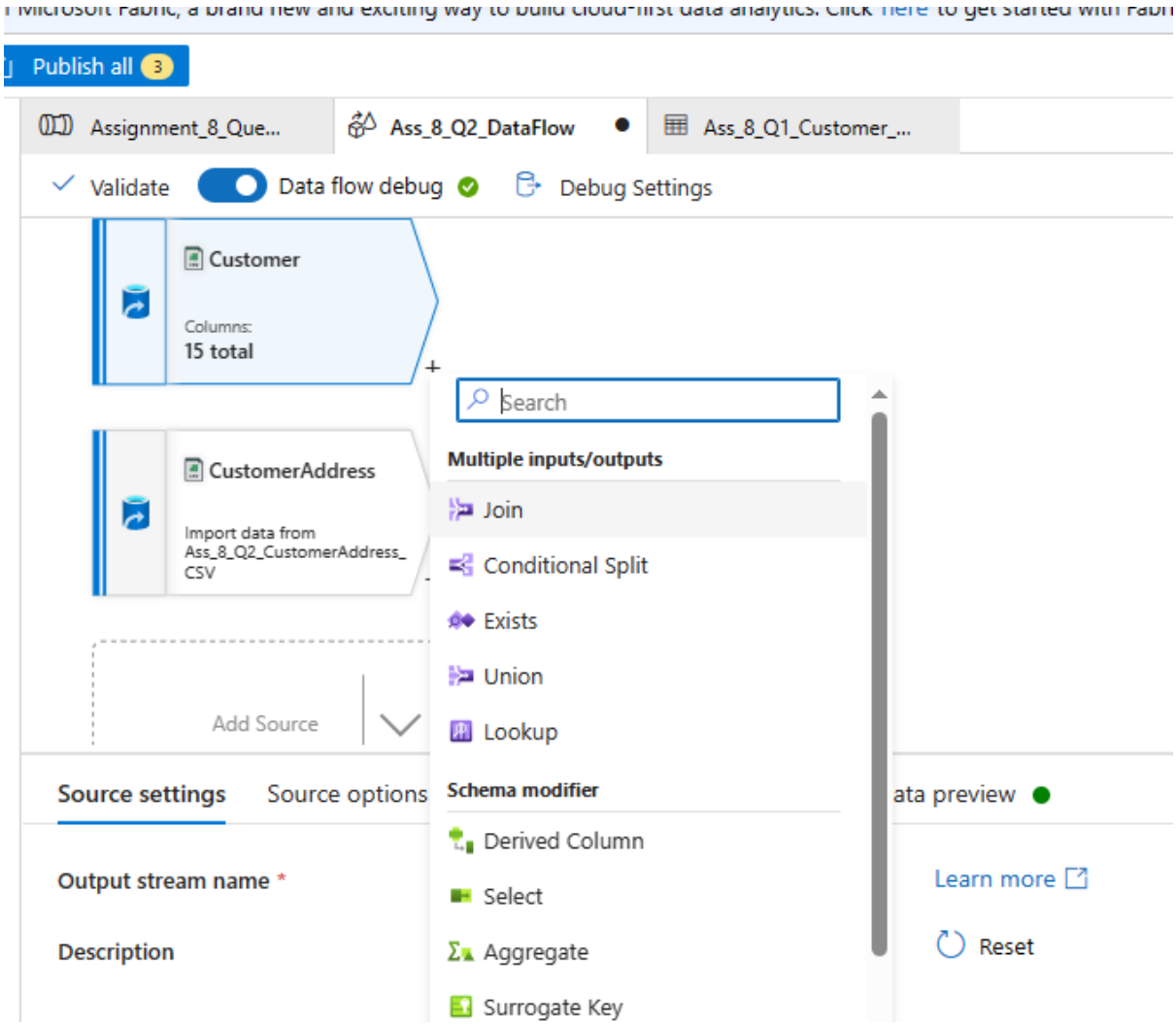

Assignment_8_Que...
Ass_8_Q2_DataFlow
Ass_8_Q1_Customer_...
Validate
Data flow debug
Debug Settings
CustomerAddress
Columns:
5 total
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Define default format
Detect data type
Import projection
Reset schema
Column name
Type
Format
CustomerID string Specify format
AddressID string Specify format
AddressType string Specify format
rowguid string Specify format
ModifiedDate string Specify format

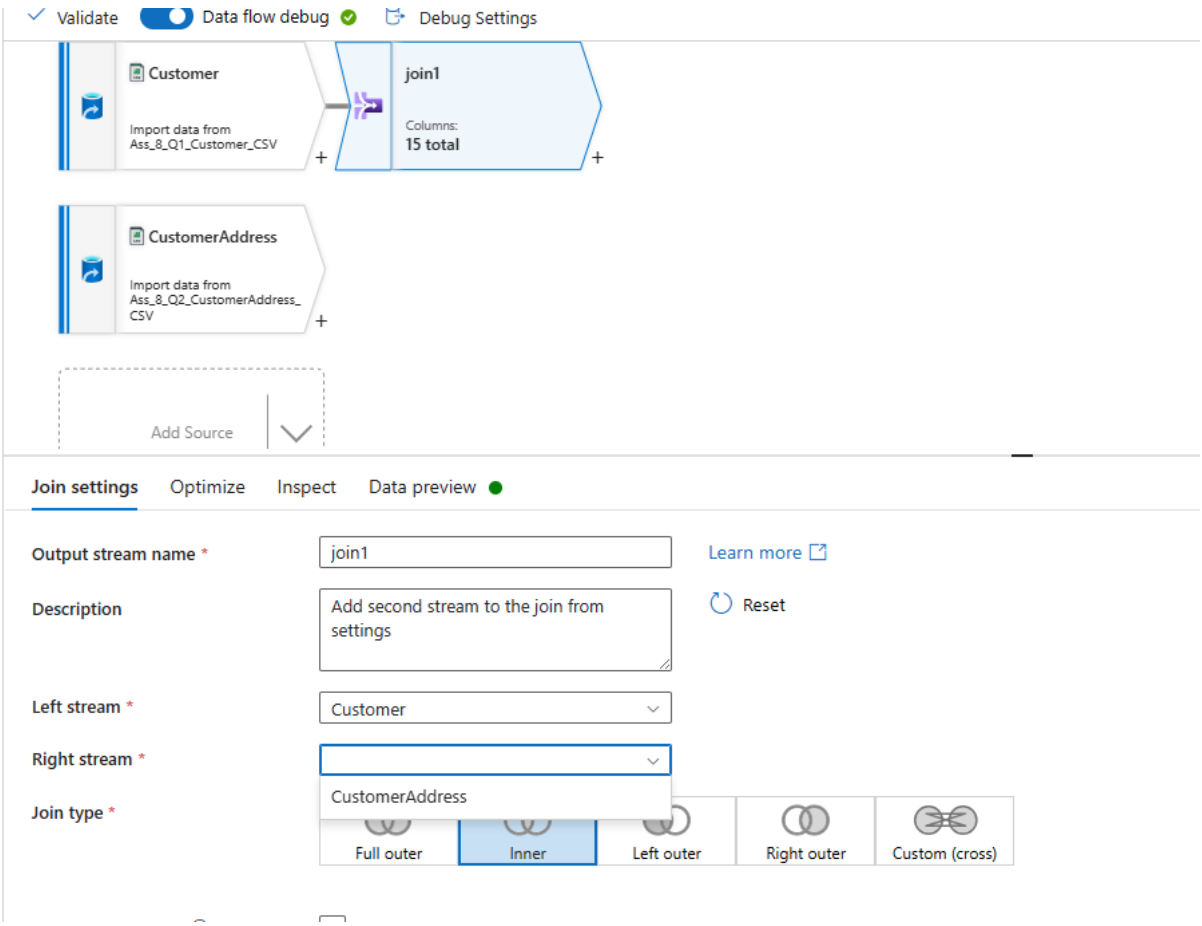


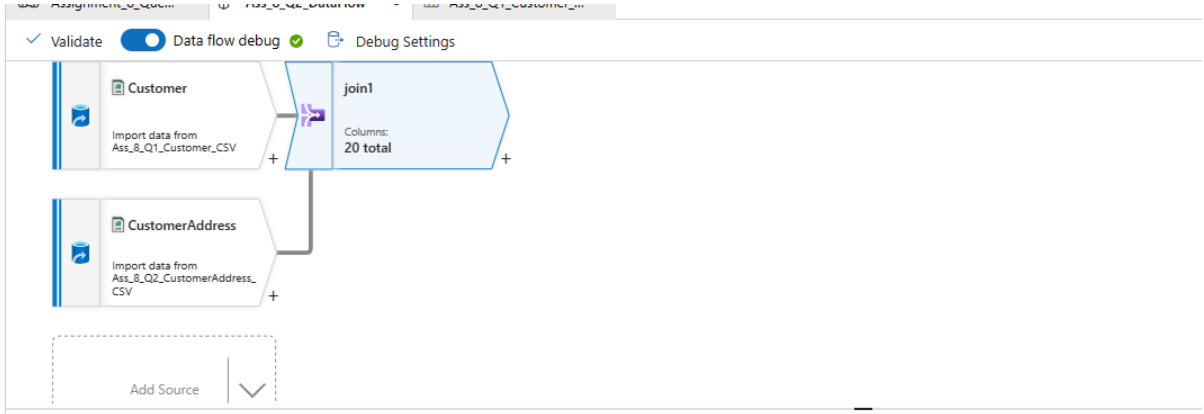


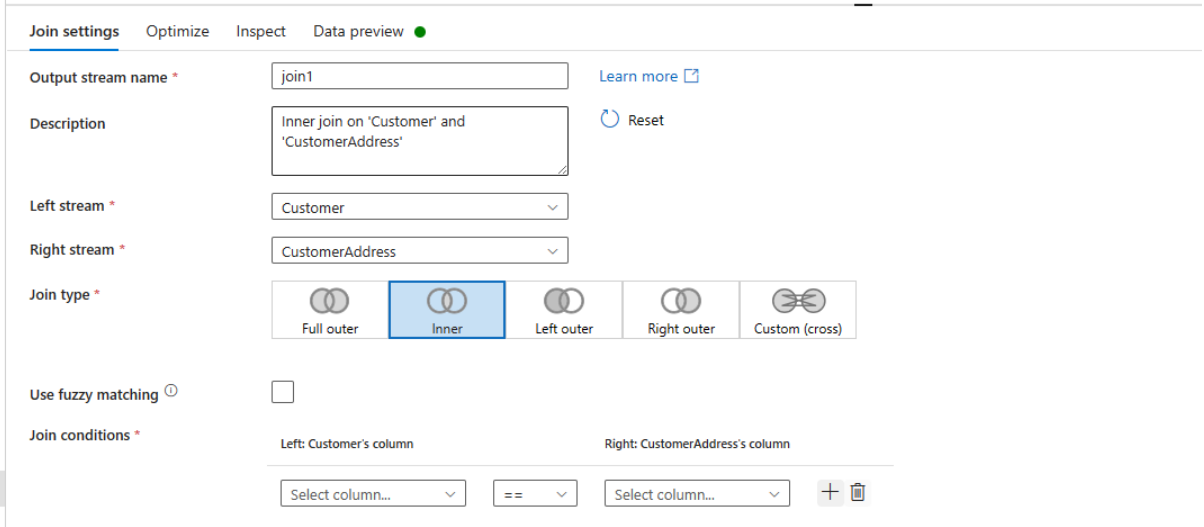
	Custom...	NameSt...	Title	FirstName	Middle...	LastName	Suffix	Compan...
+	1	False	Mr.	Orlando	N.	Gee	NULL	A Bike St...
+	2	False	Mr.	Keith	NULL	Harris	NULL	Progress...
+	3	False	Ms.	Donna	F.	Carreras	NULL	Advance...
+	4	False	Ms.	Janet	M.	Gates	NULL	Modular...
+	5	False	Mr.	Lucy	NULL	Harringt...	NULL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

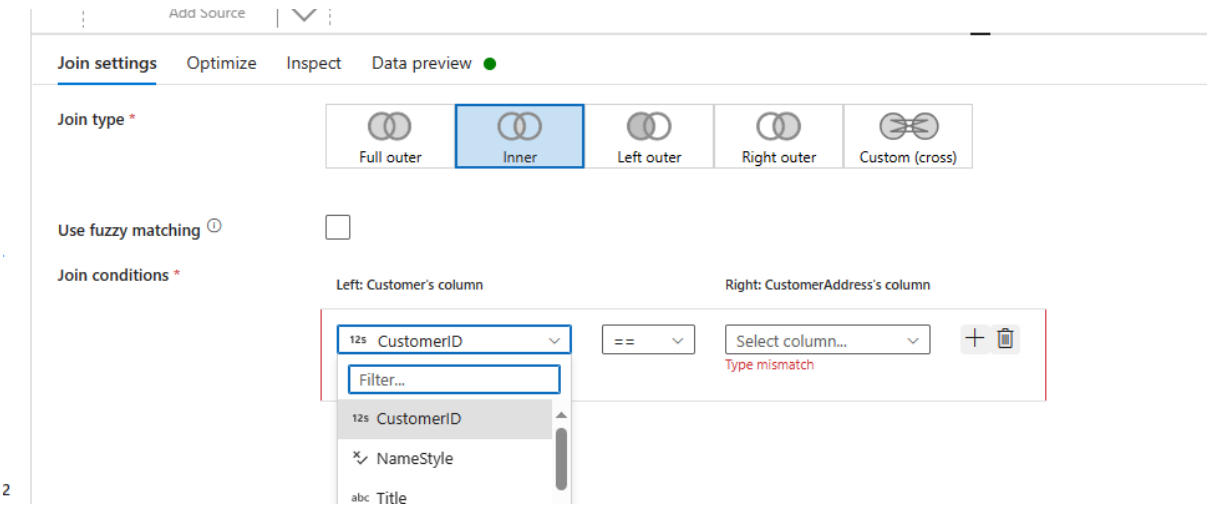
2. Create a pipeline to join the two files (Customer, Customer Address) based on customer id and save the result as a JSON file.

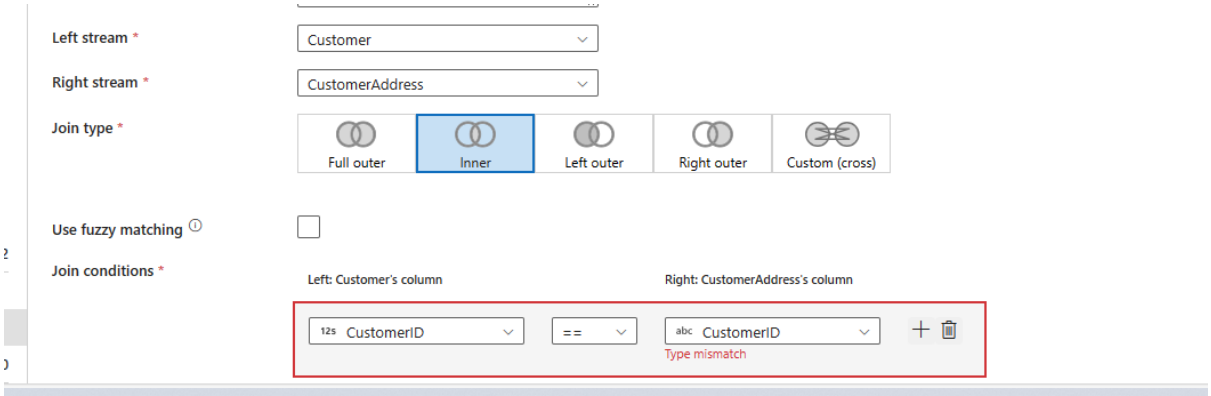


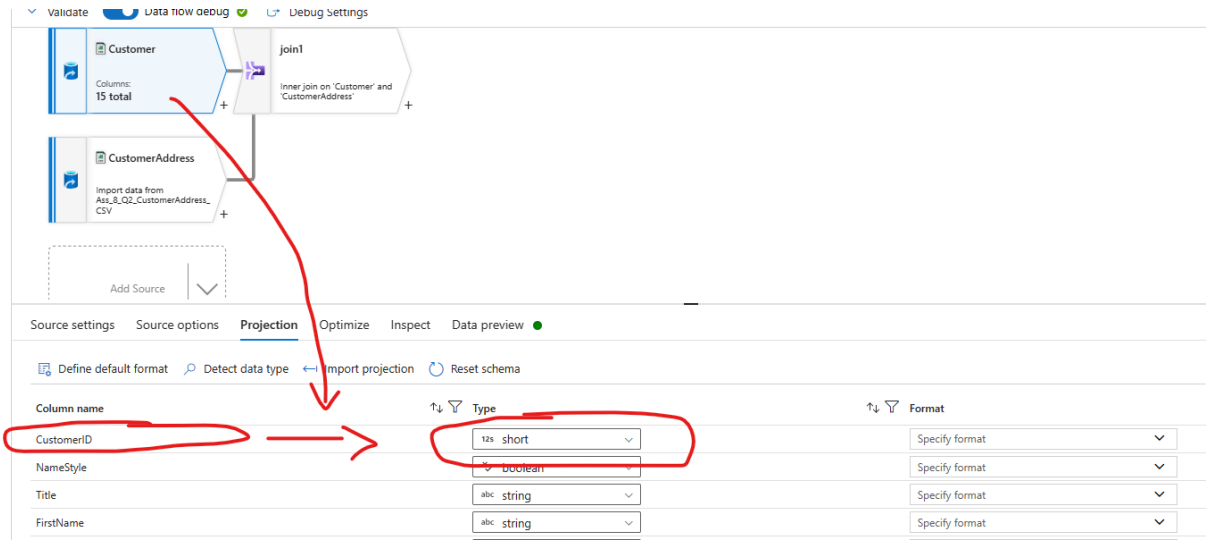

Validate
Data flow debug
Debug Settings
Customer
Import data from Ass_8_Q1_Customer_CSV
join1
Columns:
15 total
CustomerAddress
Import data from Ass_8_Q2_CustomerAddress_CSV
Add Source
Join settings
Optimize
Inspect
Data preview
Output stream name *
join1
Learn more
Description
Add second stream to the join from settings
Reset
Left stream *
Customer
Right stream *
CustomerAddress
Join type *
Full outer
Inner
Left outer
Right outer
Custom (cross)

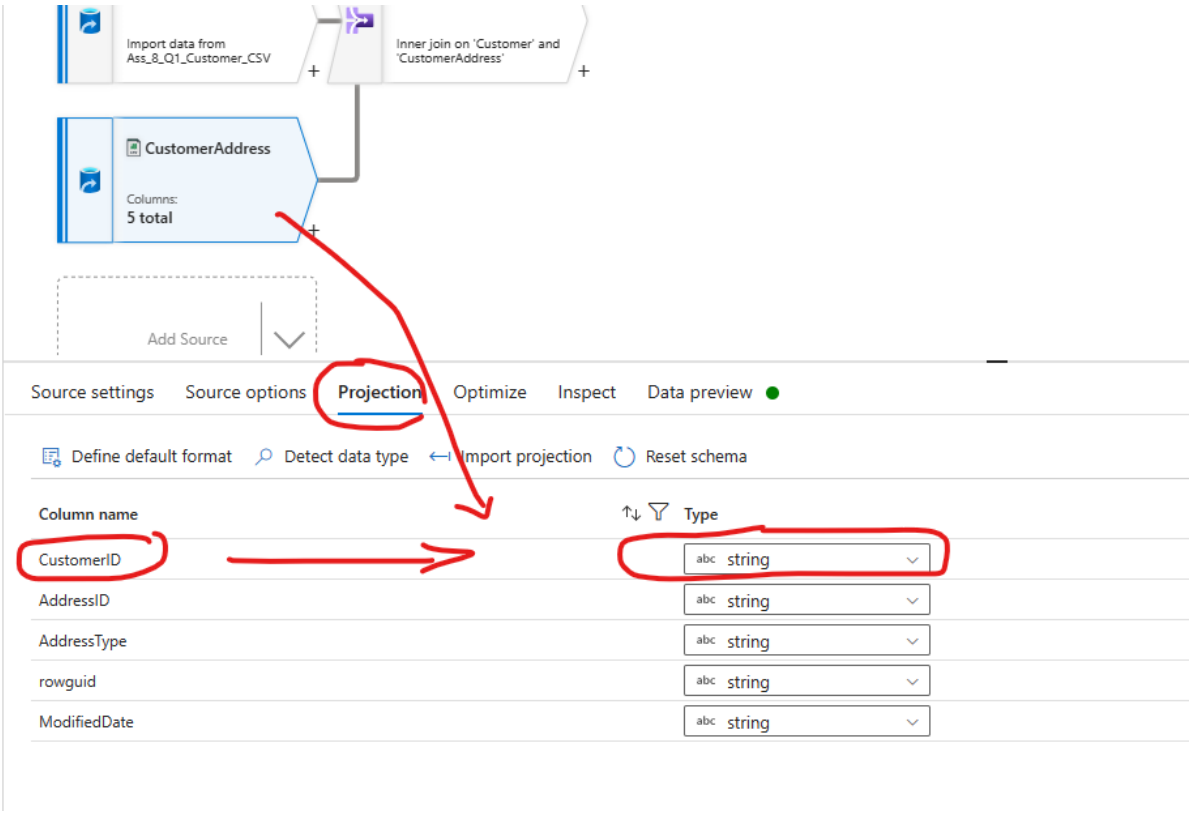

Validate
Data flow debug
Debug Settings
Customer
Import data from Ass_8_Q1_Customer_CSV
join1
Columns:
20 total
CustomerAddress
Import data from Ass_8_Q2_CustomerAddress_CSV
Add Source

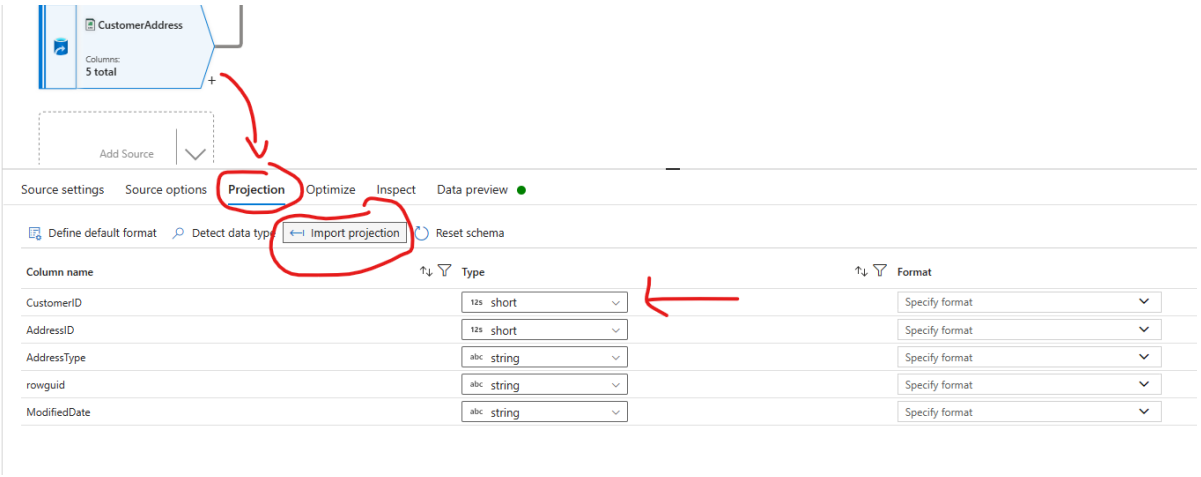

Join settings
Optimize
Inspect
Data preview
Output stream name *
join1
Learn more
Description
Inner join on 'Customer' and 'CustomerAddress'
Reset
Left stream *
Customer
Right stream *
CustomerAddress
Join type *
Full outer
Inner
Left outer
Right outer
Custom (cross)
Use fuzzy matching
Join conditions *
Left: Customer's column
Right: CustomerAddress's column
Select column...
==
Select column...

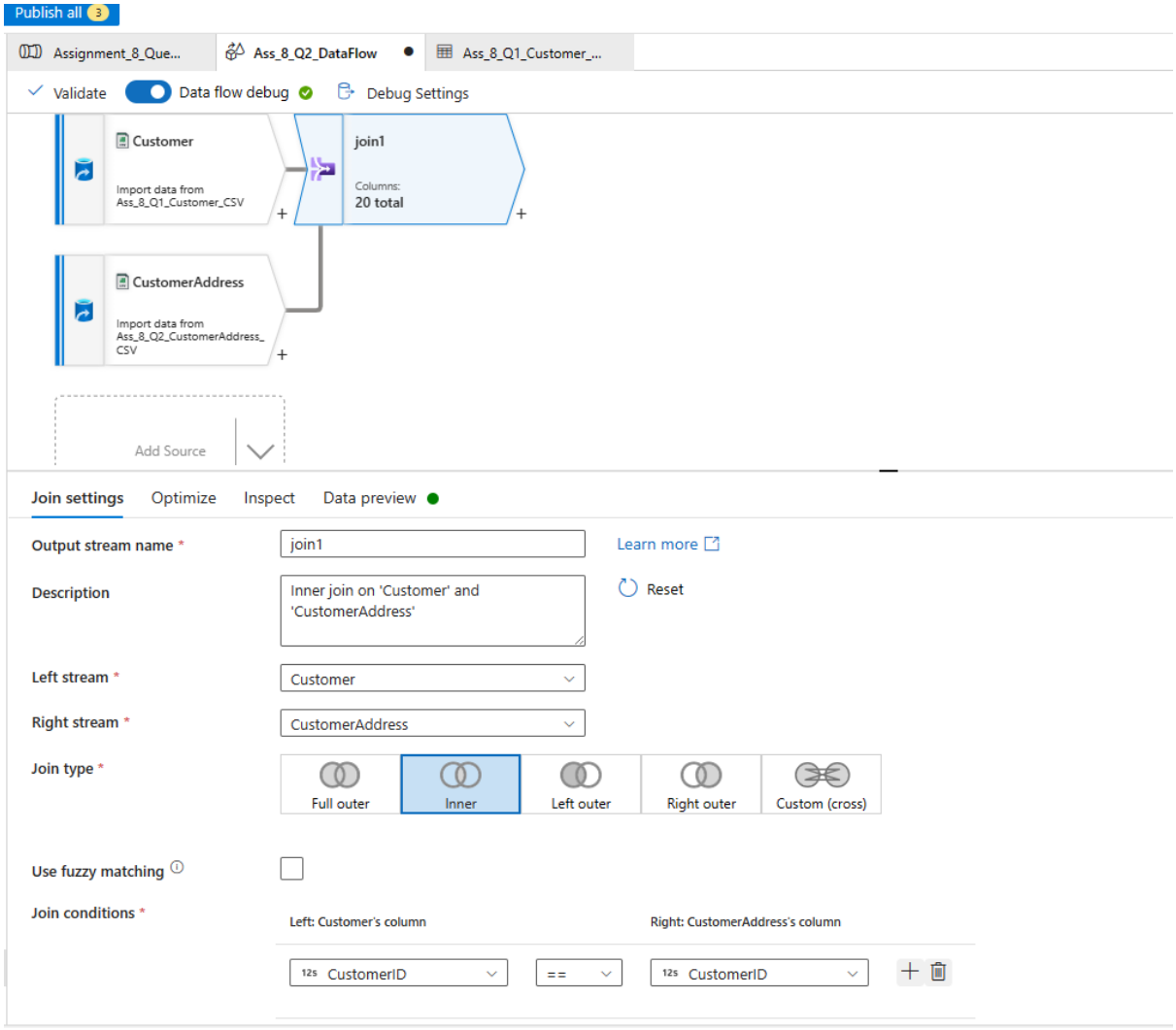

Add Source
Join settings
Optimize
Inspect
Data preview
Join type *
Full outer
Inner
Left outer
Right outer
Custom (cross)
Use fuzzy matching
Join conditions *
Left: Customer's column
Right: CustomerAddress's column
CustomerID
==
Select column...
Type mismatch
Filter...
CustomerID
NameStyle
Title

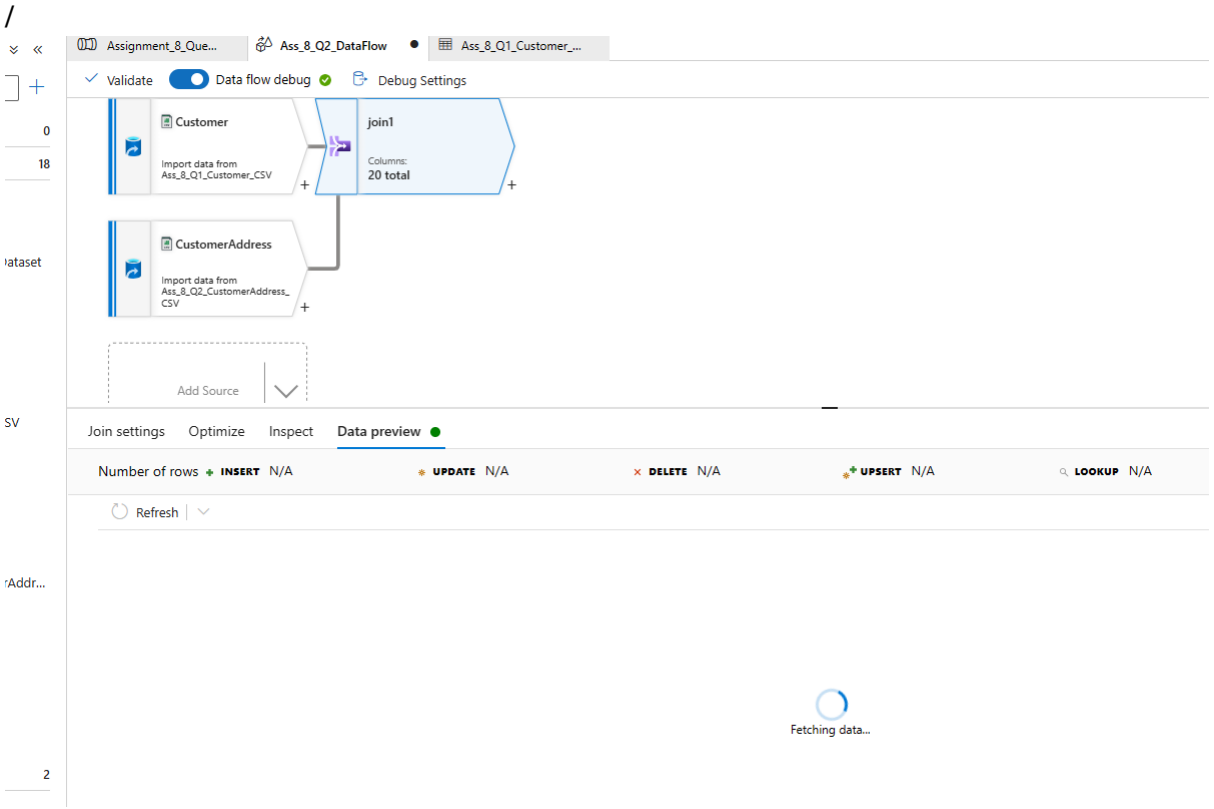

Left stream *
Customer
Right stream *
CustomerAddress
Join type *
Full outer
Inner
Left outer
Right outer
Custom (cross)
Use fuzzy matching
Join conditions *
Left: Customer's column
Right: CustomerAddress's column
CustomerID
==
CustomerID
Type mismatch

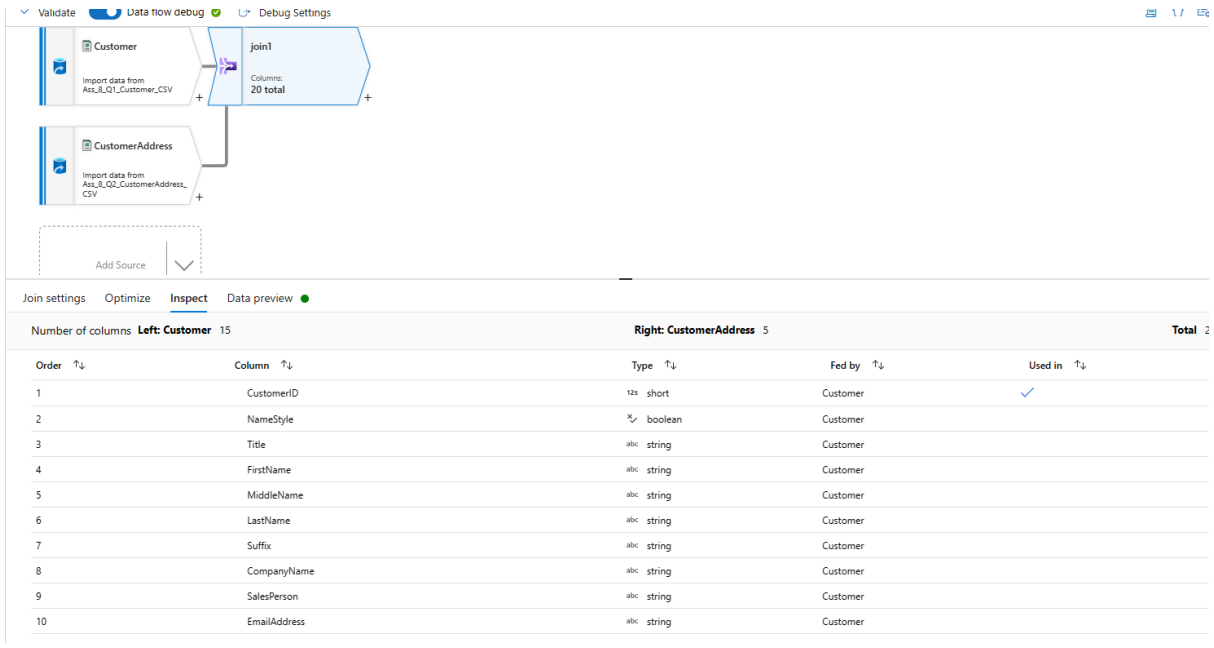

Validate
Data flow debug
Debug Settings
Customer
Columns:
15 total
join1
Inner join on 'Customer' and 'CustomerAddress'
CustomerAddress
Import data from Ass_8_Q2_CustomerAddress_CSV
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Define default format
Detect data type
Import projection
Reset schema
Column name
Type
Format
CustomerID
short
NameStyle
boolean
Title
string
FirstName
string
Specify format


Import data from Ass_8_Q1_Customer_CSV
Inner join on 'Customer' and 'CustomerAddress'
CustomerAddress
Columns:
5 total
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Define default format
Detect data type
Import projection
Reset schema
Column name
Type
CustomerID
AddressID
AddressType
rowguid
ModifiedDate
abc string


CustomerAddress
Columns:
5 total
Add Source
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Define default format
Detect data type
Import projection
Reset schema
Column name
Type
Format
CustomerID
AddressID
AddressType
rowguid
ModifiedDate
12s short
abc string
Specify format

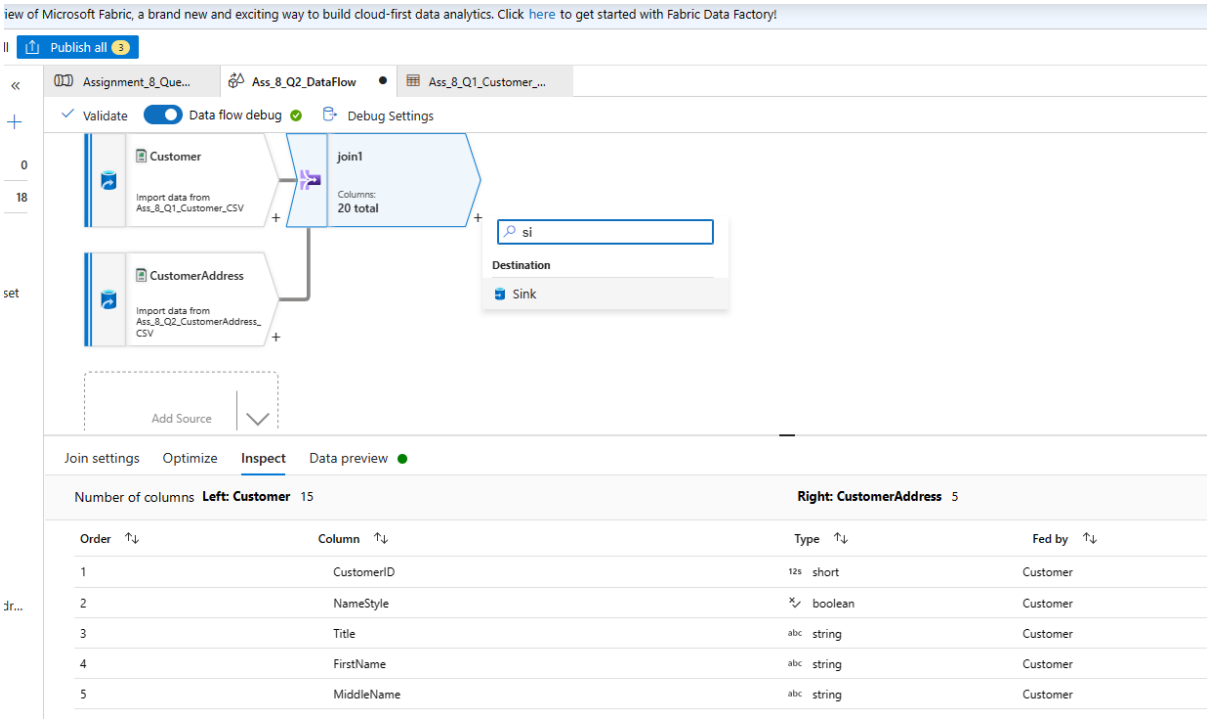

Publish all 3
Assignment_8_Que...
Ass_8_Q2_DataFlow
Ass_8_Q1_Customer_...
Validate
Data flow debug
Debug Settings
Customer
Import data from Ass_8_Q1_Customer_CSV
join1
Columns:
20 total
CustomerAddress
Import data from Ass_8_Q2_CustomerAddress_CSV
Add Source
Join settings
Optimize
Inspect
Data preview
Output stream name *
join1
Learn more
Description
Inner join on 'Customer' and 'CustomerAddress'
Reset
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
CustomerID
==
CustomerID

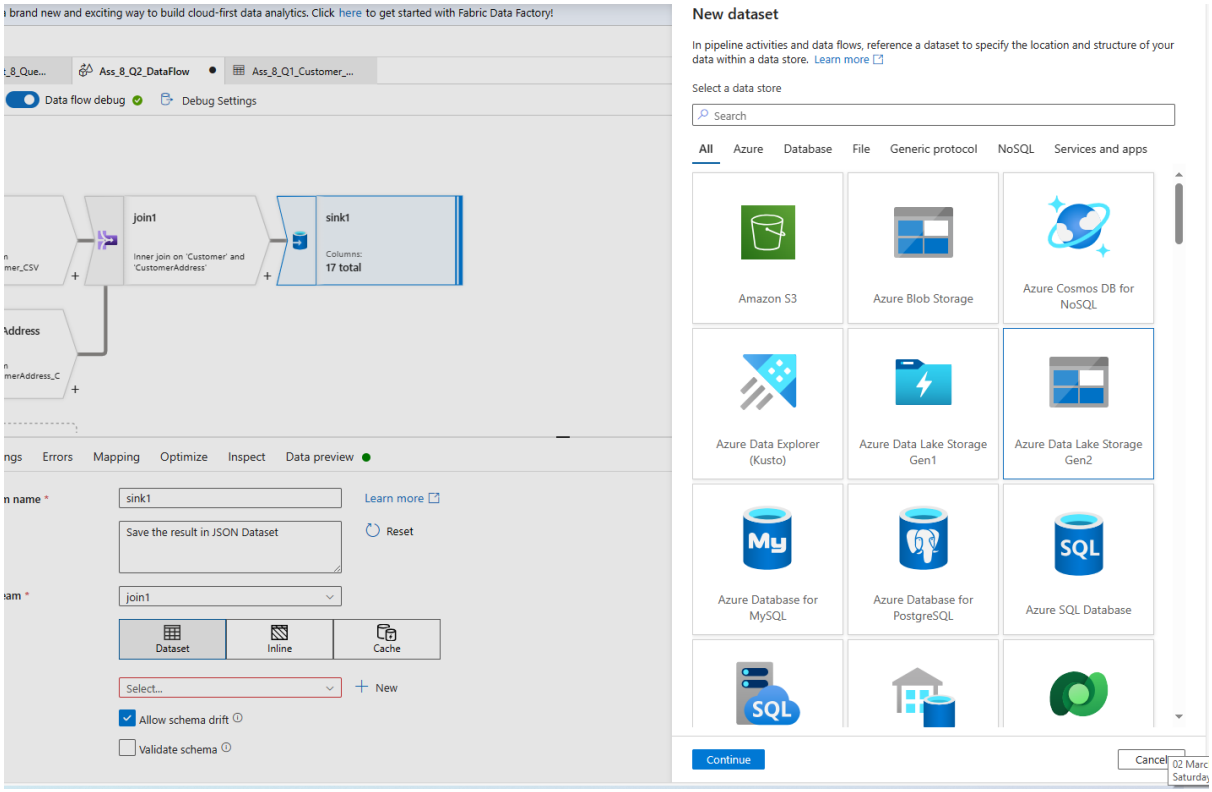


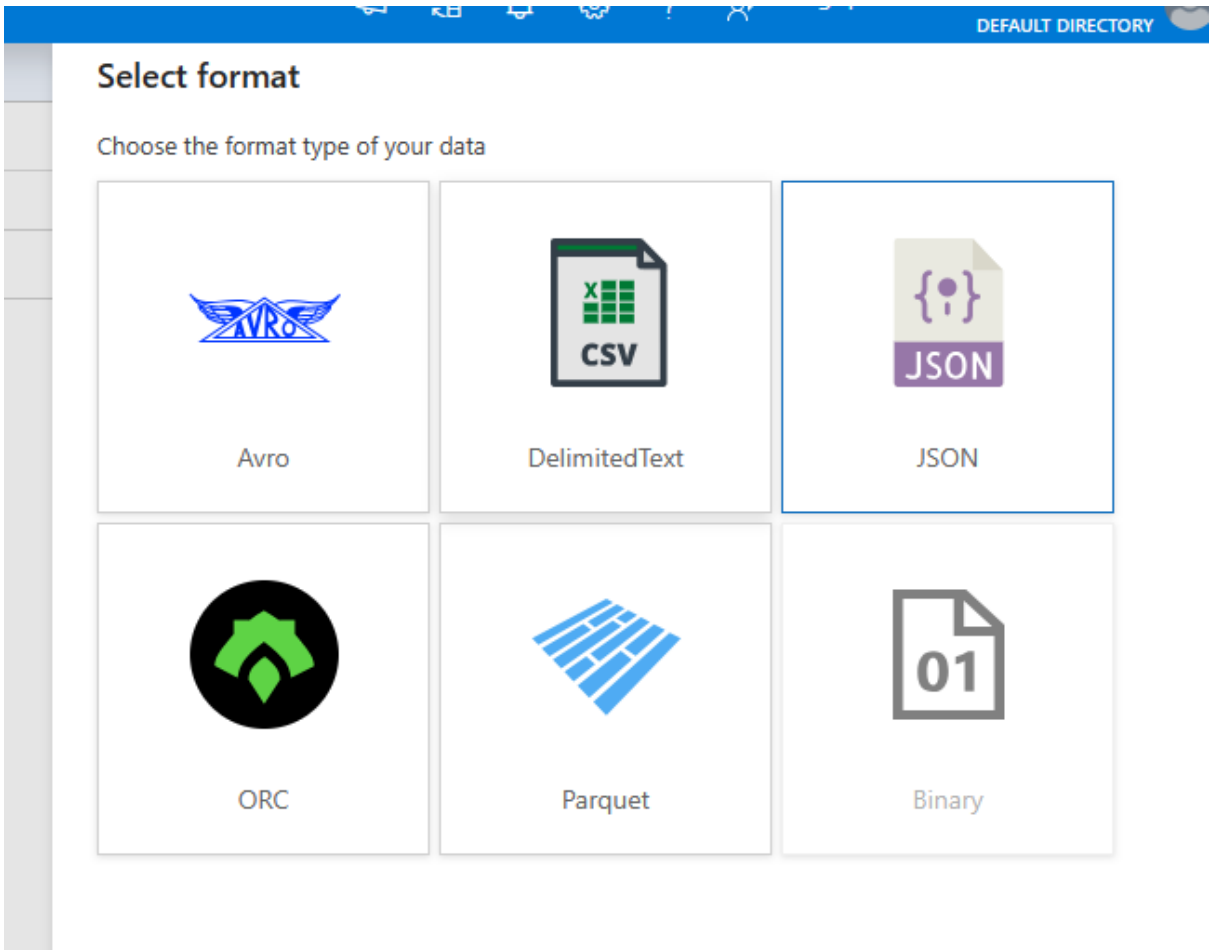


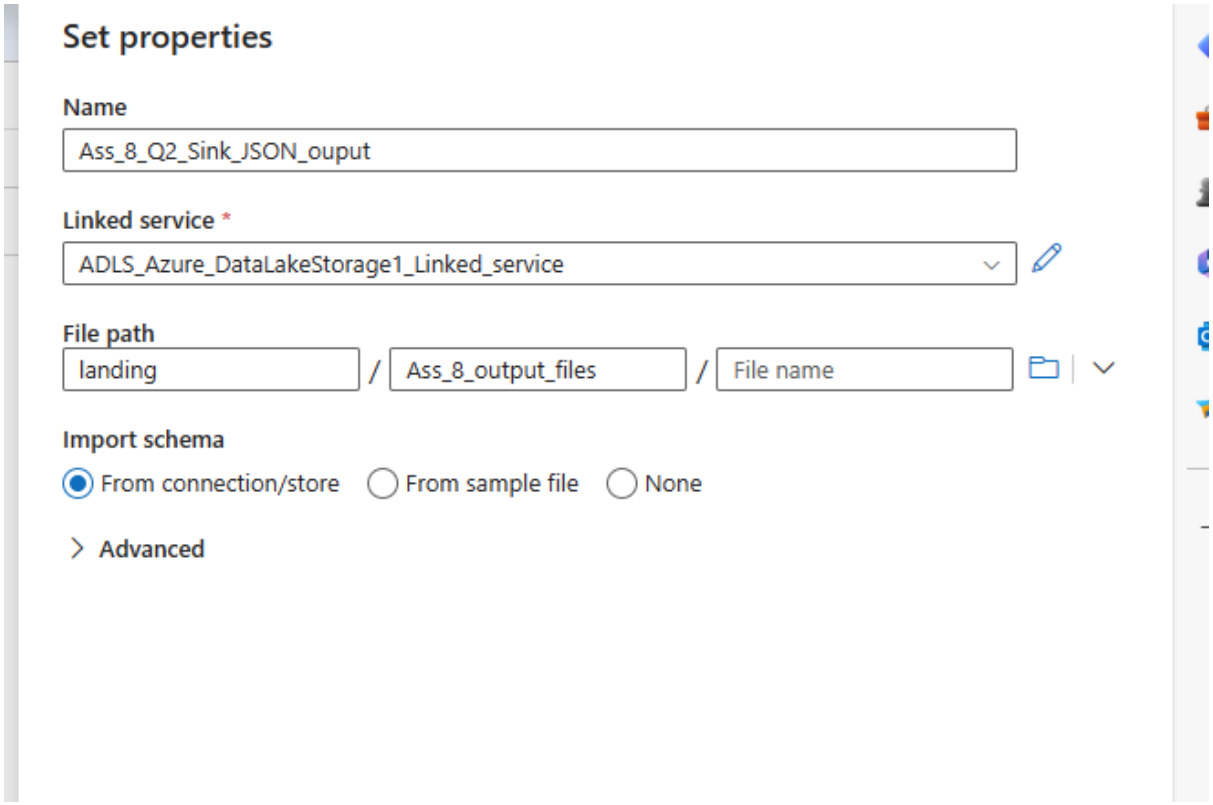
Order	Column	Type	Fed by	Used in
1	CustomerID	short	Customer	✓
2	NameStyle	boolean	Customer	
3	Title	string	Customer	
4	FirstName	string	Customer	
5	MiddleName	string	Customer	
6	LastName	string	Customer	
7	Suffix	string	Customer	
8	CompanyName	string	Customer	
9	SalesPerson	string	Customer	
10	EmailAddress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

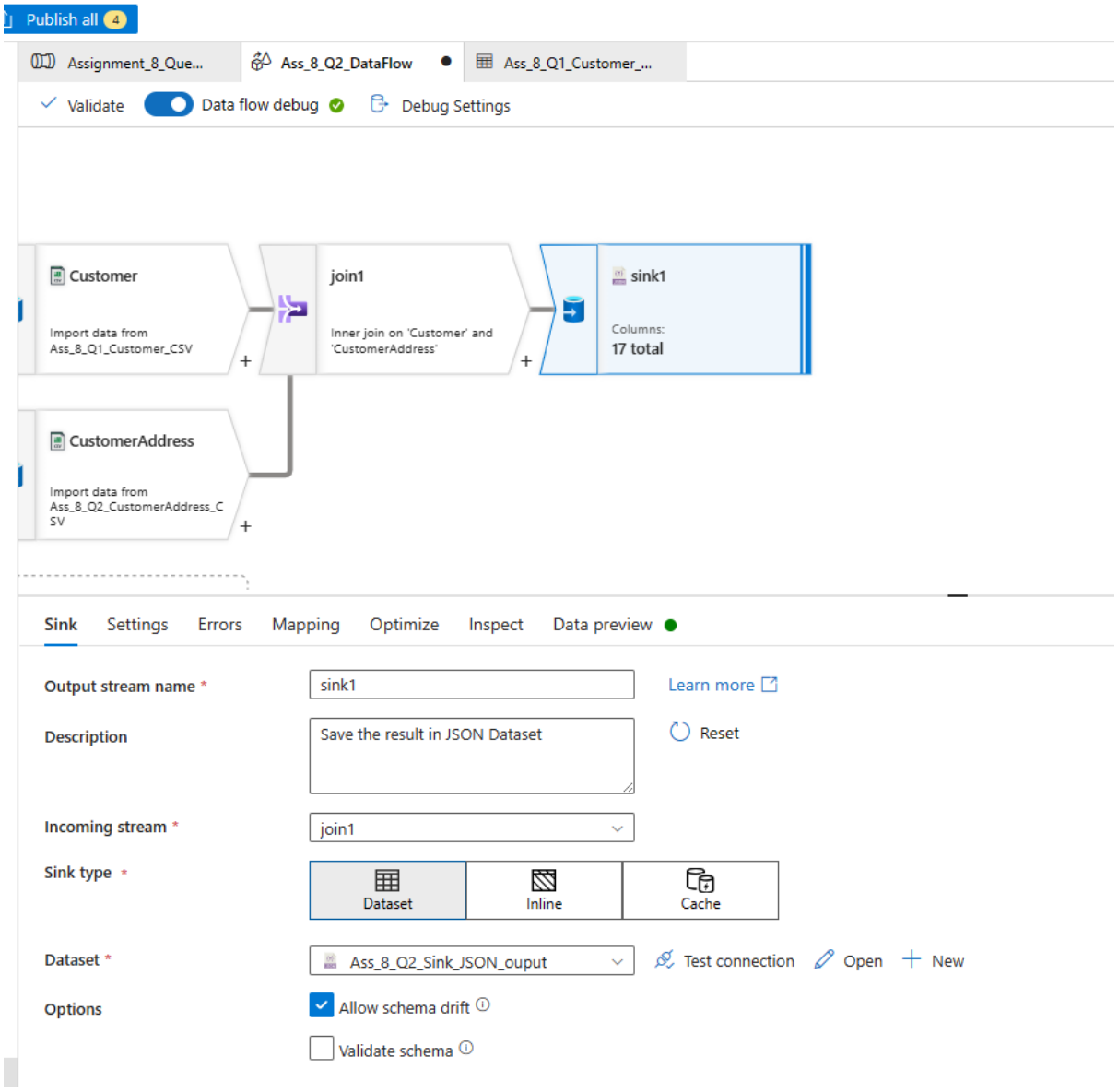
. Create a pipeline to join the two files (Custom
customer id and save the result as a JSON file.

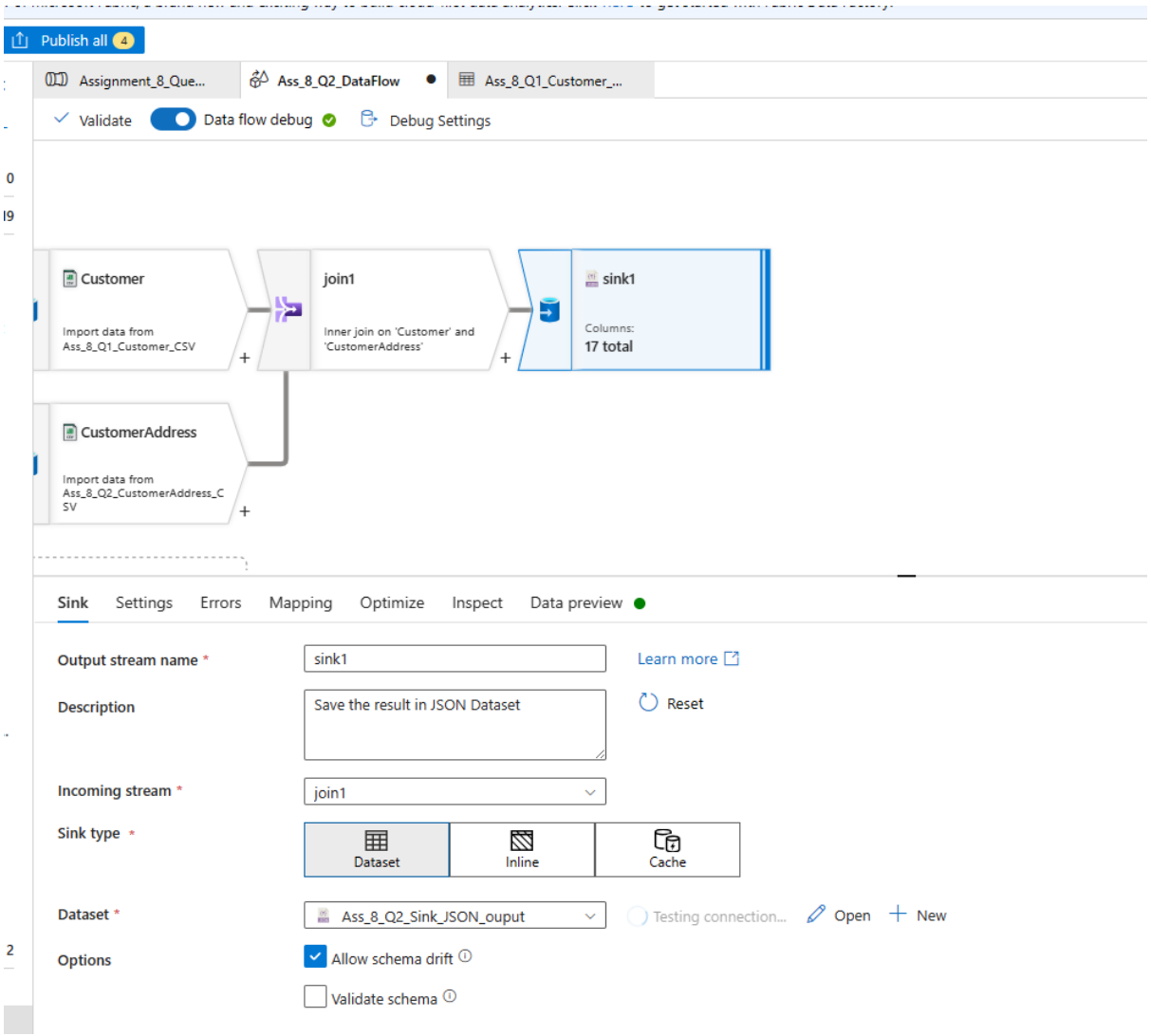


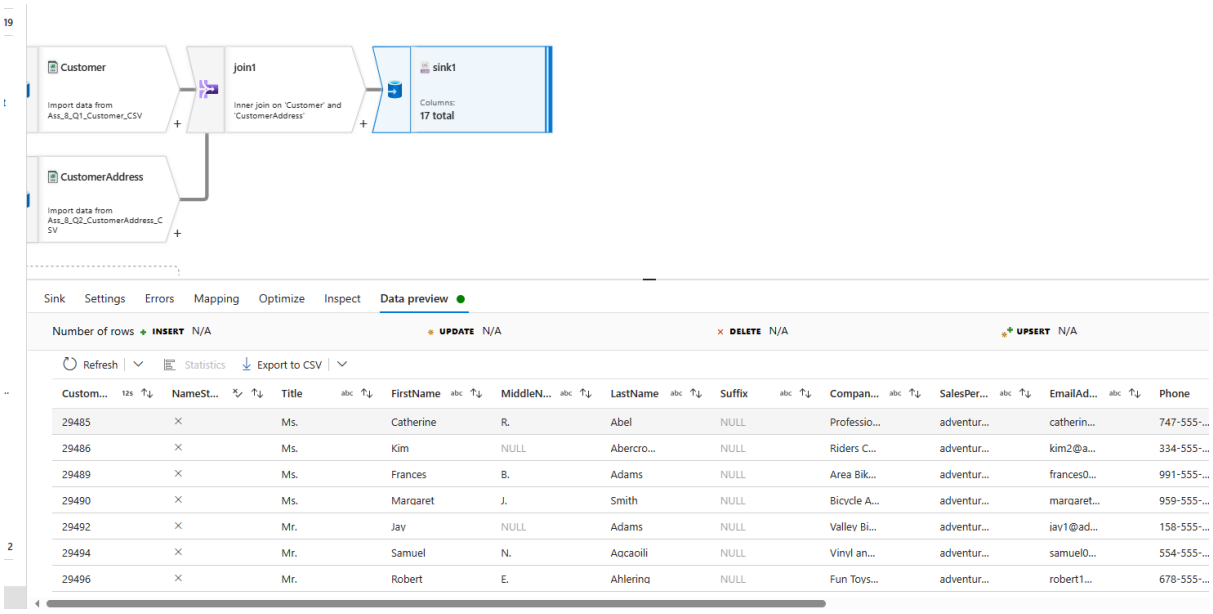


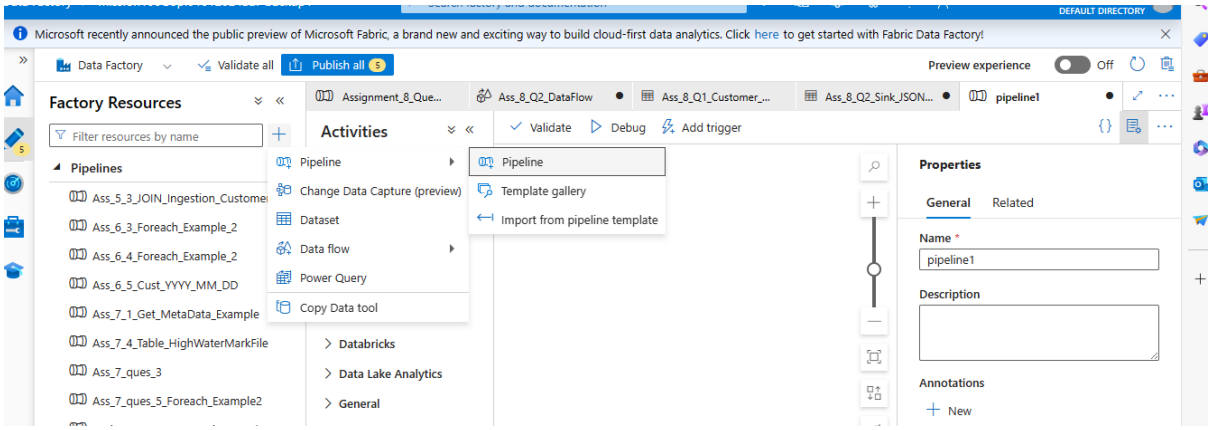

DEFAULT DIRECTORY
Select format
Choose the format type of your data
Avro
DelimitedText
JSON
ORC
Parquet
Binary

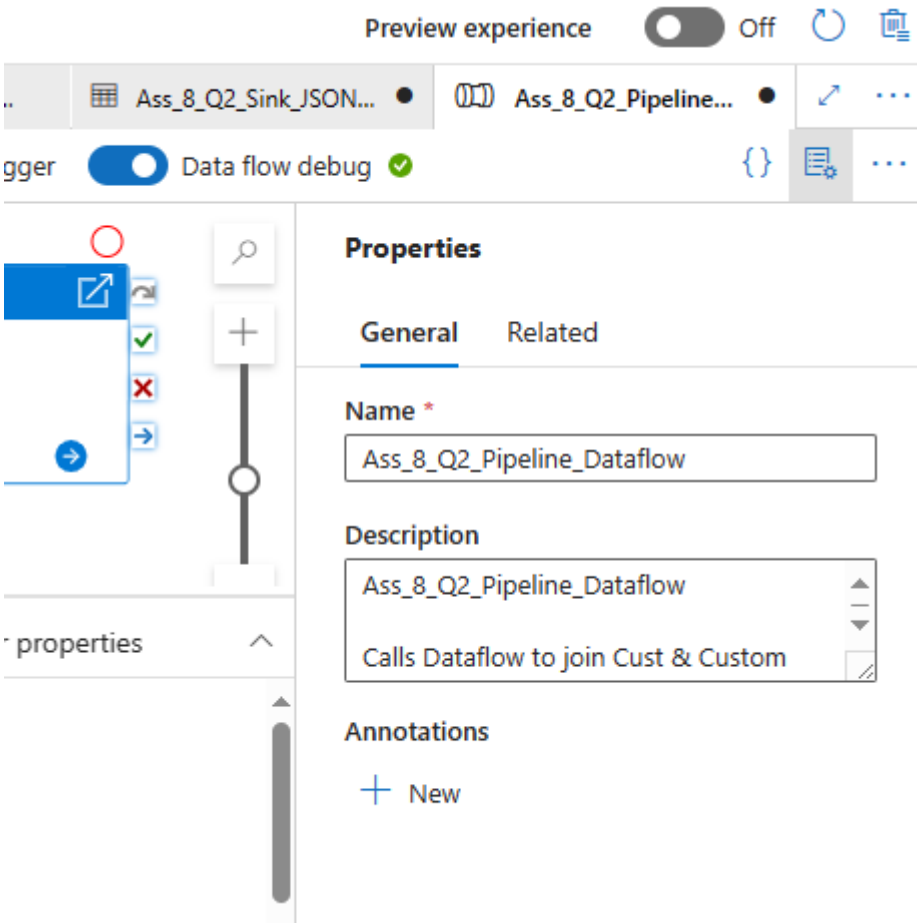

Set properties
Name
Ass_8_Q2_Sink_JSON_ouput
Linked service *
ADLS_Azure_DataLakeStorage1_Linked_service
File path
landing
/
Ass_8_output_files
/
File name
Import schema
From connection/store
From sample file
None
Advanced

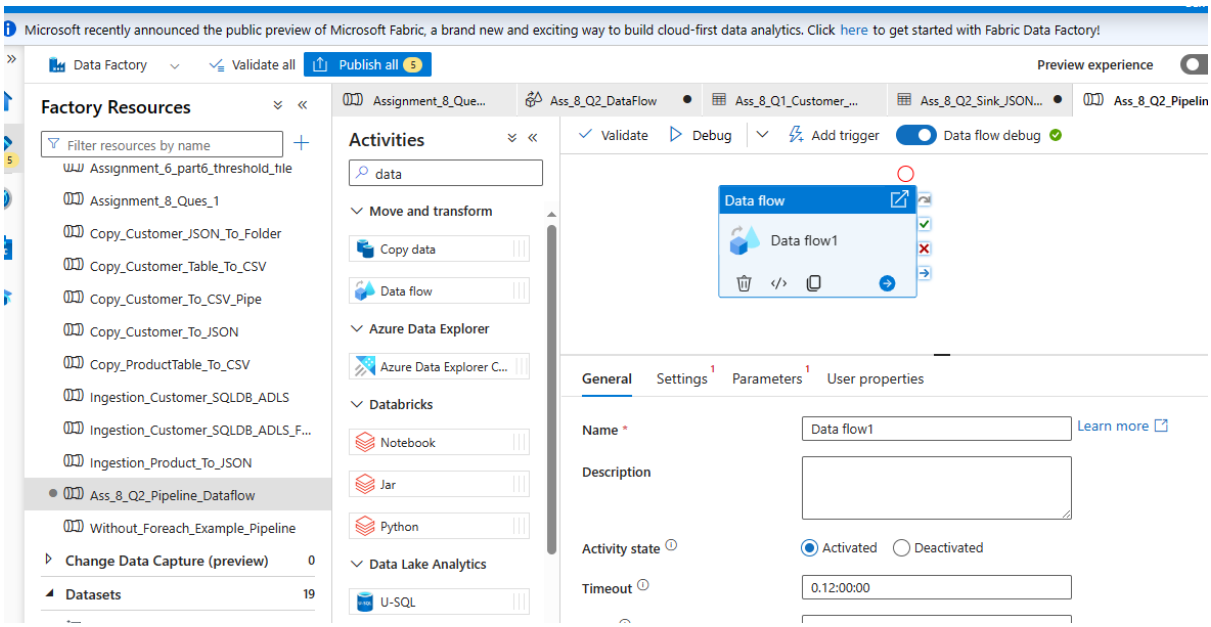

Publish all 4
Assignment_8_Que...
Ass_8_Q2_DataFlow
Ass_8_Q1_Customer_...
Validate
Data flow debug
Debug Settings
Customer
Import data from Ass_8_Q1_Customer_CSV
join1
Inner join on 'Customer' and 'CustomerAddress'
sink1
Columns:
17 total
CustomerAddress
Import data from Ass_8_Q2_CustomerAddress_CSV
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
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
Test connection
Open
New
Options
Allow schema drift
Validate schema

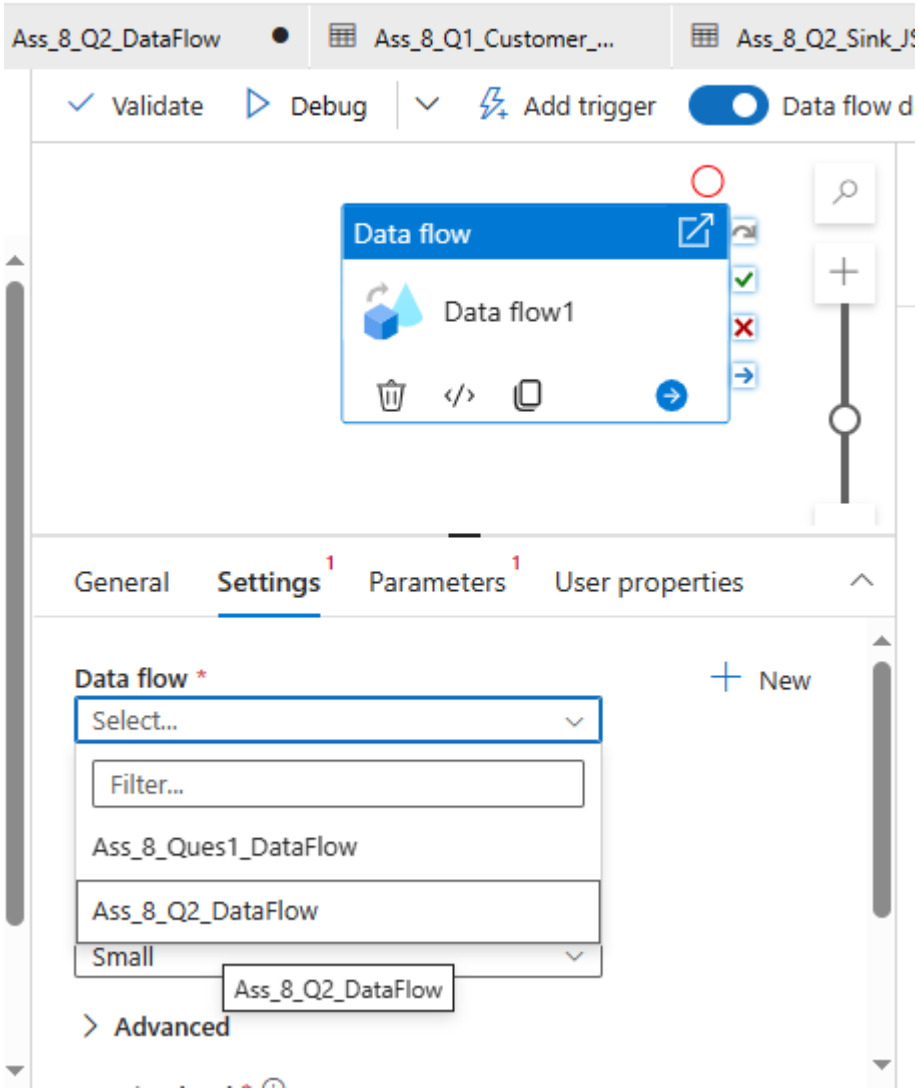

Publish all 4
Assignment_8_Que...
Ass_8_Q2_DataFlow
Ass_8_Q1_Customer_...
Validate
Data flow debug
Debug Settings
Customer
Import data from Ass_8_Q1_Customer_CSV
join1
Inner join on 'Customer' and 'CustomerAddress'
sink1
Columns:
17 total
CustomerAddress
Import data from Ass_8_Q2_CustomerAddress_CSV
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
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

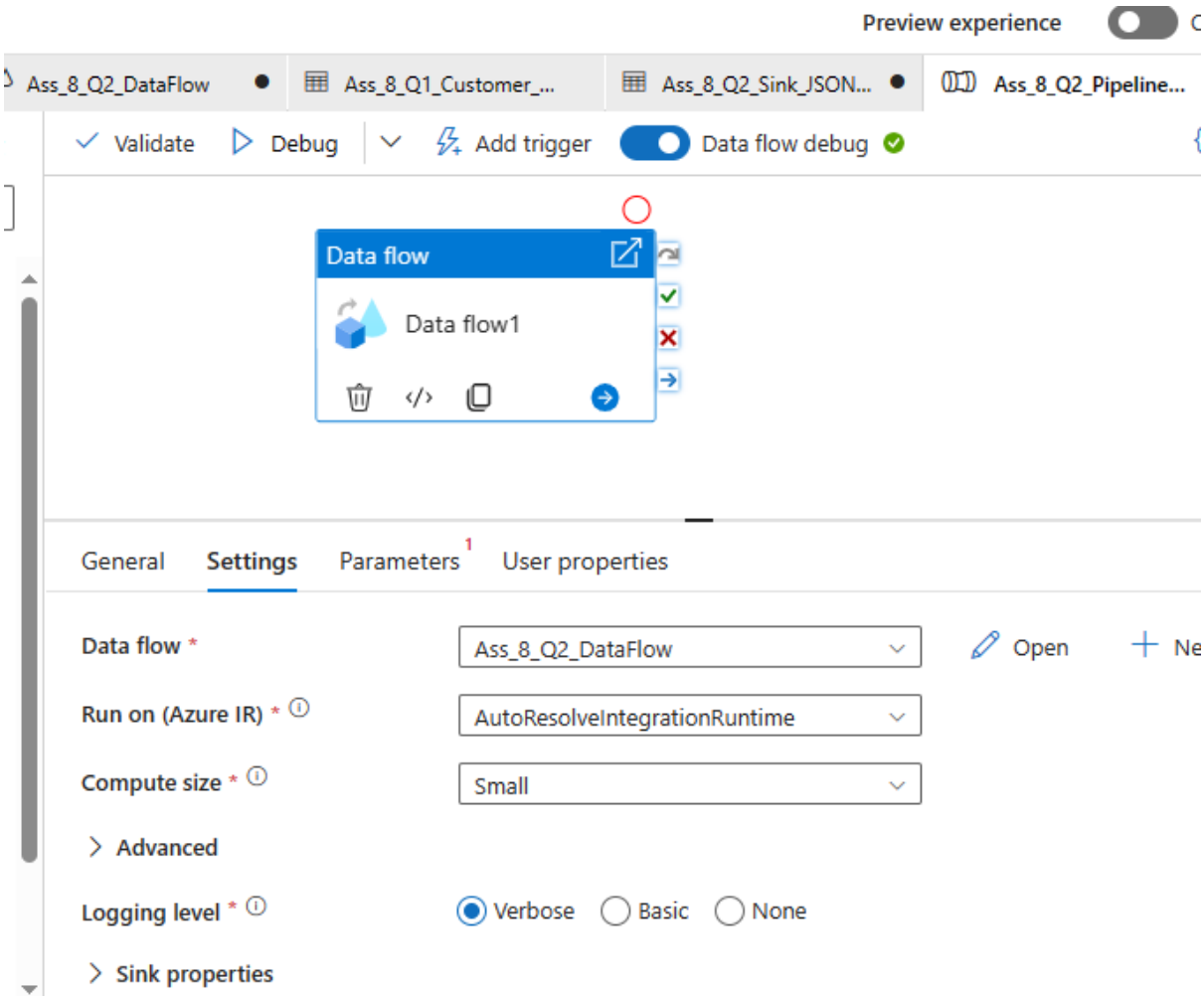

Customer
join1
sink1
CustomerAddress
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Number of rows
INSERT N/A
UPDATE N/A
DELETE N/A
UPSERT N/A
Refresh
Statistics
Export to CSV
Custom...
NameSt...
Title
FirstName
MiddleN...
LastName
Suffix
Compan...
SalesPer...
EmailAd...
Phone
29485 Ms. Catherine R. Abel NULL Professio... adventur... catherin... 747-555-...
29486 Ms. Kim NULL Abercro... NULL Riders C... adventur... kim2@a... 334-555-...
29489 Ms. Frances B. Adams NULL Area Bik... adventur... frances0... 991-555-...
29490 Ms. Margaret J. Smith NULL Bicycle A... adventur... margaret... 959-555-...
29492 Mr. Jay NULL Adams NULL Valley Bi... adventur... jay1@ad... 158-555-...
29494 Mr. Samuel N. Agcaoili NULL Vinyl an... adventur... samuel0... 554-555-...
29496 Mr. Robert E. Ahlering NULL Fun Toys... adventur... robert1... 678-555-...

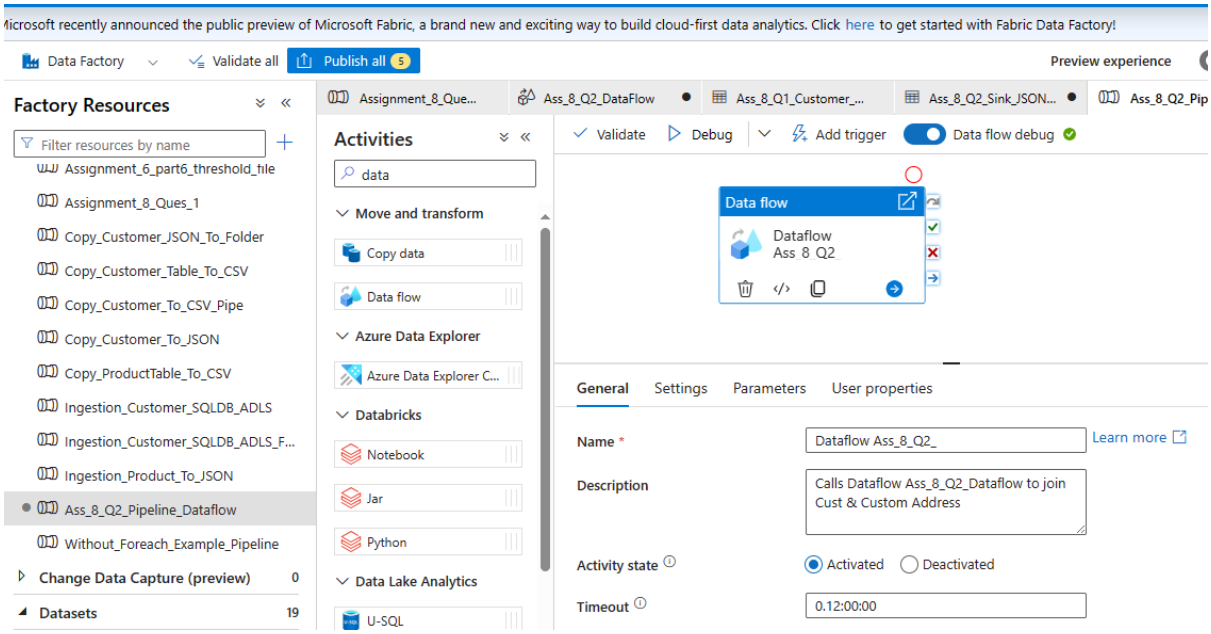

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Data Factory
Validate all
Publish all 5
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines
Ass_5_3_JOIN_Ingestion_Custome
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
Assignment_8_Que...
Ass_8_Q2_DataFlow
Ass_8_Q1_Customer_...
Ass_8_Q2_Sink_JSON...
pipeline1
Activities
Pipeline
Change Data Capture (preview)
Dataset
Data flow
Power Query
Copy Data tool
Databricks
Data Lake Analytics
General
Validate
Debug
Add trigger
Pipeline
Template gallery
Import from pipeline template
Properties
General
Related
Name *
pipeline1
Description
Annotations
New


Preview experience
Off
Ass_8_Q2_Sink_JSON...
Ass_8_Q2_Pipeline...
gger
Data flow debug
properties
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 5
Preview experience
Factory Resources
Filter resources by name
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
Assignment_8_Que...
Ass_8_Q2_DataFlow
Ass_8_Q1_Customer_...
Ass_8_Q2_Sink_JSON...
Ass_8_Q2_Pipelin
Activities
data
Move and transform
Copy data
Data flow
Azure Data Explorer
Azure Data Explorer C...
Databricks
Notebook
Jar
Python
Data Lake Analytics
U-SQL
Validate
Debug
Add trigger
Data flow debug
Data flow
Data flow1
General
Settings 1
Parameters 1
User properties
Name *
Data flow1
Learn more
Description
Activity state
Activated
Deactivated
Timeout
0.12:00:00

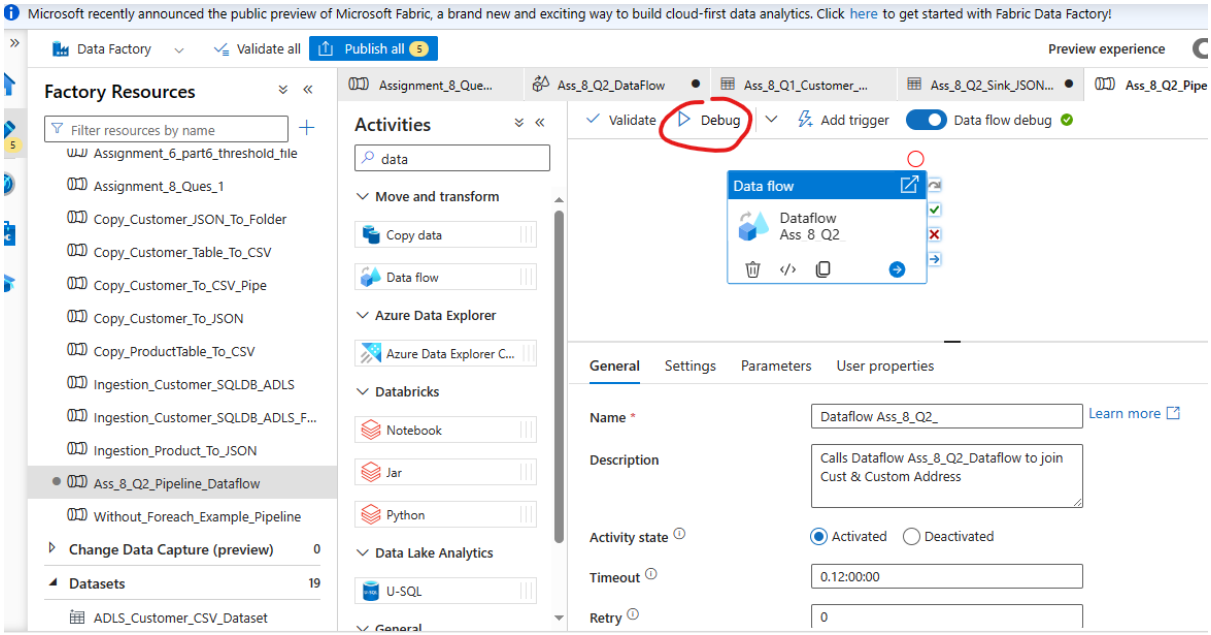

Ass_8_Q2_DataFlow
Ass_8_Q1_Customer_...
Validate
Debug
Add trigger
Data flow d
Data flow
Data flow1
General
Settings 1
Parameters 1
User properties
Data flow *
New
Select...
Filter...
Ass_8_Ques1_DataFlow
Ass_8_Q2_DataFlow
Small
Ass_8_Q2_DataFlow
Advanced

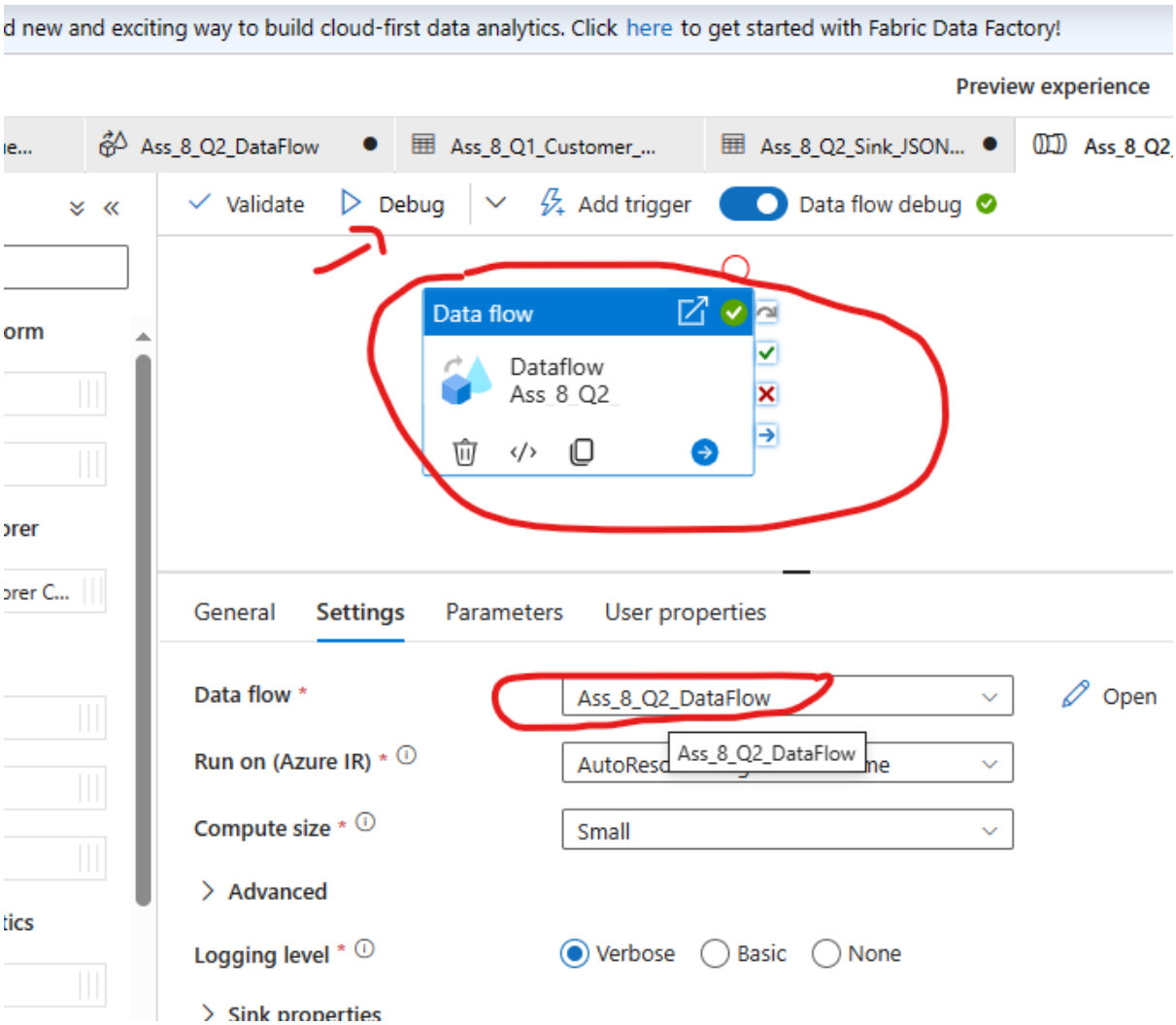

Preview experience
Ass_8_Q2_DataFlow
Ass_8_Q1_Customer_...
Ass_8_Q2_Sink_JSON...
Ass_8_Q2_Pipeline...
Validate
Debug
Add trigger
Data flow debug
Data flow
Data flow1
General
Settings
Parameters
User properties
Data flow *
Ass_8_Q2_DataFlow
Open
Run on (Azure IR) *
AutoResolveIntegrationRuntime
Compute 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
Validate all
Publish all
Preview experience
Factory Resources
Filter resources by name
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)
Datasets
Activities
data
Move and transform
Copy data
Data flow
Azure Data Explorer
Azure Data Explorer C...
Databricks
Notebook
Jar
Python
Data Lake Analytics
U-SQL
Validate
Debug
Add trigger
Data flow debug
Data flow
Dataflow Ass_8_Q2
General
Settings
Parameters
User properties
Name *
Dataflow Ass_8_Q2_
Learn more
Description
Calls Dataflow Ass_8_Q2_Dataflow to join Cust & Custom Address
Activity state
Activated
Deactivated
Timeout
0.12:00:00

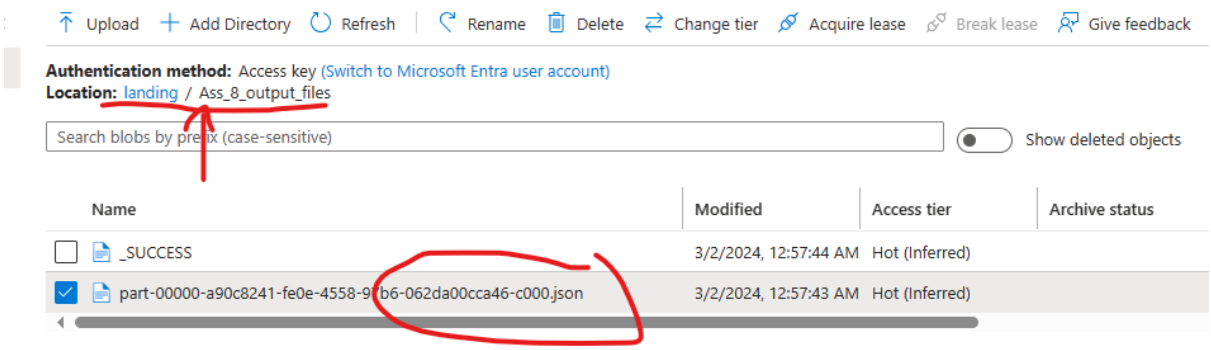
To run Dataflow,

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

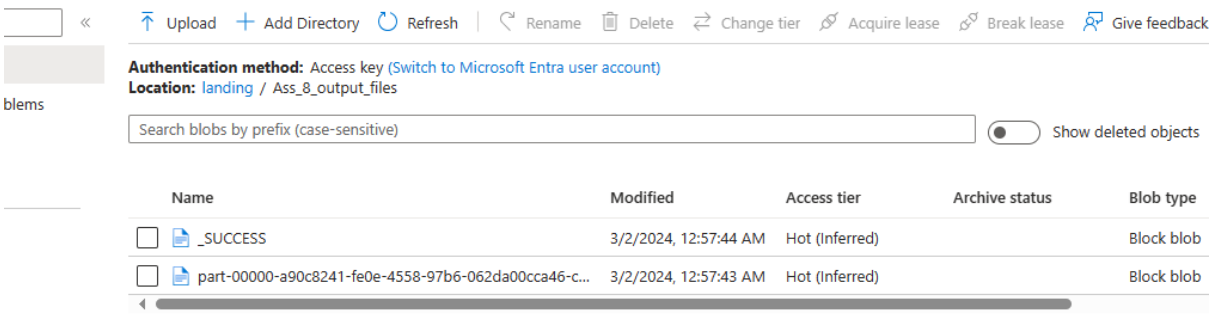


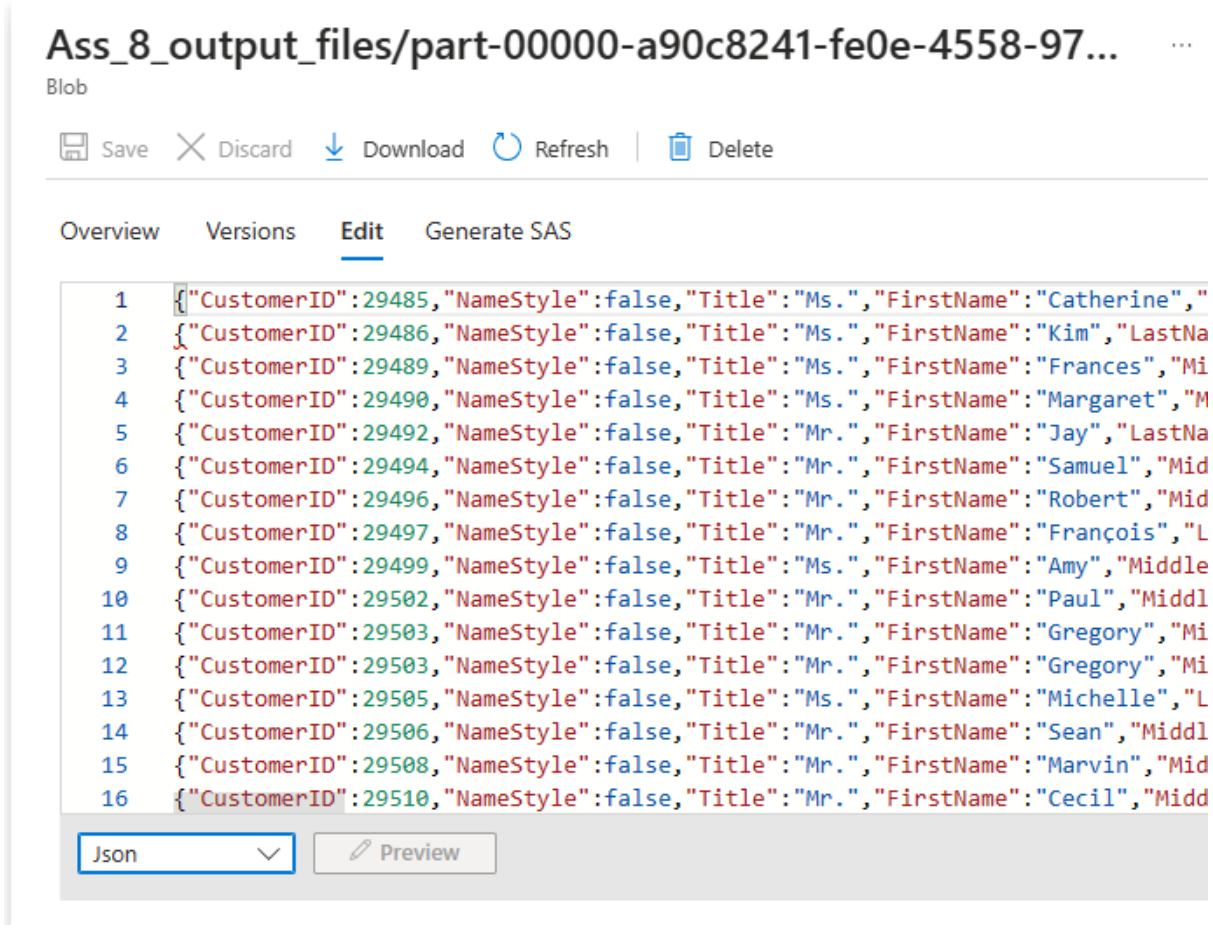


OUTPUT JSON generated in ADLS container



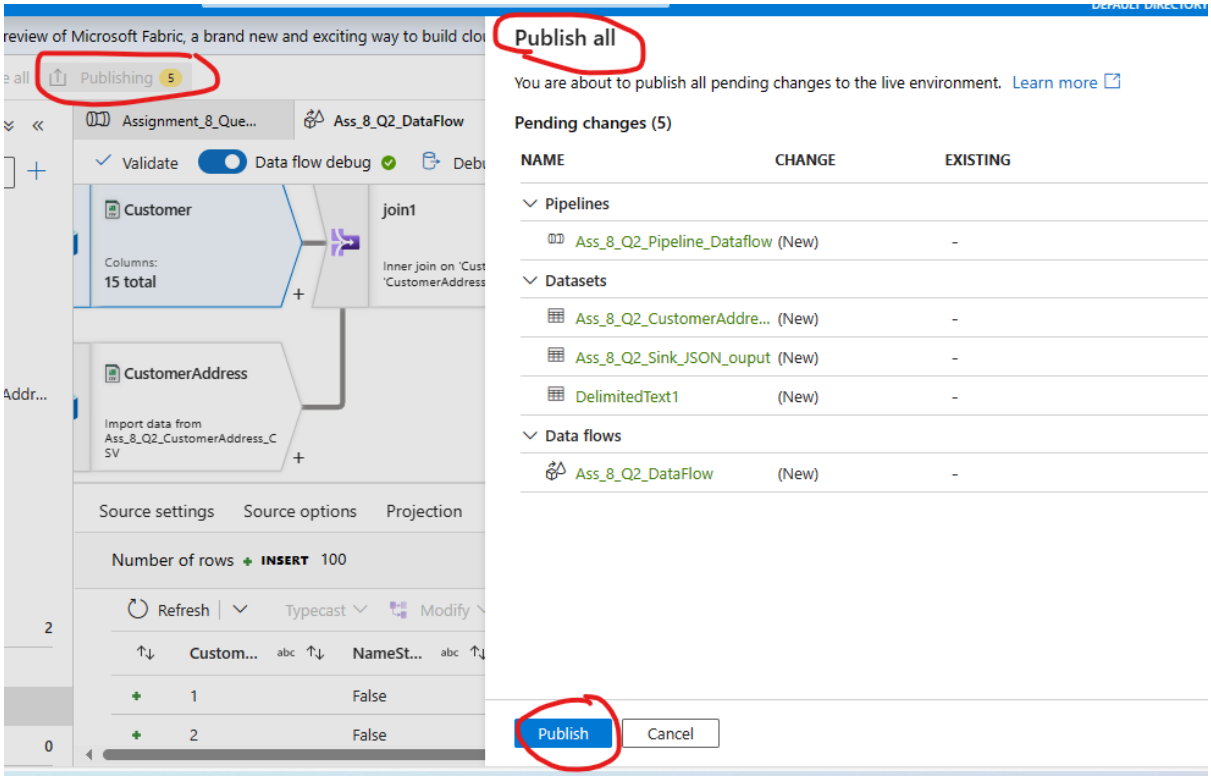
s > gopi01012024adls1 | Containers >



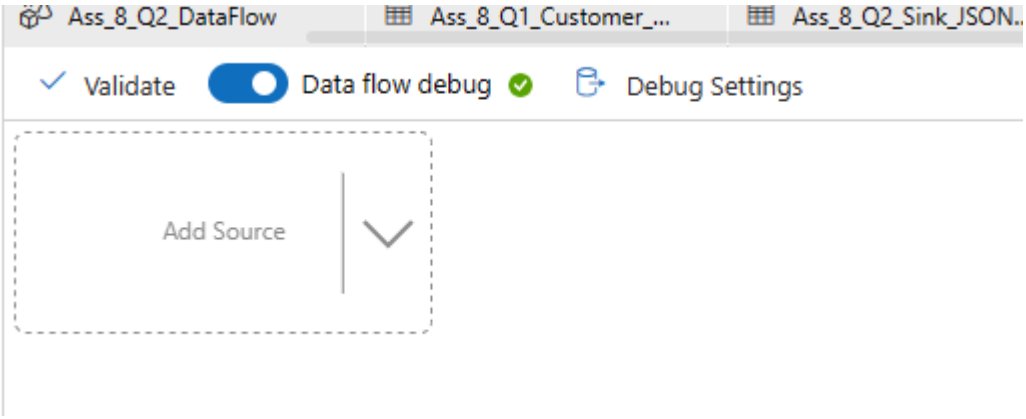


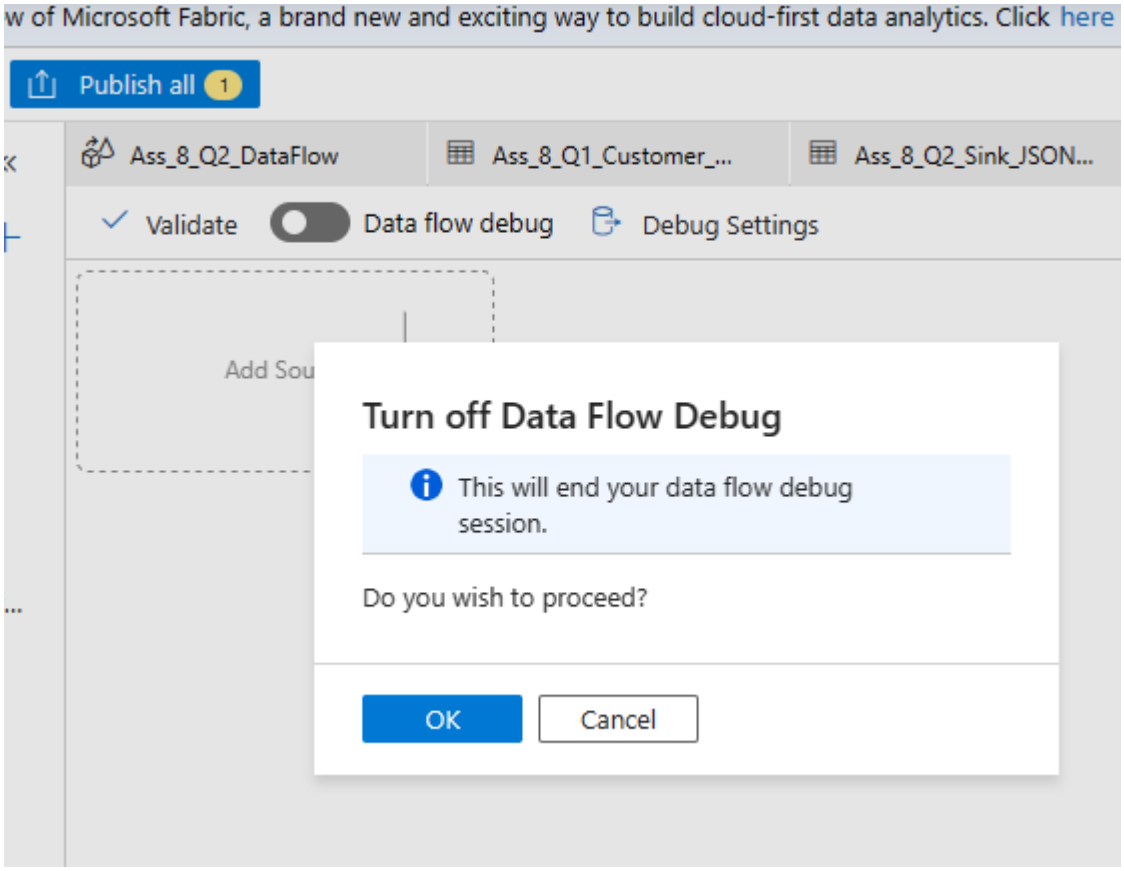
For Saving the Pipeline – Do not forget to Publish it

Publish All – will save all the output



Finally to save costs turn off your "Data flow Debug" mode





Question 3:

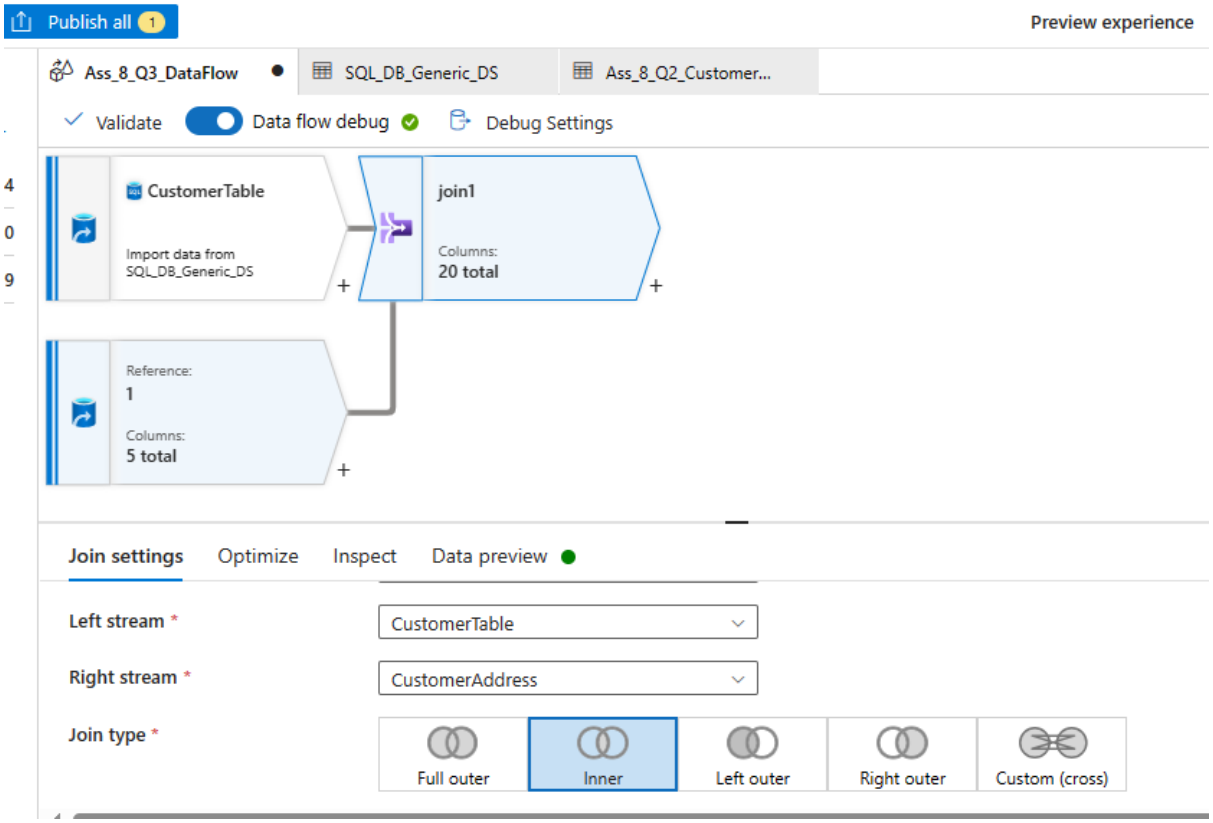
3. Create a pipeline to read the Customer table data from SQL and CustomerAddress data from CSV, join both of them,

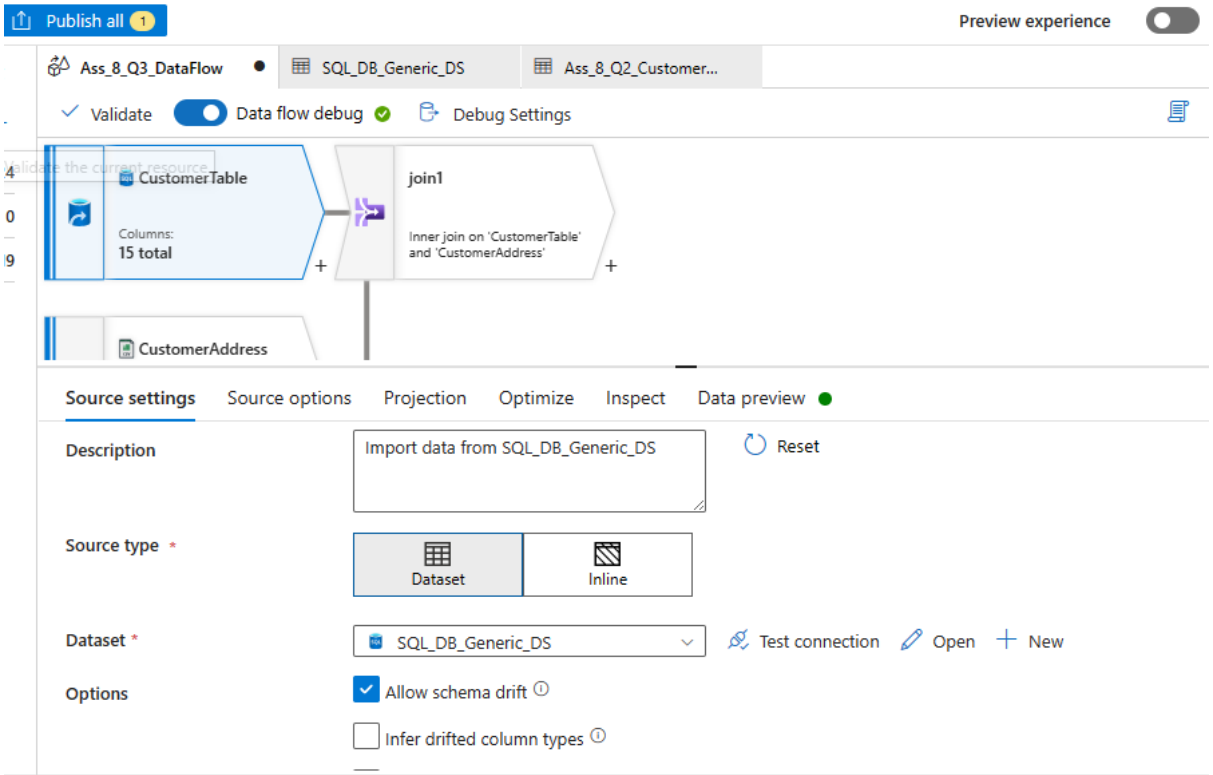
and then save the result where customer id> 1000 & Customer id <2000 in ascending order as a Parquet file.

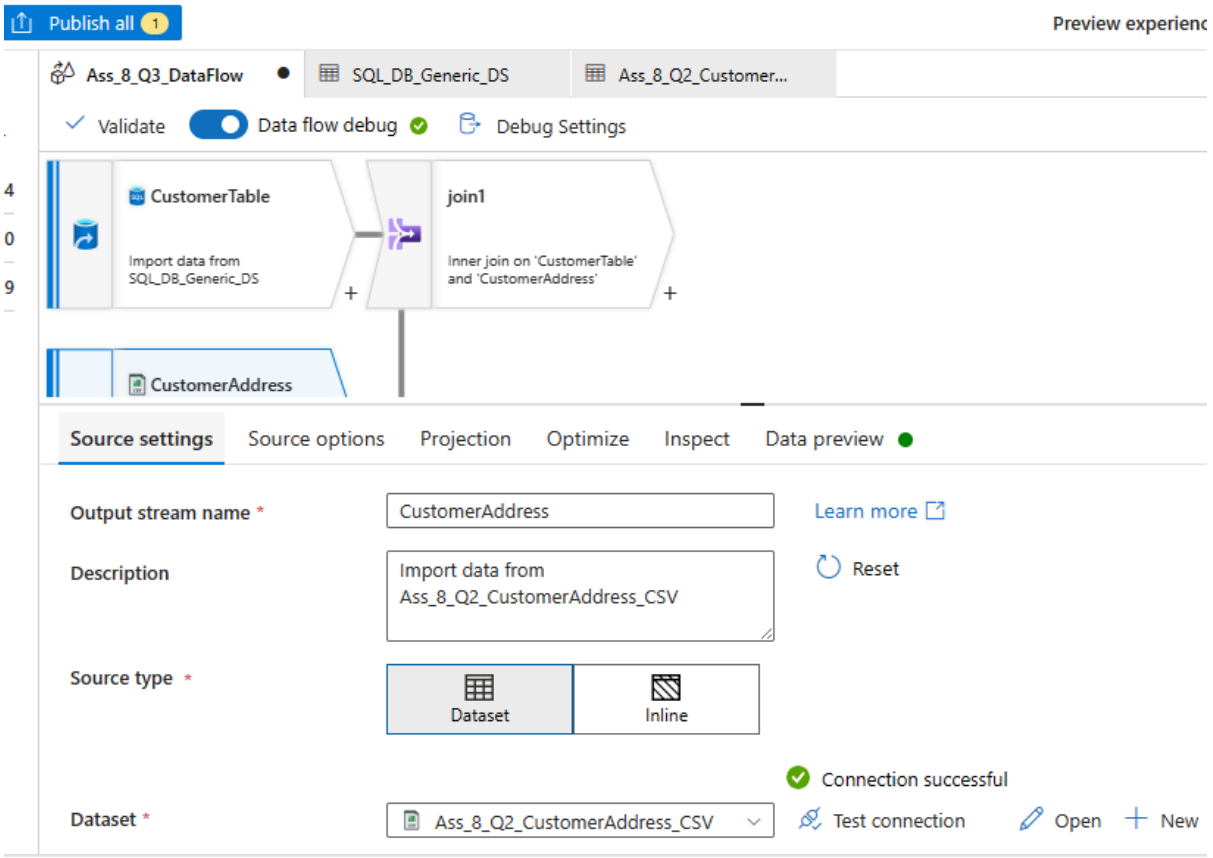
Solution 3:

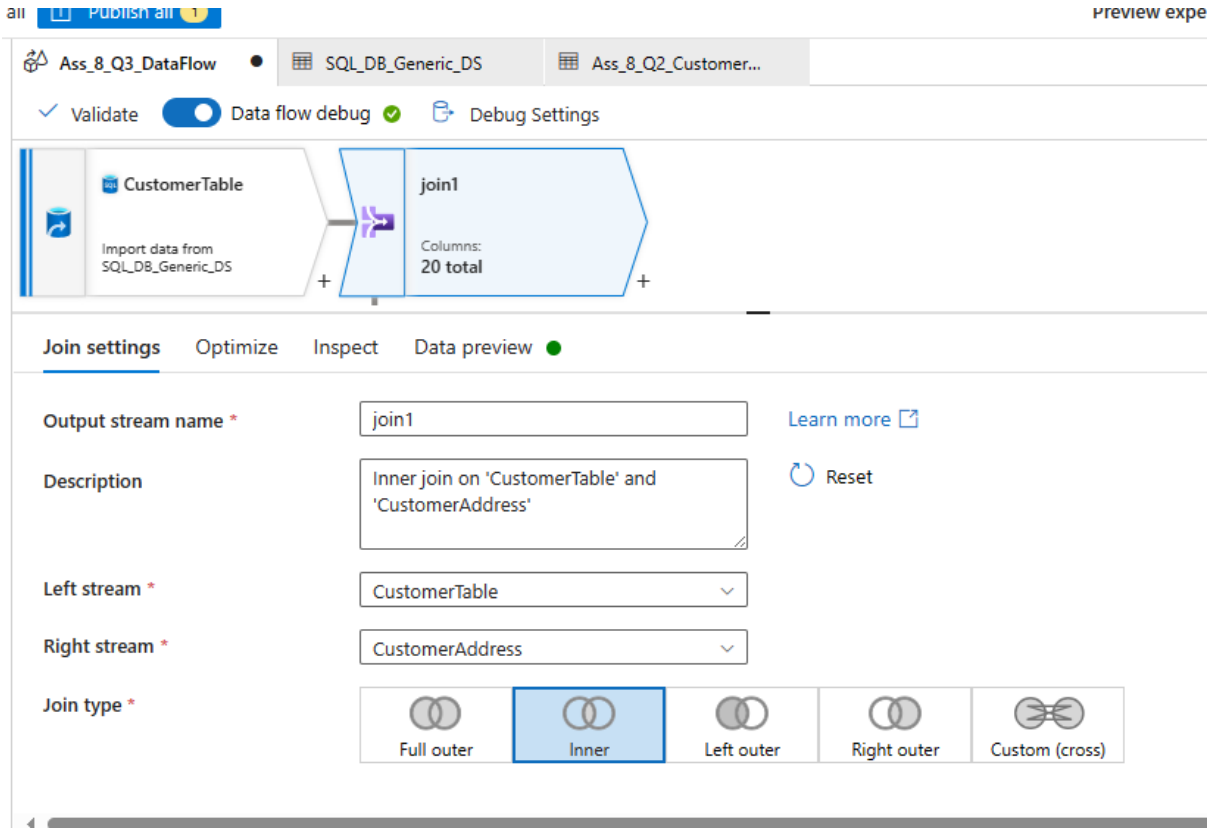
Same as Q1 & Q2 , extra thing is after join, use "Filter"

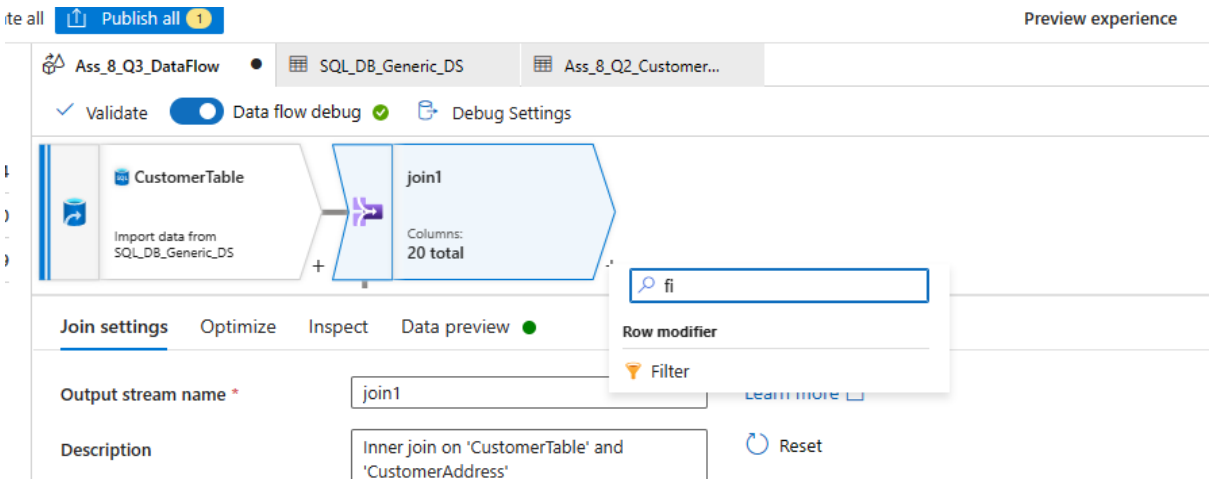
customer id> 1000 & Customer id <2000 in ascending order as a Parquet file.

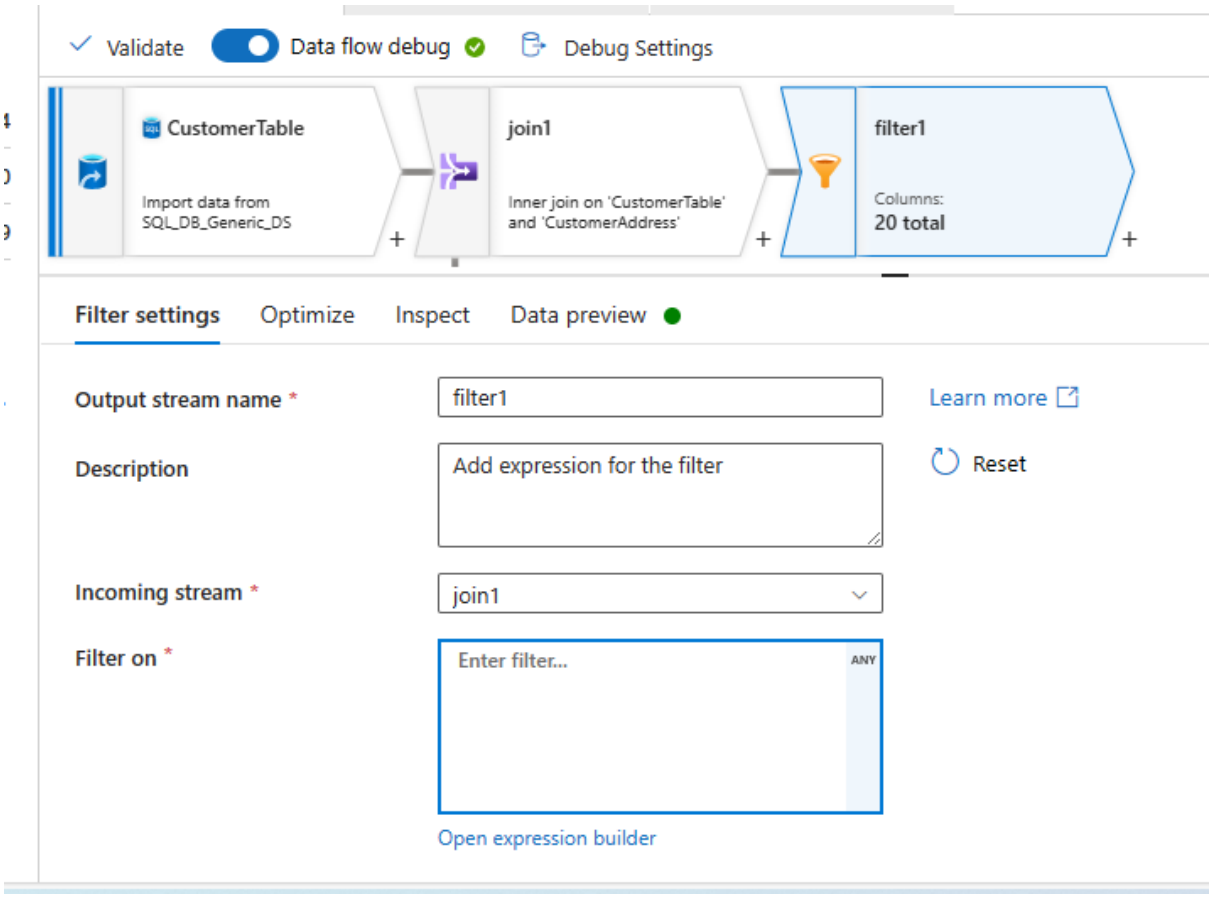

Publish all 1
Preview experience
Ass_8_Q3_DataFlow
SQL_DB_Generic_DS
Ass_8_Q2_Customer...
Validate
Data flow debug
Debug Settings
CustomerTable
Import data from SQL_DB_Generic_DS
join1
Columns:
20 total
Reference:
1
Columns:
5 total
Join settings
Optimize
Inspect
Data preview
Left stream *
CustomerTable
Right stream *
CustomerAddress
Join type *
Full outer
Inner
Left outer
Right outer
Custom (cross)

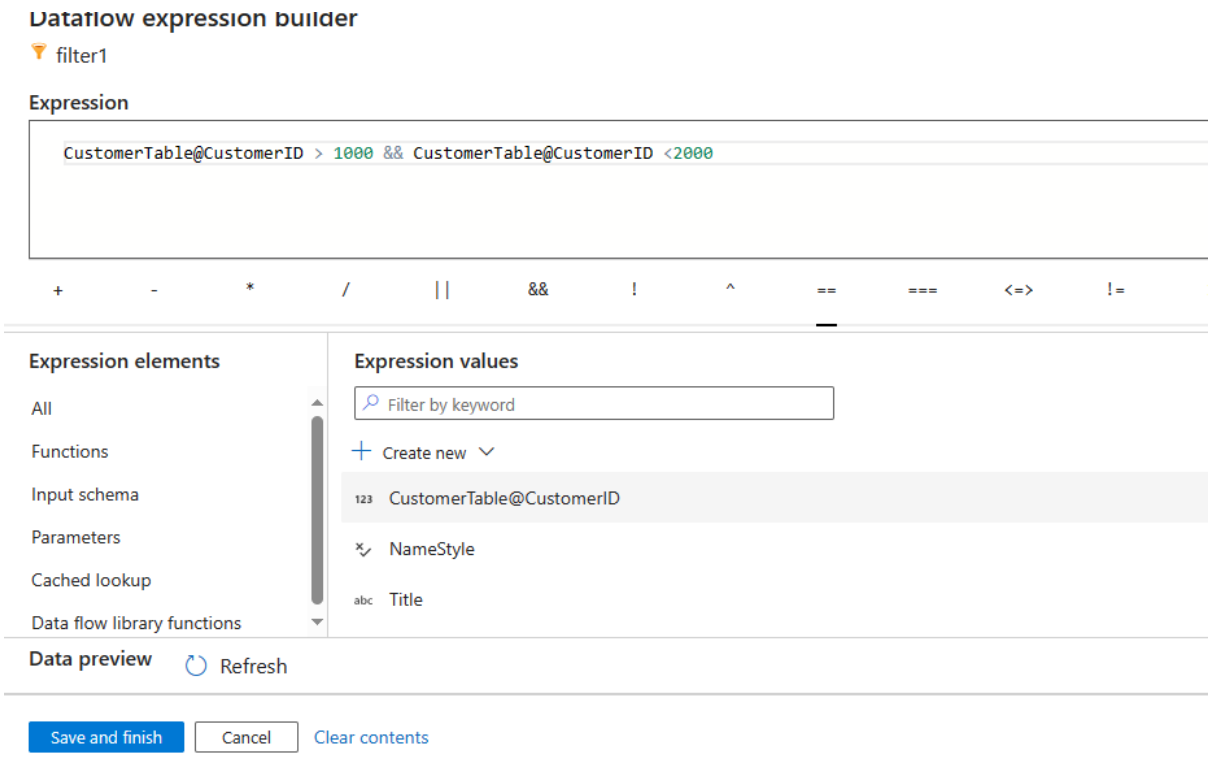

Publish all 1
Preview experience
Ass_8_Q3_DataFlow
SQL_DB_Generic_DS
Ass_8_Q2_Customer...
Validate
Data flow debug
Debug Settings
CustomerTable
Columns:
15 total
join1
Inner join on 'CustomerTable' and 'CustomerAddress'
CustomerAddress
Source settings
Source options
Projection
Optimize
Inspect
Data preview
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
Infer drifted column types

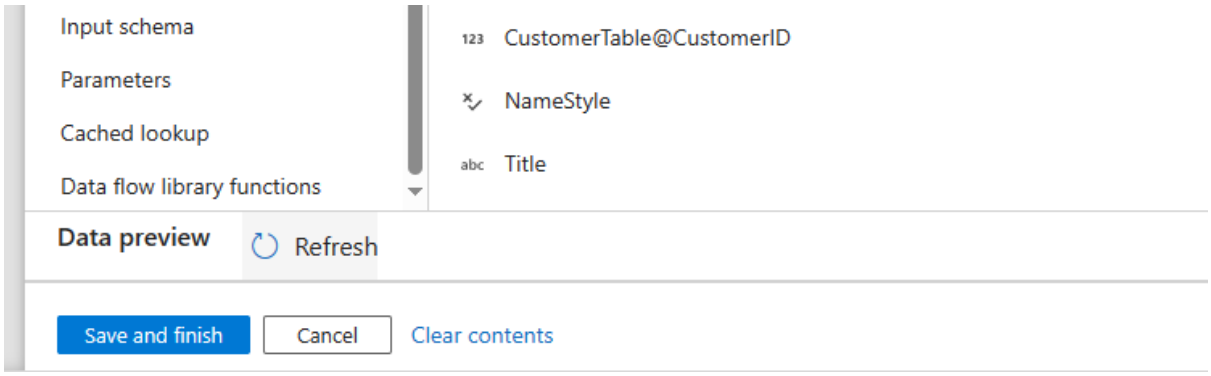

Publish all 1
Preview experienc
Ass_8_Q3_DataFlow
SQL_DB_Generic_DS
Ass_8_Q2_Customer...
Validate
Data flow debug
Debug Settings
CustomerTable
Import data from SQL_DB_Generic_DS
join1
Inner join on 'CustomerTable' and 'CustomerAddress'
CustomerAddress
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Output stream name *
CustomerAddress
Learn more
Description
Import data from Ass_8_Q2_CustomerAddress_CSV
Reset
Source type *
Dataset
Inline
Connection successful
Dataset *
Ass_8_Q2_CustomerAddress_CSV
Test connection
Open
New

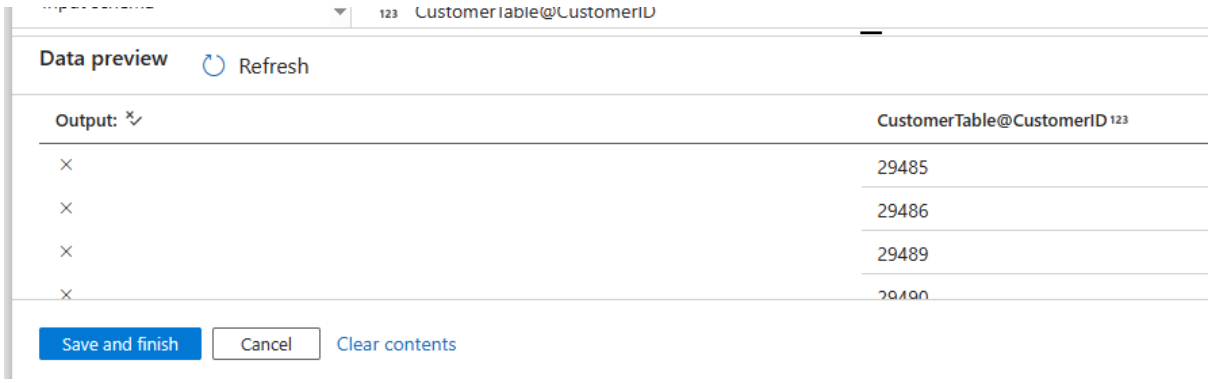

Publish all 1
Preview expe
Ass_8_Q3_DataFlow
SQL_DB_Generic_DS
Ass_8_Q2_Customer...
Validate
Data flow debug
Debug Settings
CustomerTable
Import data from SQL_DB_Generic_DS
join1
Columns:
20 total
Join settings
Optimize
Inspect
Data preview
Output stream name *
join1
Learn more
Description
Inner join on 'CustomerTable' and 'CustomerAddress'
Reset
Left stream *
CustomerTable
Right stream *
CustomerAddress
Join type *
Full outer
Inner
Left outer
Right outer
Custom (cross)


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
fi
Row modifier
Filter
Join settings
Optimize
Inspect
Data preview
Output stream name *
join1
Description
Inner join on 'CustomerTable' and 'CustomerAddress'
Reset


Validate
Data flow debug
Debug Settings
CustomerTable
Import data from SQL_DB_Generic_DS
join1
Inner join on 'CustomerTable' and 'CustomerAddress'
filter1
Columns:
20 total
Filter settings
Optimize
Inspect
Data preview
Output stream name *
filter1
Learn more
Description
Add expression for the filter
Reset
Incoming stream *
join1
Filter on *
Enter filter...
ANY
Open expression builder

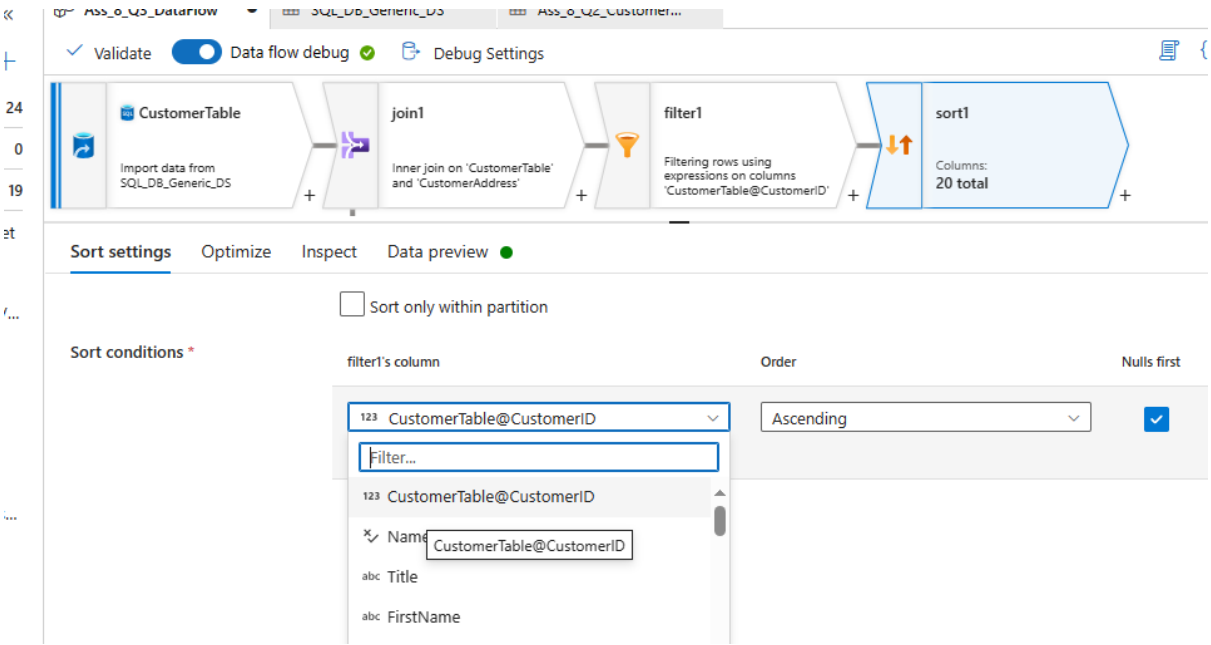

Dataflow expression builder
filter1
Expression
CustomerTable@CustomerID > 1000 && CustomerTable@CustomerID <2000
+ - * / || && ! ^ == === <=> !=
Expression elements
All
Functions
Input schema
Parameters
Cached lookup
Data flow library functions
Expression values
Filter by keyword
Create new
CustomerTable@CustomerID
NameStyle
Title
Data preview
Refresh
Save and finish
Cancel
Clear contents

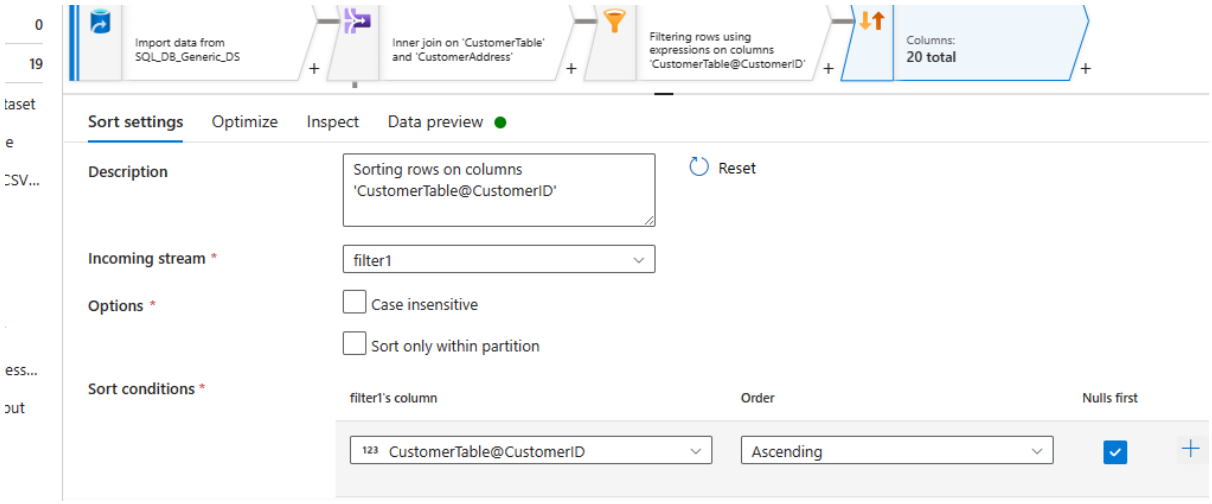




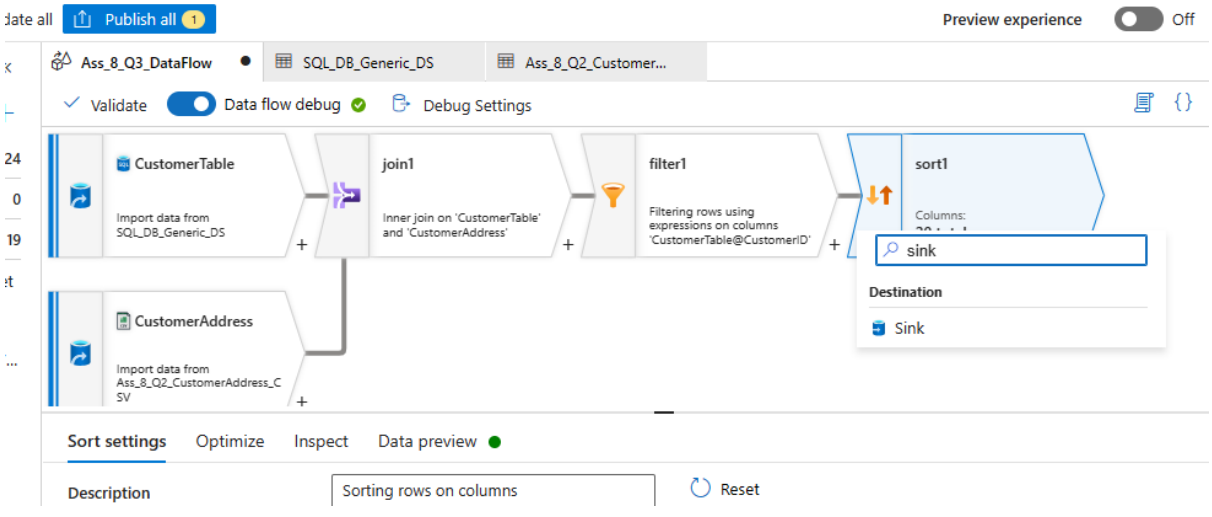
Next Step is ORDER BY → so use "Sort"

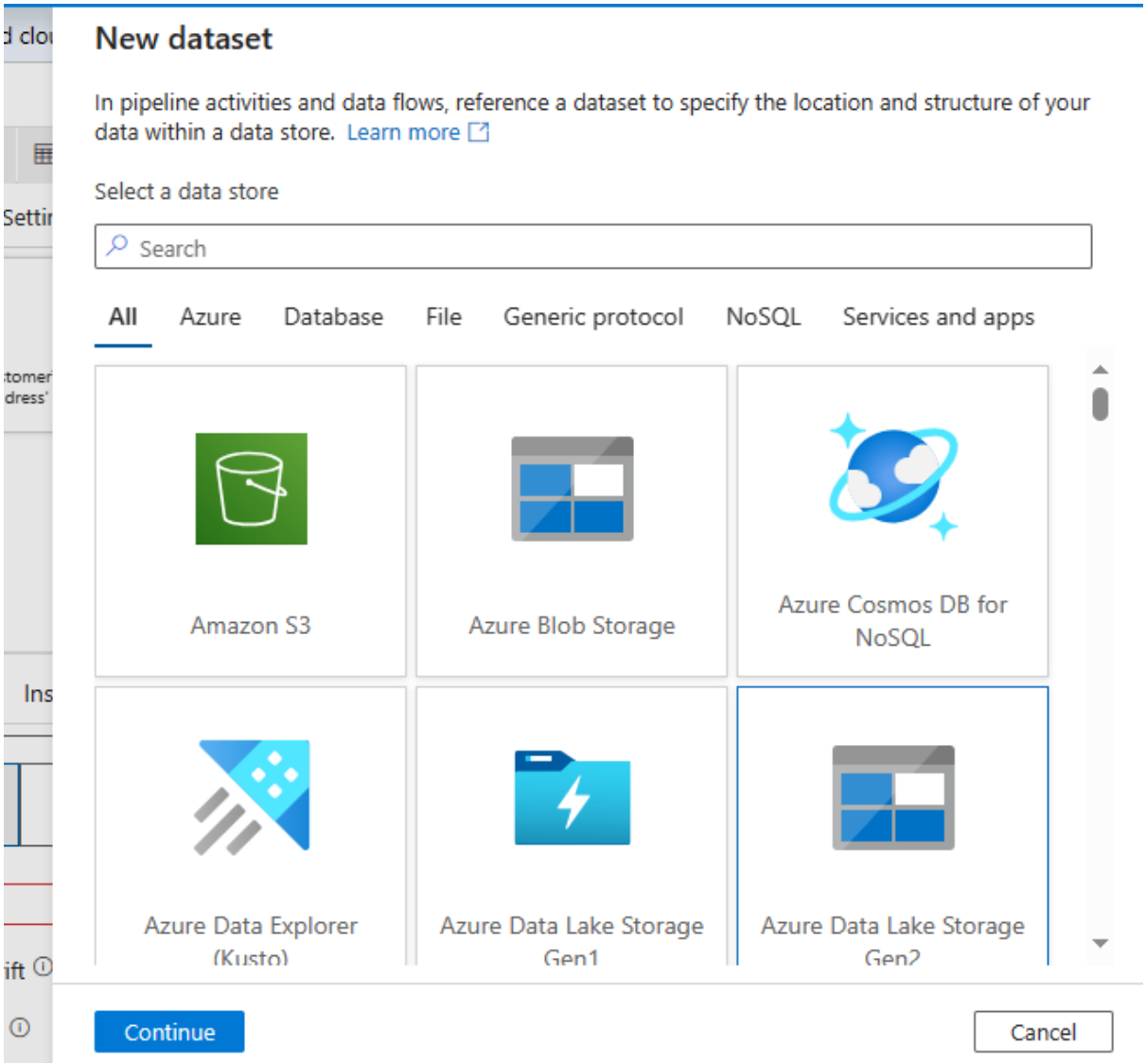
data from CSV, join b
and then save the result where customer id> 1000 & Customer id <2000 in ascending order as a Parquet file.

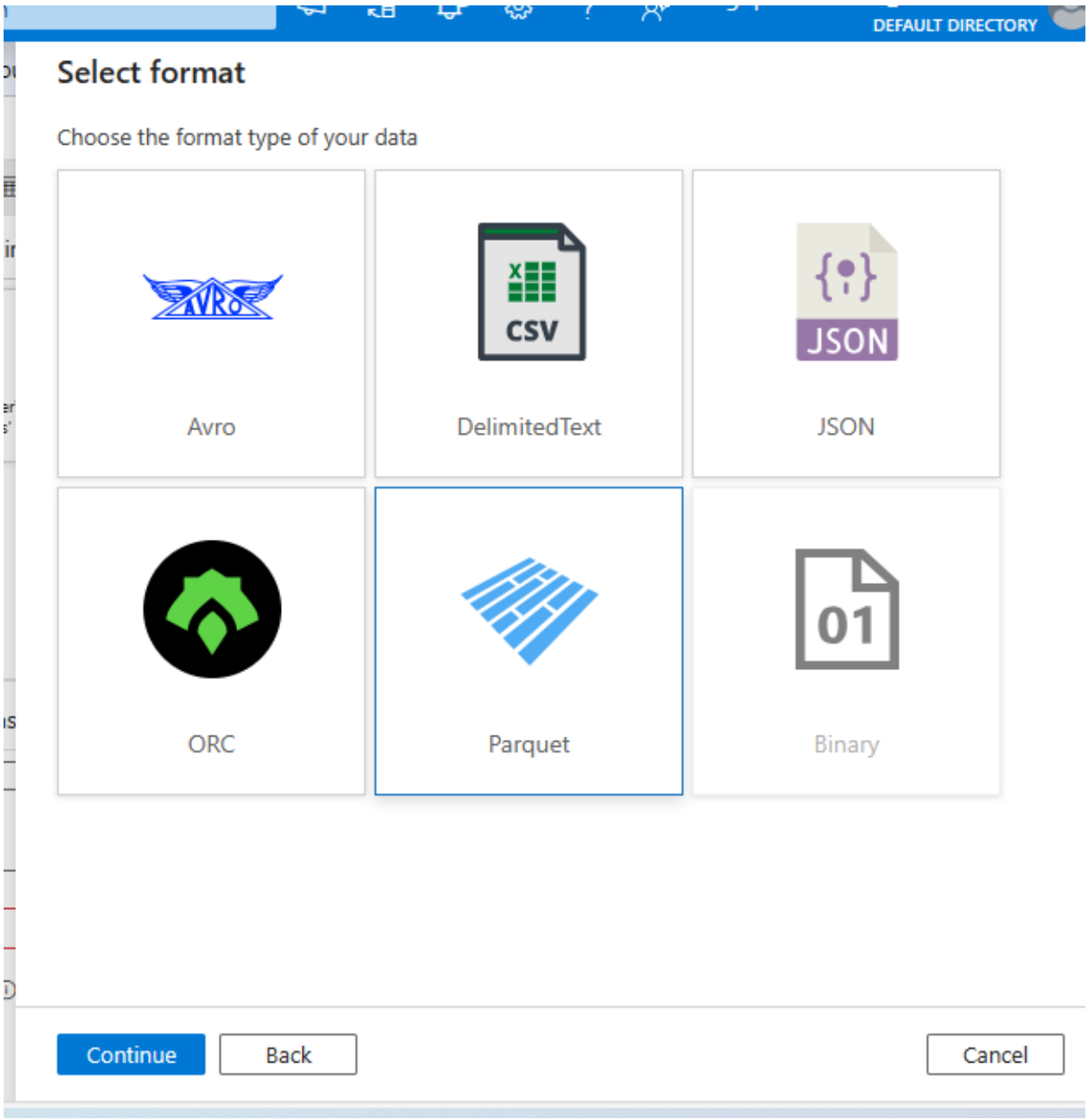


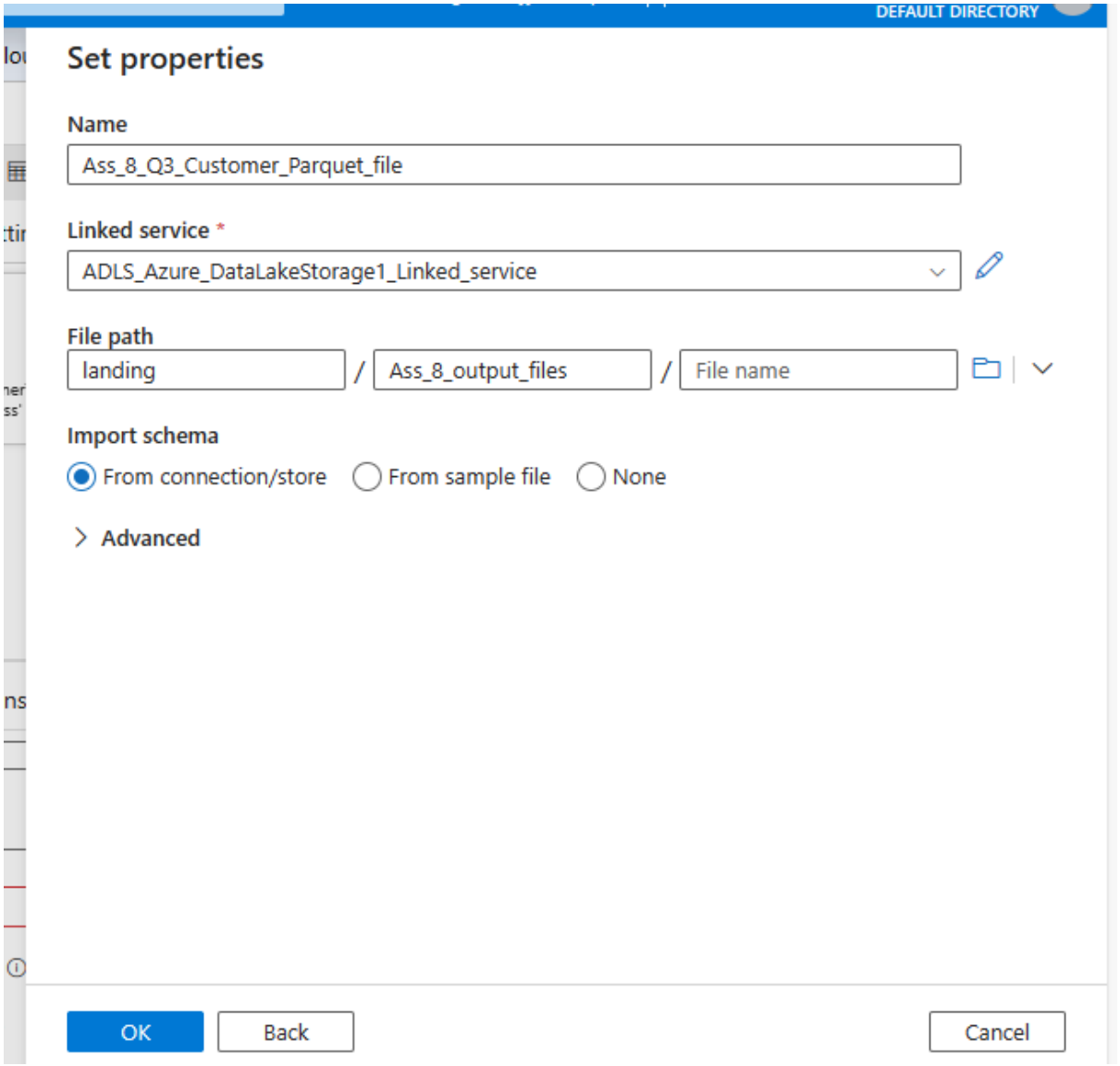


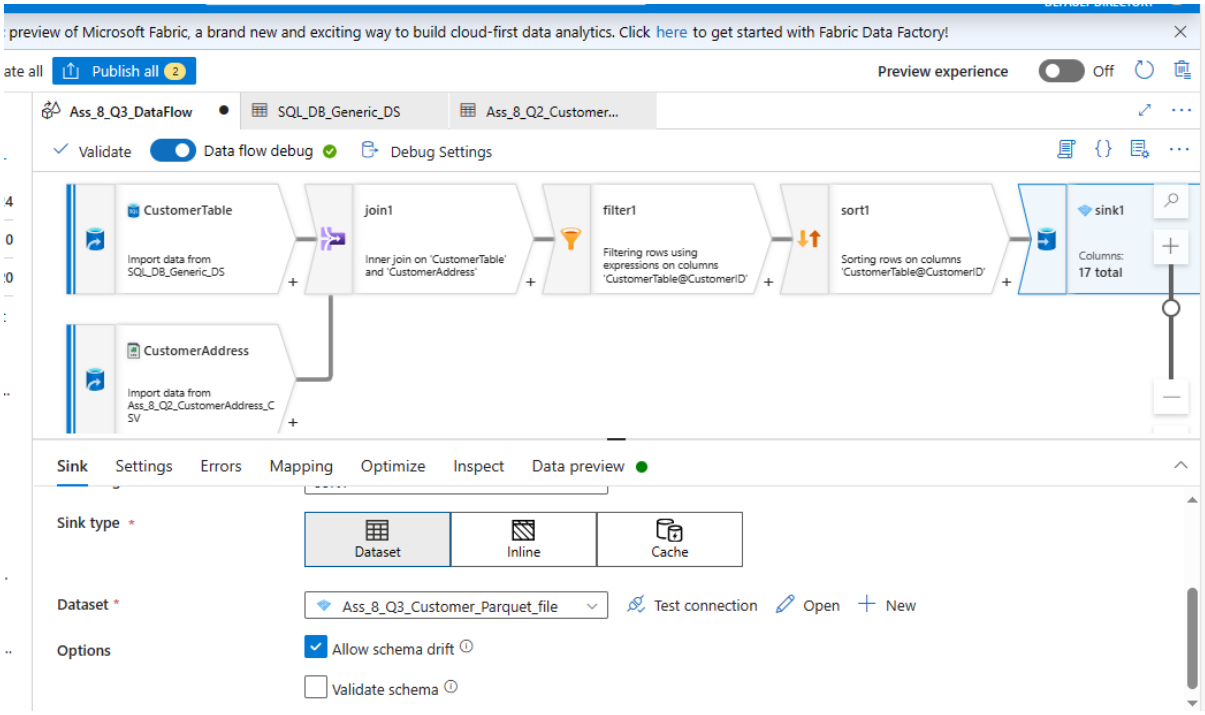
Save it as Parquet file: So use "Sink"



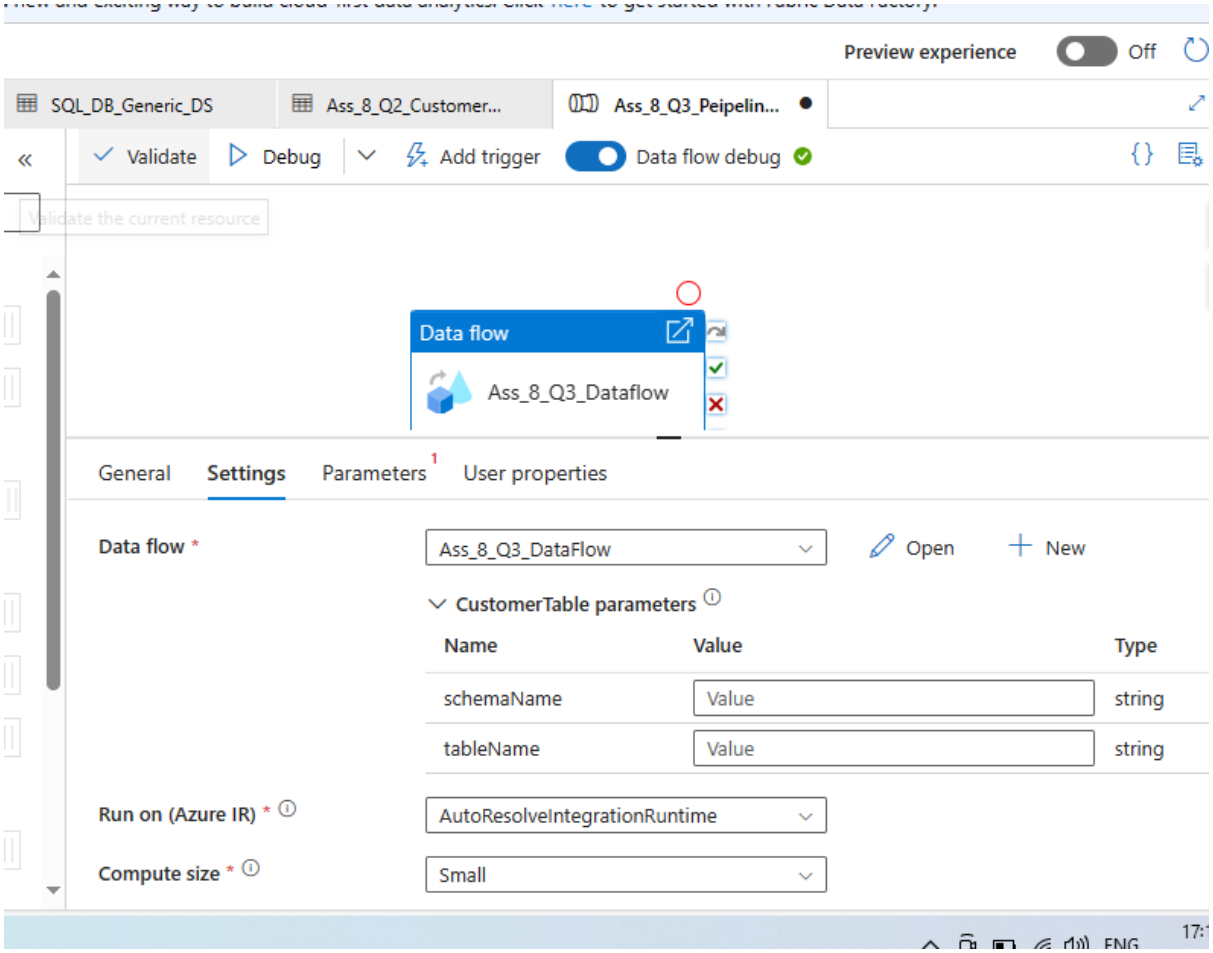

New dataset
In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more
Select a data store
Search
All
Azure
Database
File
Generic protocol
NoSQL
Services and apps
Amazon S3
Azure Blob Storage
Azure Cosmos DB for NoSQL
Azure Data Explorer (Kusto)
Azure Data Lake Storage Gen1
Azure Data Lake Storage Gen2
Continue
Cancel

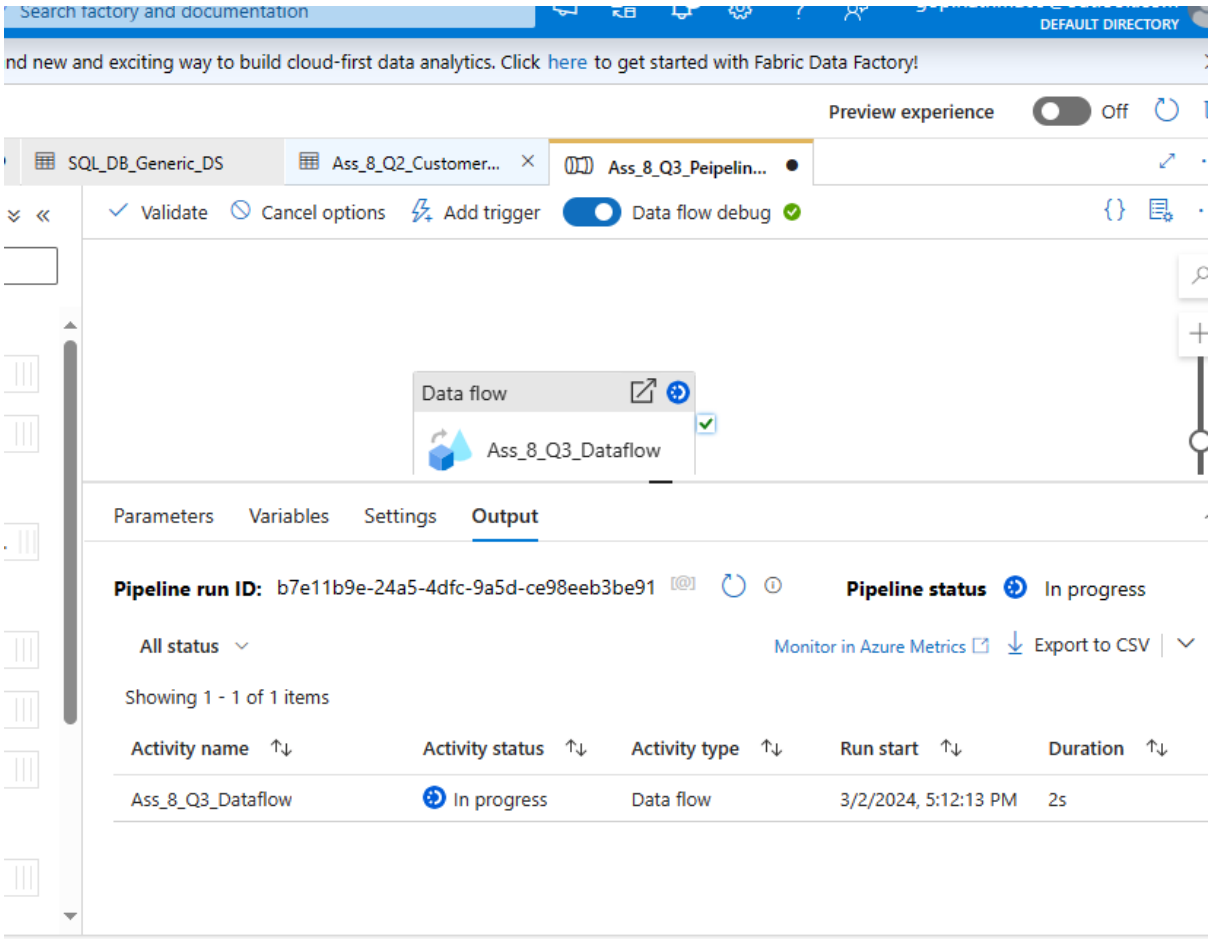

DEFAULT DIRECTORY
Select format
Choose the format type of your data
Avro
DelimitedText
JSON
ORC
Parquet
Binary
Continue
Back
Cancel

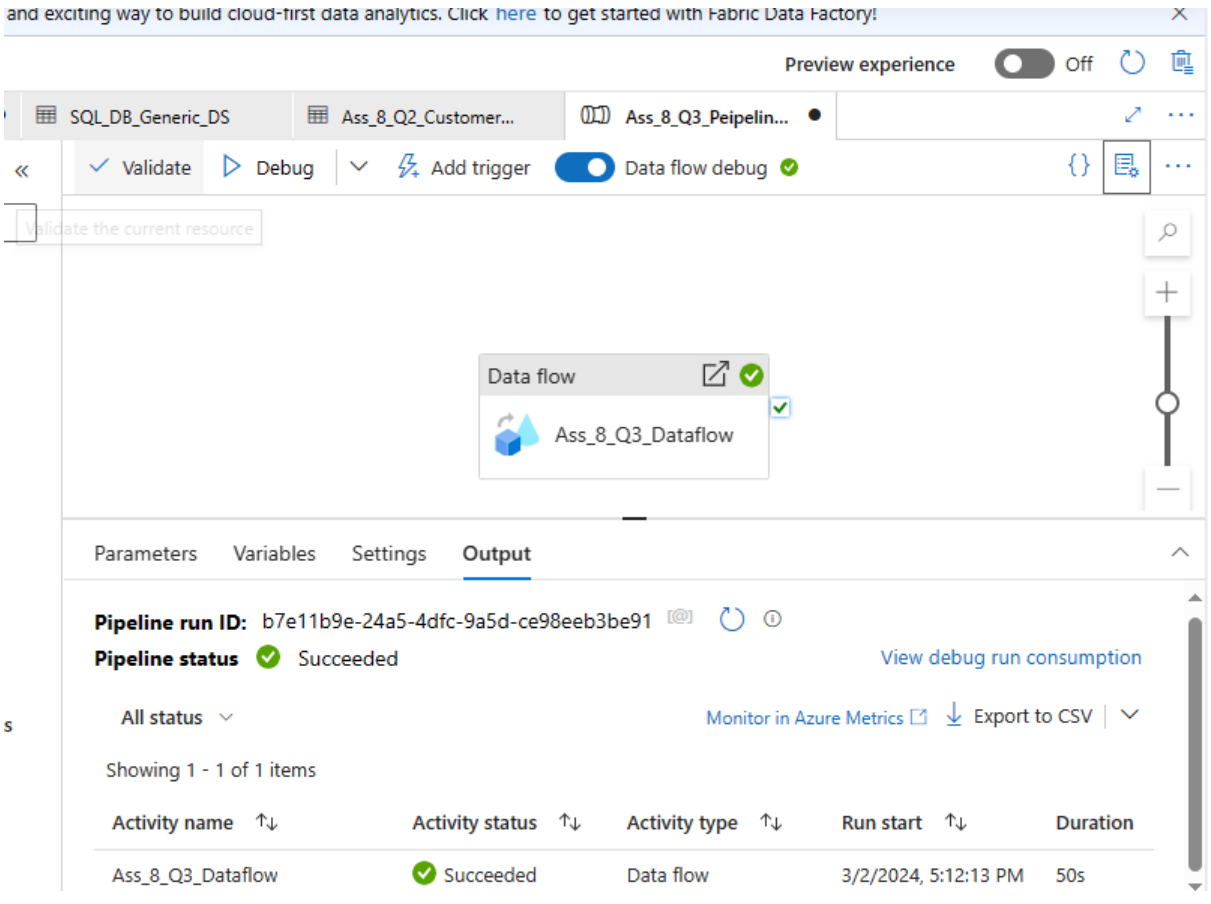

DEFAULT DIRECTORY
Set properties
Name
Ass_8_Q3_Customer_Parquet_file
Linked service *
ADLS_Azure_DataLakeStorage1_Linked_service
File path
landing
/
Ass_8_output_files
/
File name
Import schema
From connection/store
From sample file
None
Advanced
OK
Back
Cancel

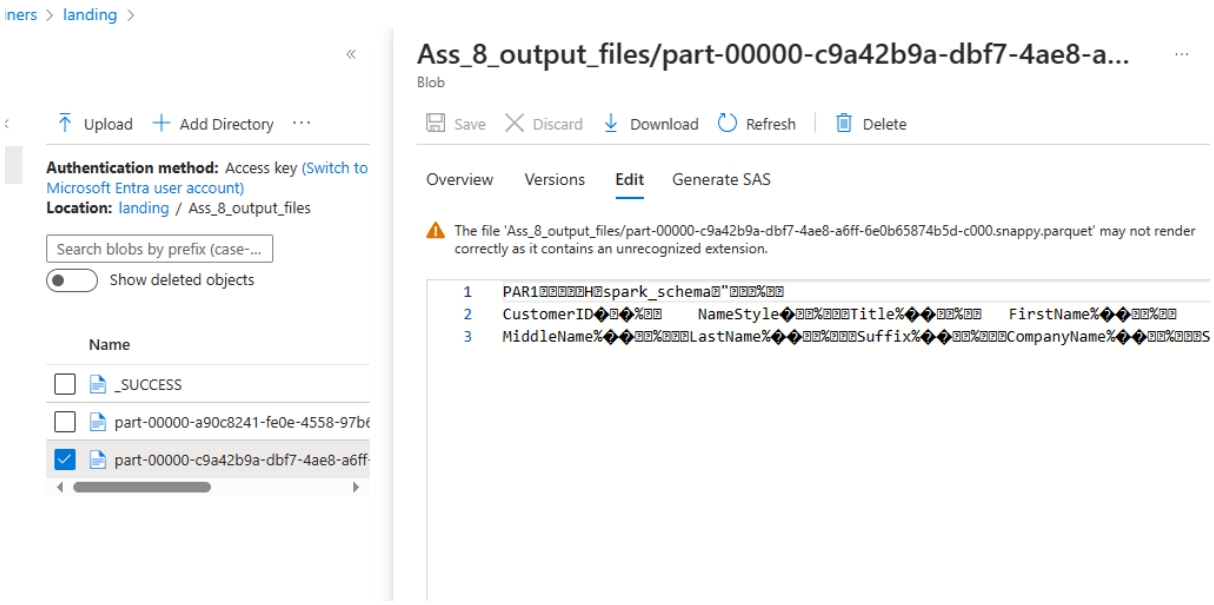


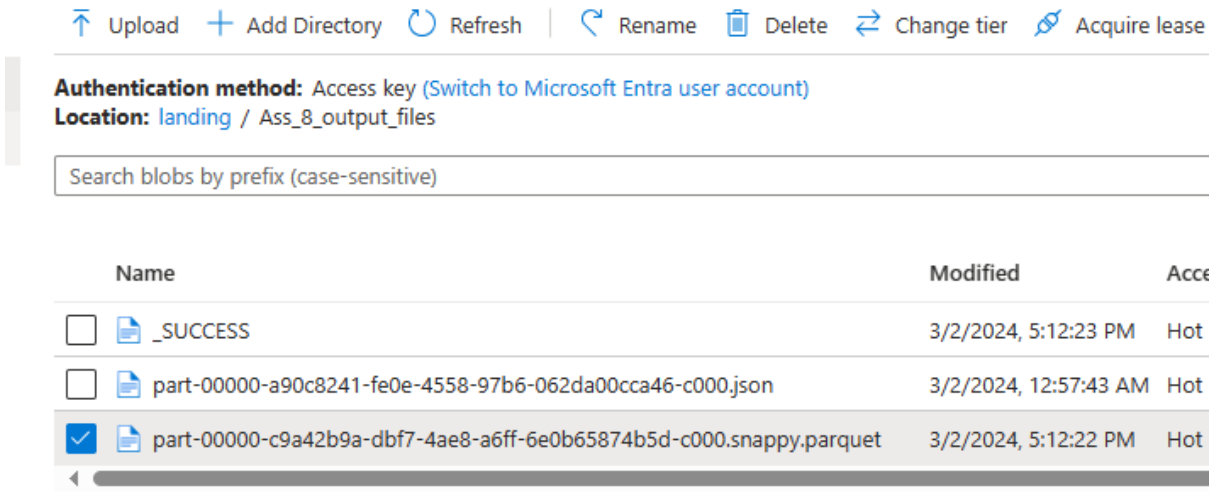
Lastly I creaete one pipeline which will call this “Ass_8_Q3_ Dataflow”

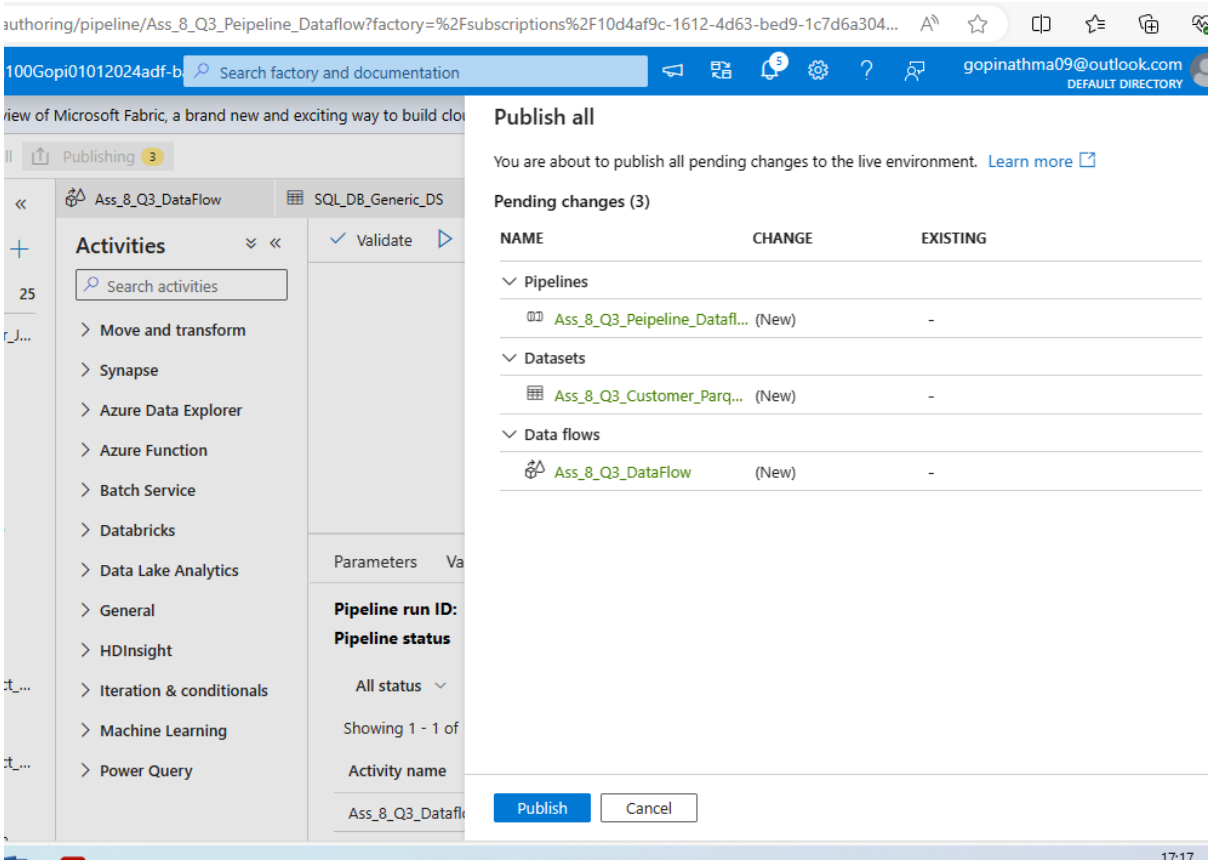

Preview experience
Off
SQL_DB_Generic_DS
Ass_8_Q2_Customer...
Ass_8_Q3_Peipelin...
Validate
Debug
Add trigger
Data flow debug
Validate the current resource
Data flow
Ass_8_Q3_Dataflow
General
Settings
Parameters 1
User properties
Data flow *
Ass_8_Q3_DataFlow
Open
New
CustomerTable parameters
Name
Value
Type
schemaName
Value
string
tableName
Value
string
Run on (Azure IR) *
AutoResolveIntegrationRuntime
Compute size *
Small
ENG

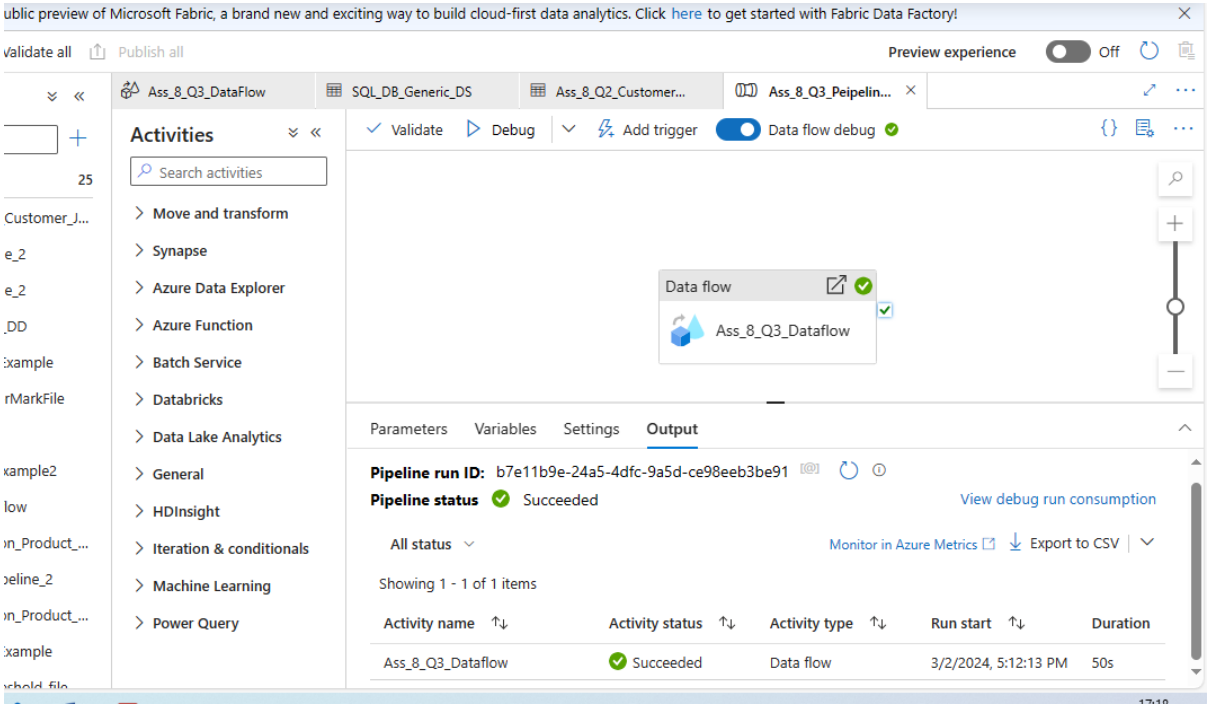

Search factory and documentation
DEFAULT DIRECTORY
nd new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Preview experience
Off
SQL_DB_Generic_DS
Ass_8_Q2_Customer...
Ass_8_Q3_Peipelin...
Validate
Cancel options
Add trigger
Data flow debug
Data 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_Peipelin...
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


iners > landing >
Upload
Add Directory
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: landing / Ass_8_output_files
Search blobs by prefix (case-...
Show deleted objects
Name
_SUCCESS
part-00000-a90c8241-fe0e-4558-97b6
part-00000-c9a42b9a-dbf7-4ae8-a6ff
Ass_8_output_files/part-00000-c9a42b9a-dbf7-4ae8-a...
Blob
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□□□□□H□spark_schema□"□□□%□□
2 CustomerID□□□%□□ NameStyle□□□%□□□Title%□□□%□□ FirstName%□□□%□□
3 MiddleName%□□□%□□□LastName%□□□%□□□Suffix%□□□%□□□CompanyName%□□□%□□□S


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
Acce
_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
part-00000-c9a42b9a-dbf7-4ae8-a6ff-6e0b65874b5d-c000.snappy.parquet
3/2/2024, 5:12:22 PM
Hot


authoring/pipeline/Ass_8_Q3_Peipeline_Dataflow?factory=%2Fsubscriptions%2F10d4af9c-1612-4d63-bed9-1c7d6a304...
100Gopi01012024adf-b
Search factory and documentation
gopinathma09@outlook.com
DEFAULT DIRECTORY
iew of Microsoft Fabric, a brand new and exciting way to build clou
Publishing 3
Ass_8_Q3_DataFlow
SQL_DB_Generic_DS
Activities
Search activities
Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query
Validate
Parameters
Pipeline run ID:
Pipeline status
All status
Showing 1 - 1 of
Activity name
Publish all
You are about to publish all pending changes to the live environment. Learn more
Pending changes (3)
NAME
CHANGE
EXISTING
Pipelines
Ass_8_Q3_Peipeline_Datafl...
(New)
Datasets
Ass_8_Q3_Customer_Parq...
(New)
Data flows
Ass_8_Q3_DataFlow
(New)
Publish
Cancel

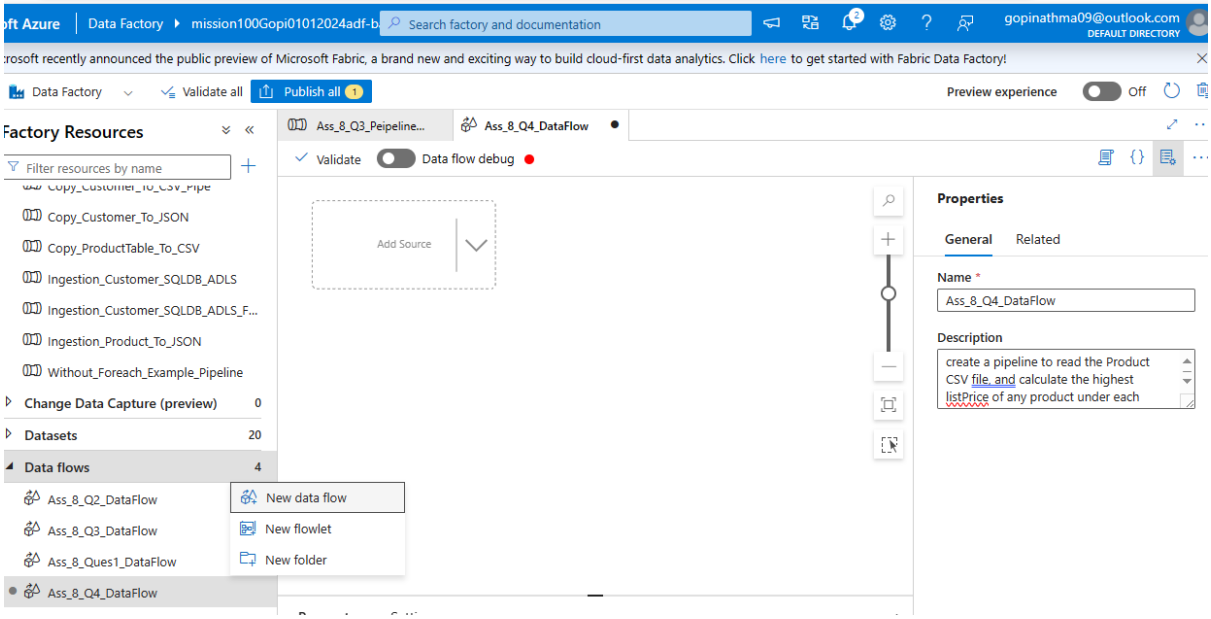

ublic preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Validate all
Publish all
Preview experience
Off
Ass_8_Q3_DataFlow
SQL_DB_Generic_DS
Ass_8_Q2_Customer...
Ass_8_Q3_Peipelin...
Activities
Search activities
Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query
Validate
Debug
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
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

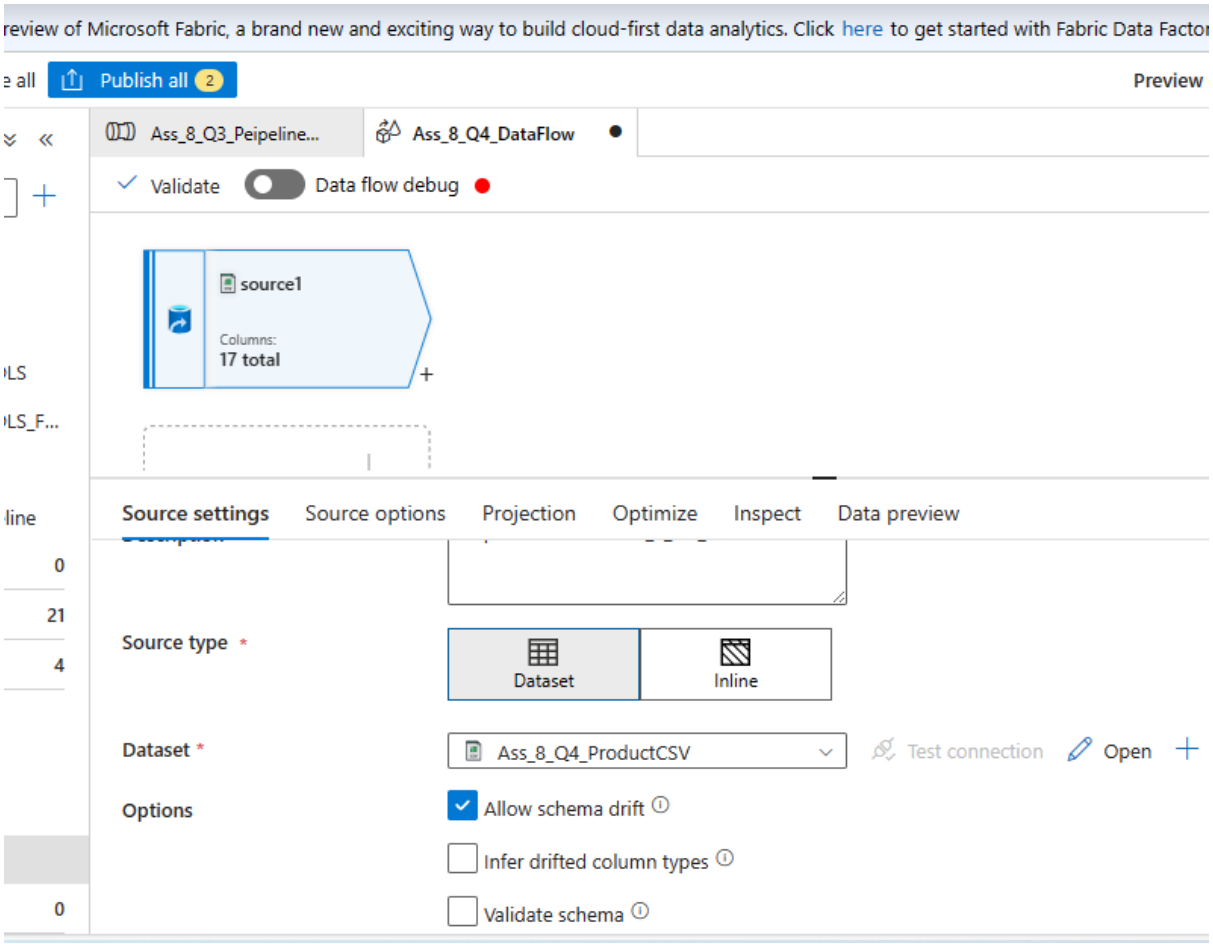
Question 4:

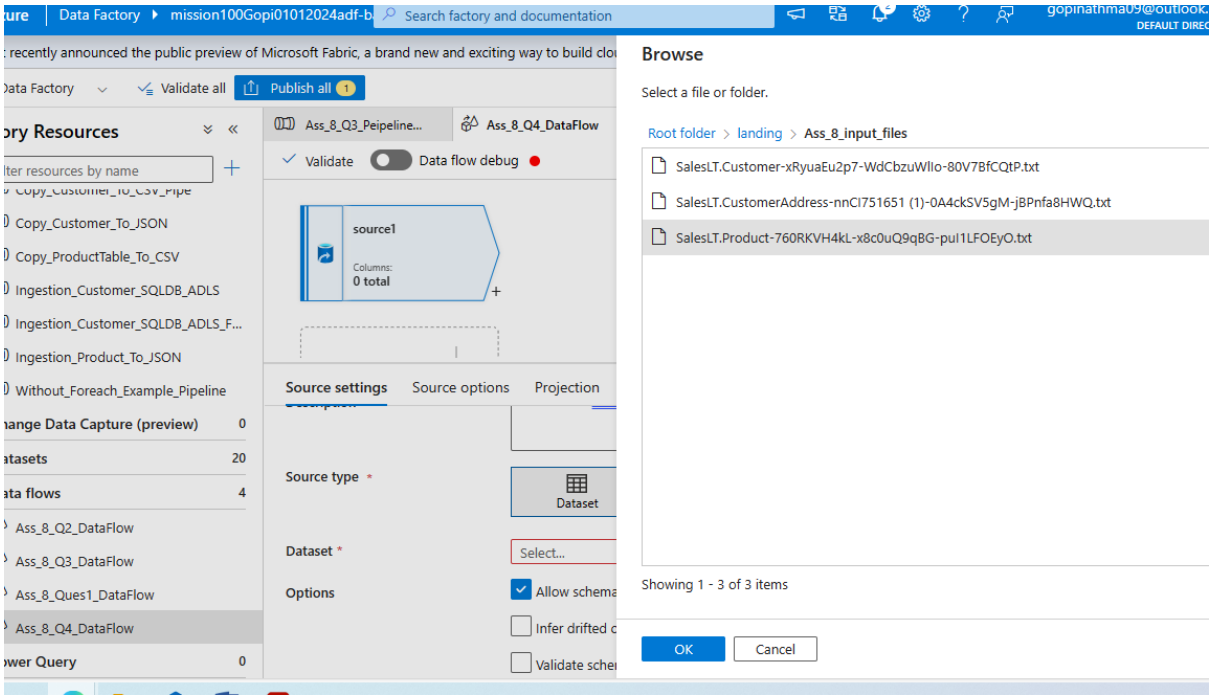
create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.
Ensure that product shouldn't be of blue in color and save the result as CSV file inside ProductResult folder.

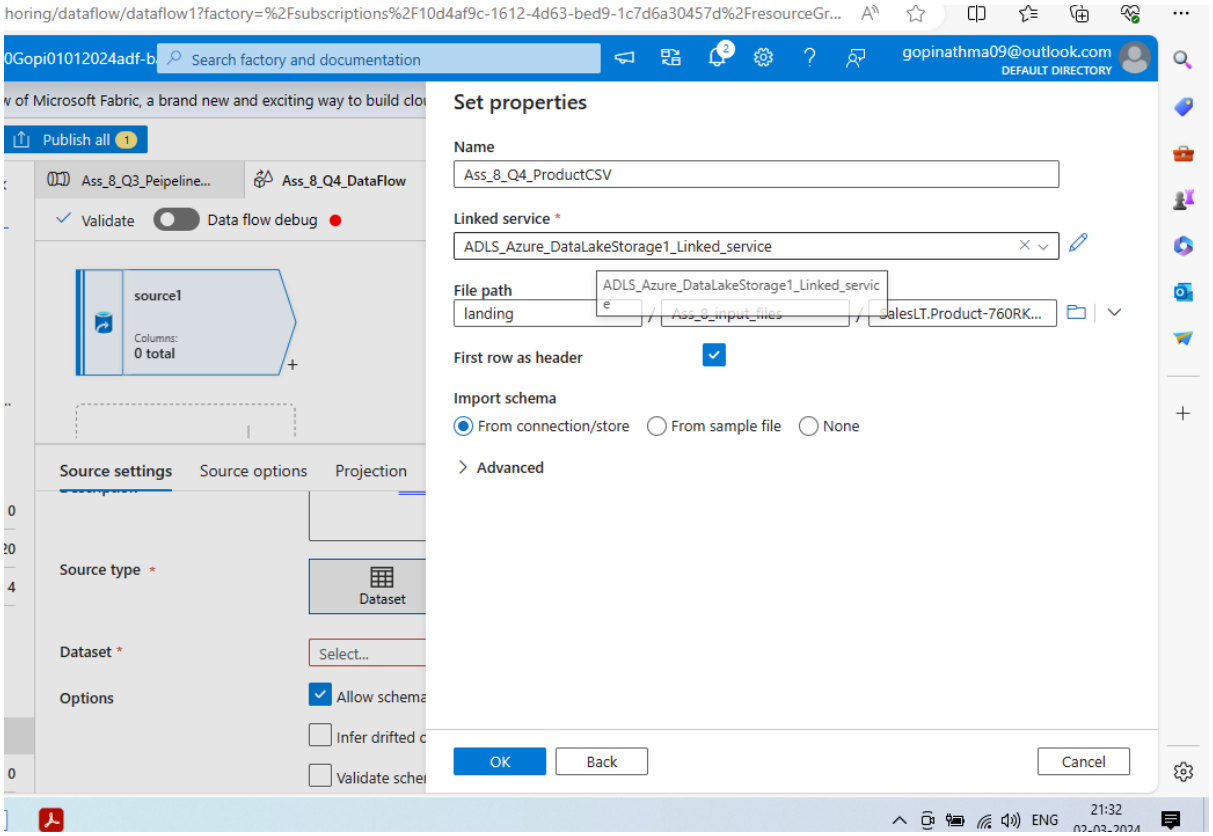
Solution 4:

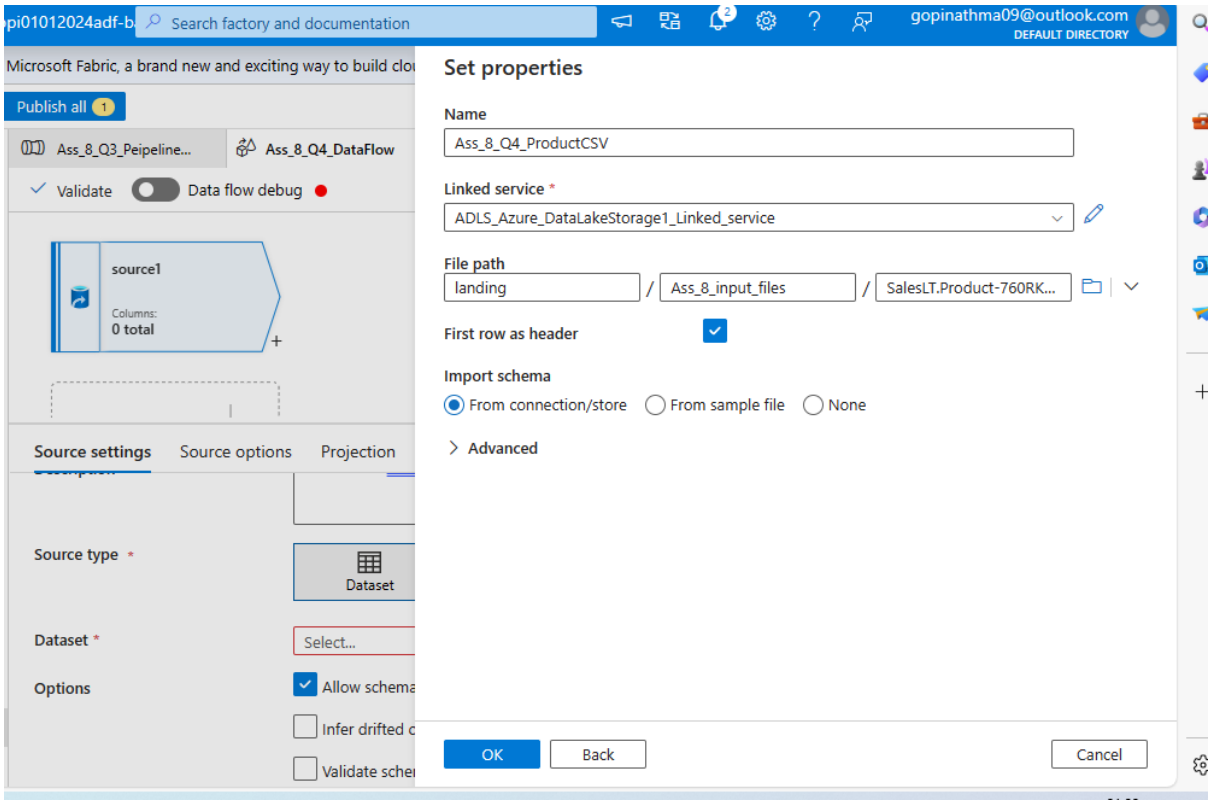
Since question has csv file, so we need to use "Data flow"

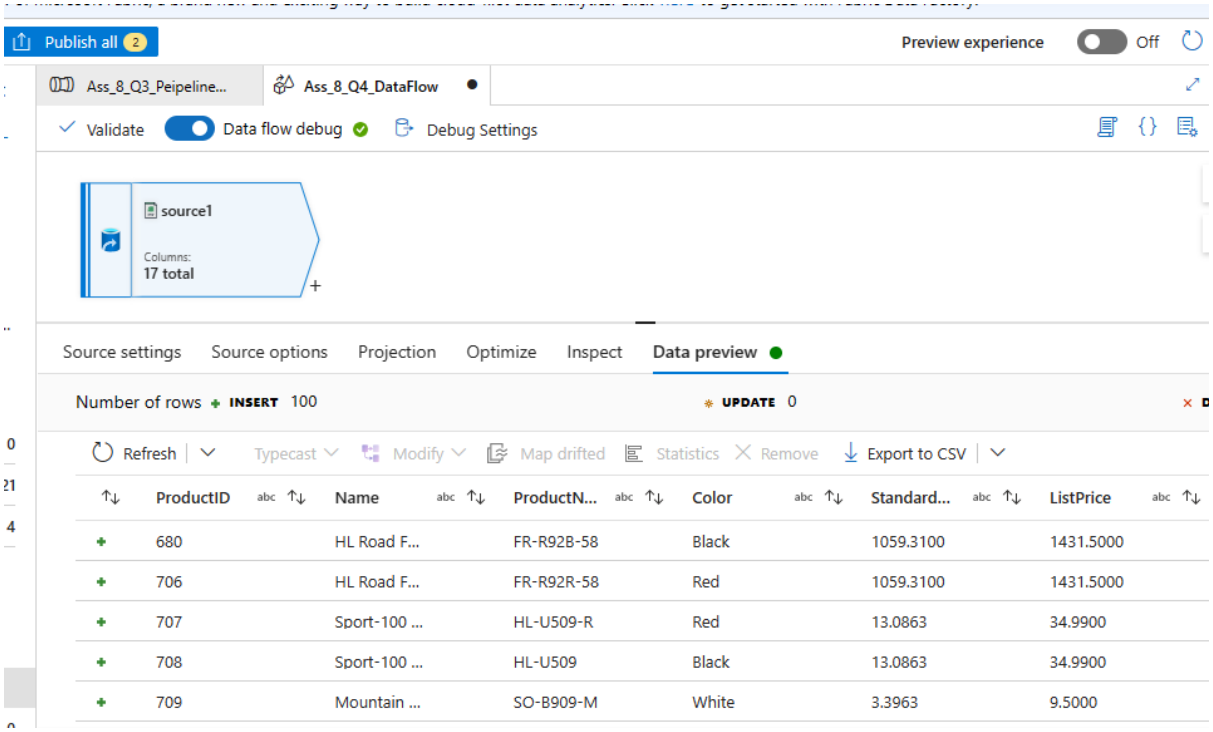


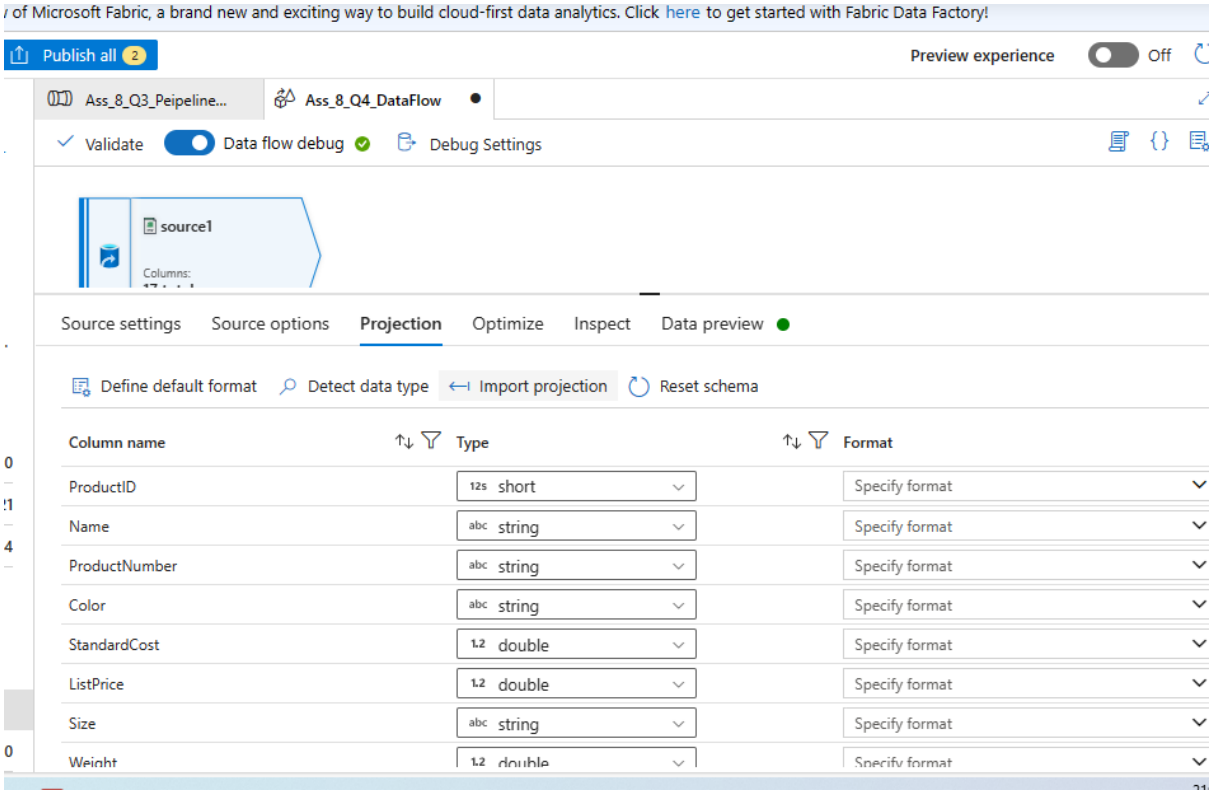

review of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factor
Publish all 2
Preview
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
source1
Columns:
17 total
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Source type *
Dataset
Inline
Dataset *
Ass_8_Q4_ProductCSV
Test connection
Open
Options
Allow schema drift
Infer drifted column types
Validate schema


Data Factory
mission100Gopi01012024adf-b
Search factory and documentation
Validate all
Publish all 1
Resources
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
Ass_8_Q2_DataFlow
Ass_8_Q3_DataFlow
Ass_8_Ques1_DataFlow
Ass_8_Q4_DataFlow
Source settings
Source options
Projection
Source type *
Dataset
Dataset *
Select...
Options
Allow schema
Browse
Select a file or folder.
Root folder > landing > Ass_8_input_files
SalesLT.Customer-xRyuaEu2p7-WdCbzuWIIo-80V7BfCQtP.txt
SalesLT.CustomerAddress-nnCI751651 (1)-0A4ckSV5gM-jBPnfa8HWQ.txt
SalesLT.Product-760RKVH4kL-x8c0uQ9qBG-pul1LFOEyO.txt
Showing 1 - 3 of 3 items
OK
Cancel

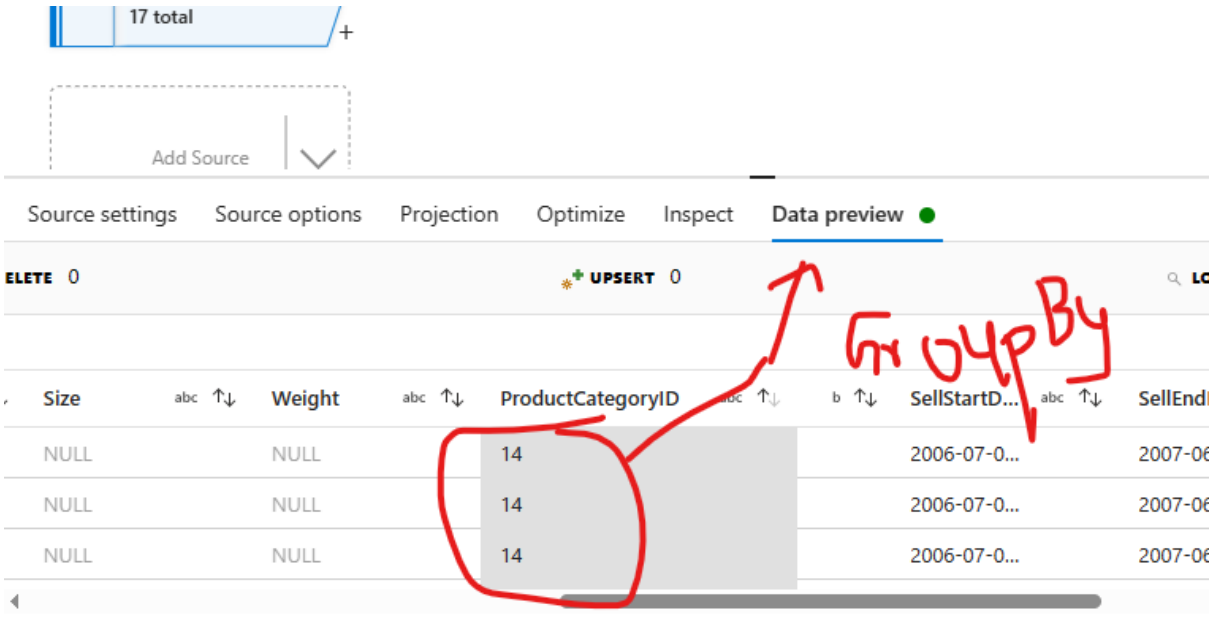

Search factory and documentation
gopinathma09@outlook.com
DEFAULT DIRECTORY
Publish all 1
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
source1
Columns:
0 total
Source settings
Source options
Projection
Source type *
Dataset
Dataset *
Select...
Options
Allow schema
Infer drifted c
Validate schei
Set properties
Name
Ass_8_Q4_ProductCSV
Linked service *
ADLS_Azure_DataLakeStorage1_Linked_service
File path
landing
Ass_8_input_files
SalesLT.Product-760RK...
First row as header
Import schema
From connection/store
From sample file
None
Advanced
OK
Back
Cancel
ENG
21:32


Search factory and documentation
gopinathma09@outlook.com
DEFAULT DIRECTORY
Publish all 1
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
source1
Columns:
0 total
Source settings
Source options
Projection
Source type *
Dataset
Dataset *
Select...
Options
Allow schema
Infer drifted c
Validate schei
Set properties
Name
Ass_8_Q4_ProductCSV
Linked service *
ADLS_Azure_DataLakeStorage1_Linked_service
File path
landing
Ass_8_input_files
SalesLT.Product-760RK...
First row as header
Import schema
From connection/store
From sample file
None
Advanced
OK
Back
Cancel

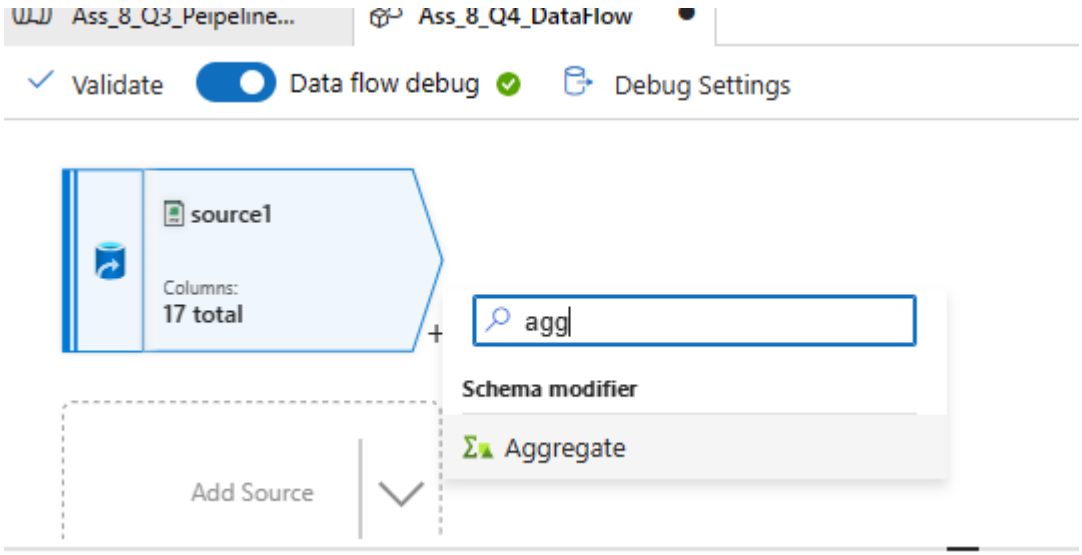

Publish all 2
Preview experience
Off
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
Debug Settings
source1
Columns:
17 total
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Number of rows
INSERT 100
UPDATE 0
Refresh
Typecast
Modify
Map drifted
Statistics
Remove
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

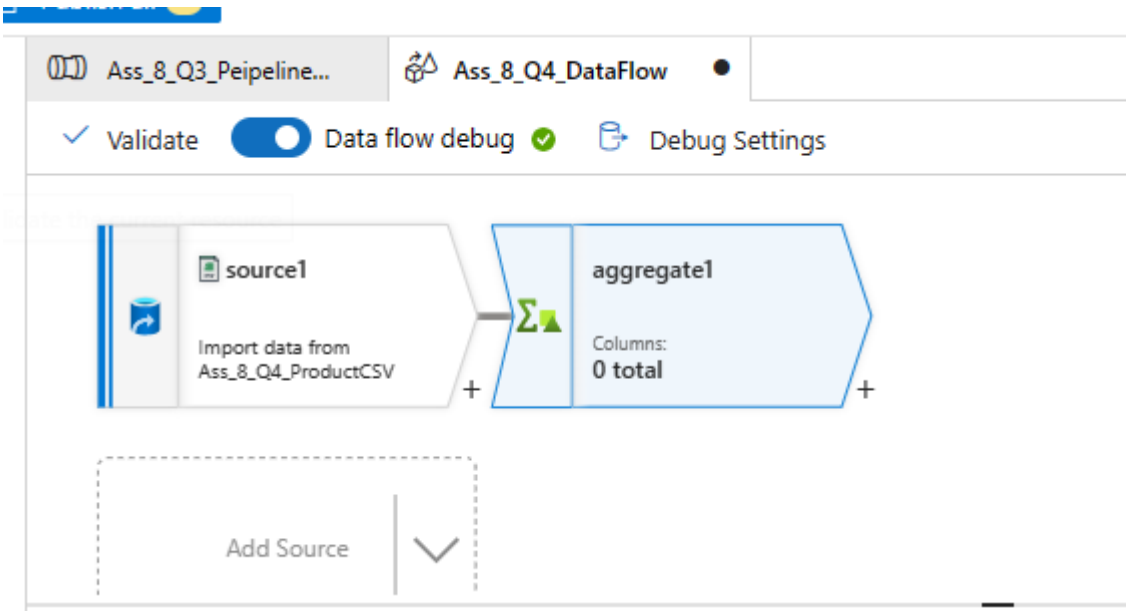

of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Publish all 2
Preview experience
Off
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
Debug Settings
source1
Columns:
Source settings
Source options
Projection
Optimize
Inspect
Data preview
Define default format
Detect data type
Import projection
Reset schema
Column name
Type
Format
ProductID short Specify format
Name string Specify format
ProductNumber string Specify format
Color string Specify format
StandardCost double Specify format
ListPrice double Specify format
Size string Specify format
Weight double Specify format

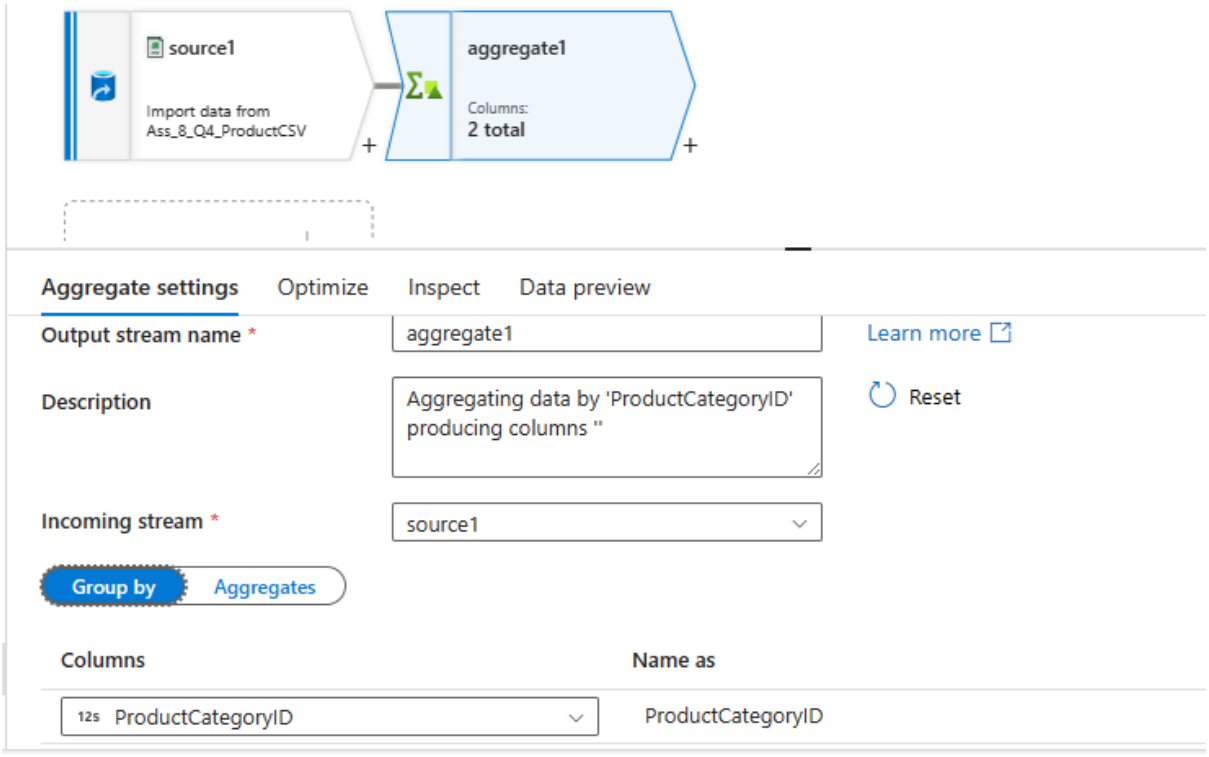
4. create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.
Ensure that product shouldn't be of blue in color and save the result as CSV file inside ProductResult folder.

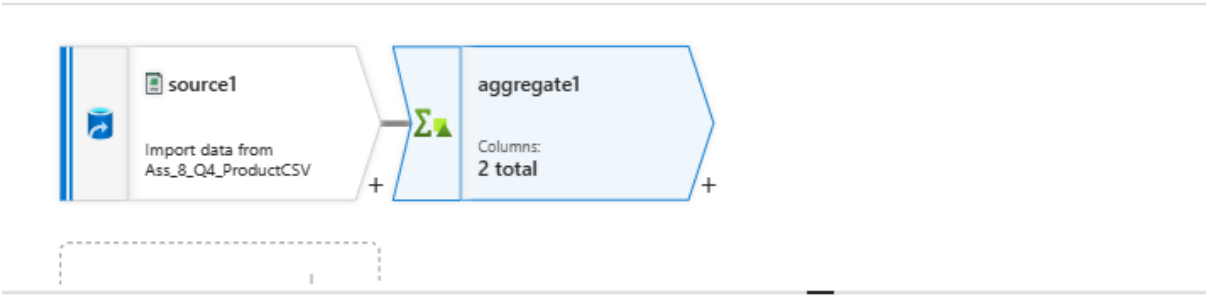


Do Aggreagate function to get groupby by category id




Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
Debug Settings
source1
Import data from
Ass_8_Q4_ProductCSV
aggregate1
Columns:
0 total
Add Source


source1
Import data from
Ass_8_Q4_ProductCSV
aggregate1
Columns:
2 total
Aggregate settings
Optimize
Inspect
Data preview
Output stream name *
aggregate1
Learn more
Description
Aggregating data by 'ProductCategoryID' producing columns ''
Reset
Incoming stream *
source1
Group by
Aggregates
Columns
Name as
12s ProductCategoryID
ProductCategoryID



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(ProductCategoryID)

Column name *

HighestListPrice

Expression

```
max()
```

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

Expression elements

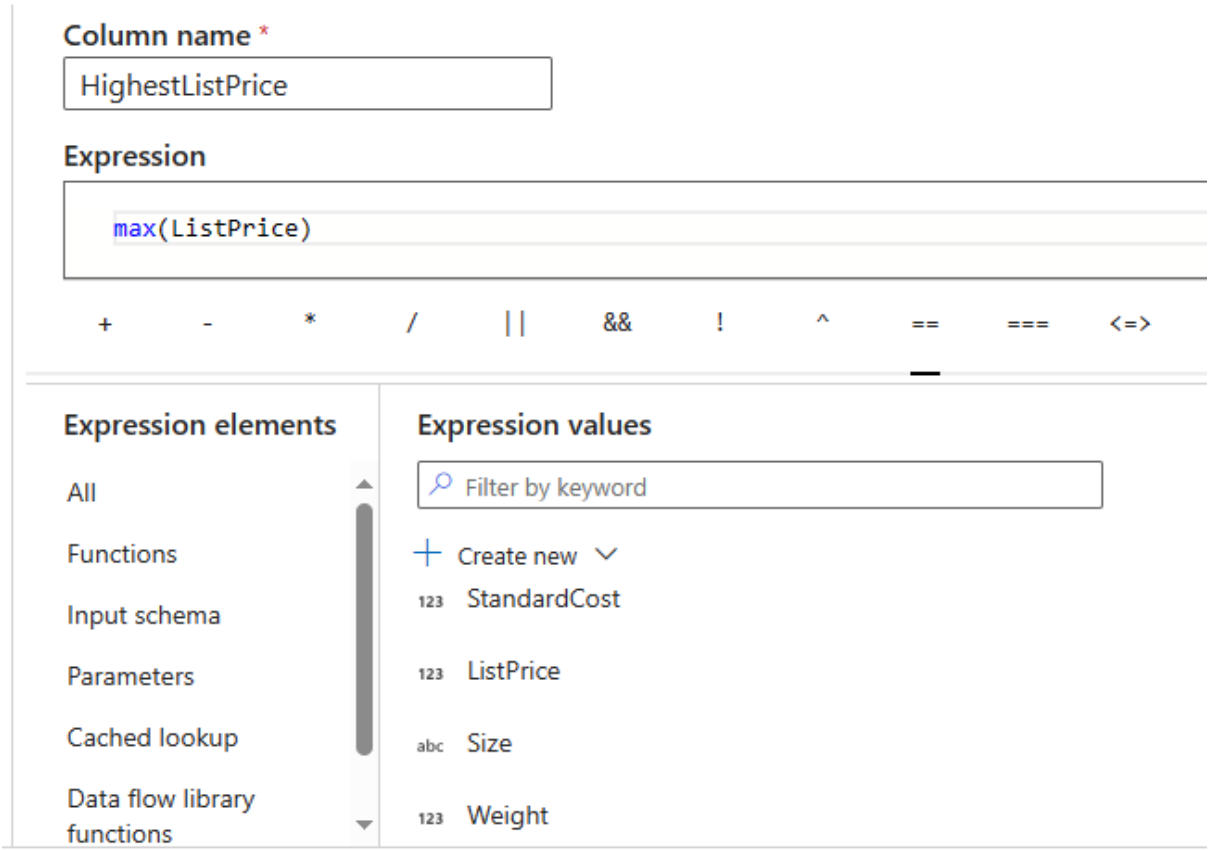
- All
- Functions
- Input schema
- Parameters
- Cached lookup
- Data flow library functions

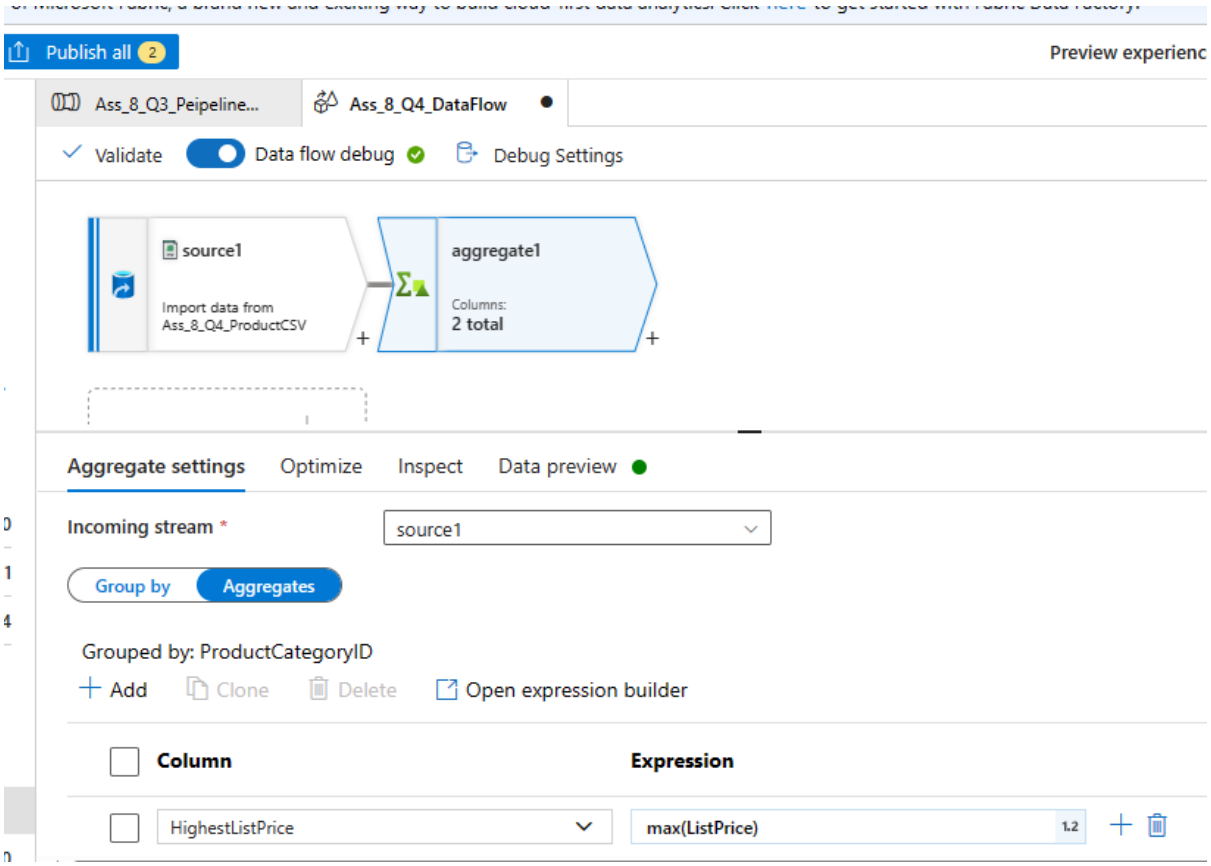
Expression values

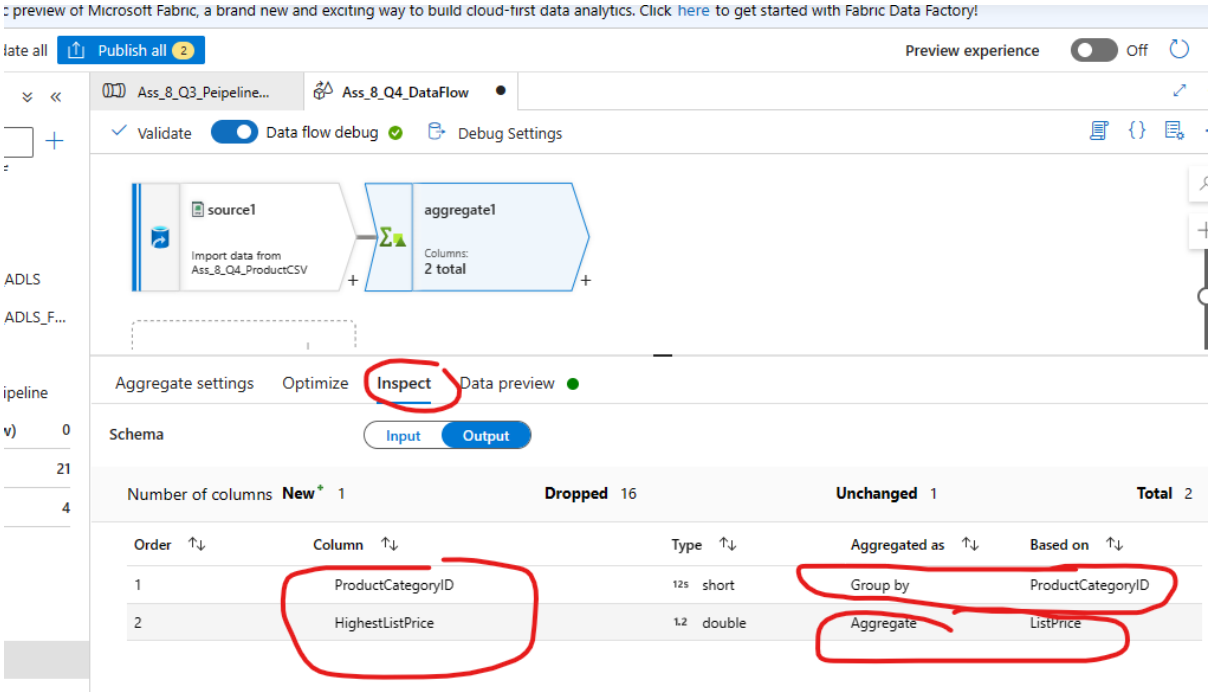
Filter by keyword

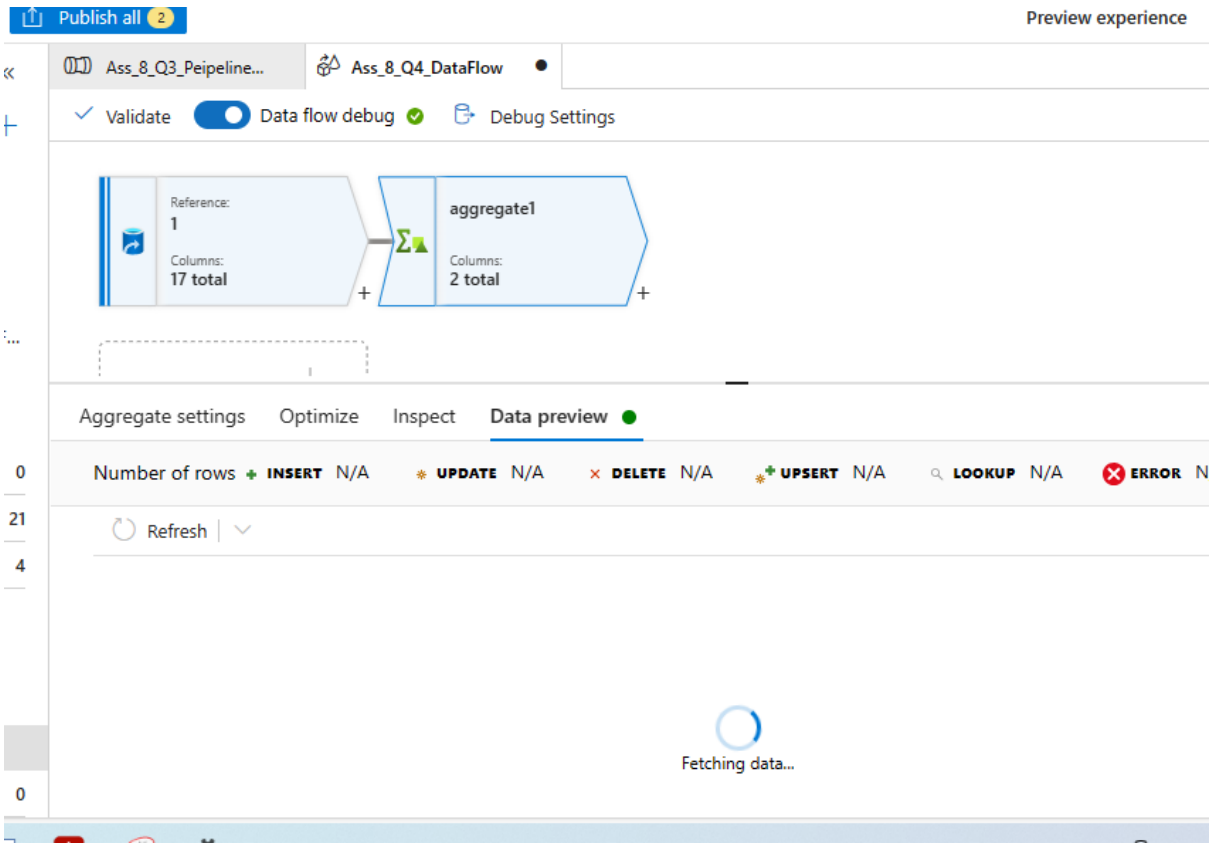
+ Create new

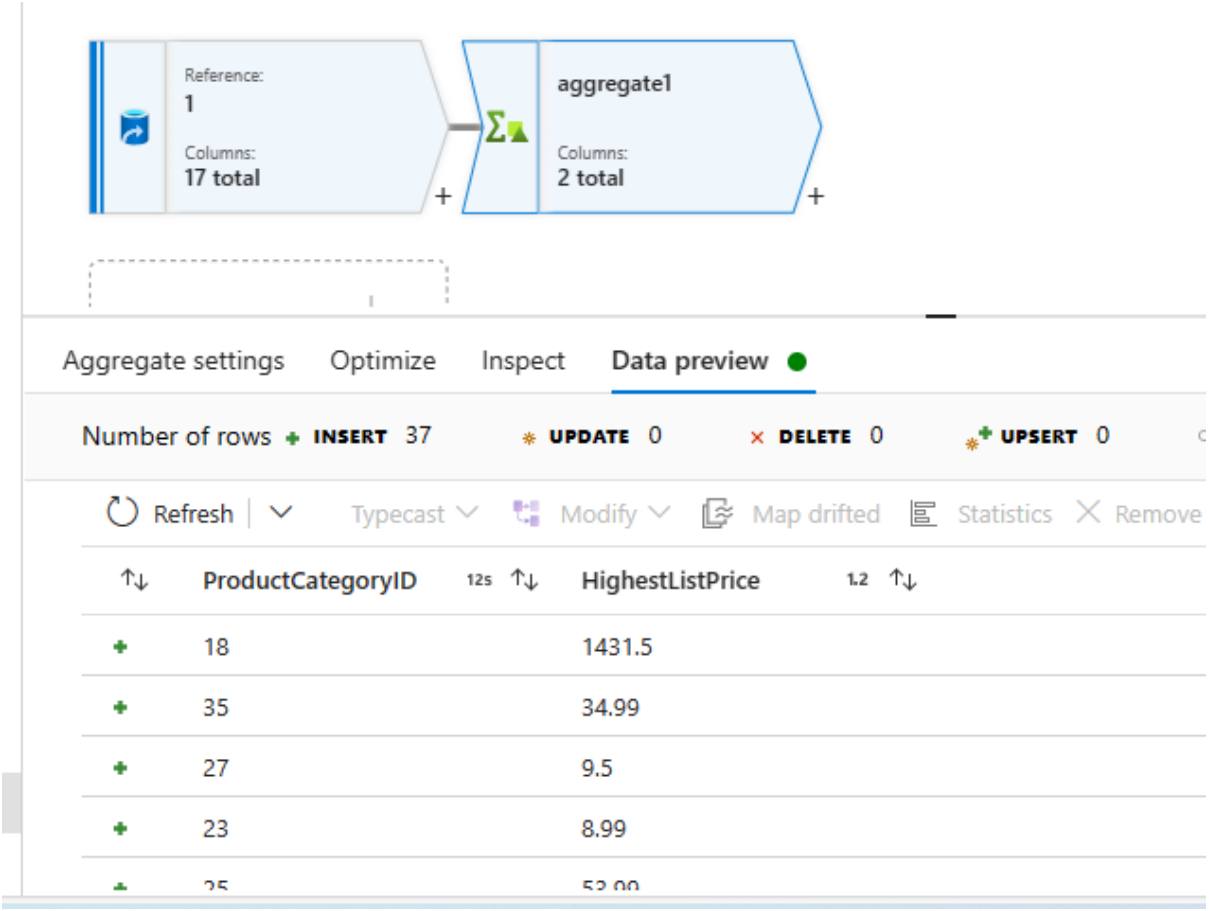
- 123 StandardCost
- 123 ListPrice
- abc Size
- 123 Weight


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

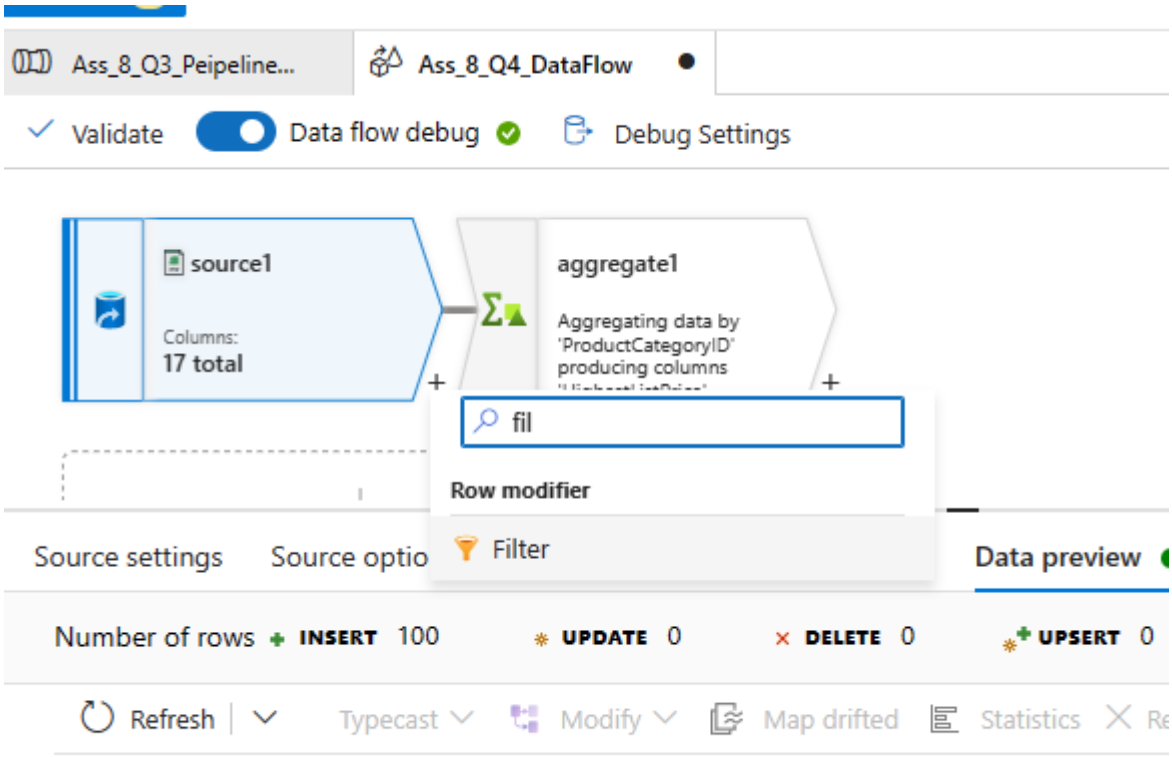

Publish all 2
Preview experienc
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
Debug Settings
source1
Import data from Ass_8_Q4_ProductCSV
aggregate1
Columns:
2 total
Aggregate settings
Optimize
Inspect
Data preview
Incoming stream *
source1
Group by
Aggregates
Grouped by: ProductCategoryID
Add
Clone
Delete
Open expression builder
Column
Expression
HighestListPrice
max(ListPrice)

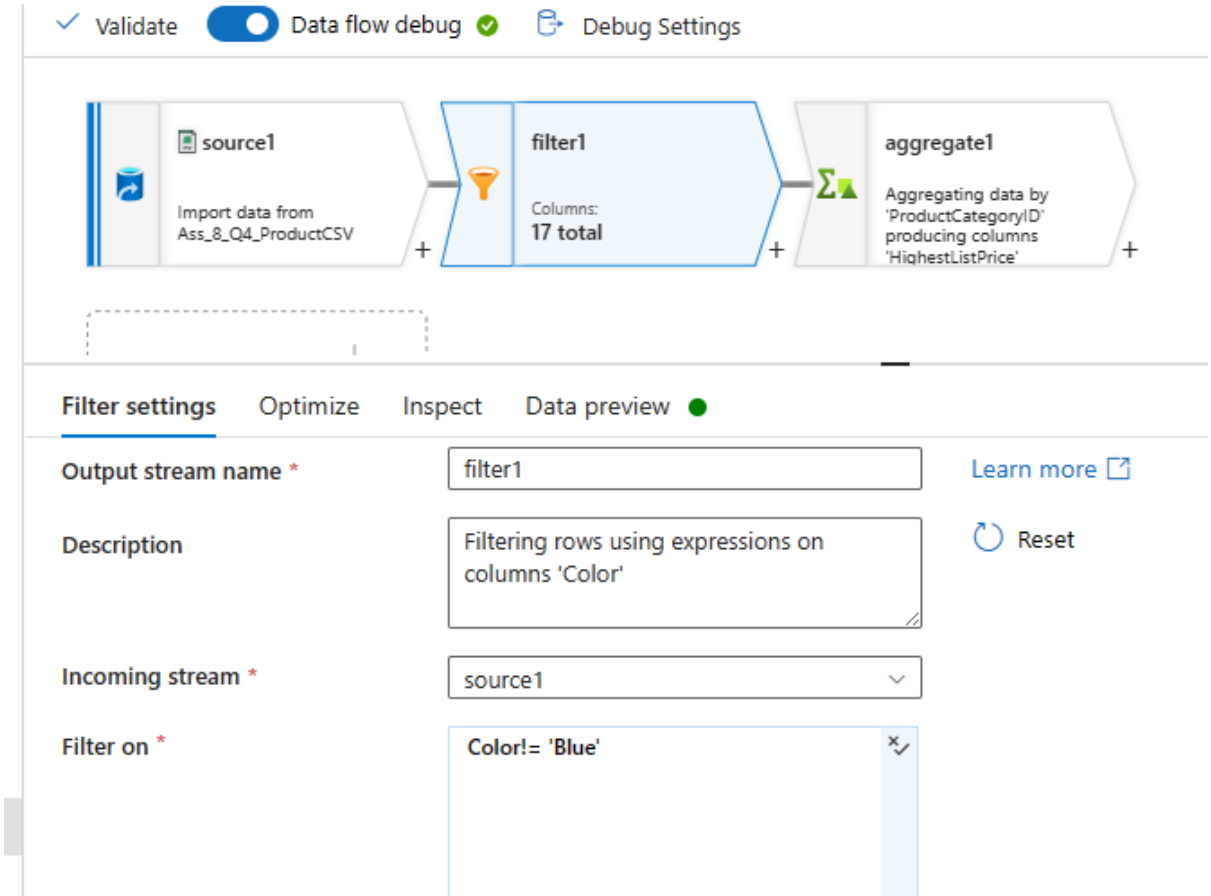

preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Publish all 2
Preview experience
Off
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
Debug Settings
source1
Import data from Ass_8_Q4_ProductCSV
aggregate1
Columns:
2 total
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
short
Group by
ProductCategoryID
2
HighestListPrice
double
Aggregate
ListPrice

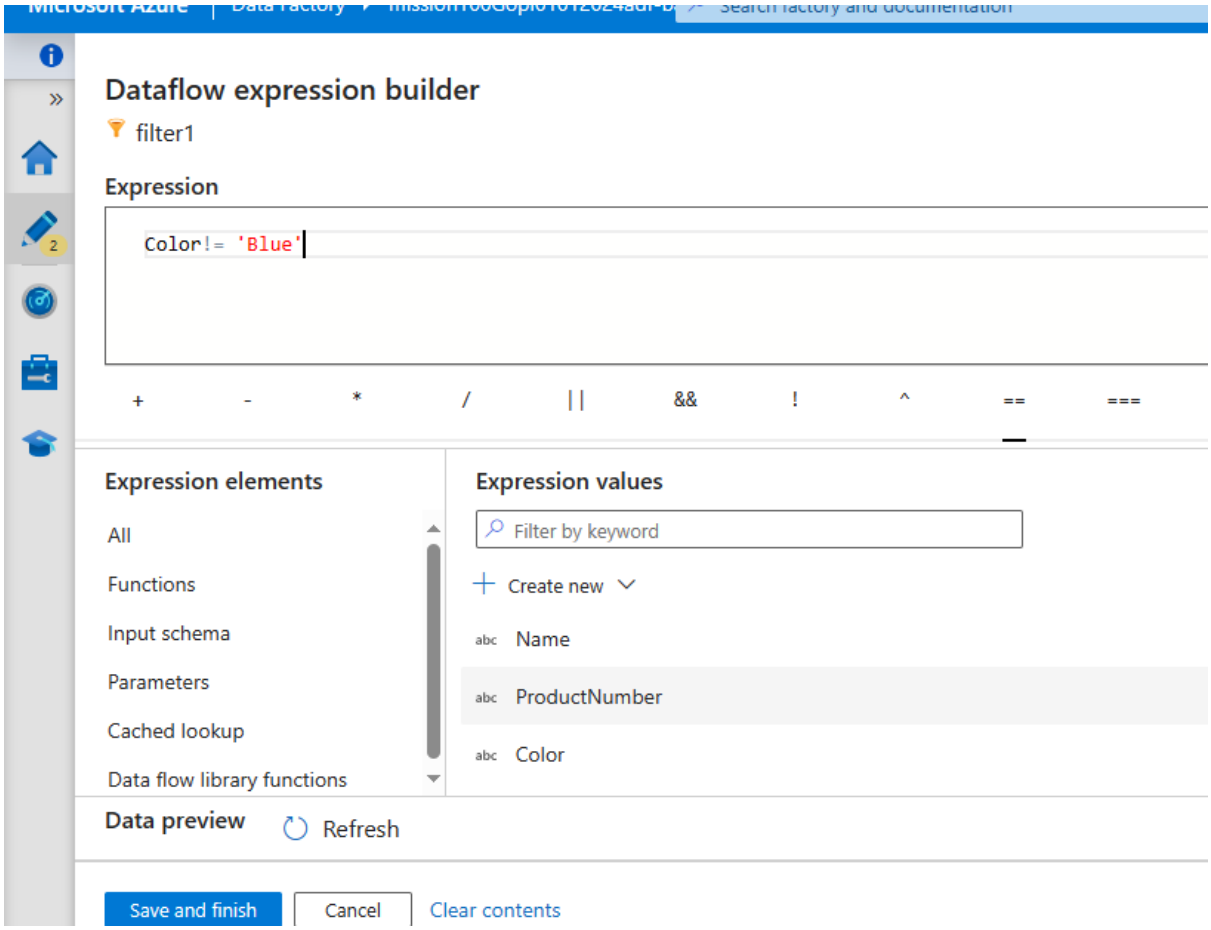

Publish all 2
Preview experience
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
Debug Settings
Reference:
1
Columns:
17 total
aggregate1
Columns:
2 total
Aggregate settings
Optimize
Inspect
Data preview
Number of rows
INSERT N/A
UPDATE N/A
DELETE N/A
UPSERT N/A
LOOKUP N/A
ERROR N
Refresh
Fetching data...



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


Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
Debug Settings
source1
Columns:
17 total
aggregate1
Aggregating data by 'ProductCategoryID' producing columns
fil
Row modifier
Filter
Source settings
Source optio
Data preview
Number of rows
INSERT 100
UPDATE 0
DELETE 0
UPSERT 0
Refresh
Typecast
Modify
Map drifted
Statistics


Validate
Data flow debug
Debug Settings
source1
Import data from Ass_8_Q4_ProductCSV
filter1
Columns:
17 total
aggregate1
Aggregating data by 'ProductCategoryID' producing columns 'HighestListPrice'
Filter settings
Optimize
Inspect
Data preview
Output stream name *
filter1
Learn more
Description
Filtering rows using expressions on columns 'Color'
Reset
Incoming stream *
source1
Filter on *
Color!= 'Blue'


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

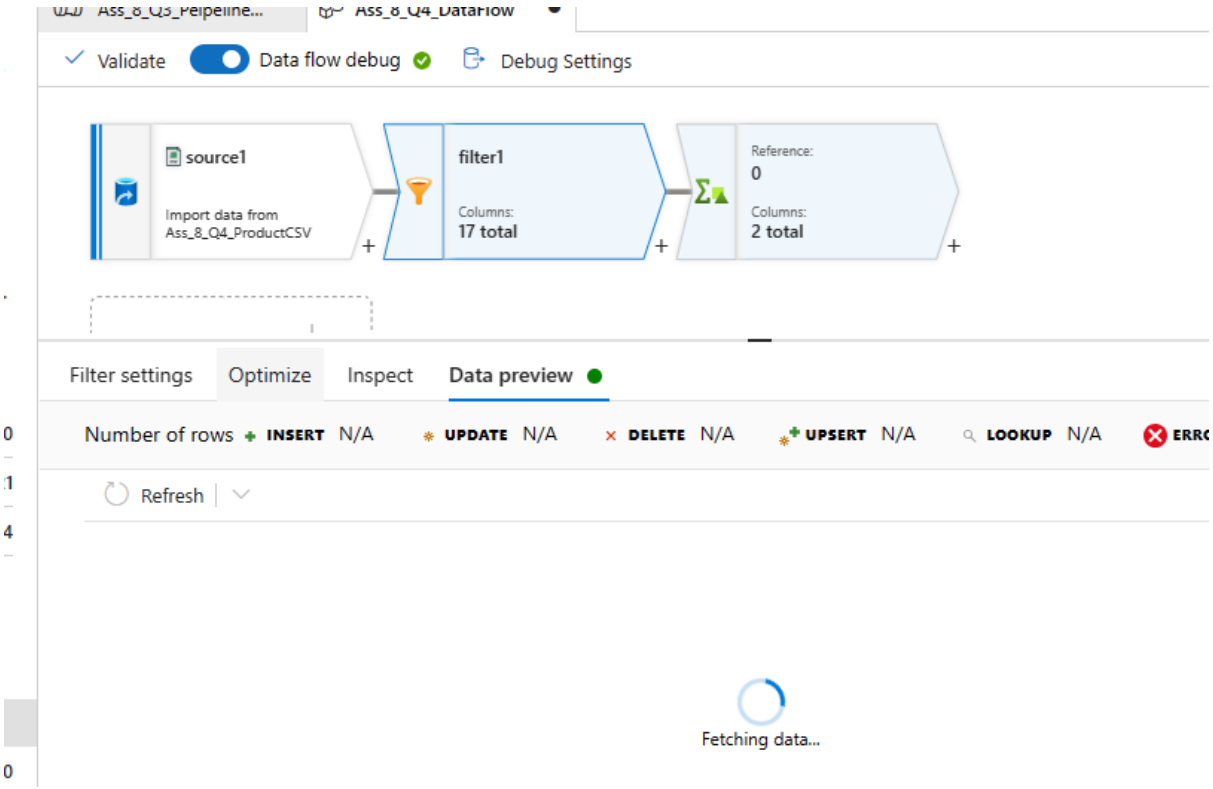
Create new

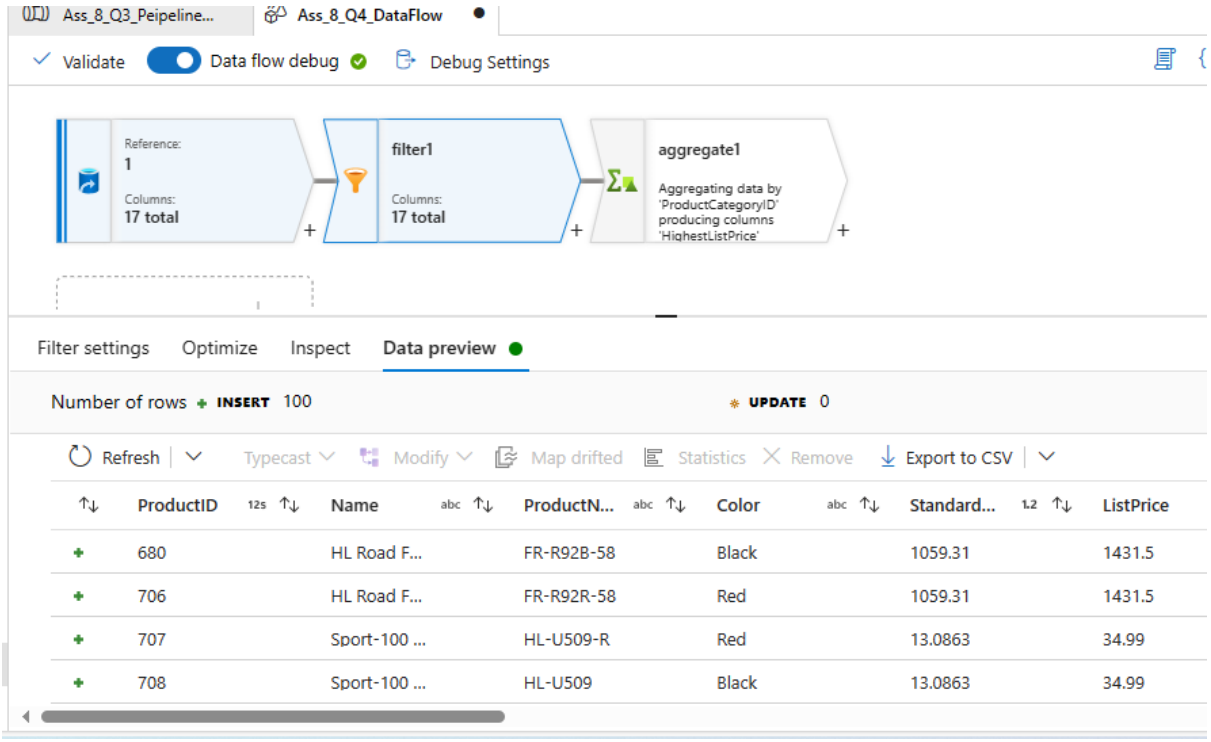
123 ProductID

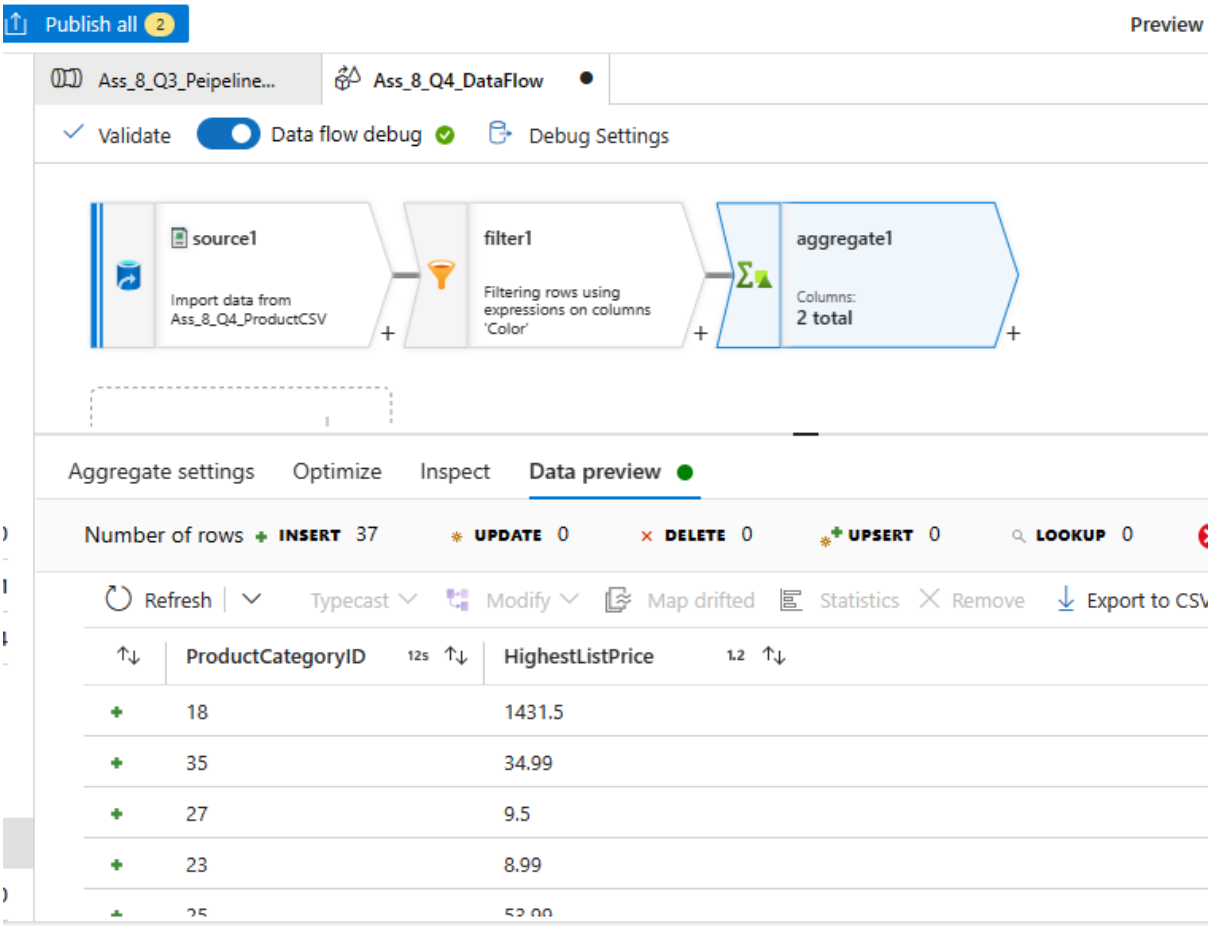
Data preview Refresh

✓	White
✓	White
×	Blue
✓	Multi
✓	Multi

Save and finish | Cancel | Clear contents

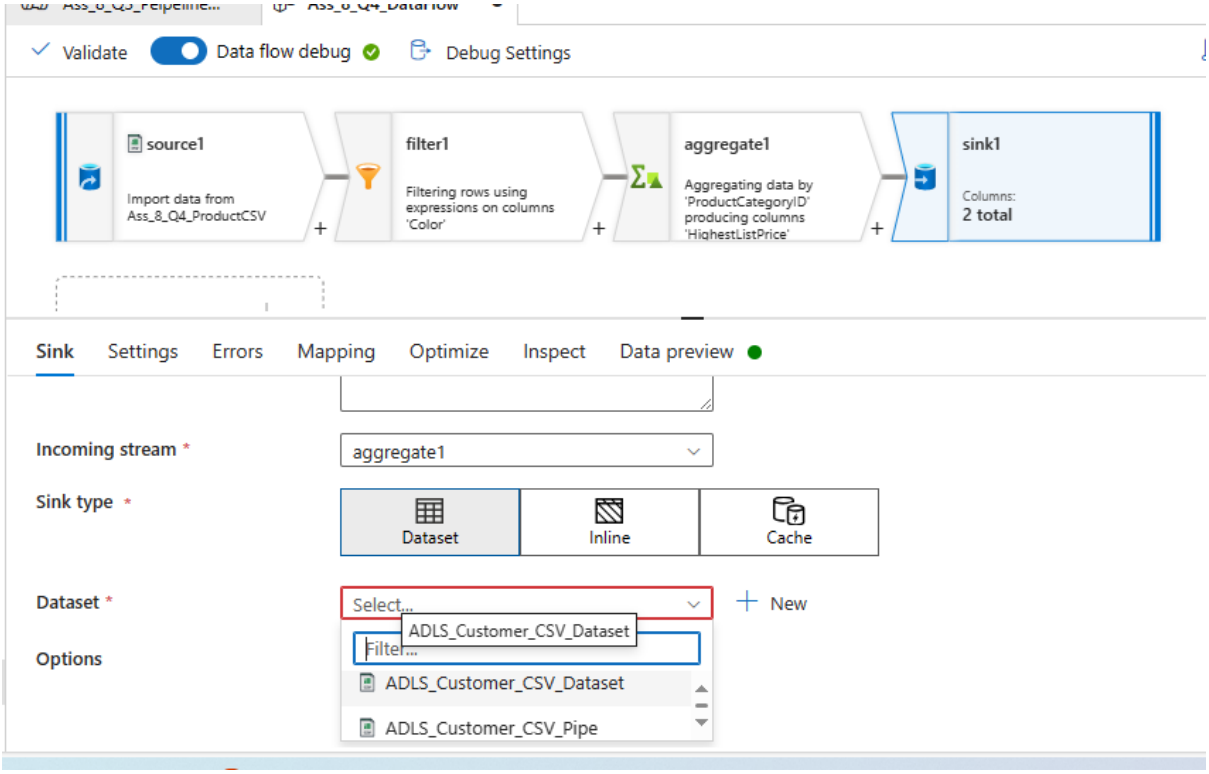


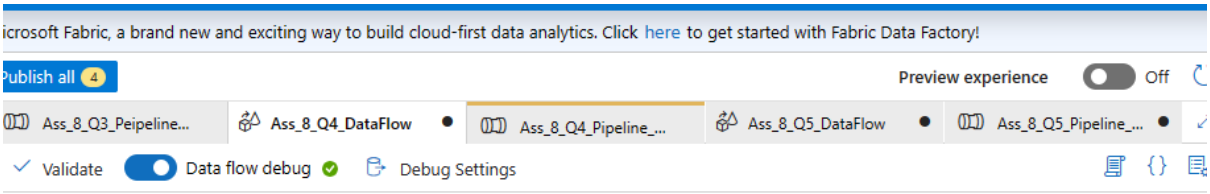

Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
Debug Settings
Reference:
1
Columns:
17 total
filter1
Columns:
17 total
aggregate1
Aggregating data by 'ProductCategoryID' producing columns 'HighestListPrice'
Filter settings
Optimize
Inspect
Data preview
Number of rows INSERT 100
UPDATE 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 1059.31 1431.5
706 HL Road F... FR-R92R-58 Red 1059.31 1431.5
707 Sport-100 ... HL-U509-R Red 13.0863 34.99
708 Sport-100 ... HL-U509 Black 13.0863 34.99

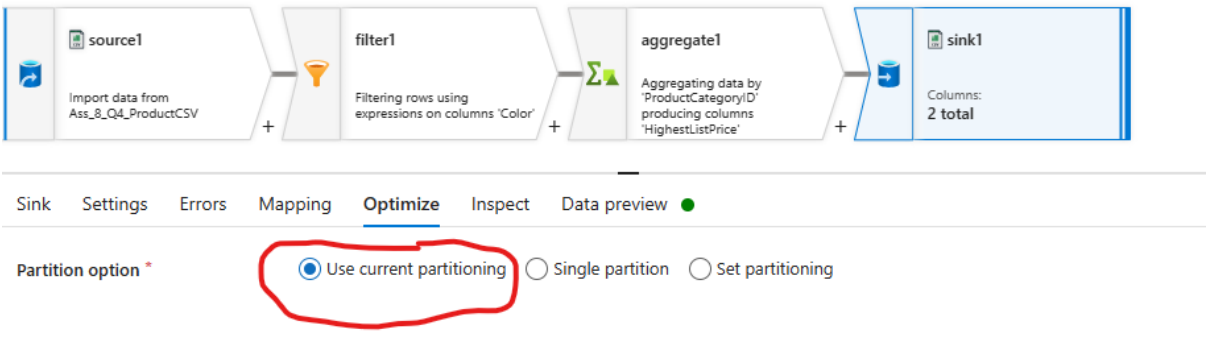

Publish all 2
Preview
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
Validate
Data flow debug
Debug Settings
source1
Import data from Ass_8_Q4_ProductCSV
filter1
Filtering rows using expressions on columns 'Color'
aggregate1
Columns:
2 total
Aggregate settings
Optimize
Inspect
Data preview
Number of rows INSERT 37
UPDATE 0
DELETE 0
UPSERT 0
LOOKUP 0
Refresh
Typecast
Modify
Map drifted
Statistics
Remove
Export to CSV
ProductCategoryID
HighestListPrice
18 1431.5
35 34.99
27 9.5
23 8.99

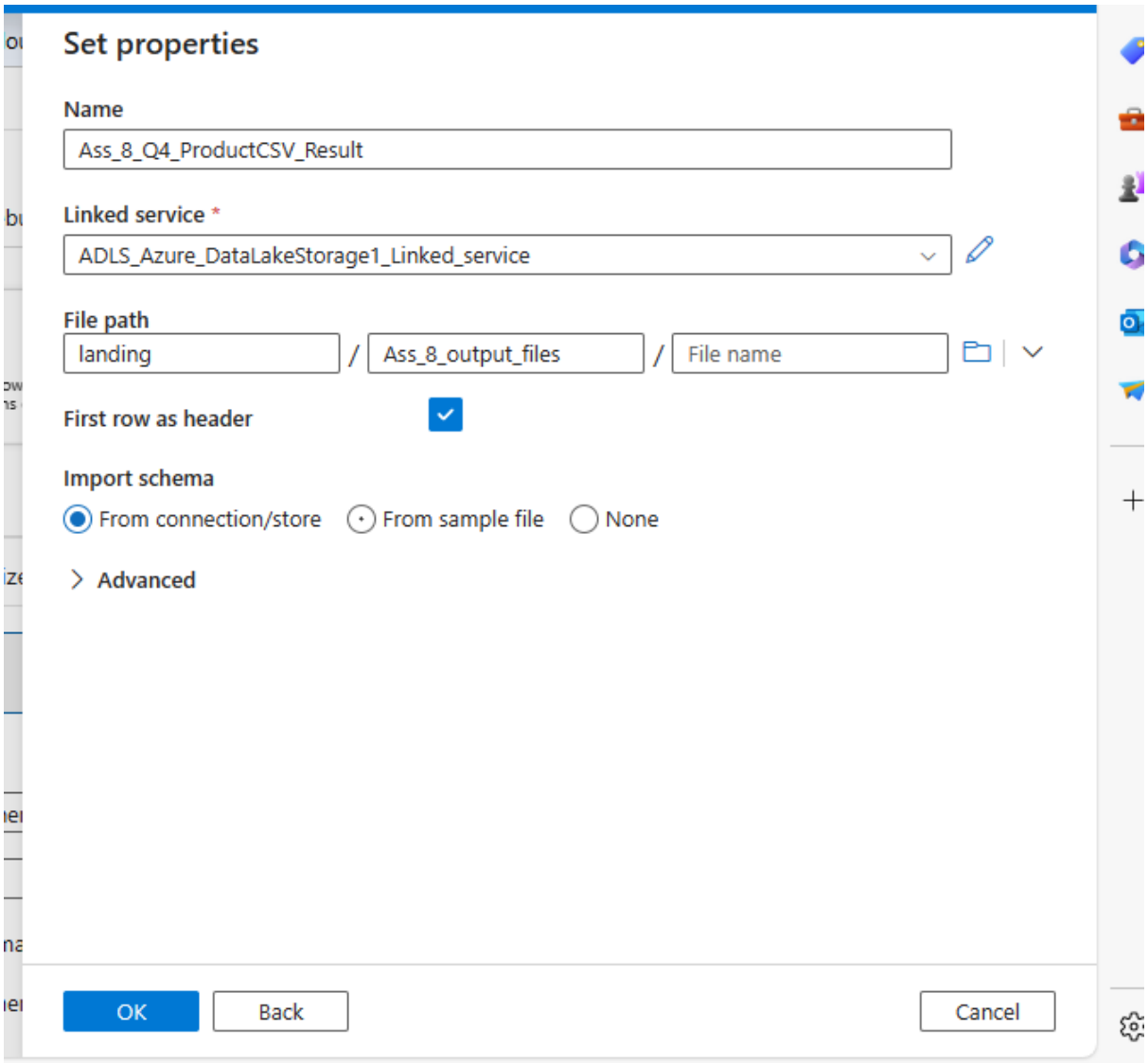
save the result as CSV file inside ProductResult folder.

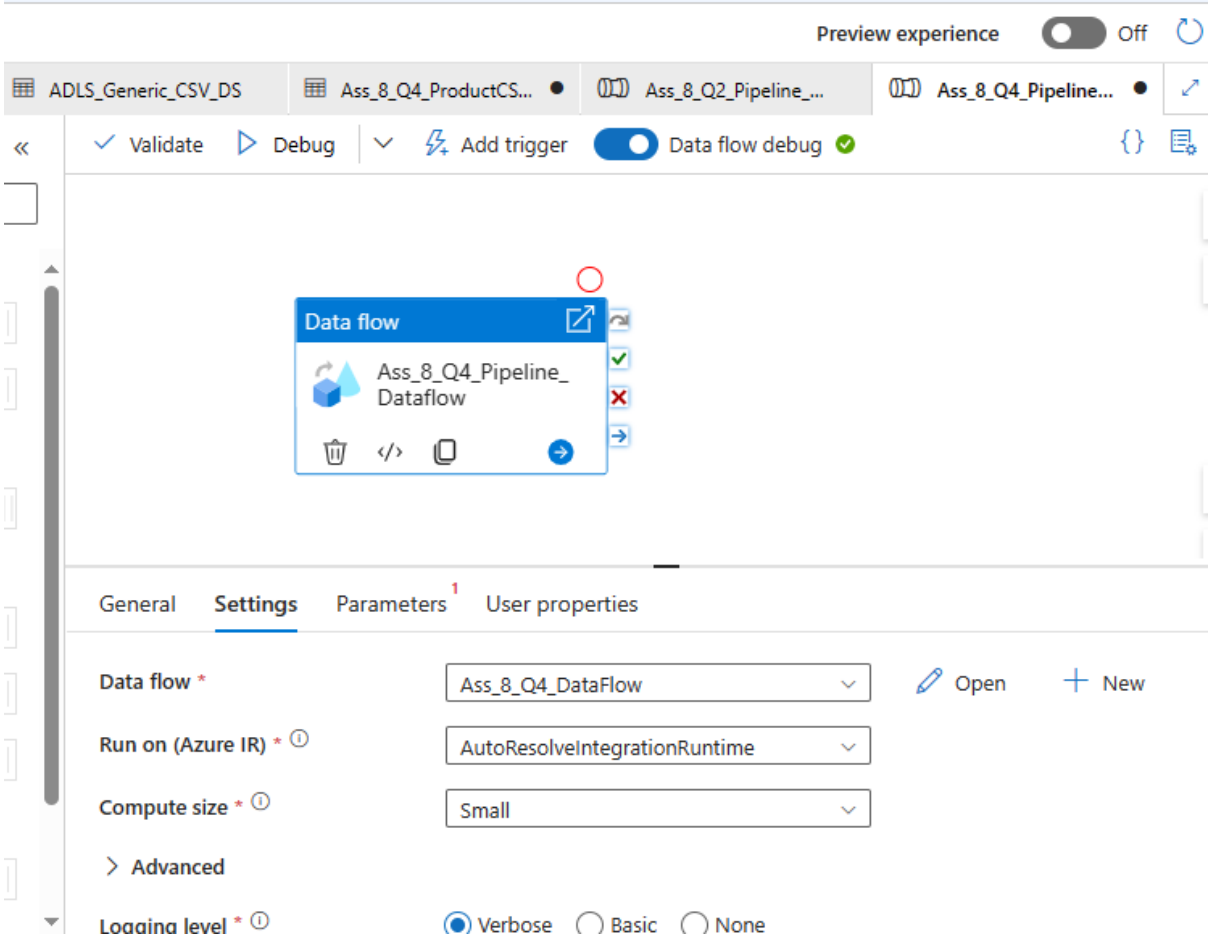
➔ So save in “Sink” file

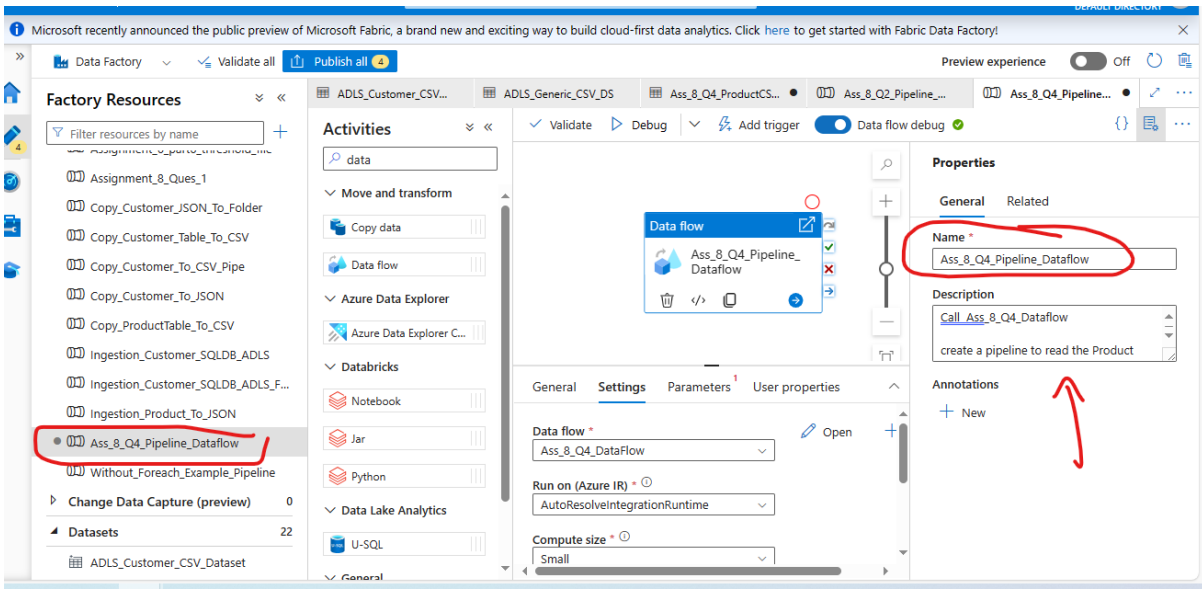


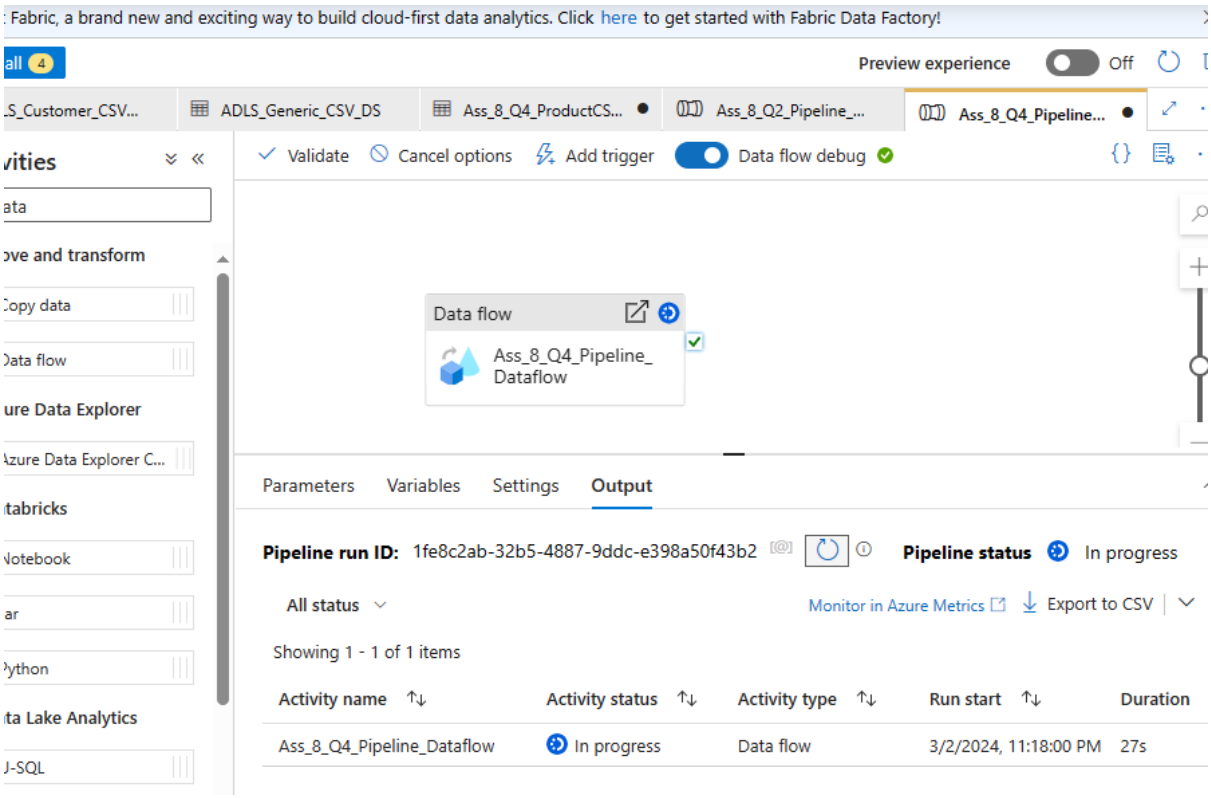


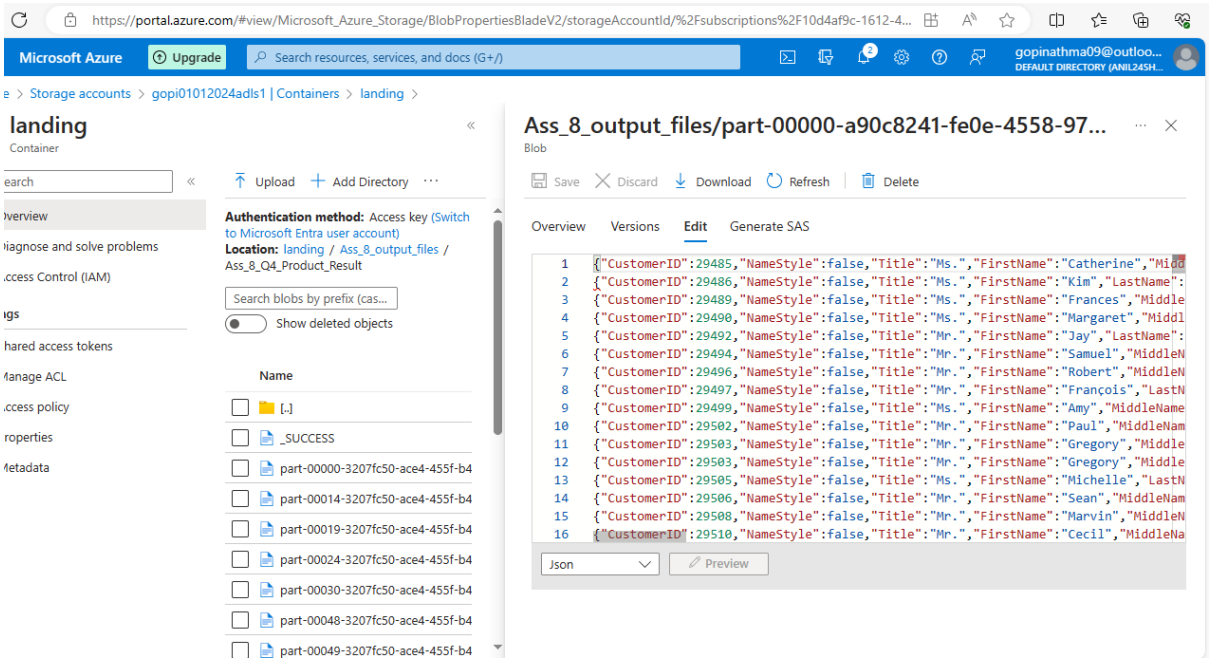


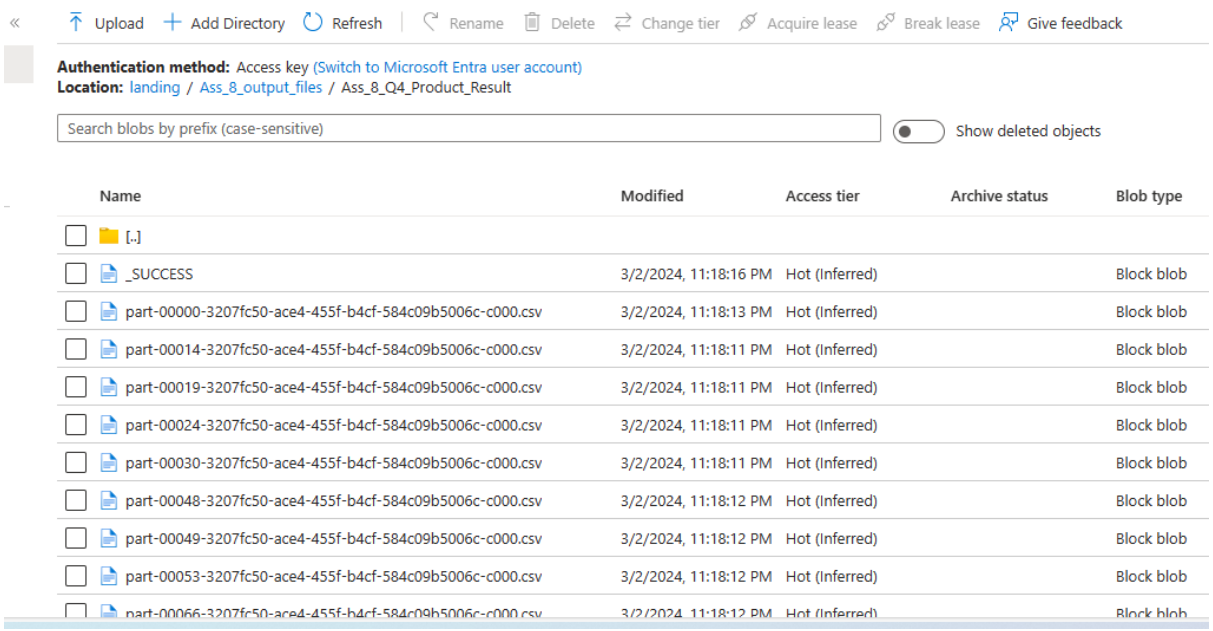

Set properties
Name
Ass_8_Q4_ProductCSV_Result
Linked service *
ADLS_Azure_DataLakeStorage1_Linked_service
File path
landing
Ass_8_output_files
File name
First row as header
Import schema
From connection/store
From sample file
None
Advanced
OK
Back
Cancel

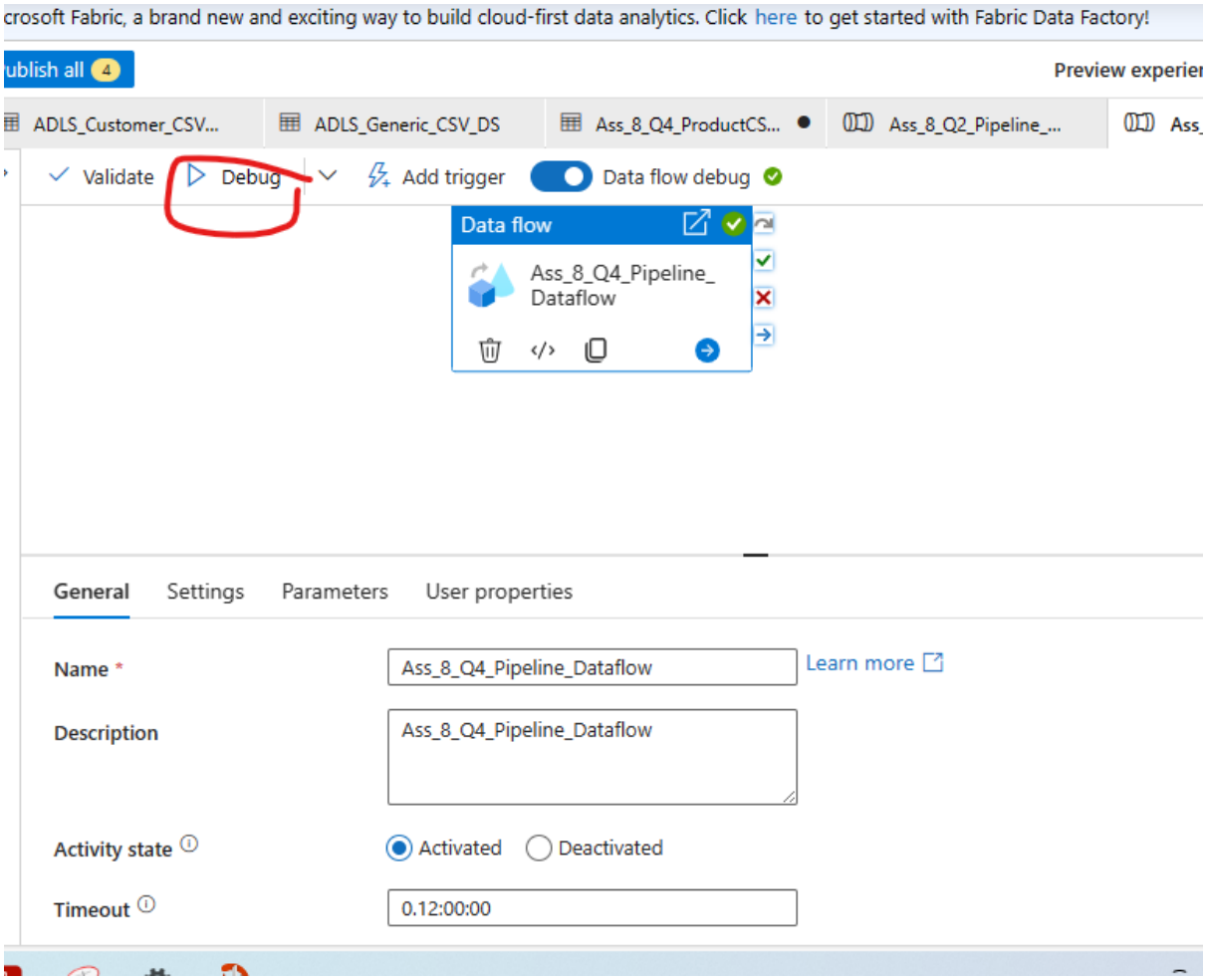

Preview experience
Off
ADLS_Generic_CSV_DS
Ass_8_Q4_ProductCS...
Ass_8_Q2_Pipeline_...
Ass_8_Q4_Pipeline...
Validate
Debug
Add trigger
Data flow debug
Data flow
Ass_8_Q4_Pipeline_
Dataflow
General
Settings
Parameters
1
User properties
Data flow *
Ass_8_Q4_DataFlow
Open
New
Run on (Azure IR) *
AutoResolveIntegrationRuntime
Compute size *
Small
Advanced
Logging level *
Verbose
Basic
None

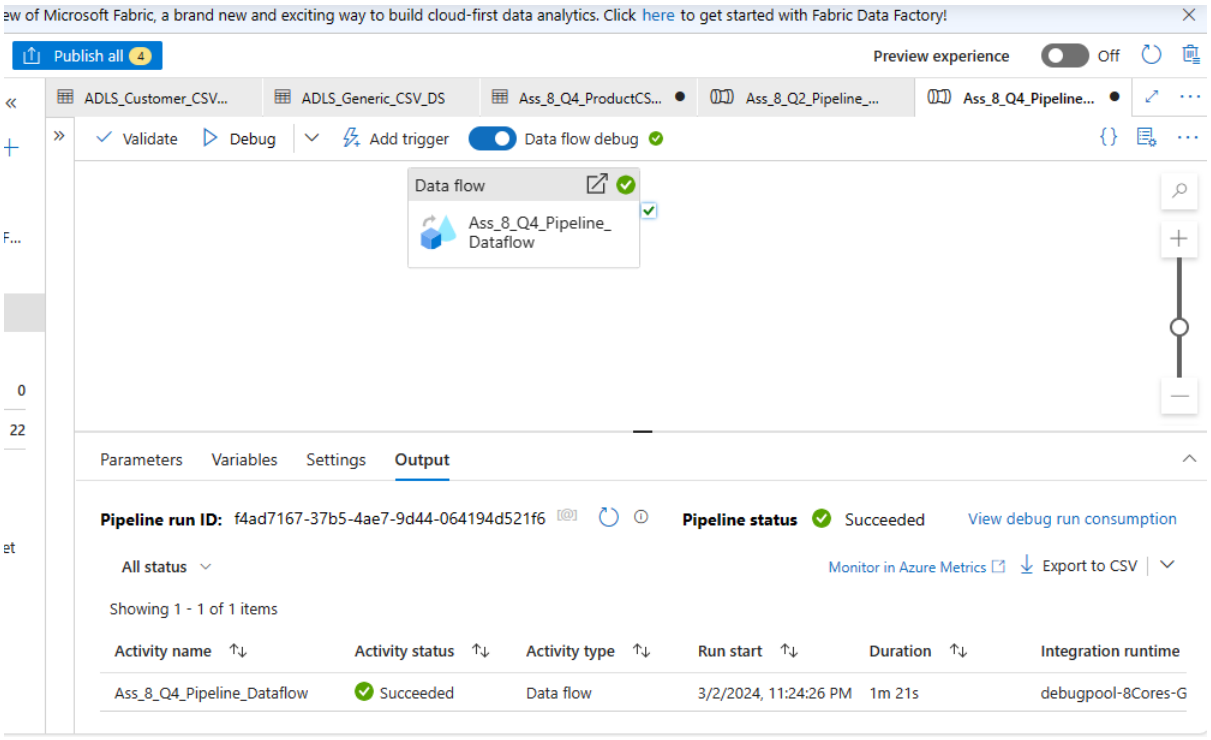

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Data Factory
Validate all
Publish all 4
Preview experience
Off
Factory Resources
Filter resources by name
Assignment_8_Ques_1
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_Q4_Pipeline_Dataflow
Without_Foreach_Example_Pipeline
Change Data Capture (preview)
Datasets
ADLS_Customer_CSV_Dataset
Activities
data
Move and transform
Copy data
Data flow
Azure Data Explorer
Azure Data Explorer C...
Databricks
Notebook
Jar
Python
Data Lake Analytics
U-SQL
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
1
User properties
Data flow *
Ass_8_Q4_DataFlow
Open
Run on (Azure IR) *
AutoResolveIntegrationRuntime
Compute size *
Small
Properties
General
Related
Name *
Ass_8_Q4_Pipeline_Dataflow
Description
Call Ass_8_Q4_Dataflow
create a pipeline to read the Product
Annotations
New

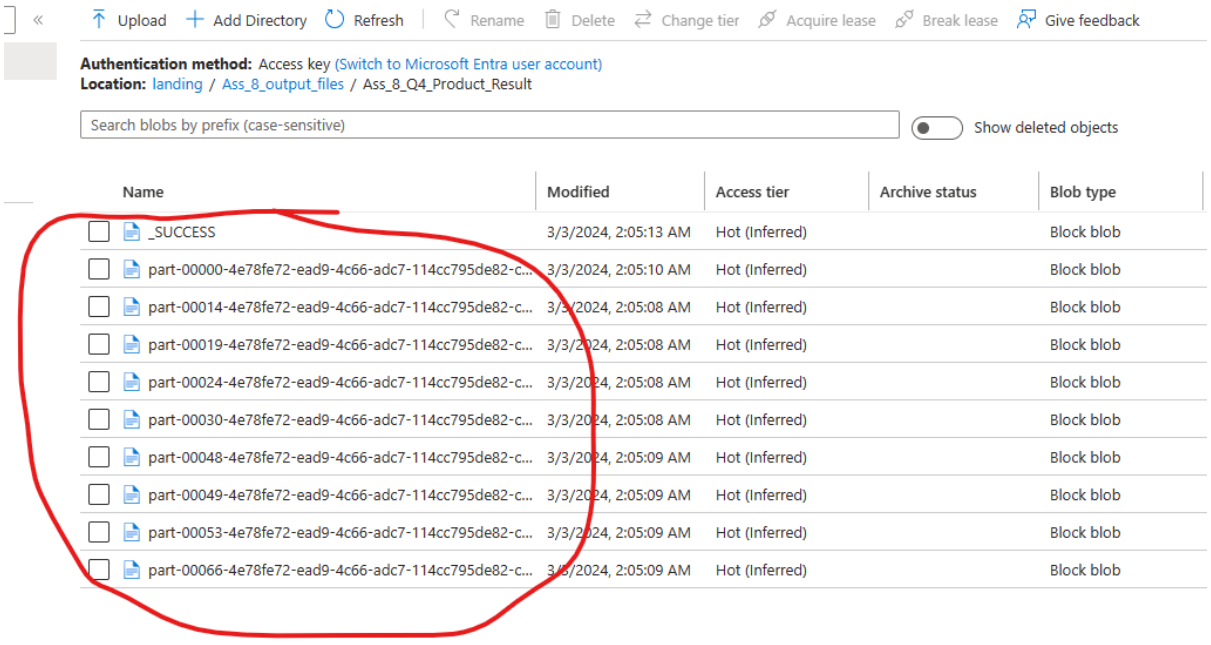

Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Preview experience
Off
ADLS_Generic_CSV_DS
Ass_8_Q4_ProductCS...
Ass_8_Q2_Pipeline_...
Ass_8_Q4_Pipeline...
Validate
Cancel options
Add trigger
Data flow debug
Data flow
Ass_8_Q4_Pipeline_ Dataflow
Parameters
Variables
Settings
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

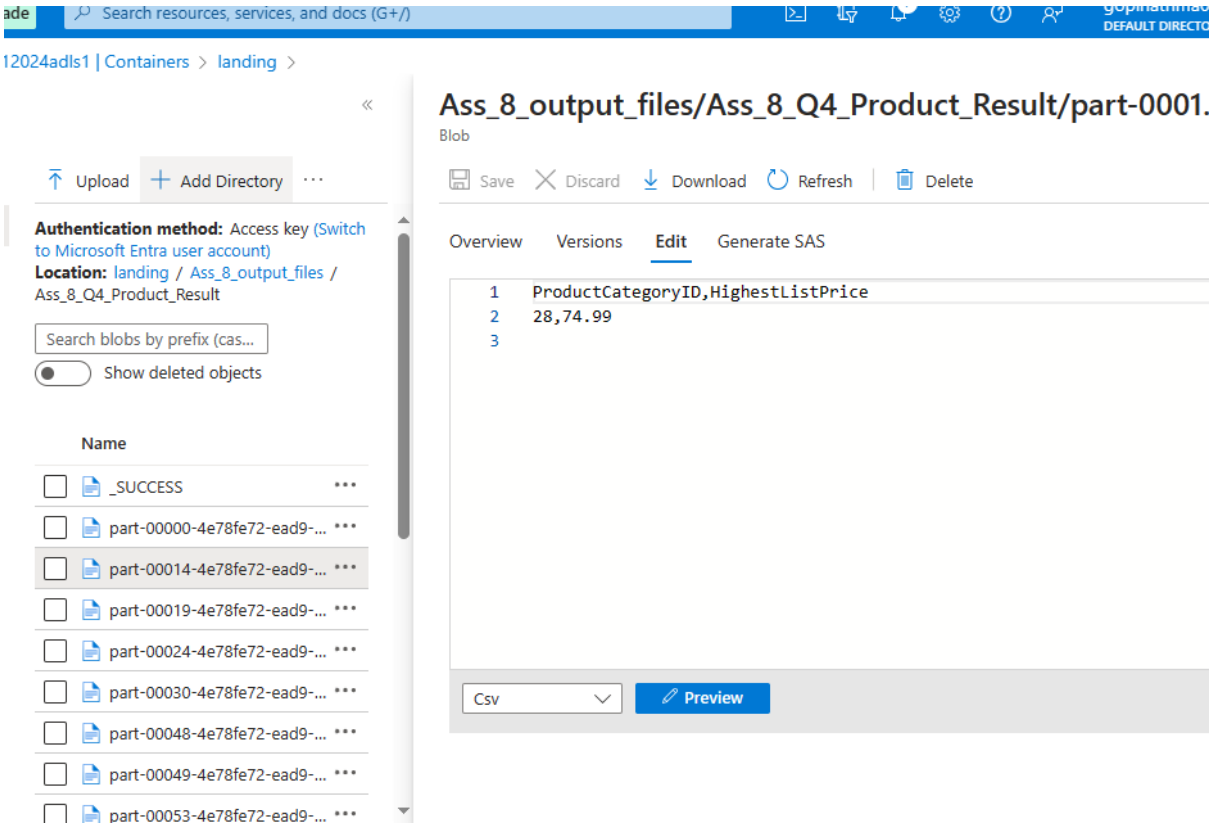

Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Storage accounts > gopi01012024adls1 | Containers > landing
landing
Container
Upload
Add Directory
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: landing / Ass_8_output_files / Ass_8_Q4_Product_Result
Search blobs by prefix (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...
Blob
Save
Discard
Download
Refresh
Delete
Overview
Versions
Edit
Generate SAS
{"CustomerID":29485,"NameStyle":false,"Title":"Ms.","FirstName":"Catherine","Midd
{"CustomerID":29486,"NameStyle":false,"Title":"Ms.","FirstName":"Kim","LastName":
{"CustomerID":29489,"NameStyle":false,"Title":"Ms.","FirstName":"Frances","Middle
{"CustomerID":29490,"NameStyle":false,"Title":"Ms.","FirstName":"Margaret","Middl
{"CustomerID":29492,"NameStyle":false,"Title":"Mr.","FirstName":"Jay","LastName":
{"CustomerID":29494,"NameStyle":false,"Title":"Mr.","FirstName":"Samuel","MiddleN
{"CustomerID":29496,"NameStyle":false,"Title":"Mr.","FirstName":"Robert","MiddleN
{"CustomerID":29497,"NameStyle":false,"Title":"Mr.","FirstName":"François","LastN
{"CustomerID":29499,"NameStyle":false,"Title":"Ms.","FirstName":"Amy","MiddleName
{"CustomerID":29502,"NameStyle":false,"Title":"Mr.","FirstName":"Paul","MiddleNam
{"CustomerID":29503,"NameStyle":false,"Title":"Mr.","FirstName":"Gregory","Middle
{"CustomerID":29503,"NameStyle":false,"Title":"Mr.","FirstName":"Gregory","Middle
{"CustomerID":29505,"NameStyle":false,"Title":"Ms.","FirstName":"Michelle","LastN
{"CustomerID":29506,"NameStyle":false,"Title":"Mr.","FirstName":"Sean","MiddleNam
{"CustomerID":29508,"NameStyle":false,"Title":"Mr.","FirstName":"Marvin","MiddleN
{"CustomerID":29510,"NameStyle":false,"Title":"Mr.","FirstName":"Cecil","MiddleNa
Json
Preview

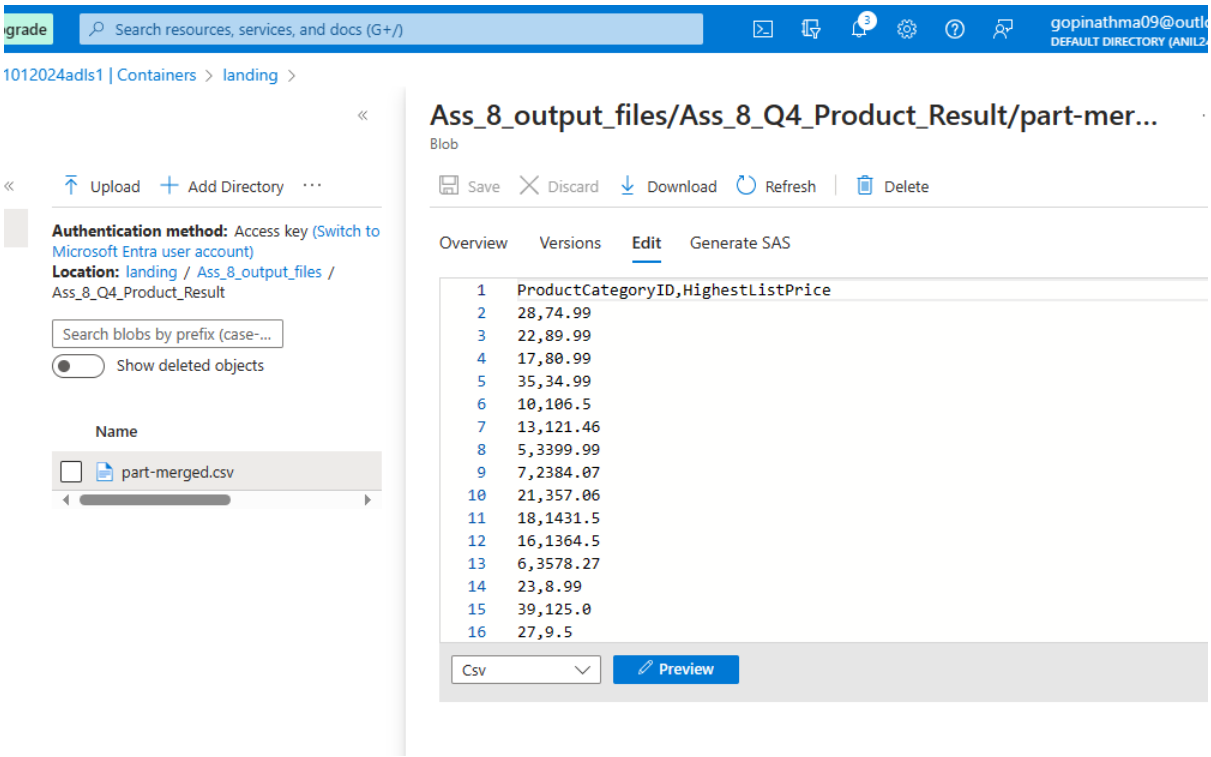

Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: landing / Ass_8_output_files / Ass_8_Q4_Product_Result
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
[..]
_SUCCESS 3/2/2024, 11:18:16 PM Hot (Inferred) Block blob
part-00000-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv 3/2/2024, 11:18:13 PM Hot (Inferred) Block blob
part-00014-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv 3/2/2024, 11:18:11 PM Hot (Inferred) Block blob
part-00019-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv 3/2/2024, 11:18:11 PM Hot (Inferred) Block blob
part-00024-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv 3/2/2024, 11:18:11 PM Hot (Inferred) Block blob
part-00030-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv 3/2/2024, 11:18:11 PM Hot (Inferred) Block blob
part-00048-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv 3/2/2024, 11:18:12 PM Hot (Inferred) Block blob
part-00049-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv 3/2/2024, 11:18:12 PM Hot (Inferred) Block blob
part-00053-3207fc50-ace4-455f-b4cf-584c09b5006c-c000.csv 3/2/2024, 11:18:12 PM Hot (Inferred) Block blob

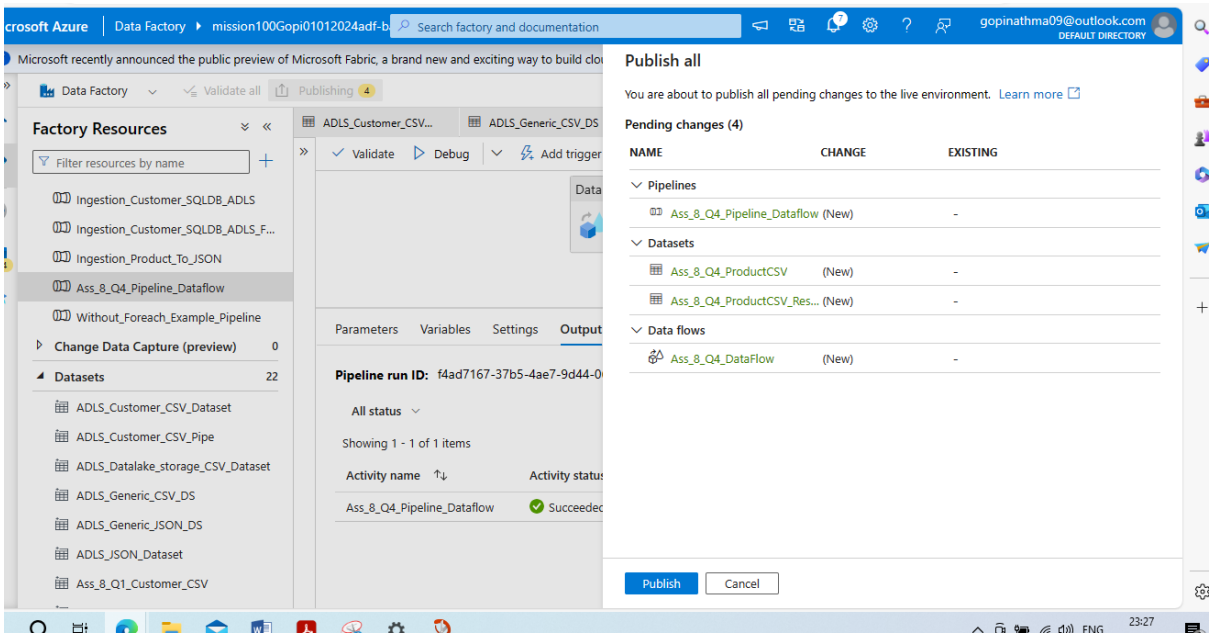

crosoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
ublish all 4
Preview experier
ADLS_Customer_CSV...
ADLS_Generic_CSV_DS
Ass_8_Q4_ProductCS...
Ass_8_Q2_Pipeline_...
Ass_
Validate
Debug
Add trigger
Data flow debug
Data 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

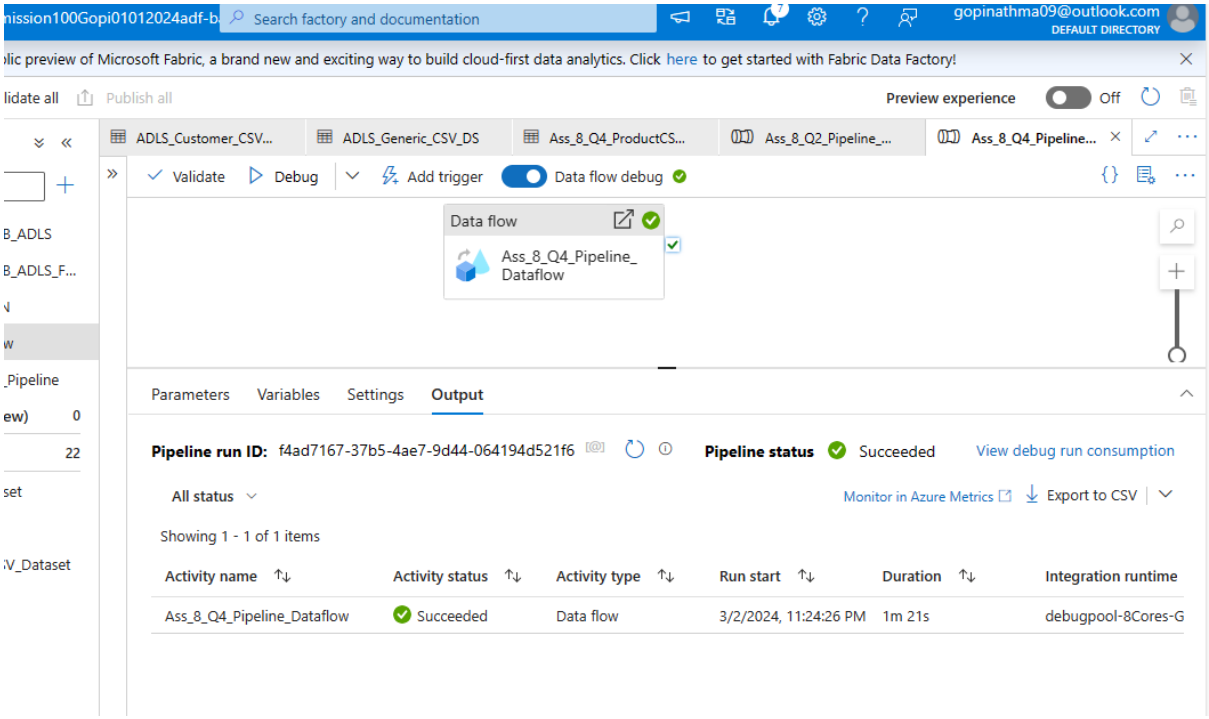

ew of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Publish all 4
Preview experience
Off
ADLS_Customer_CSV...
ADLS_Generic_CSV_DS
Ass_8_Q4_ProductCS...
Ass_8_Q2_Pipeline_...
Ass_8_Q4_Pipeline...
Validate
Debug
Add 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
_SUCCESS 3/3/2024, 2:05:13 AM Hot (Inferred) Block blob
part-00000-4e78fe72-ead9-4c66-adc7-114cc795de82-c... 3/3/2024, 2:05:10 AM Hot (Inferred) Block blob
part-00014-4e78fe72-ead9-4c66-adc7-114cc795de82-c... 3/3/2024, 2:05:08 AM Hot (Inferred) Block blob
part-00019-4e78fe72-ead9-4c66-adc7-114cc795de82-c... 3/3/2024, 2:05:08 AM Hot (Inferred) Block blob
part-00024-4e78fe72-ead9-4c66-adc7-114cc795de82-c... 3/3/2024, 2:05:08 AM Hot (Inferred) Block blob
part-00030-4e78fe72-ead9-4c66-adc7-114cc795de82-c... 3/3/2024, 2:05:08 AM Hot (Inferred) Block blob
part-00048-4e78fe72-ead9-4c66-adc7-114cc795de82-c... 3/3/2024, 2:05:09 AM Hot (Inferred) Block blob
part-00049-4e78fe72-ead9-4c66-adc7-114cc795de82-c... 3/3/2024, 2:05:09 AM Hot (Inferred) Block blob
part-00053-4e78fe72-ead9-4c66-adc7-114cc795de82-c... 3/3/2024, 2:05:09 AM Hot (Inferred) Block blob
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.
Blob
Upload
Add Directory
Save
Discard
Download
Refresh
Delete
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-00000-4e78fe72-ead9-...
part-00014-4e78fe72-ead9-...
part-00019-4e78fe72-ead9-...
part-00024-4e78fe72-ead9-...
part-00030-4e78fe72-ead9-...
part-00048-4e78fe72-ead9-...
part-00049-4e78fe72-ead9-...
part-00053-4e78fe72-ead9-...
Overview
Versions
Edit
Generate SAS
1 ProductCategoryID,HighestListPrice
2 28,74.99
3
Csv
Preview


Search resources, services, and docs (G+/)
gopinathma09@outl
DEFAULT DIRECTORY (ANIL2
1012024adls1 | Containers > landing >
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-...
Show deleted objects
Name
part-merged.csv
Ass_8_output_files/Ass_8_Q4_Product_Result/part-mer...
Blob
Save
Discard
Download
Refresh
Delete
Overview
Versions
Edit
Generate SAS
1 ProductCategoryID,HighestListPrice
2 28,74.99
3 22,89.99
4 17,80.99
5 35,34.99
6 10,106.5
7 13,121.46
8 5,3399.99
9 7,2384.07
10 21,357.06
11 18,1431.5
12 16,1364.5
13 6,3578.27
14 23,8.99
15 39,125.0
16 27,9.5
Csv
Preview


Microsoft Azure
Data Factory
Search factory and documentation
gopinathma09@outlook.com
DEFAULT DIRECTORY
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build clo
Factory Resources
Filter resources by name
Ingestion_Customer_SQLDB_ADLS
Ingestion_Customer_SQLDB_ADLS_F...
Ingestion_Product_To_JSON
Ass_8_Q4_Pipeline_Dataflow
Without_Foreach_Example_Pipeline
Change Data Capture (preview)
Datasets
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
Validate
Debug
Add trigger
Parameters
Variables
Settings
Output
Pipeline run ID:
All status
Showing 1 - 1 of 1 items
Activity name
Ass_8_Q4_Pipeline_Dataflow
Publish all
You are about to publish all pending changes to the live environment. Learn more
Pending changes (4)
NAME
CHANGE
EXISTING
Pipelines
Ass_8_Q4_Pipeline_Dataflow (New)
Datasets
Ass_8_Q4_ProductCSV (New)
Ass_8_Q4_ProductCSV_Res... (New)
Data flows
Ass_8_Q4_DataFlow (New)
Publish
Cancel



Question 5:

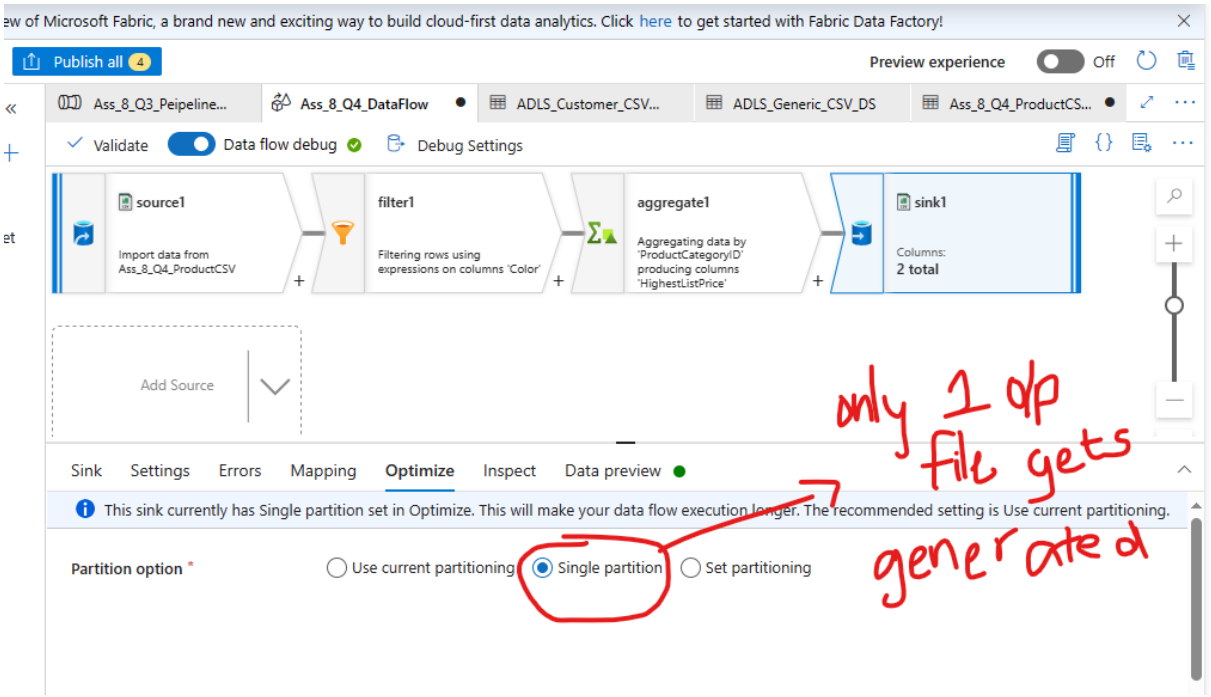
create a pipeline to read the Product CSV file, and calculate the highest listPrice of any product under each productcategory.
Ensure that product shouldn't be of blue in color and save the result as a SINGLE CSV file inside ProductSingleResult folder.

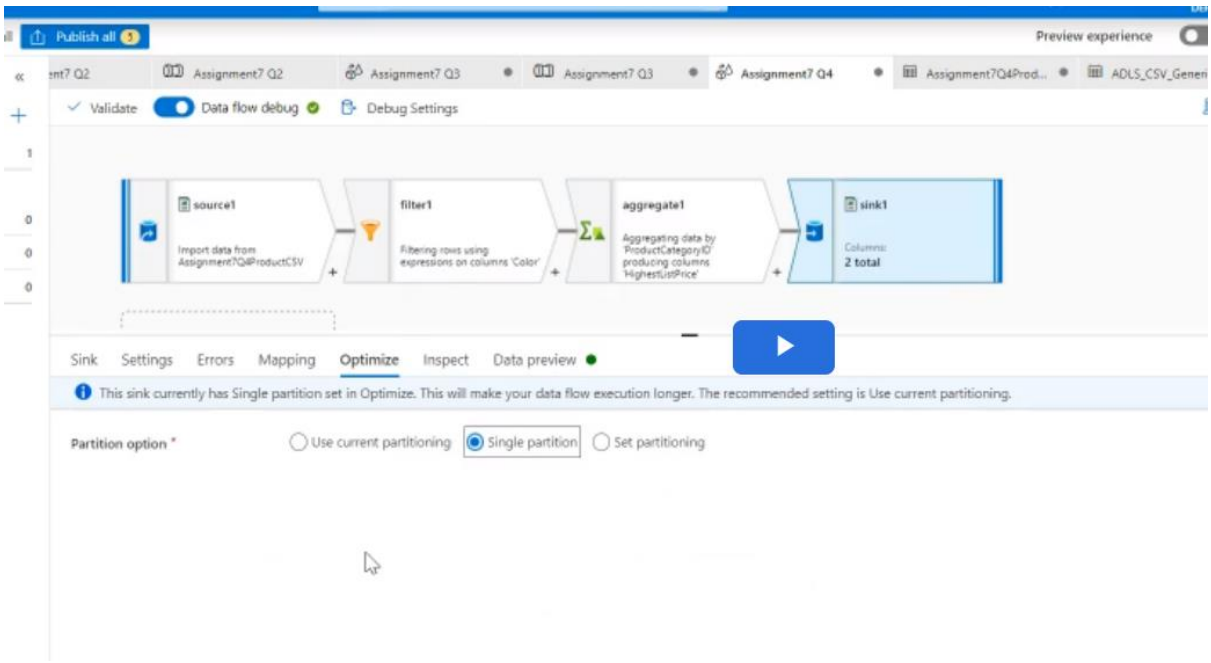
Same Solution as Question 4:

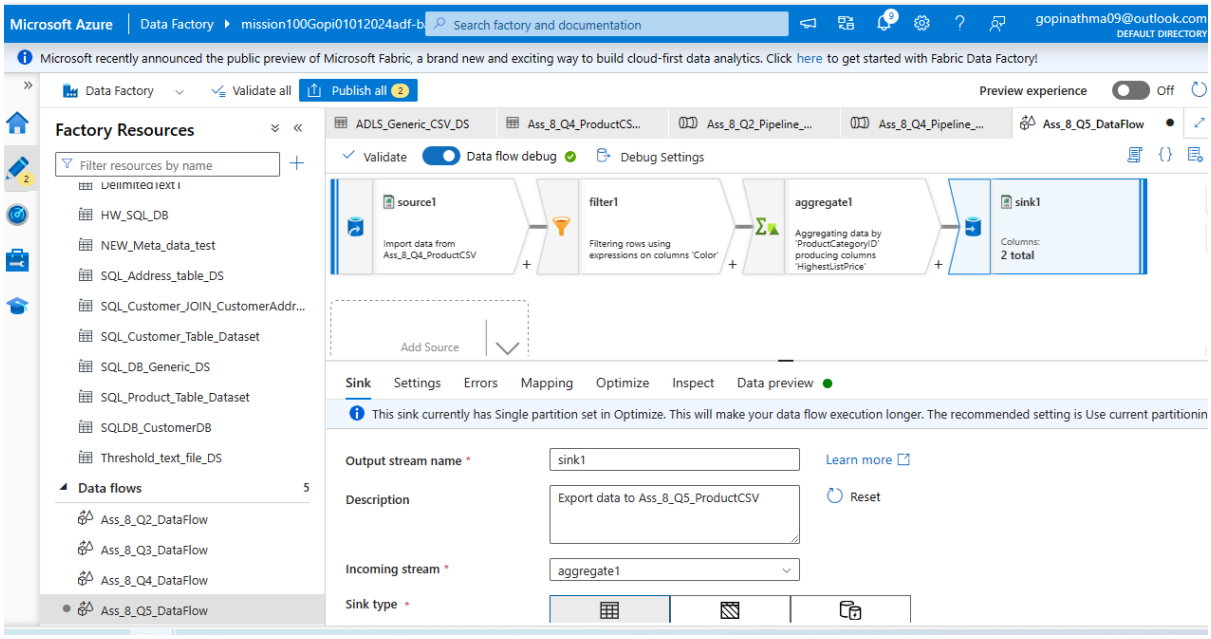
Only difference is in the Sink "Optimize " tab instead of "use current partitioning"

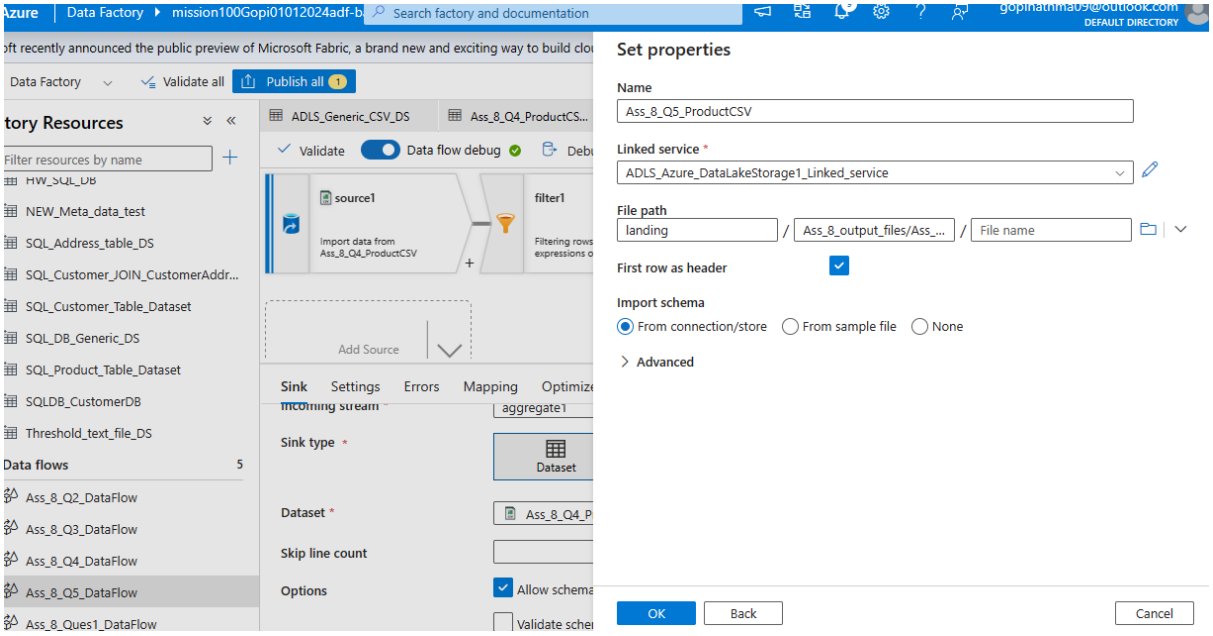
Select "Single partition" that's it == All output will be saved in SINGLE CSV file

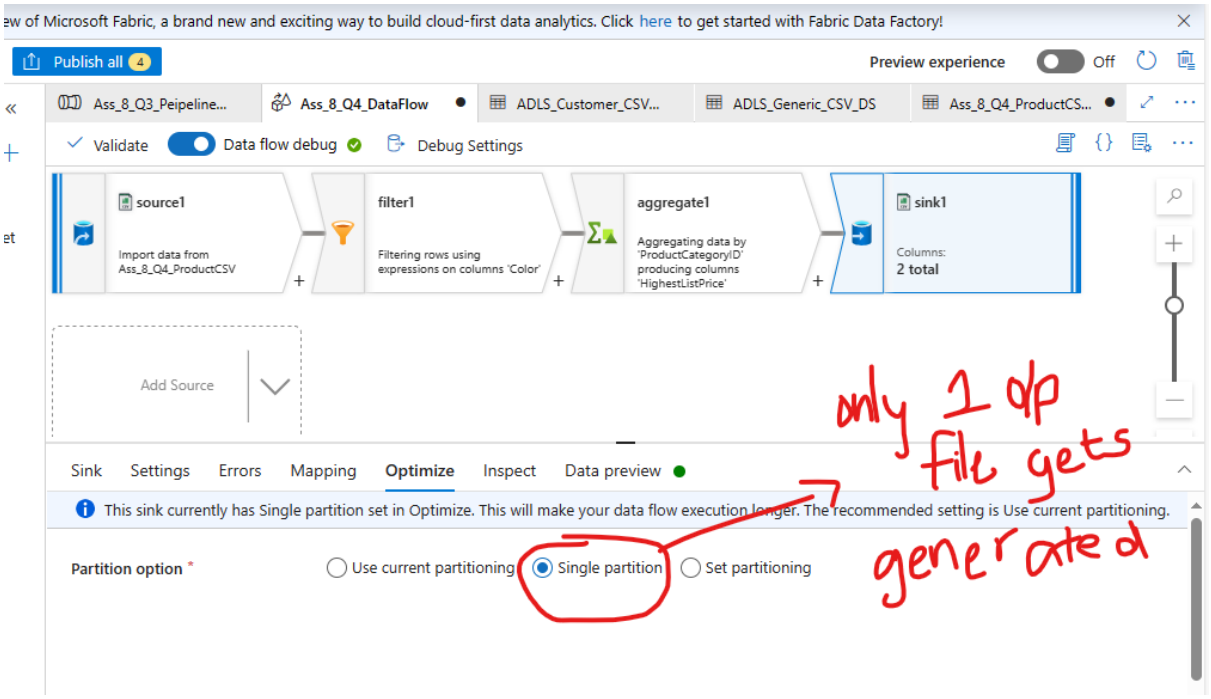
Earlier it was multi partitioning (so multiple processing and multipleoutput csv files) – Now only single CSV file

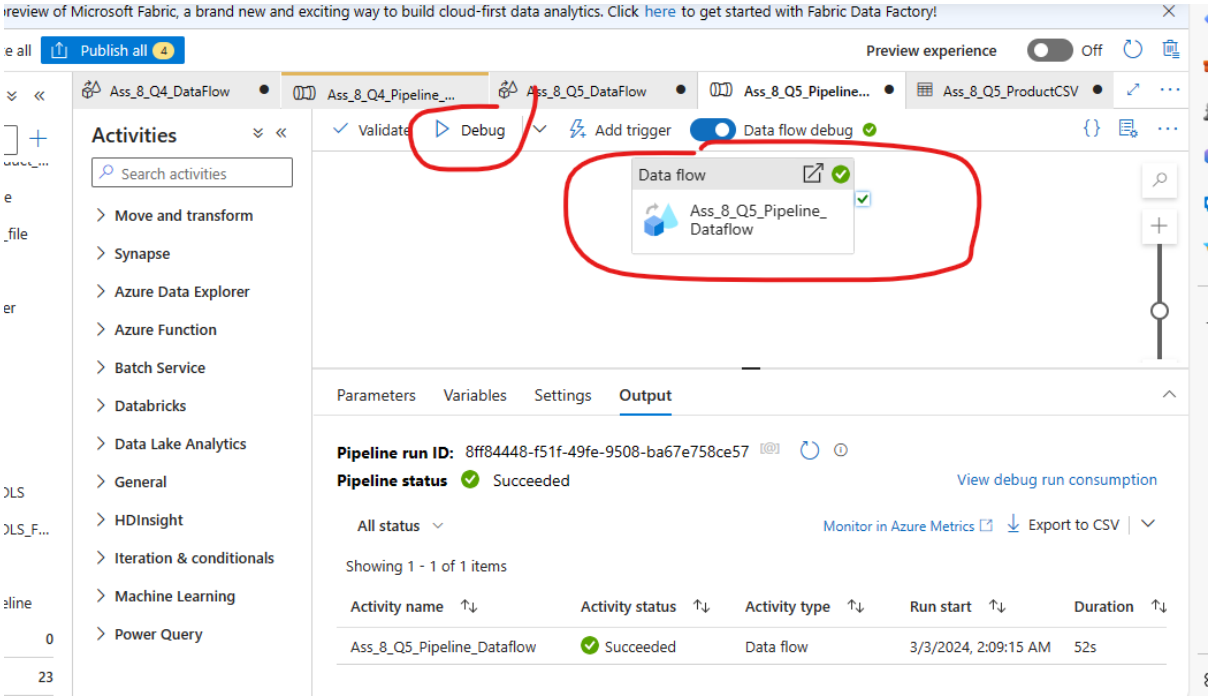

ew of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Publish all 4
Preview experience
Off
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
ADLS_Customer_CSV...
ADLS_Generic_CSV_DS
Ass_8_Q4_ProductCS...
Validate
Data flow debug
Debug Settings
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
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 op file gets generated

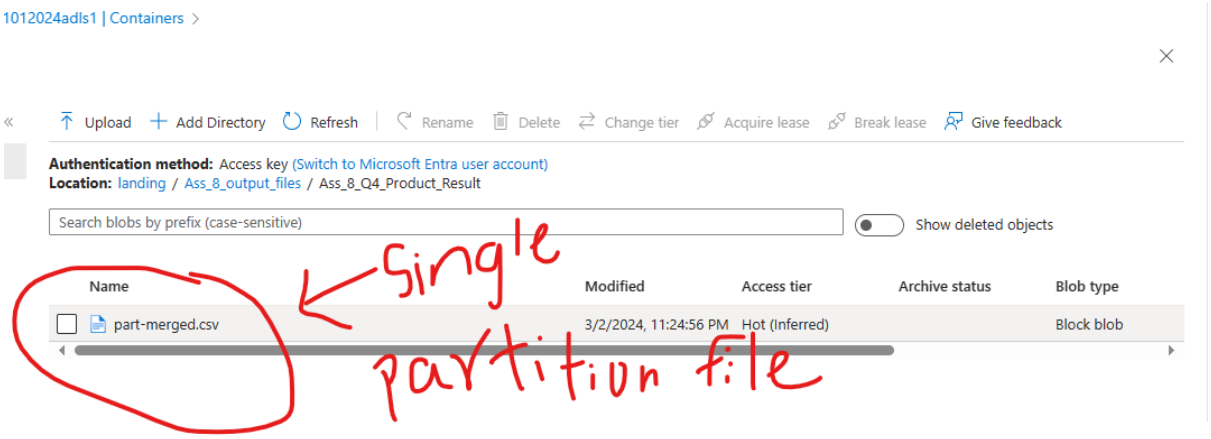

Publish all 3
Preview experience
Assignment7 Q2
Assignment7 Q3
Assignment7 Q4
Validate
Data flow debug
Debug Settings
source1
filter1
aggregate1
sink1
2 total
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
This sink currently has Single partition set in Optimize. This will make your data flow execution longer. The recommended setting is Use current partitioning.
Partition option *
Use current partitioning
Single partition
Set partitioning

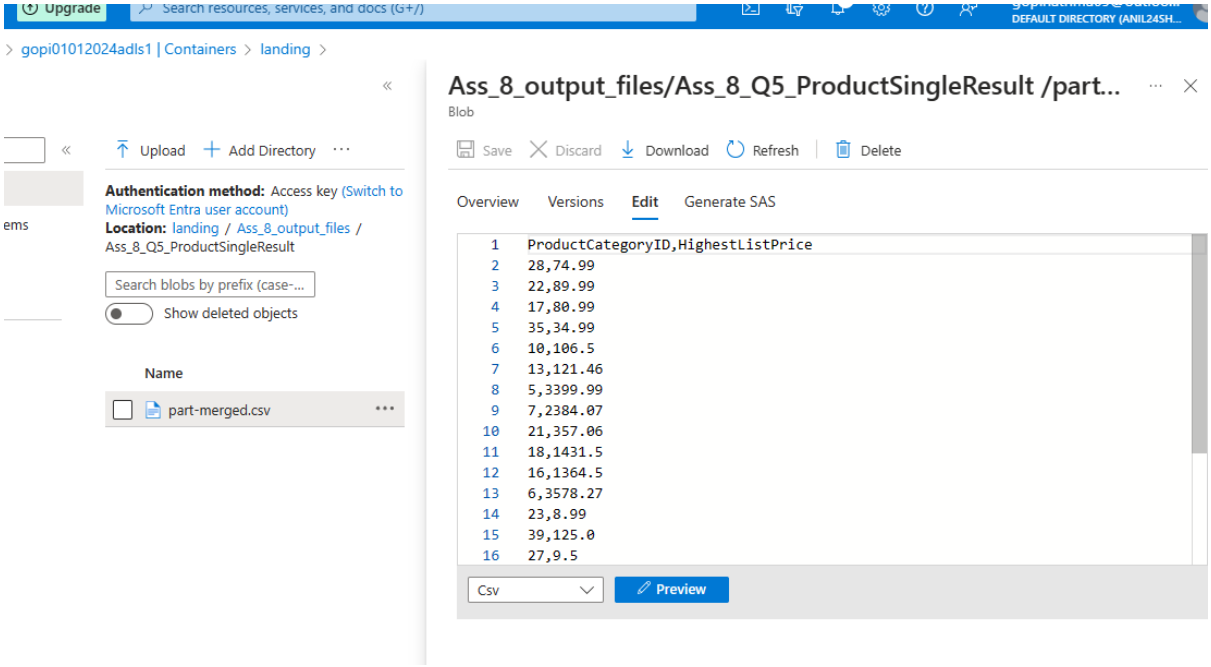

Microsoft Azure
Data Factory
mission100Gopi01012024adf-b
Search factory and documentation
gopinathma09@outlook.com
DEFAULT DIRECTORY
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Data Factory
Validate all
Publish all 2
Preview experience
Off
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
ADLS_Generic_CSV_DS
Ass_8_Q4_ProductCS...
Ass_8_Q2_Pipeline_...
Ass_8_Q4_Pipeline_...
Ass_8_Q5_DataFlow
Validate
Data flow debug
Debug Settings
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
This sink currently has Single partition set in Optimize. This will make your data flow execution longer. The recommended setting is Use current partitionin
Output stream name *
sink1
Learn more
Description
Export data to Ass_8_Q5_ProductCSV
Reset
Incoming stream *
aggregate1
Sink type *

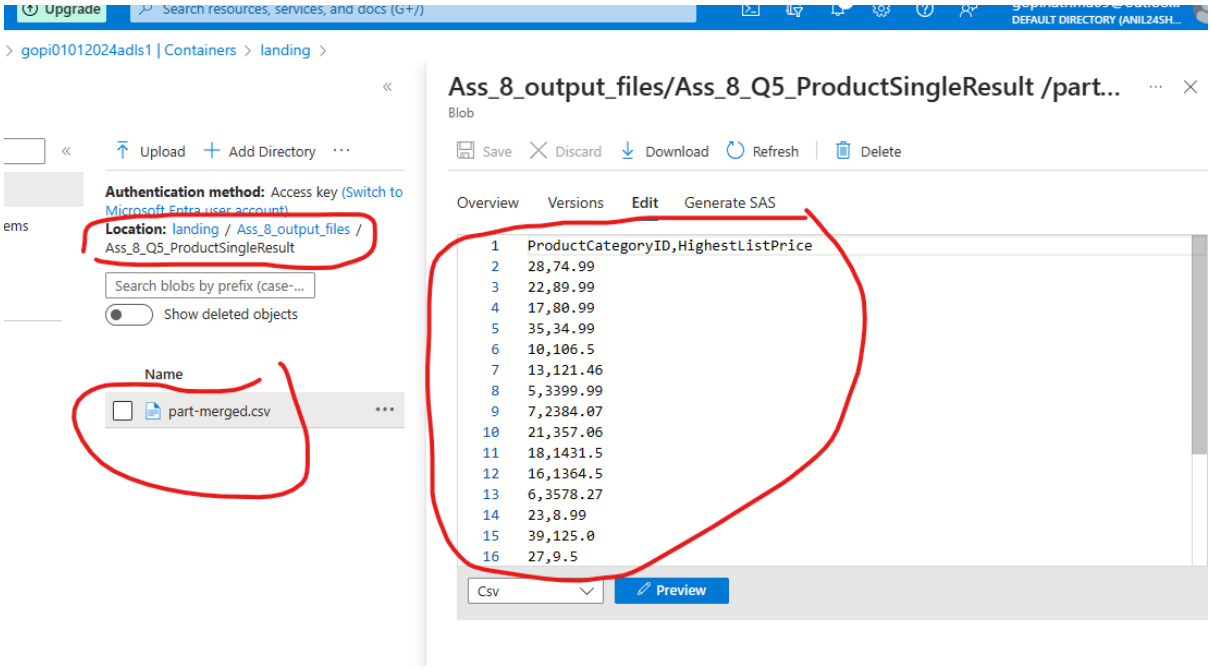

Set properties
Name
Ass_8_Q5_ProductCSV
Linked service *
ADLS_Azure_DataLakeStorage1_Linked_service
File path
landing
Ass_8_output_files/Ass_...
File name
First row as header
Import schema
From connection/store
From sample file
None
Advanced
OK
Back
Cancel
Sink type *
Dataset
Dataset *
Skip line count
Options
Allow schema
Validate schei


ew 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 4
Preview experience
Off
Ass_8_Q3_Peipeline...
Ass_8_Q4_DataFlow
ADLS_Customer_CSV...
ADLS_Generic_CSV_DS
Ass_8_Q4_ProductCS...
Validate
Data flow debug
Debug Settings
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
only 1 dp file gets generated
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
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


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!
Publish all 4
Preview experience
Off
Ass_8_Q4_DataFlow
Ass_8_Q4_Pipeline_...
Ass_8_Q5_DataFlow
Ass_8_Q5_Pipeline...
Ass_8_Q5_ProductCSV
Activities
Search activities
Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query
Validate
Debug
Add trigger
Data flow debug
Data flow
Ass_8_Q5_Pipeline_ Dataflow
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


1012024adls1 | Containers
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
part-merged.csv
3/2/2024, 11:24:56 PM
Hot (Inferred)
Block blob
Single partition file


Upgrade
Search resources, services, and docs (G+/)
gopi01012024adls1 | Containers
landing
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-...
Show deleted objects
Name
part-merged.csv
Ass_8_output_files/Ass_8_Q5_ProductSingleResult /part...
Blob
Save
Discard
Download
Refresh
Delete
Overview
Versions
Edit
Generate SAS
1 ProductCategoryID,HighestListPrice
2 28,74.99
3 22,89.99
4 17,80.99
5 35,34.99
6 10,106.5
7 13,121.46
8 5,3399.99
9 7,2384.07
10 21,357.06
11 18,1431.5
12 16,1364.5
13 6,3578.27
14 23,8.99
15 39,125.0
16 27,9.5
Csv
Preview



Finally Publish all