

ETL - extract, transform, load

New runtime to perform transformations

manage --> Integration runtime --> new --> Azure self hosted --> Azure --> name,region(auto resolve) --> create

for Real time - self hosted integration runtime is mostly used.

Create new pipeline - employee_blob_to_adls

create dataset and linked service

copy activity

2. create a new copy activity to transfer data from blob to cosmodb

datasets/retail/lineitem.csv --> target --> cosmodb table/storage account --> table

open source file --> check column delimiter, SCHEMA --> CLEAR --> import schema

dataflow - perform actual etl process

NEW Dataflow

Add source

select dataset

nation key --> integer in projection

+ --> filter --> filter on --> open expression builder

+ --> sink --> add ADLS --> browse --> tgt/2023/nation_filter.csv

activity --> data flow

name change as per data flow

add trigger --> new --> type -> schedule

publish

add trigger

monitor --> pipeline runs

data flow --> data flow debug --> sink --> data preview

get metadata size

in array

@array()--> filter items ==> @array(activity('Get Metadata1').output.size)]

condition

variables -

size integer 200

set variables

value -- get metadata size

control flow - copy data --> blob - source --> adls - sink

get metadata --> file --> size, column, child items

Trigger --> schedule, tumbling window, storage, event

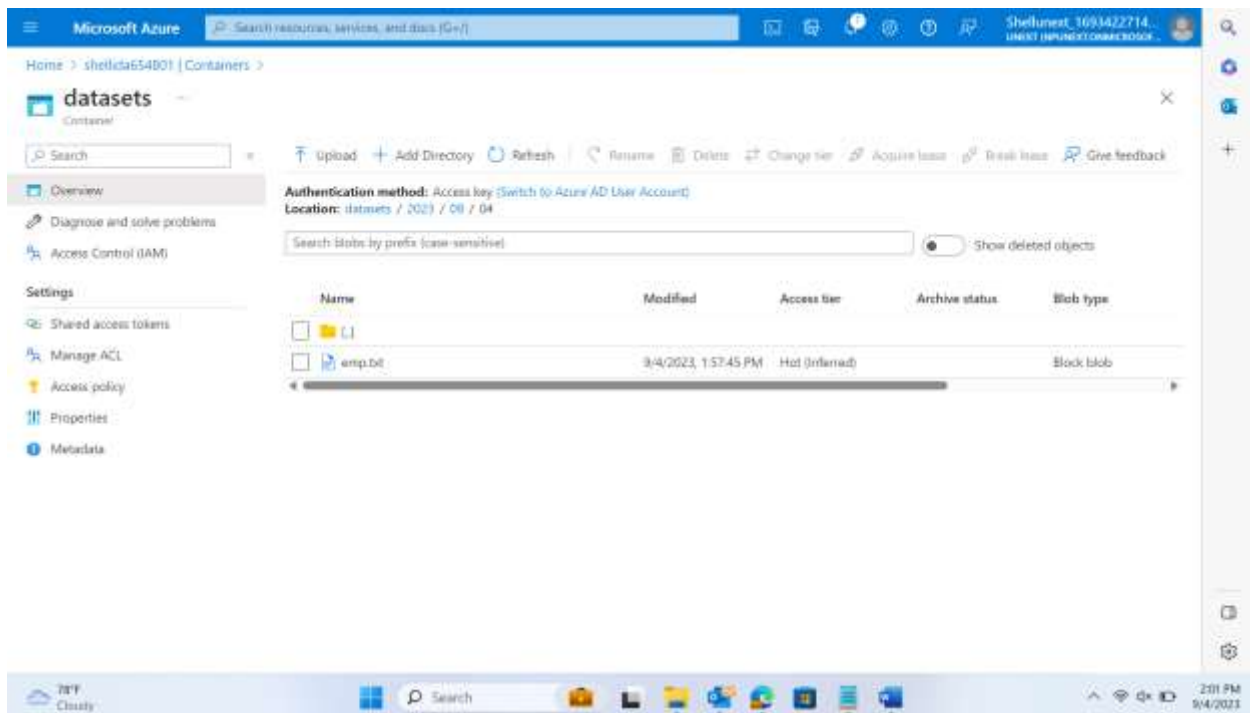
Monitor --> debug

resource group

blob storage

key vault --> secrets --> blob key

adls storage



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

Home > shellsta654801blob | Containers >

datasets

Container

Search

Upload | Change access level | Refresh | Delete | Change file | Acquire lease | Break lease | View snapshots

Authentication method: Access key (Switch to Azure AD User Account)
Location: datasets

Search blobs by prefix (case-sensitive)

Show deleted blobs

Add filter

Name	Modified	Access tier	Archive status	Blob type
emp.txt	8/4/2023, 12:24:44 PM	Hot (Inferred)		Block blob

https://portal.azure.com/#

78°F Cloudy

Search

2:01 PM 8/4/2023

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

Home > staticwebsitehost654801_1693817449962 | Overview > staticwebsitehost654801

staticwebsitehost654801 | Static website

Storage account

Search

Save | Discard | Give feedback

Overview | Activity log | Tags | Diagnose and solve problems | Access Control (IAM) | Data migration | Events | Storage browser | Storage Mover

Data storage

Containers | File shares | Queues | Tables

Security + networking

Enabling static websites on the blob service allows you to host static content. Webpages may include static content and client-side scripts. Server-side scripting is not supported. As data is replicated asynchronously from primary to secondary regions, files at the secondary endpoint may not be immediately available or in sync with files at the primary endpoint. [Learn more](#)

Static website

Disabled **Enabled**

An Azure Storage container has been created to host your static website. [View](#)

Primary endpoint [🔗](#)

<https://staticwebsitehost654801.z13.web.core.windows.net/>

Secondary endpoint [🔗](#)

<https://staticwebsitehost654801-secondary.z13.web.core.windows.net/>

Index document name [🔗](#)

[index.html](#)

Error document path [🔗](#)

[error.html](#)

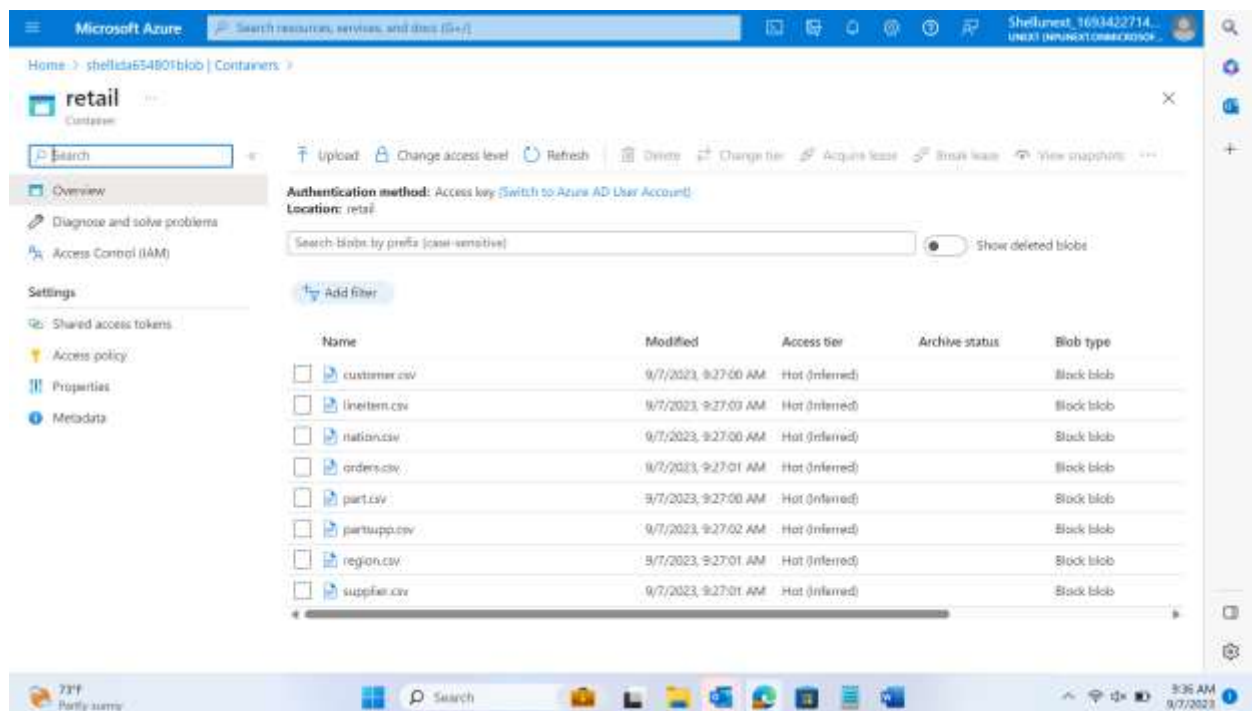
78°F Cloudy

Search

2:31 PM 8/4/2023



Blob storage



Key vault

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the 'Microsoft Azure' logo, a search bar, and user information. The main content area is titled 'shell801kv | Secrets'. On the left, a sidebar lists various Azure services, with 'Secrets' highlighted. The main panel displays a table of secrets. A notification at the top states: 'The secret blobRetailisk has been successfully created.' The table has columns for Name, Type, Status, and Expiration date. One secret, 'blobRetailisk', is listed with a status of 'Enabled'.

Name	Type	Status	Expiration date
blobRetailisk		✓ Enabled	

Adls storage

The screenshot shows the Microsoft Azure portal interface for an ADLS storage account named 'shellida654801adls'. The top navigation bar includes the 'Microsoft Azure' logo, a search bar, and user information. The main content area is titled 'shellida654801adls | Overview'. On the left, a sidebar lists various Azure services, with 'Storage browser' highlighted. The main panel displays the 'Essentials' section, which includes a table of account details. Below this, the 'Properties' section is expanded, showing 'Data Lake Storage' and 'Security' settings.

Property	Value
Resource group	shellida654801-rg
Location	East US
Subscription	msunest-1693422714
Subscription ID	a451199f-8896-42M-a175-69f12a316916
Disk state	Available
Performance	Standard
Replication	Locally-redundant storage (LRS)
Account kind	StorageV2 (general purpose v2)
Provisioning state	Succeeded
Created	9/7/2023, 9:28:17 AM

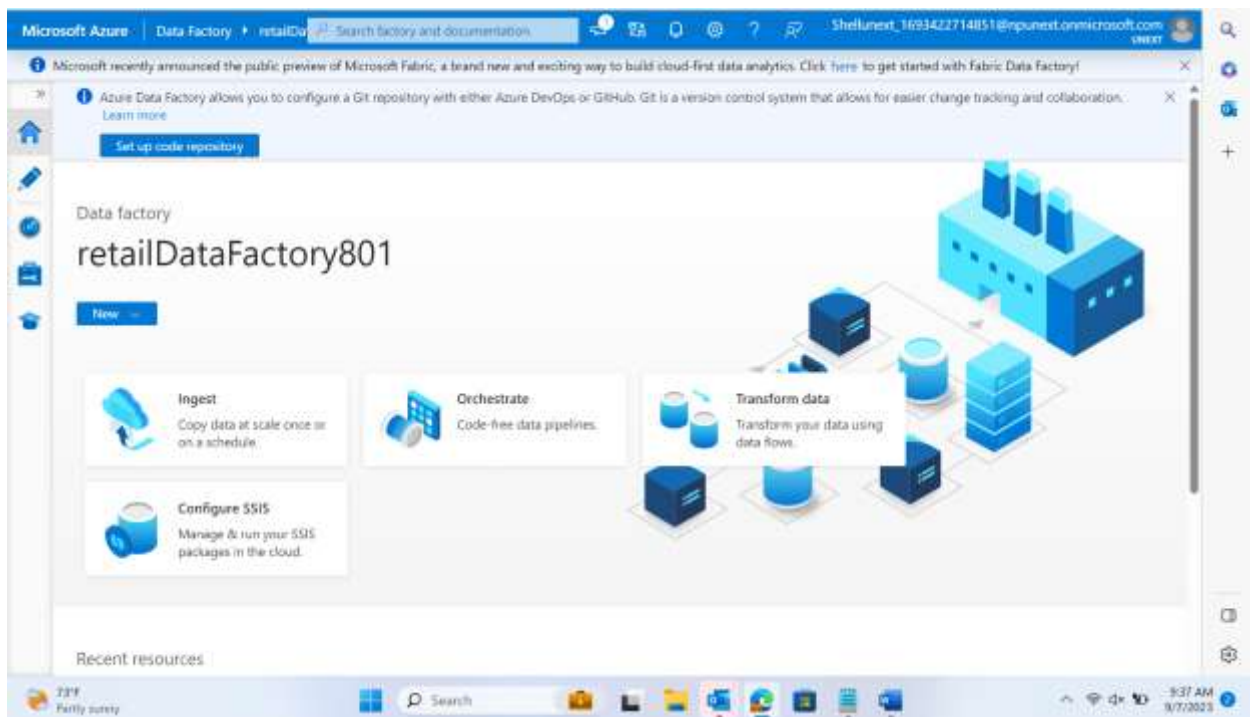
Data Lake Storage

Property	Value
Hierarchical namespace	Enabled
Default access tier	Hot
Blob anonymous access	Disabled
Blob soft delete	Enabled (7 days)
Container soft delete	Enabled (7 days)
Versioning	Disabled

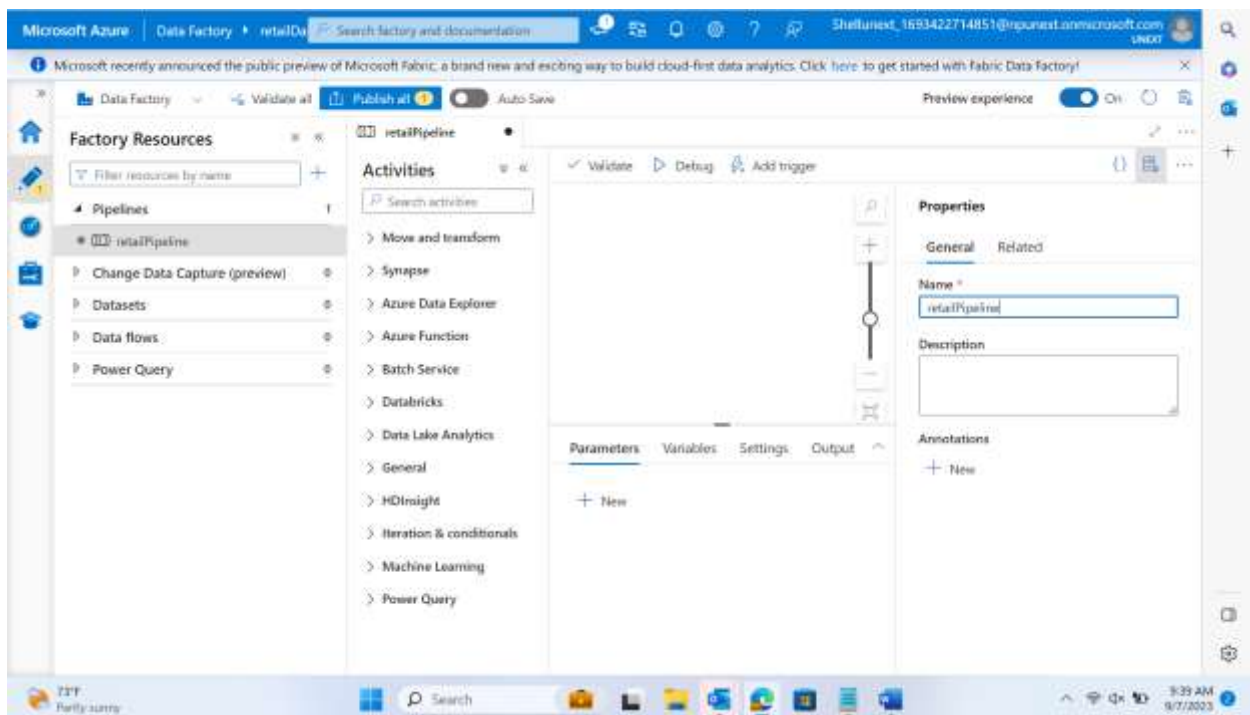
Security

Property	Value
Require secure transfer for REST API operations	Enabled
Storage account key access	Enabled
Minimum TLS version	Version 1.2
Infrastructure encryption	Disabled

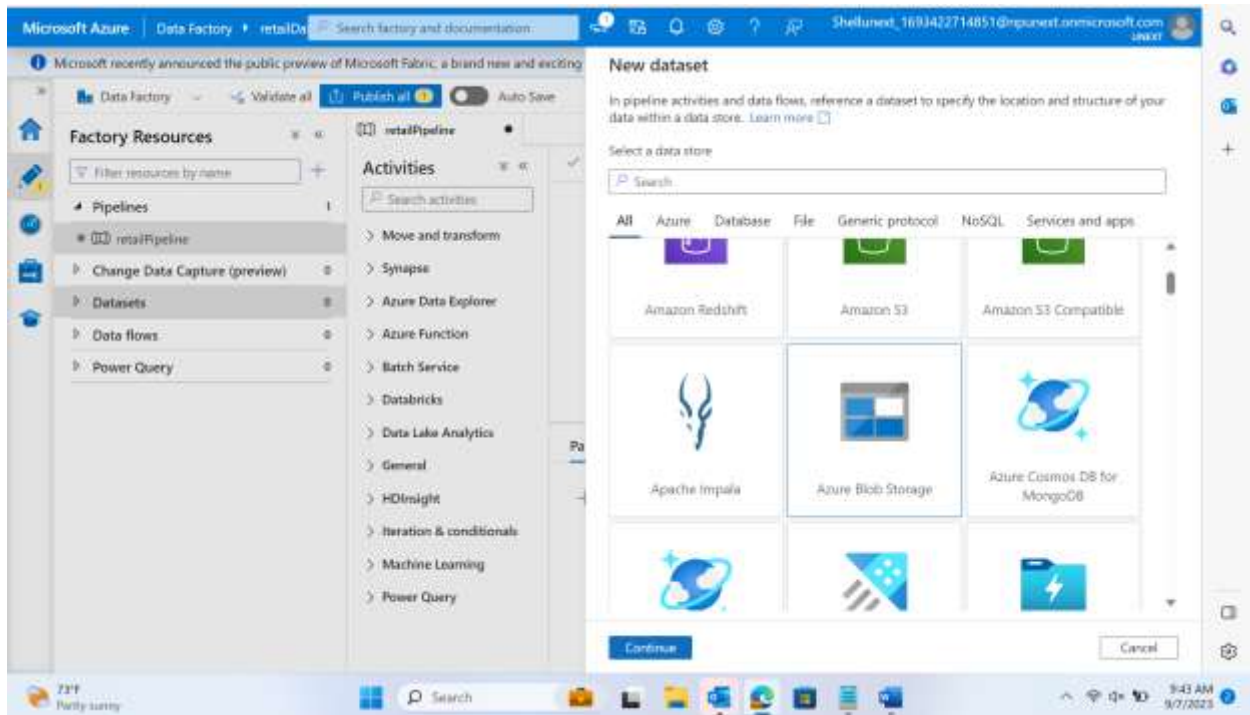
Data factory



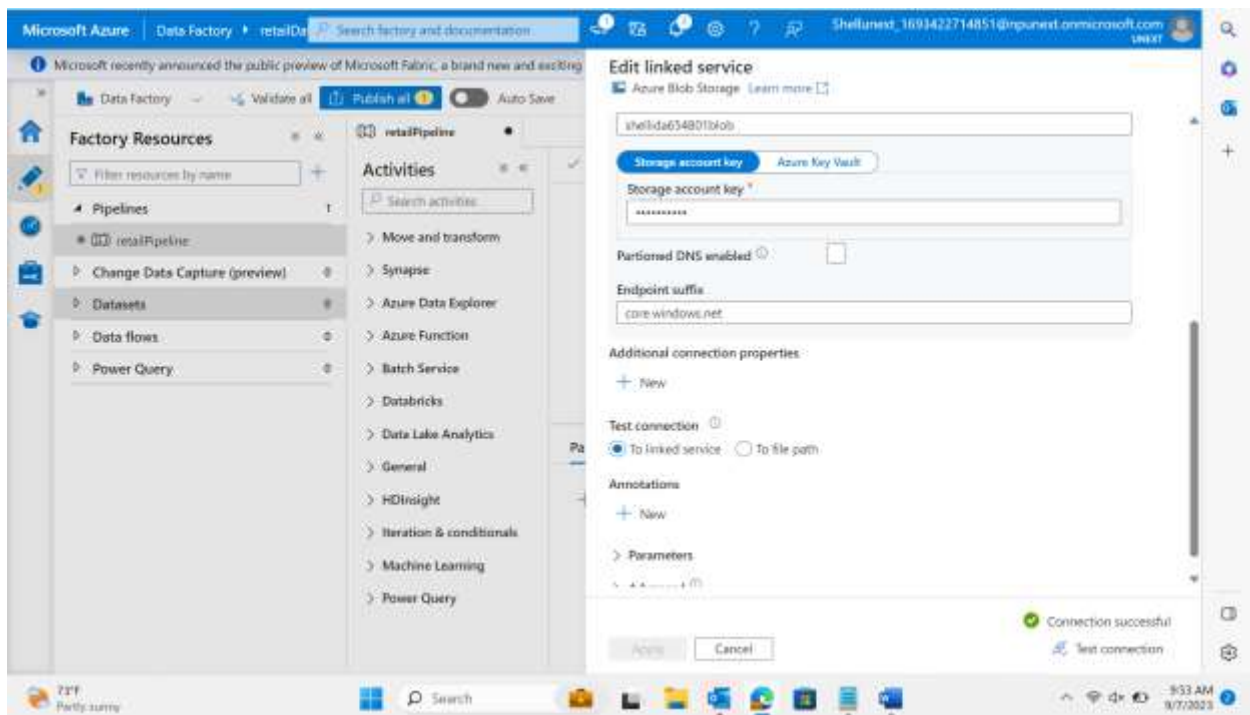
Pipeline

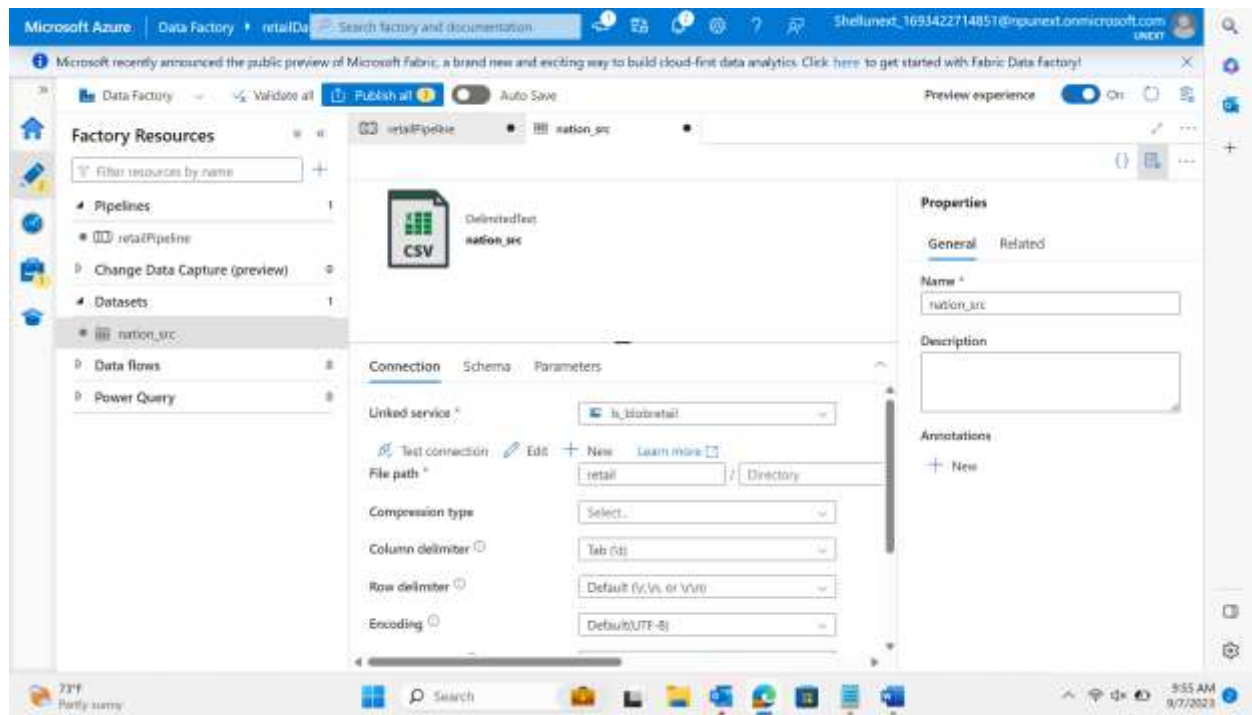


Dataset



LS



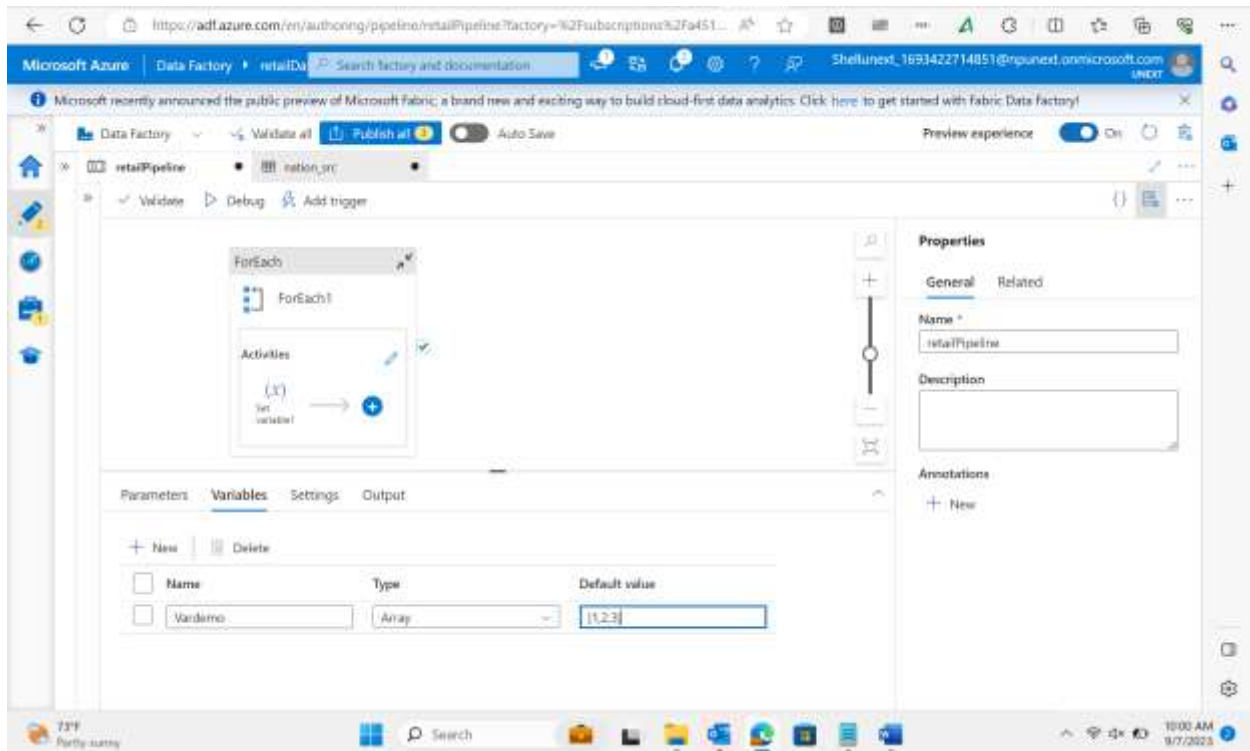


For each activity

Activity + set variable

Click outside

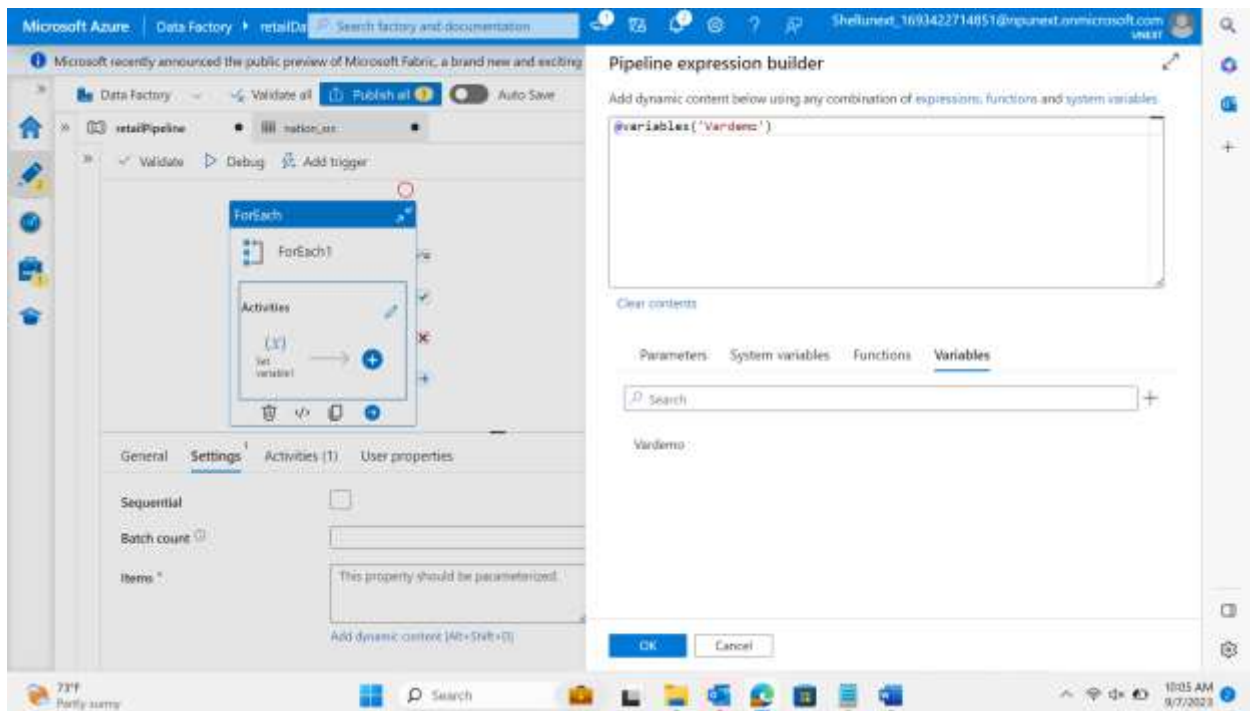
Variables → Vardemo, Array, [1,2,3]



Parameters → dynamic content -value -runtime

Variables → pipeline → append variable

Double click on foreach → settings → items → add dynamic content



Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunext_1693422714851@ngunext.onmicrosoft.com

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 | Auto Save | Preview experience | On

retailPipeline | nation_src

Validate | Debug | Add trigger

ForEach

ForEach1

Activities

(X) Set variable1

General | Settings | Activities (1) | User properties

Case	Activity
ForEach	(X) Set variable1 1 Activity

Properties

General | Related

Name *
retailPipeline

Description

Annotations
+ New

73°F
Partly sunny

Search

10:06 AM
9/7/2023

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunext_1693422714851@ngunext.onmicrosoft.com

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 | Auto Save | Preview experience | On

retailPipeline | nation_src

Validate | Debug | Add trigger

retailPipeline > ForEach1

(X) Set variable1

General | Settings | User properties

Variable type ☒ Pipeline variable ☐ Pipeline return value

Name * + New

Properties

General | Related

Name *
retailPipeline

Description

Annotations
+ New

73°F
Partly sunny

Search

10:07 AM
9/7/2023

Microsoft Azure | Data Factory | retailPipeline | Search factory and documentation

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

retailPipeline | nation_scs | Validate | Publish all | Auto Save

retailPipeline >ForEach1

(X) Set variable1

Variable type: Pipeline variable (selected) | Pipeline return value

Name: Vardemo

Value: @array(item())

Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables.

@array(item())

Clear contents

ForEach iterator | Parameters | System variables | Functions | Variables

Search

ForEach1
Current item

OK Cancel

72°F Partly sunny 10:12 AM 9/7/2023

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!

retailPipeline | nation_scs | Validate | Publish all | Auto Save

Preview experience: On

ForEach

ForEach1

Activities

(X) Set variable1

Parameters | Variables | Settings | Output

Showing 1 - 4 of 4 items

Activity name	Activity status	Run start	Duration	Integration runtime
Set variable1	Succeeded	9/7/2023, 10:10:56 AM	Less than 1s	
Set variable1	Succeeded	9/7/2023, 10:10:55 AM	Less than 1s	
Set variable1	Succeeded	9/7/2023, 10:10:55 AM	Less than 1s	
ForEach1	Succeeded	9/7/2023, 10:10:55 AM	3s	

Properties

General | Related

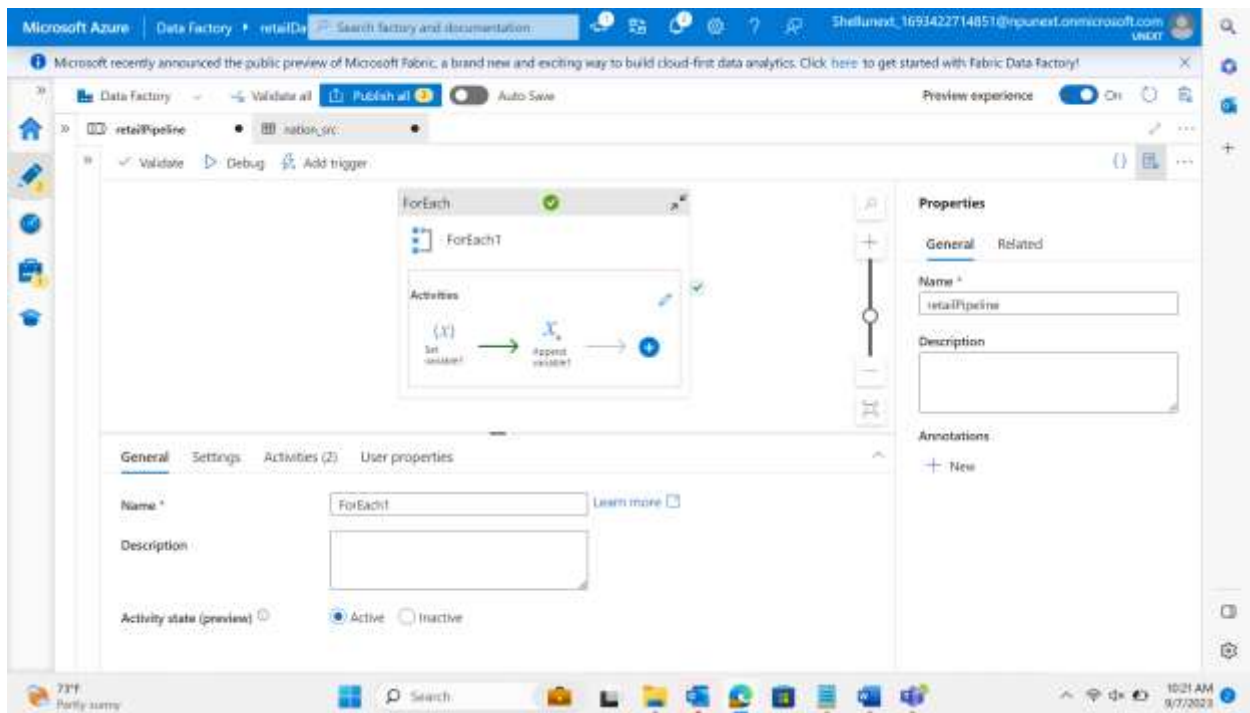
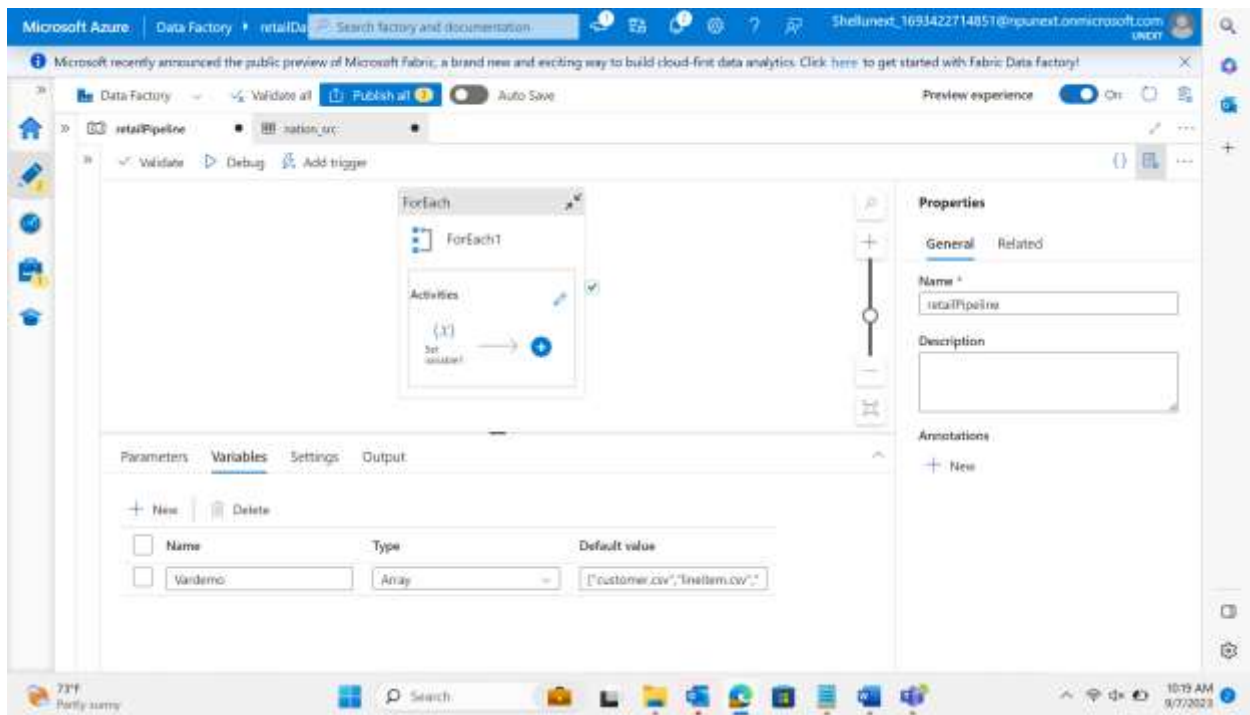
Name: retailPipeline

Description

Annotations

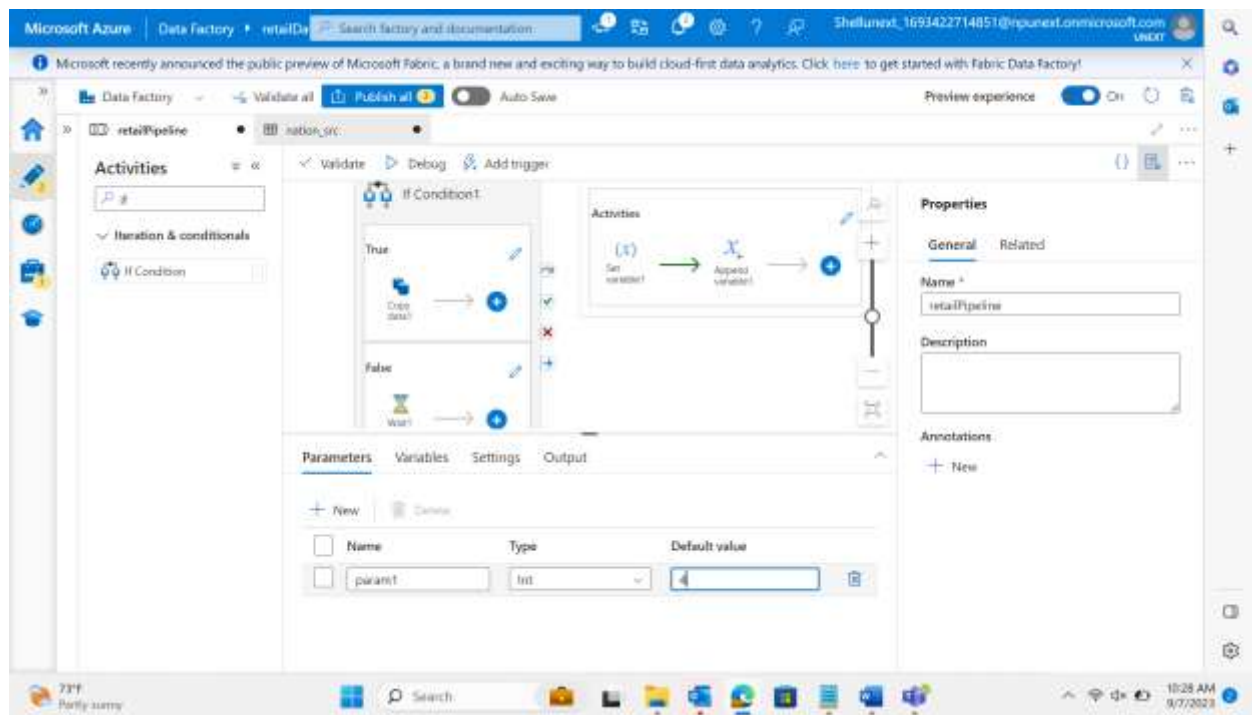
+ New

72°F Partly sunny 10:11 AM 9/7/2023



If condition

Pipeline parameter



Lubricant retail sales data :

Data Flow -> transformation

Line item --> column →

Df1 →

Calculate something which is not present in source table

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunet.1691422714051@rgueast.onmicrosoft.com | UNEXT

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!

Factory Resources

- Pipelines: 1
 - retailPipeline
- Change Data Capture (preview): 0
- Datasets: 2
 - initem_src
 - nation_src
- Data flows: 1
 - dataflow1
- Power Query: 0

retailPipeline | nation_src | dataflow1 | initem_src

Validate | Data flow debug

source1
Columns: 16 total

Add Source

Properties

General | Related

Name: dataflow1

Description:

Source settings | Source options | Projection

Source settings	Source options	Projection
L_ORDERKEY	str: string	Specify format
L_PARTKEY	str: string	Specify format
L_SUPPKEY	str: string	Specify format
L_LINENUMBER	str: string	Specify format
L_QUANTITY	int: integer	Specify format
L_EXTENDEDPRICE	str: string	Specify format
L_DISCOUNT	dt: double	Specify format

73°F Partly sunny | 11:10 AM 9/7/2023

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunet.1691422714051@rgueast.onmicrosoft.com | UNEXT

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!

Factory Resources

- Pipelines: 1
 - retailPipeline
- Change Data Capture (preview): 0
- Datasets: 2
 - initem_src
 - nation_src
- Data flows: 1
 - dataflow1
- Power Query: 0

retailPipeline | nation_src | dataflow1 | initem_src

Validate | Data flow debug

source1
Columns: 16 total

Add Source

Schema modifier
Derived Column

Properties

General | Related

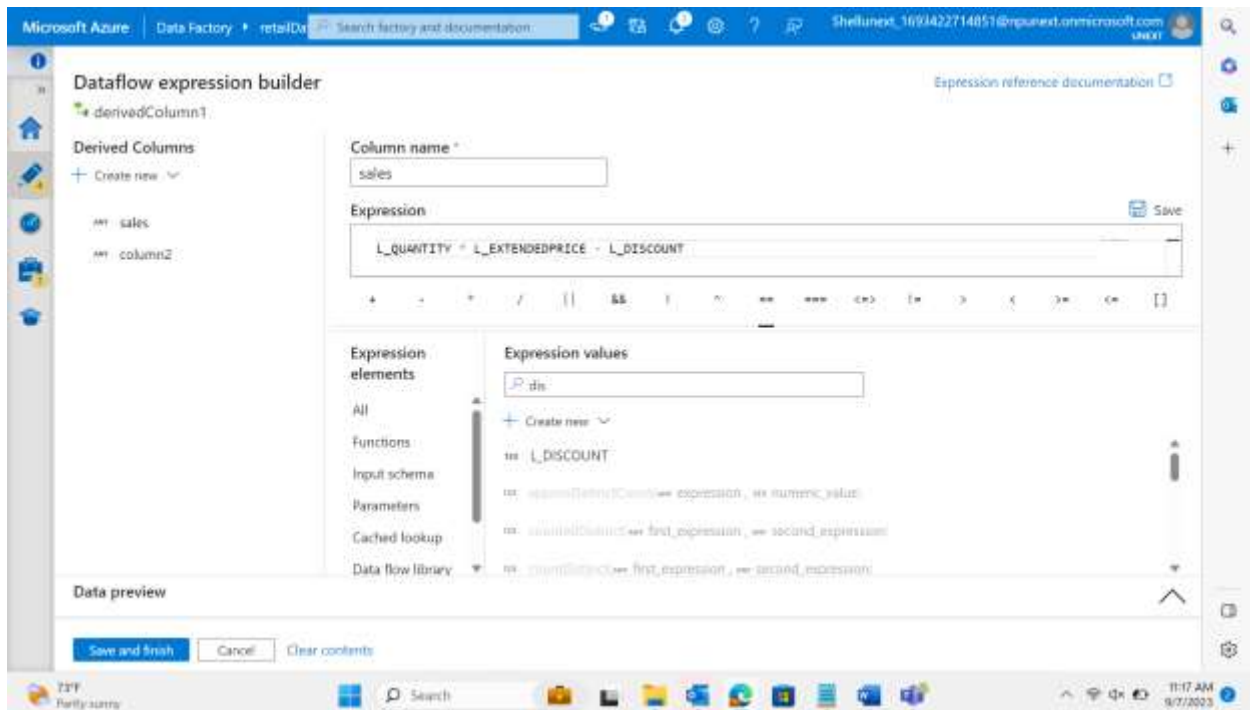
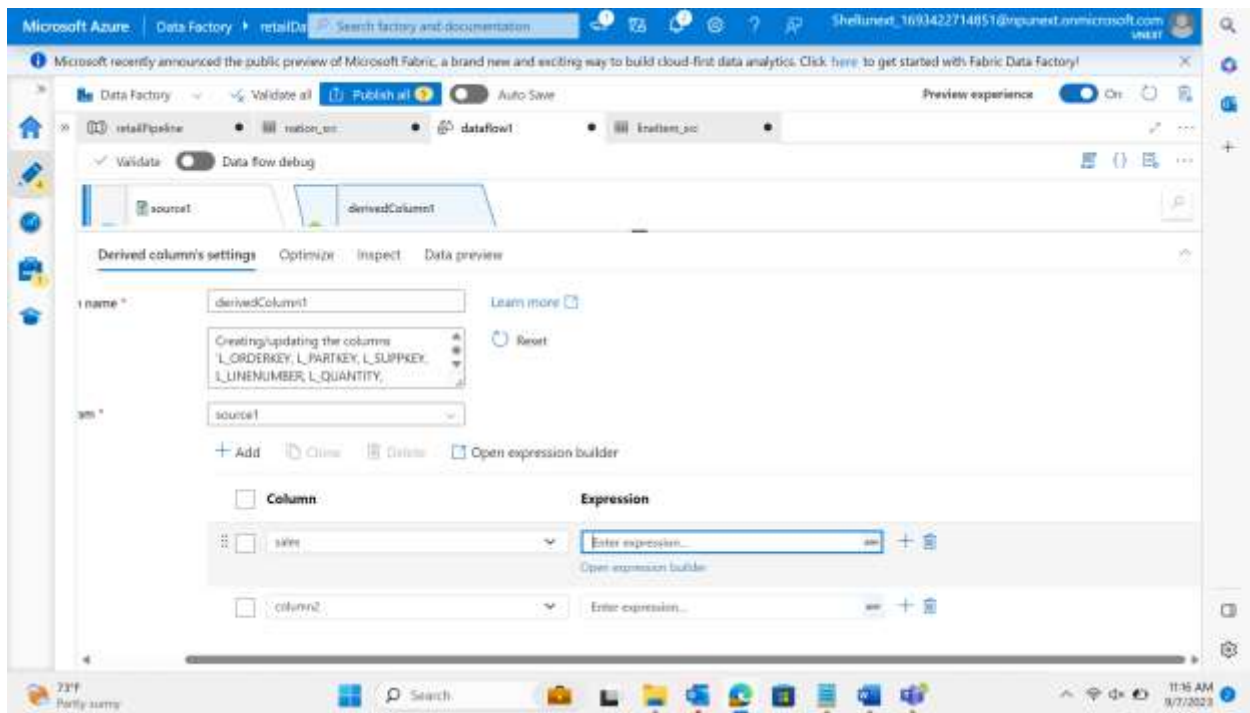
Name: dataflow1

Description:

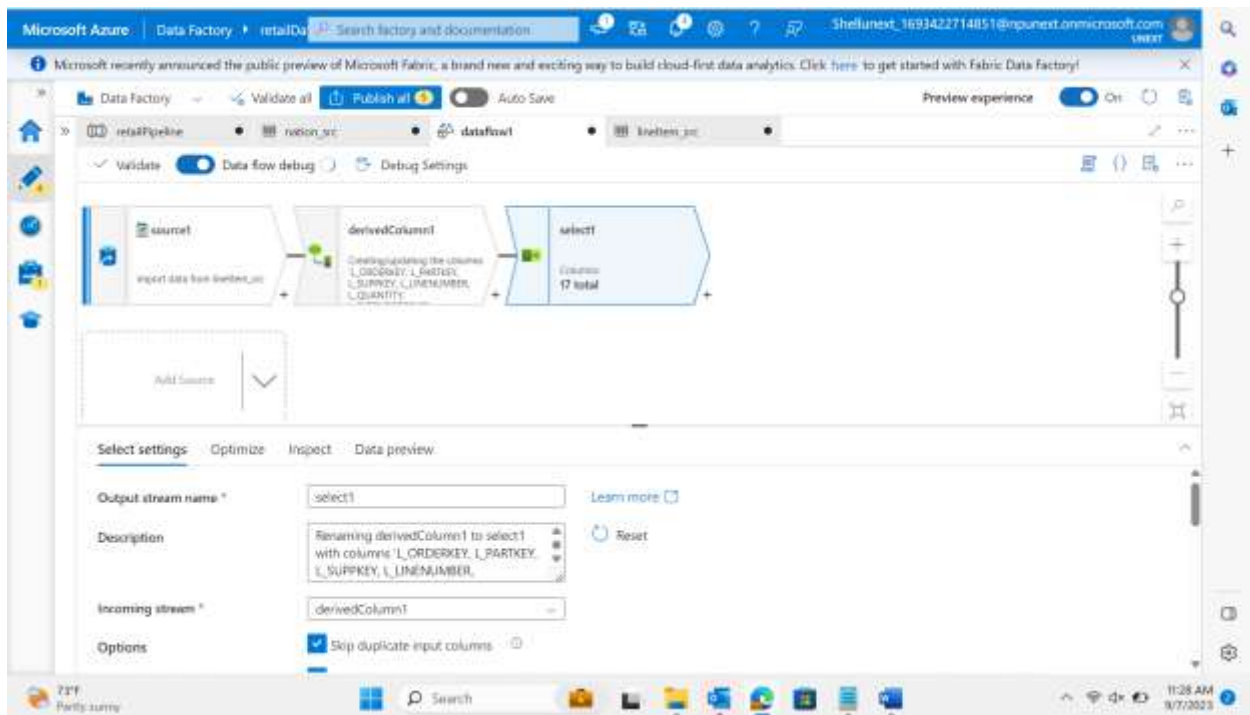
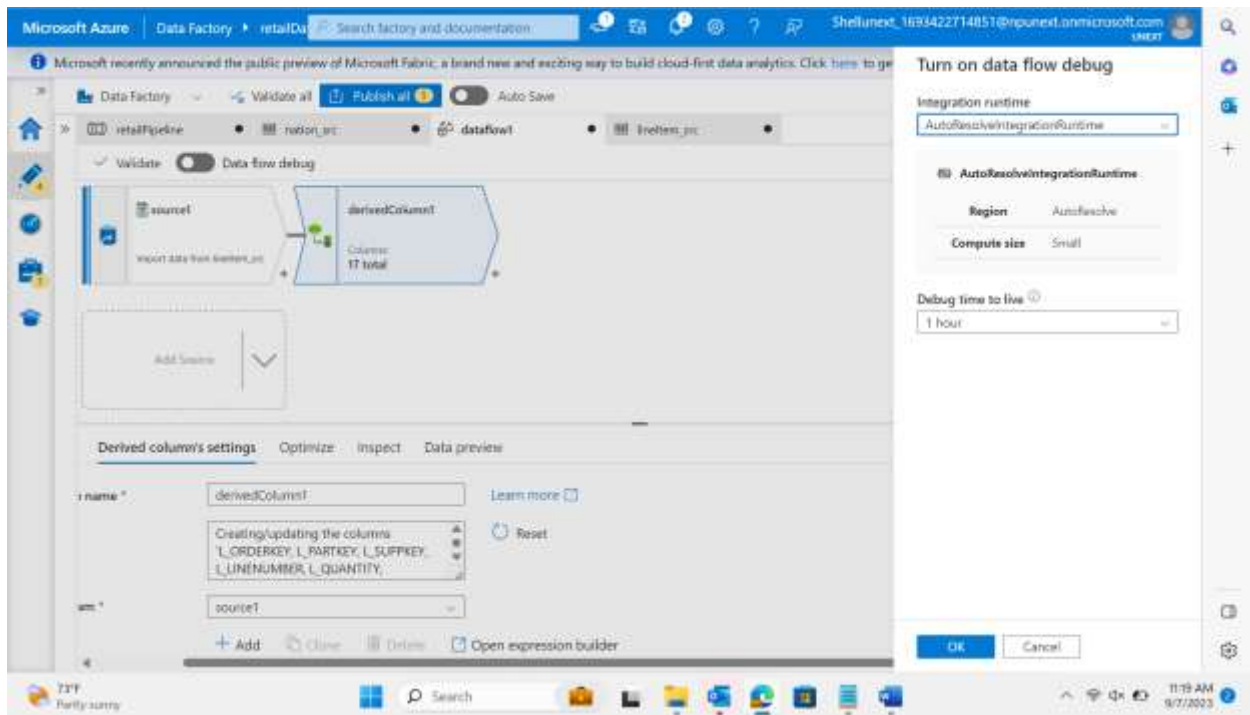
Source settings | Source options | Projection

Source settings	Source options	Projection
L_ORDERKEY	str: string	Specify format
L_PARTKEY	str: string	Specify format
L_SUPPKEY	str: string	Specify format
L_LINENUMBER	str: string	Specify format
L_QUANTITY	int: integer	Specify format
L_EXTENDEDPRICE	dt: double	Specify format
L_DISCOUNT	dt: double	Specify format

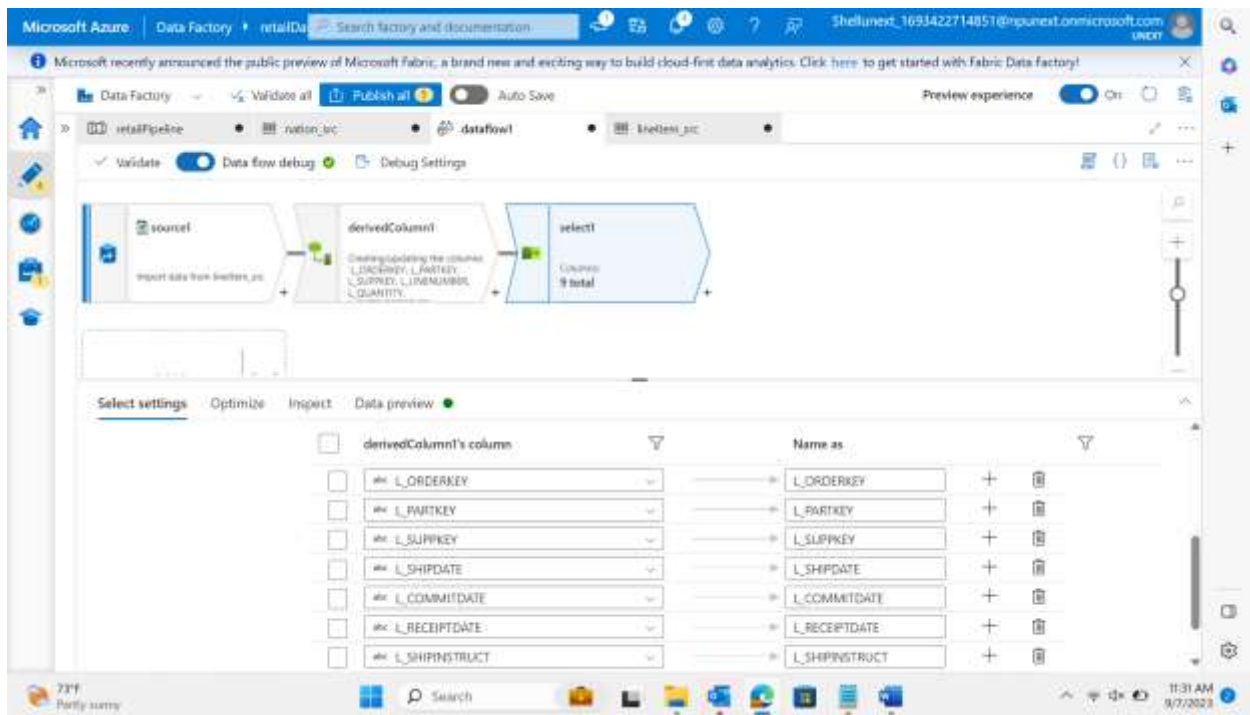
73°F Partly sunny | 11:14 AM 9/7/2023



Enable data flow debug



Selected only the required future columns



Conditional split – route your data to multiple destinations/sinks


Microsoft Azure | Data Factory | retailOut | Search factory and documentation | Shellunext_1633422714851@vpxunext.anmicrosoft.com | 1/1/23

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!

38 Data Factory | Validate all | Publish all | Auto Save | Preview experience | On

retailPipeline | nation_jrc | dataflow1 | Inetrem_jrc

Validate | Data flow debug | Debug Settings



Conditional split settings | Optimize | Inspect | Data preview

Output stream name * | split1 | Learn more

Description | Add cases to distribute the data in multiple groups | Reset

Incoming stream * | select1

Split on | ☒ First matching condition | ☐ All matching conditions

Split condition

Stream names	Condition
Enter output stream name...	

73°F Partly sunny | 11:34 AM 9/7/2023


Microsoft Azure | Data Factory | retailOut | Search factory and documentation | Shellunext_1633422714851@vpxunext.anmicrosoft.com | 1/1/23

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!

38 Data Factory | Validate all | Publish all | Auto Save | Preview experience | On

retailPipeline | nation_jrc | dataflow1 | Inetrem_jrc

Validate | Data flow debug | Debug Settings



Conditional split settings | Optimize | Inspect | Data preview

Output stream name * | shipmentmode | Learn more

Description | Conditionally distributing the data in L_SHIPMODE groups, based on column '11' | Reset

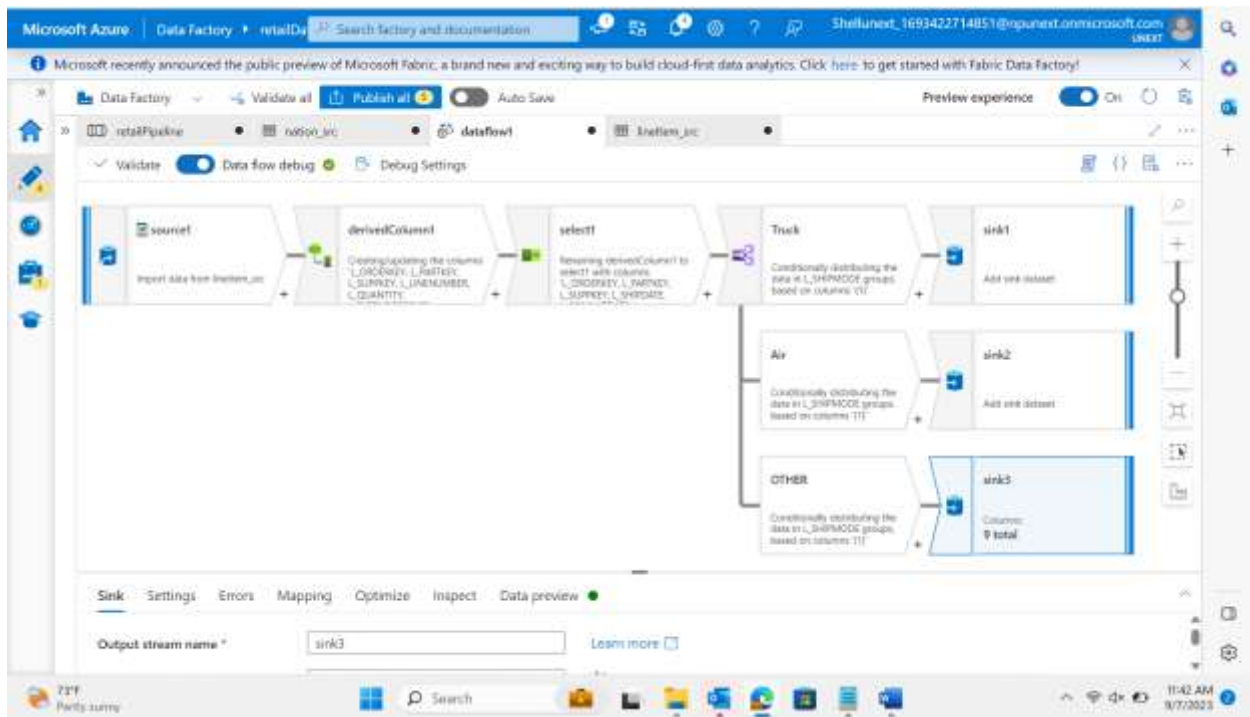
Incoming stream * | select1

Split on | ☒ First matching condition | ☐ All matching conditions

Split condition

Stream names	Condition
Truck	L_SHIPMODE == "TRUCK"
Air	L_SHIPMODE == "AIR"
OTHER	Rows that do not meet any condition will use this output stream

73°F Partly sunny | 11:42 AM 9/7/2023



ADLS2 → json

The screenshot shows the Microsoft Azure Data Factory interface. The pipeline is named 'retailPipeline' and is currently in the 'Data flow debug' mode. The pipeline consists of the following steps:

- source1**: Import data from 'retail_warehouse'.
- derivedColumn1**: Copying/adding the columns L_ORDERKEY, L_PARTKEY, L_SUPPKEY, L_LINENUMBER, L_QUANTITY.
- select1**: Returning derivedColumn1 to select1 with columns L_ORDERKEY, L_PARTKEY, L_SUPPKEY, L_SHIPDATE.
- Truck**: Conditionally distributing the data in L_SHIPMODE groups based on column ITT.
- Other**: Conditionally distributing the data in L_SHIPMODE groups based on column ITT.
- Other**: Conditionally distributing the data in L_SHIPMODE groups based on column ITT.
- sink1**: Add sink dataset.
- sink2**: Add sink dataset.
- sink3**: Column: 9 total.

The 'Data preview' tab is selected, showing the output of the pipeline. The output is a table with columns: L_ORDERKEY, L_PARTKEY, L_SUPPKEY, L_SHIPDATE, L_COMM1, L_RECEIPT, L_SHIPINS, L_SHIPMO. The data is as follows:

L_ORDERKEY	L_PARTKEY	L_SUPPKEY	L_SHIPDATE	L_COMM1	L_RECEIPT	L_SHIPINS	L_SHIPMO
1	1552	93	1996-03-13	1996-02-12	1996-03-22	DELIVER IN...	TRUCK
3	294	22	1993-12-04	1994-01-07	1994-01-01	NONE	TRUCK
6	1397	36	1992-04-27	1992-05-15	1992-05-02	TAKE BACK...	TRUCK
7	1519	40	1996-02-11	1996-02-24	1996-02-18	DELIVER IN...	TRUCK
32	828	95	1995-10-23	1995-06-27	1995-10-26	TAKE BACK...	TRUCK

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunext_1693422714851@ipunext.onmicrosoft.com | VM007

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 | Auto Save | Preview experience | On

retailPipeline | nation_jrc | dataflow1 | iretrem_jrc

Validate | Data flow debug | Debug Settings

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, ERROR

Refresh | Instance | Export to CSV

L_ORDERKEY...	L_PARTKEY...	L_SUPPKEY...	L_SHIPDATE...	L_COMML...	L_RECEIPT...	L_SHIPINS...	L_SHIPMO...
1	22	48	1996-04-21	1996-03-30	1996-05-16	NONE	AIR
3	43	19	1994-02-02	1994-01-04	1994-02-23	NONE	AIR
5	1086	67	1994-10-31	1994-06-31	1994-11-20	NONE	AIR
5	375	5	1994-08-08	1994-10-13	1994-08-26	DELIVER IN...	AIR

73°F Partly sunny | Search | 11:51 AM 9/7/2023

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunext_1693422714851@ipunext.onmicrosoft.com | VM007

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 | Auto Save | Preview experience | On

retailPipeline | nation_jrc | dataflow1 | iretrem_jrc

Validate | Data flow debug | Debug Settings

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, ERROR

Refresh | Instance | Export to CSV

L_ORDERKEY...	L_PARTKEY...	L_SUPPKEY...	L_SHIPDATE...	L_COMML...	L_RECEIPT...	L_SHIPINS...	L_SHIPMO...
1	674	75	1996-04-12	1996-02-28	1996-04-30	TAKE BACK...	MAIL
1	637	38	1996-01-29	1996-03-05	1996-01-31	TAKE BACK...	REG AIR
1	341	23	1996-03-30	1996-03-14	1996-04-01	NONE	FOB
1	157	10	1996-01-30	1996-02-07	1996-02-03	DELIVER IN...	MAIL

73°F Partly sunny | Search | 11:51 AM 9/7/2023

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunet_1693422714851@mgunest.onmicrosoft.com

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!

retailPipeline | nation_src | dataflow1 | intellem_src | ds_smAir_adj

Activities: Search activities, Move and transform, Copy data, Data flow, Synapse, Azure Data Explorer, Azure Function, Batch Service, Databricks, Data Lake Analytics, General, HDInsight, Iteration & conditionals, Machine Learning.

Properties: General, Related, Name: retailPipeline, Description, Annotations: + New.

Parameters Variables Settings Output

Showing 1 - 3 of 3 items

Activity name	Activity status	Run start	Duration
Data flow 1	In progress	9/7/2023, 11:52:15 AM	1m 4s
If Condition1	Inactive	9/7/2023, 11:52:16 AM	Less than 1
ForEach1	Inactive	9/7/2023, 11:52:16 AM	Less than 1

73°F Partly sunny | 11:54 AM 9/7/2023

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunet_1693422714851@mgunest.onmicrosoft.com

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!

Factory Resources: Filter resources by name, Pipelines (1), Datasets (5), Data flows (1), Power Query (0).

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.

Parameters Variables Settings Output

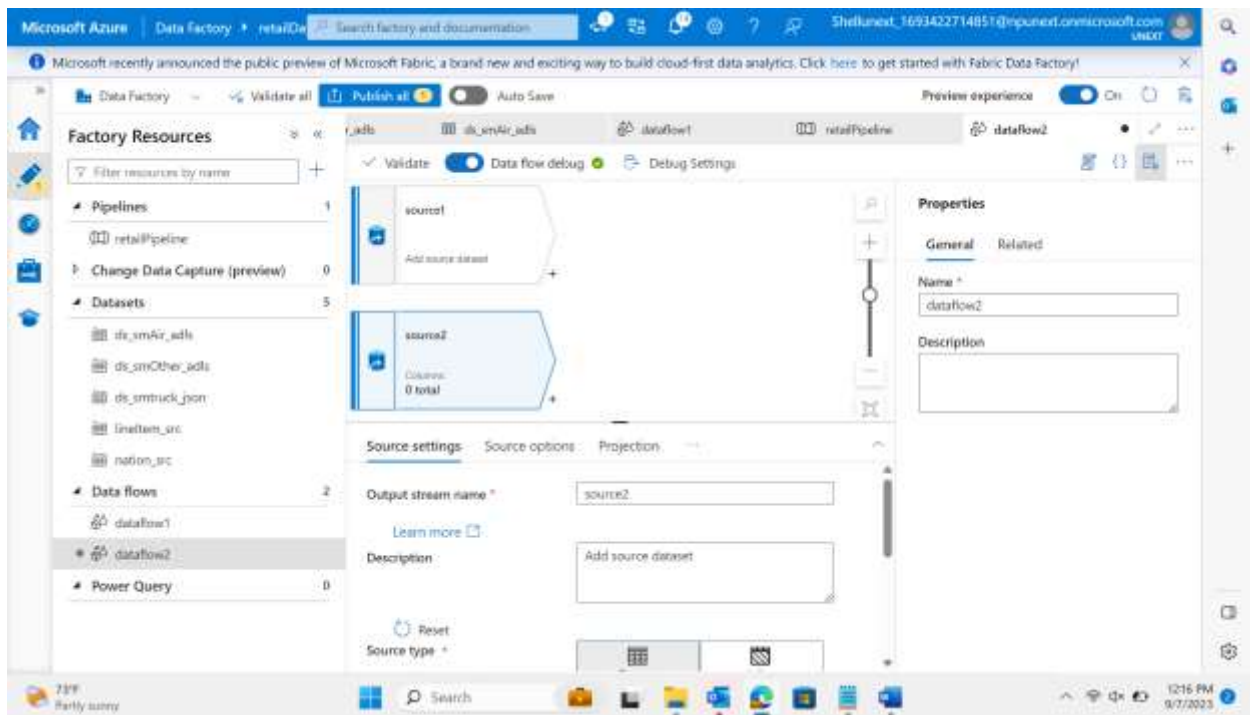
Pipeline run ID: ba764ad9-a5af-4ed5-9314-abd3ed87b974

Showing 1 - 3 of 3 items

Activity name	Activity status	Run start	Duration
Data flow 1	Succeeded	9/7/2023, 12:13:45 PM	1m 8s

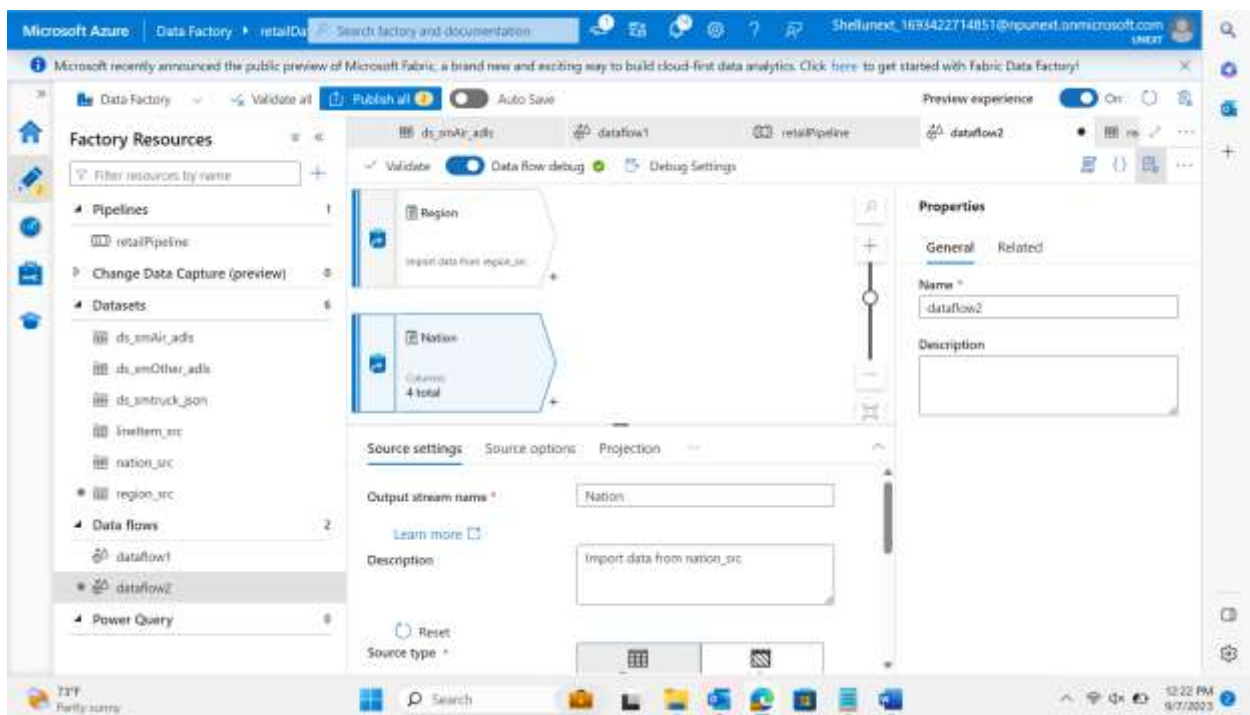
73°F Partly sunny | 12:15 PM 9/7/2023

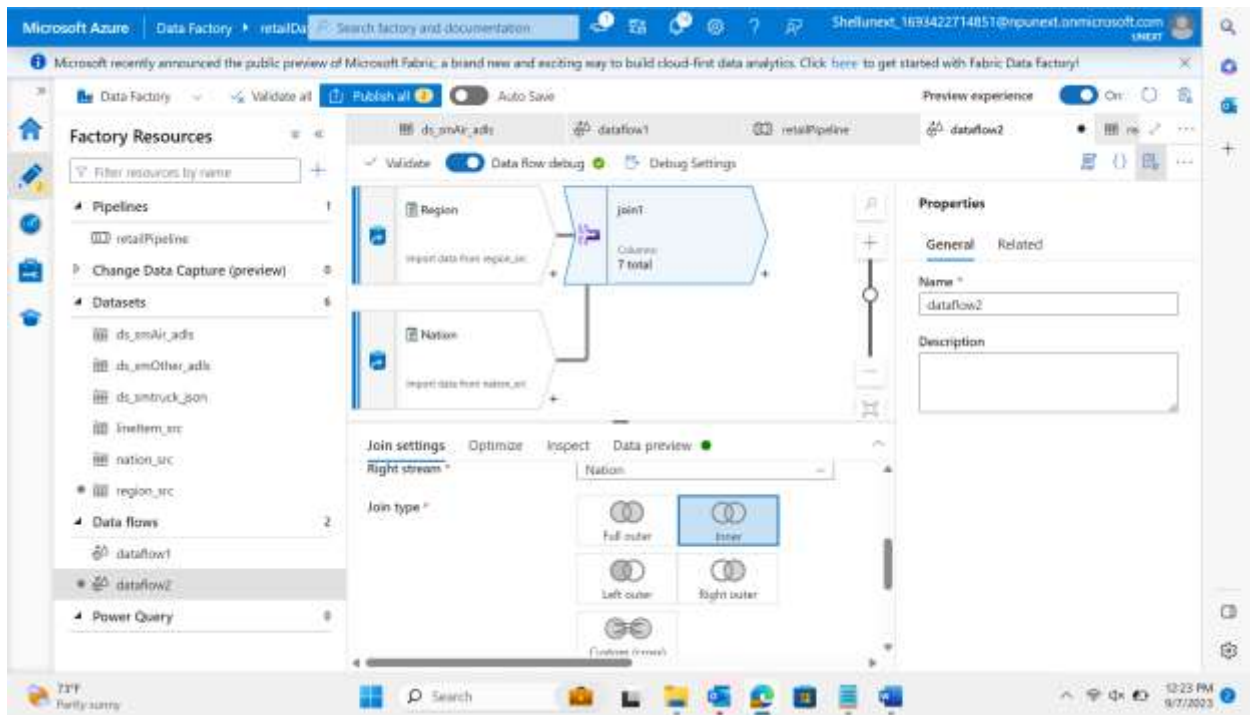
Another Data flow task



Region source

Nation second source





Join settings Optimize Inspect Data preview ●

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

Use fuzzy matching ⓘ ☐

Join conditions *

Left: Region's column		Right: Nation's column
abc R_REGIONKEY	==	abc N_REGIONKEY

+ -

The screenshot shows the Microsoft Azure Data Factory interface. At the top, there's a navigation bar with 'Data Factory', 'retailData', and a search bar. Below this, a banner mentions Microsoft Fabric. The main workspace shows a data pipeline with a 'Region' activity (input data from regionList) and a 'Join' activity (columns: 7 total). The 'Data preview' tab is active, displaying a table with 25 rows and 12 columns. The table contains data for various regions and their corresponding sales data.

T4	R_REGION...	R_NAME	R_COMME...	N_NATION...	N_NAME	N_REGION...	N_COMME...
*	0	AFRICA	lar deposit...	0	ALGERIA	0	hazale ch...
*	1	AMERICA	hs use iron...	1	ARGENTINA	1	al boxes pr...
*	1	AMERICA	hs use iron...	2	BSAZIL	1	y alongsd...
*	1	AMERICA	hs use iron...	3	CANADA	1	was hang it...
*	4	MIDDLE EA...	uckly spec...	4	EGYPT	4	y above th...
*	0	AFRICA	lar deposit...	5	ETHIOPIA	0	ven packag...
*	3	EUROPE	le final cou...	6	FRANCE	3	refute final...

Copy Data: Ingest the raw sales data from various sources (e.g., CSV files, databases) into Azure Data Lake Storage.

****Ingesting multiple files all at once**

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunet, 1693422714851@punext.onmicrosoft.com | UNLOCK

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 | Auto Save | Preview experience: On

Factory Resources

- Pipelines: 2
 - caseStudySalesPipeline
 - retailPipeline
- Change Data Capture (preview): 0
- Datasets: 7
 - ds_smAir_ads
 - ds_smOther_ads
 - ds_smtruck_json
 - inselltem_src
 - nation_src**
 - region_src
 - sales_src
- Data flows: 1
 - dataflow1
- Power Query: 0

nation_src | inselltem_src | ds_smtruck_json | ds_smOther_ads | ds_smAir_ads

Delimited text nation_src

Connection | Schema | Parameters

Linked service: ls_blobretail

File path: retail / Directory / File name Browse

Compression type: Select

Column delimiter: Tab (t)

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

Encoding: Default(UTF-8)

82°F Mostly cloudy | Search | 2:34 PM 9/7/2023

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunet, 1693422714851@punext.onmicrosoft.com | UNLOCK

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 | Auto Save | Preview experience: On

Factory Resources

- Pipelines: 2
 - caseStudySalesPipeline**
 - retailPipeline
- Change Data Capture (preview): 0
- Datasets: 7
 - ds_smAir_ads
 - ds_smOther_ads
 - ds_smtruck_json
 - inselltem_src
 - nation_src
 - region_src
 - sales_src
- Data flows: 1
 - dataflow1
- Power Query: 0

dataflow1 | retailPipeline | region_src | caseStudySalesPipeline

Validate | Debug | Add trigger

Get Metadata | Get Metadata1

ForEach | ForEach1

Activities | Copy data1

Properties

General | Related

Name: caseStudySalesPipeline

Description:

Annotations: + New

Parameters | Variables | Settings | **Output**

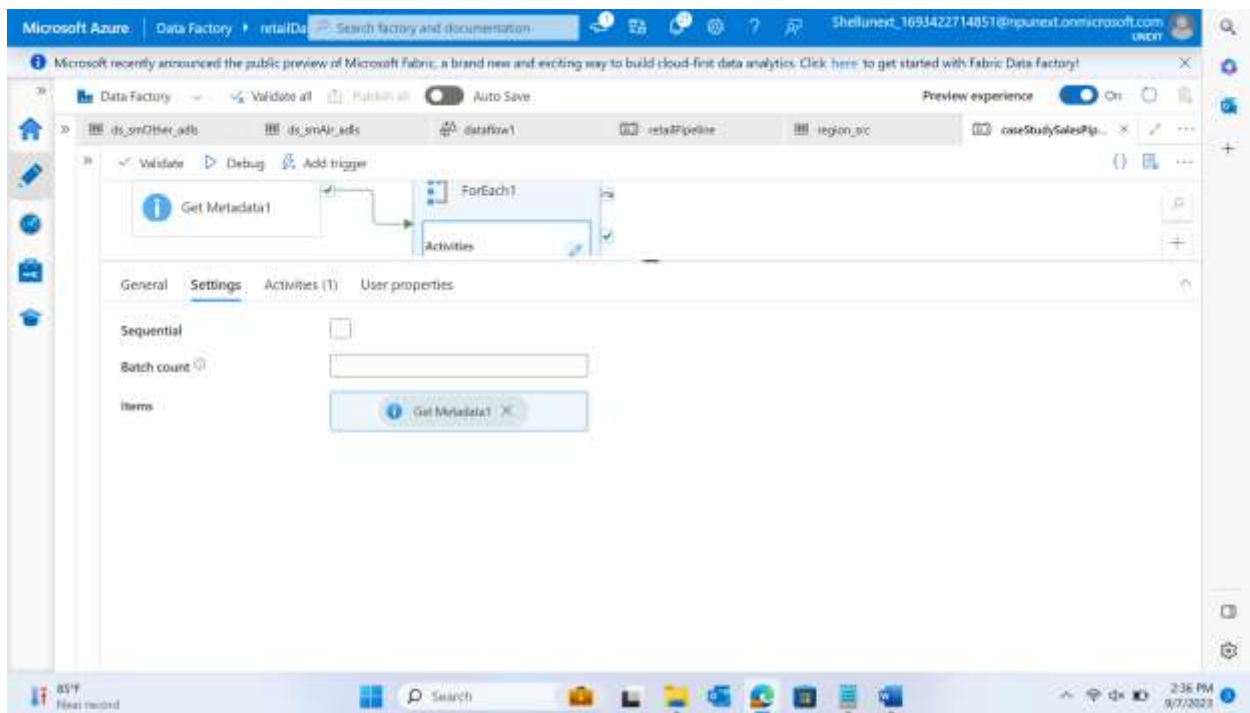
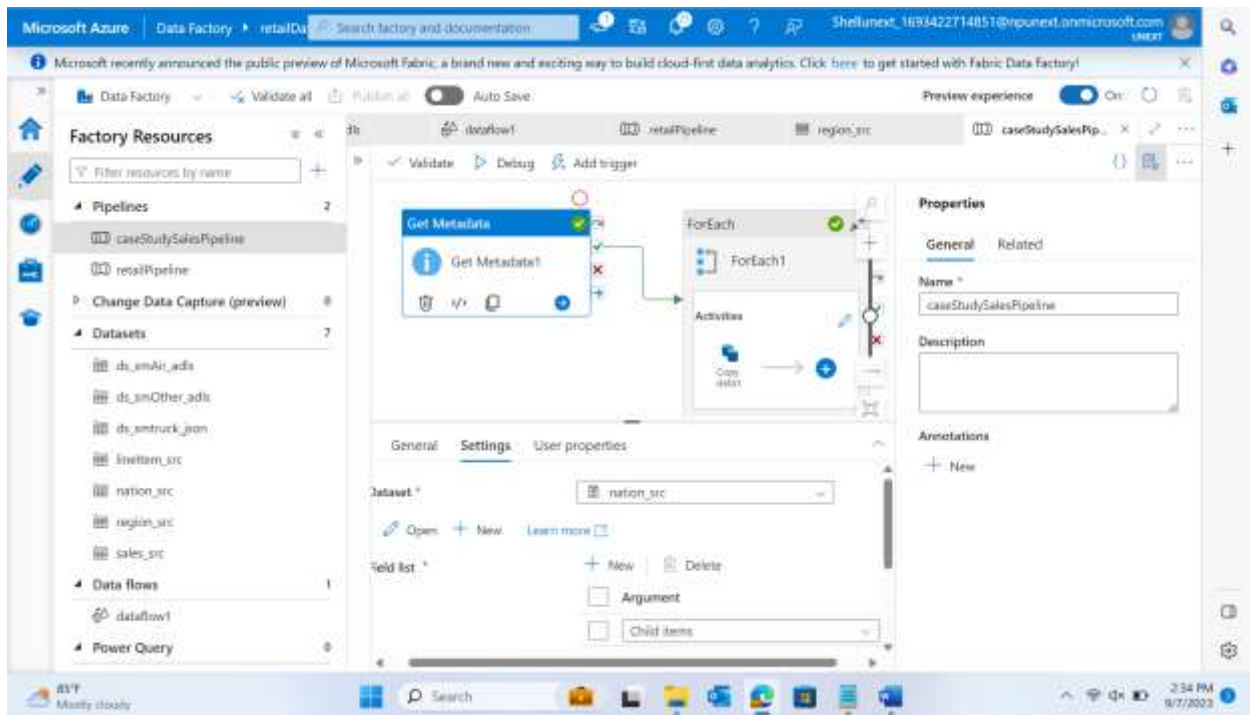
Pipeline run ID: c7ee5d34-a895-44fa-bdc2-3b67072e18d9 | View debug run | consumption

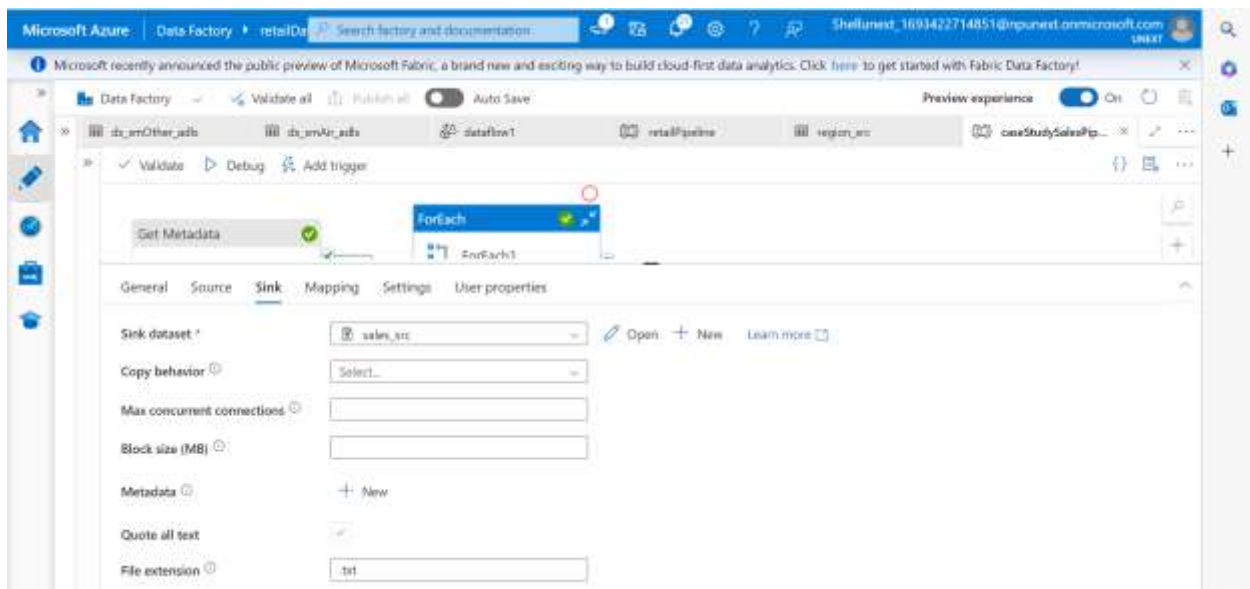
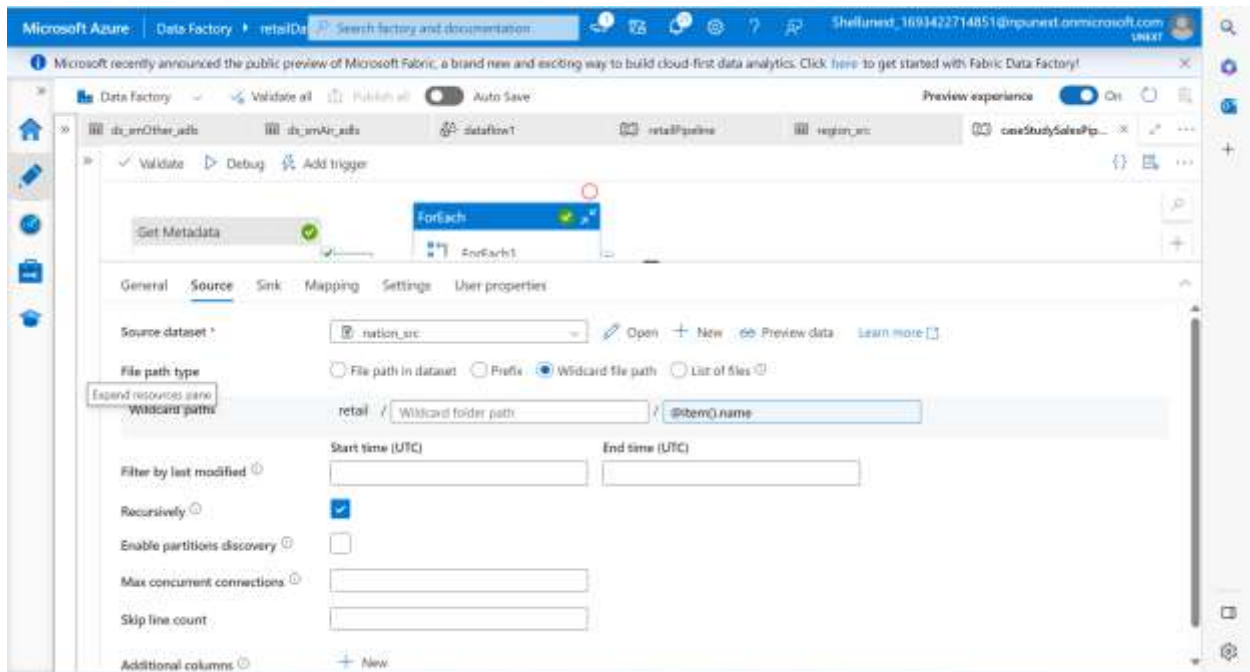
All status: List | Monitor in Azure Metrics | Export to CSV

Showing 1 - 10 of 10 items

Activity name	Activity status	Run start
Copy data1	Succeeded	9/7/2023, 2:33:47 PM

83°F Mostly cloudy | Search | 2:34 PM 9/7/2023





Microsoft Azure | Data Factory | retailData | Search factory and documentation

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

Data Factory | Validate all | Publish all | Auto Save | Preview experience: On

ds_omOther_ads | ds_omAir_ads | dataflow1 | retailPipeline | region_ads | caseStudySalesOp...

✓ Validate | Debug | Add trigger

Parameters | Variables | Settings | **Output**

Pipeline run ID: c7ee5d34-af06-44fe-bdc2-3b67072e18d9 | Pipeline status: Succeeded | View debug run consumption | Monitor in Azure Metrics | Export to CSV

All status | List

Showing 1 - 10 of 10 items

Activity name	Activity status	Run start	Duration	Integration runtime	User properties	Activity run ID
Copy data1	Succeeded	9/7/2023, 2:33:47 PM	10s	AutoResolveIntegration		999b3771-4488-4839-8db
Copy data1	Succeeded	9/7/2023, 2:33:47 PM	9s	AutoResolveIntegration		47885b3c-bcd0-4246-b42
Copy data1	Succeeded	9/7/2023, 2:33:47 PM	9s	AutoResolveIntegration		65dec10a-9544-45a8-b89f
Copy data1	Succeeded	9/7/2023, 2:33:47 PM	9s	AutoResolveIntegration		3d581add-4diff-40b8-a15
Copy data1	Succeeded	9/7/2023, 2:33:47 PM	10s	AutoResolveIntegration		43d84b05-7a06-407b-9f1
Copy data1	Succeeded	9/7/2023, 2:33:47 PM	9s	AutoResolveIntegration		64f6a768-776e-43ab-a72f
Copy data1	Succeeded	9/7/2023, 2:33:47 PM	10s	AutoResolveIntegration		3c5ad7d8-1464-4666-b0w
Copy data1	Succeeded	9/7/2023, 2:33:47 PM	9s	AutoResolveIntegration		e6e23dc5-3a92-4e74-b44
ForEach1	Succeeded	9/7/2023, 2:33:47 PM	14s			a45067a6-8eb2-494d-a0e1

85°F | Mostly cloudy | Search | 2:36 PM 9/7/2023

API-based integration

reqres.in/api/users?page=2

Microsoft Azure | Data Factory | retailData | Search factory and documentation

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

Data Factory | Validate all | Publish all | Auto Save | Preview experience: On

ds_omOther_ads | ds_omAir_ads | dataflow1 | retailPipeline | region_ads | caseStudySalesOp...

✓ Validate | ✓ Validate copy runtime | Debug | Add trigger

Activities | Search activities

Move and transform | Copy data | Data flow | Synapse | Azure Data Explorer | Azure Function | Batch Service | Databricks | Data Lake Analytics | General | HDInsight | Iteration & conditionals | Machine Learning

Copy data1

General | **Source** | Sink | Mapping

Source dataset: Select...

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

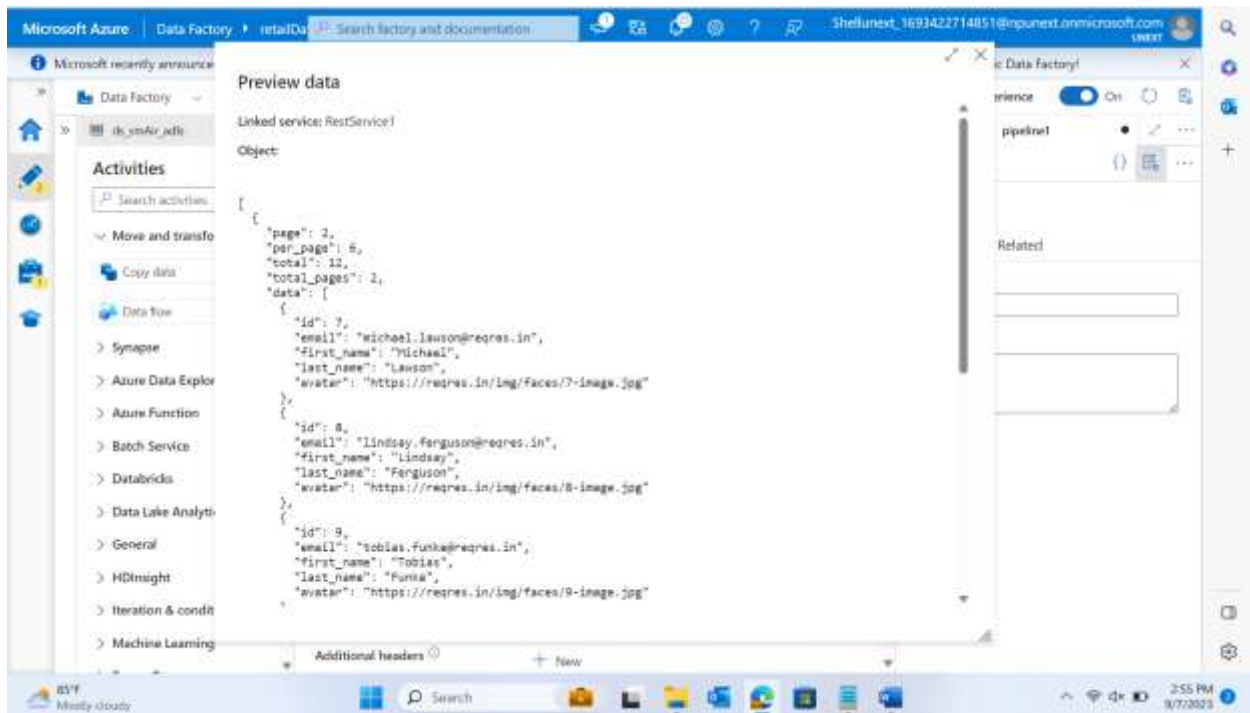
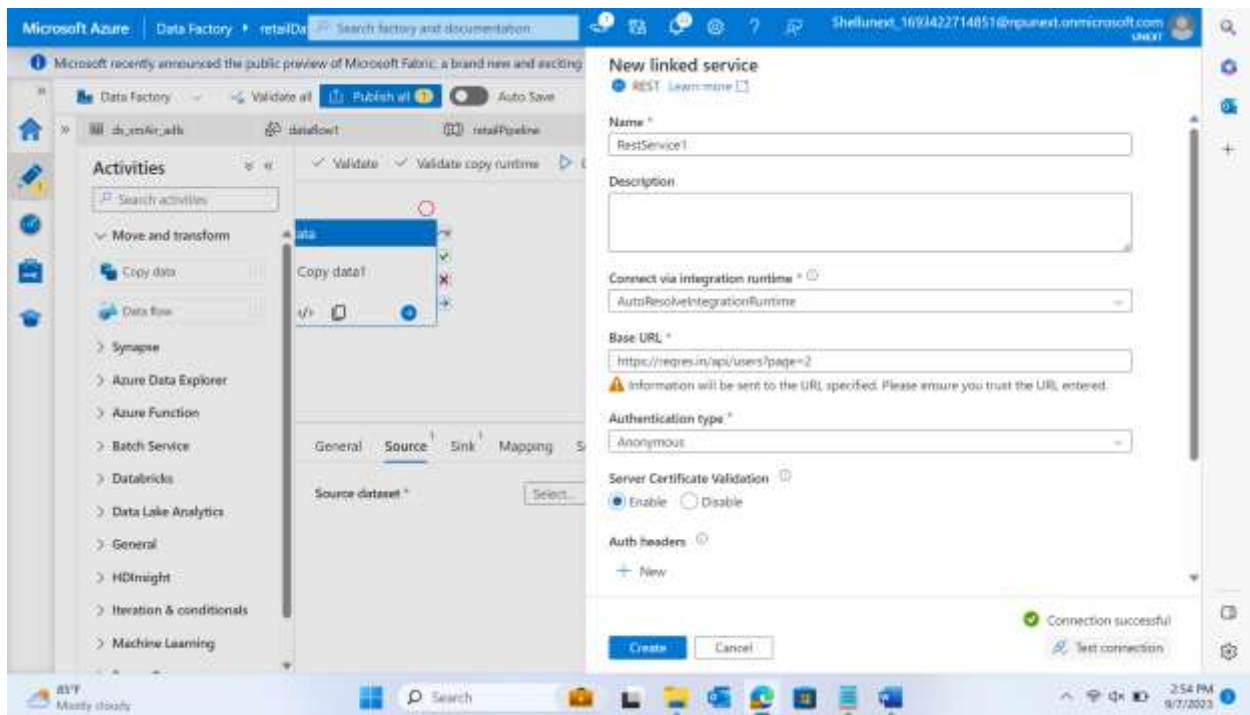
rest

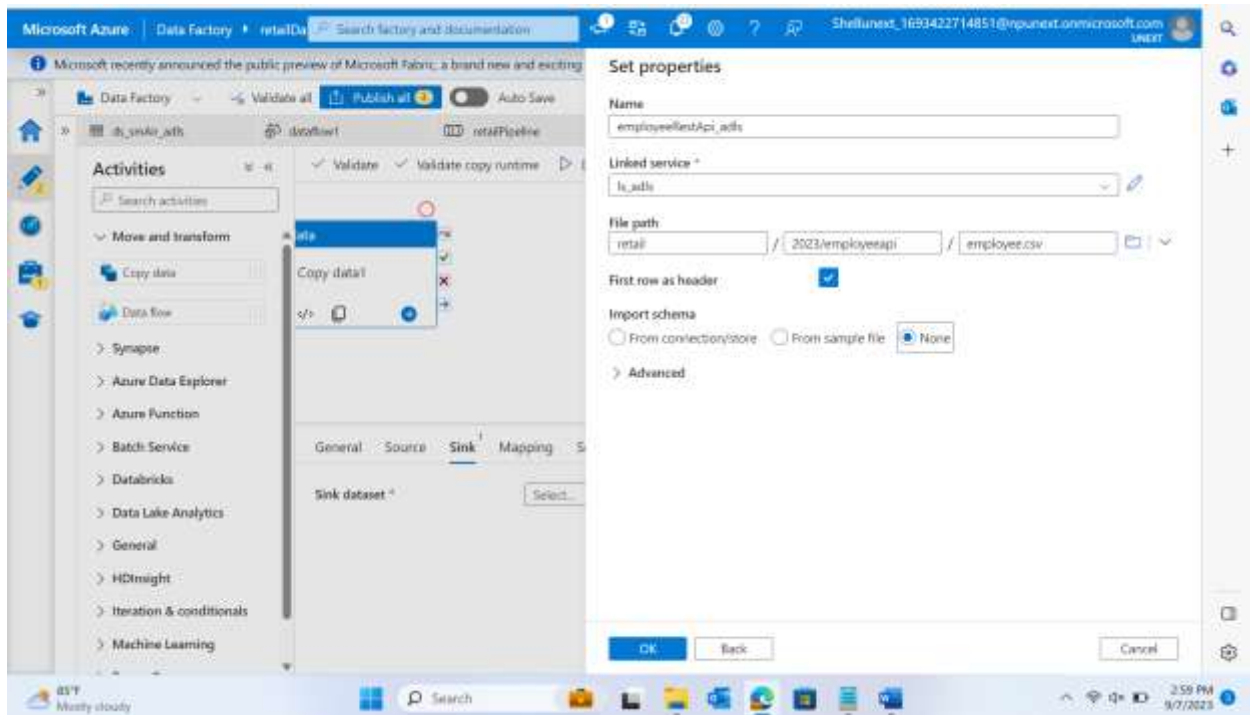
All | Azure | Database | File | Generic protocol | NoSQL | Services and apps

Presto | REST

Continue | Cancel

85°F | Mostly cloudy | Search | 2:36 PM 9/7/2023





General Source Sink **Mapping** Settings User properties

Import schemas

+ New mapping Clear ^① Delete ☐ Advanced editor

Collection reference ^①

Map complex values to string ☐

Name	Type	Collection reference	Column name
ANV	ANV		

General
Source
Sink
Mapping
Settings
User properties

Import schemas

+ New mapping
Clear ⓘ
Delete
☒ Advanced editor

Collection reference ⓘ

\$['data']

Map complex values to string
☐

Name	Type	Collection reference	Column name	Type	<input checked="" type="checkbox"/> Include
Add dynamic content [Alt+Shift+D]					

Import schemas

The screenshot displays the Microsoft Fabric Data Factory Mapping view. The top navigation bar includes tabs for General, Source, Sink, Mapping (selected), Settings, and User properties. The Mapping view shows a pipeline with two source datasets, 'vsa' and 'vab', which are joined by a 'union1' operation. The 'vsa' dataset is described as 'import data from vsa_blob' and the 'vab' dataset as 'import data from vab_blob'. The 'union1' operation is described as 'Combining rows from transformation vsa_sab'. The output of the union is a 'surrogateKey1' operation, described as 'Customer YS total'. The left sidebar shows the 'Factory Resources' section with a list of pipelines and datasets. The bottom status bar shows the current time as 4:25 PM on 8/7/2023.

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunet_1693422714851@rpunext.onmicrosoft.com | UNEXT

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!

Factory Resources

- Pipelines
 - caseStudySalesPipeline
 - pipeline1
 - retailPipeline
- Change Data Capture (preview)
- Datasets
 - ds_smAr_ads
 - ds_smOther_ads
 - ds_smtruck_json
 - employeeFeedApi
 - employeeFeedApi_ads
 - lineItem_src
 - nation_src
 - region_src
 - sales_src

dataflow1 | retailPipeline | region_src | caseStudySalesPipe... | pipeline1

Validate | Data flow debug | Debug Settings

Surrogate key settings | Optimize | Inspect | Data preview

Output stream name * surrogateKey1 [Learn more](#)

Description Adding new key sk_name starting from 1 with step 1 [Reset](#)

Incoming stream * union1

Key column * sk_name

Start value * 1000

Step value 1

83°F Mostly cloudy | Search | 4:26 PM 9/7/2023

Microsoft Azure | Data Factory | retailData | Search factory and documentation | Shellunet_1693422714851@rpunext.onmicrosoft.com | UNEXT

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!

Factory Resources

- Pipelines
 - caseStudySalesKpeline
 - pipeline1
 - retailPipeline
- Change Data Capture (preview)
- Datasets
 - ds_smAr_ads
 - ds_smOther_ads
 - ds_smtruck_json
 - employeeFeedApi
 - employeeFeedApi_ads
 - lineItem_src
 - nation_src
 - region_src
 - sales_src

dataflow1 | retailPipeline | region_src | caseStudySalesPipe... | pipeline1

Validate | Data flow debug | Debug Settings

Surrogate key settings | Optimize | Inspect | Data preview

LOOKUP 0 ERROR 0 TOTAL 1999

den	cr	crim...	crim...	0	2012	5	sk_name
		NULL	NULL	NULL	NULL	NULL	1000
		NULL	NULL	NULL	NULL	NULL	1001
		NULL	NULL	NULL	NULL	NULL	1002
		NULL	NULL	NULL	NULL	NULL	1003
		NULL	NULL	NULL	NULL	NULL	1004
		NULL	NULL	NULL	NULL	NULL	1005

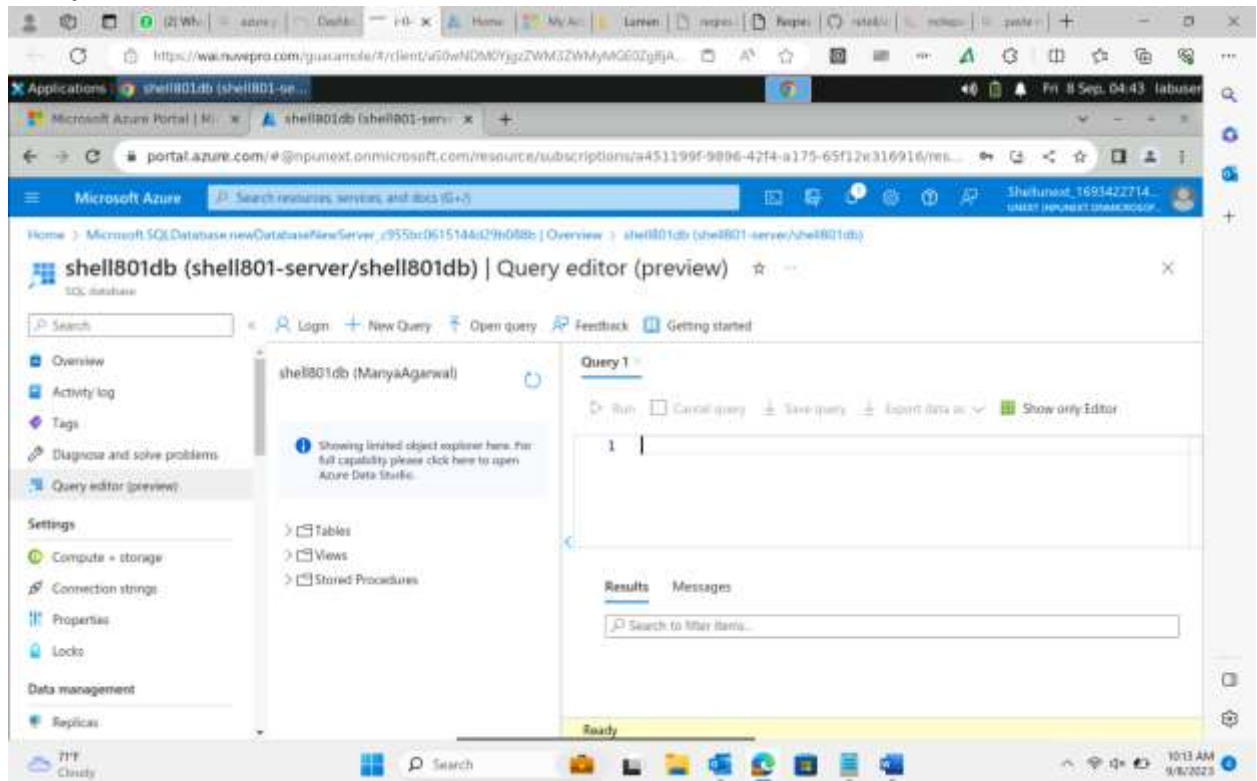
83°F Mostly cloudy | Search | 4:27 PM 9/7/2023

Friday – 8/9/23

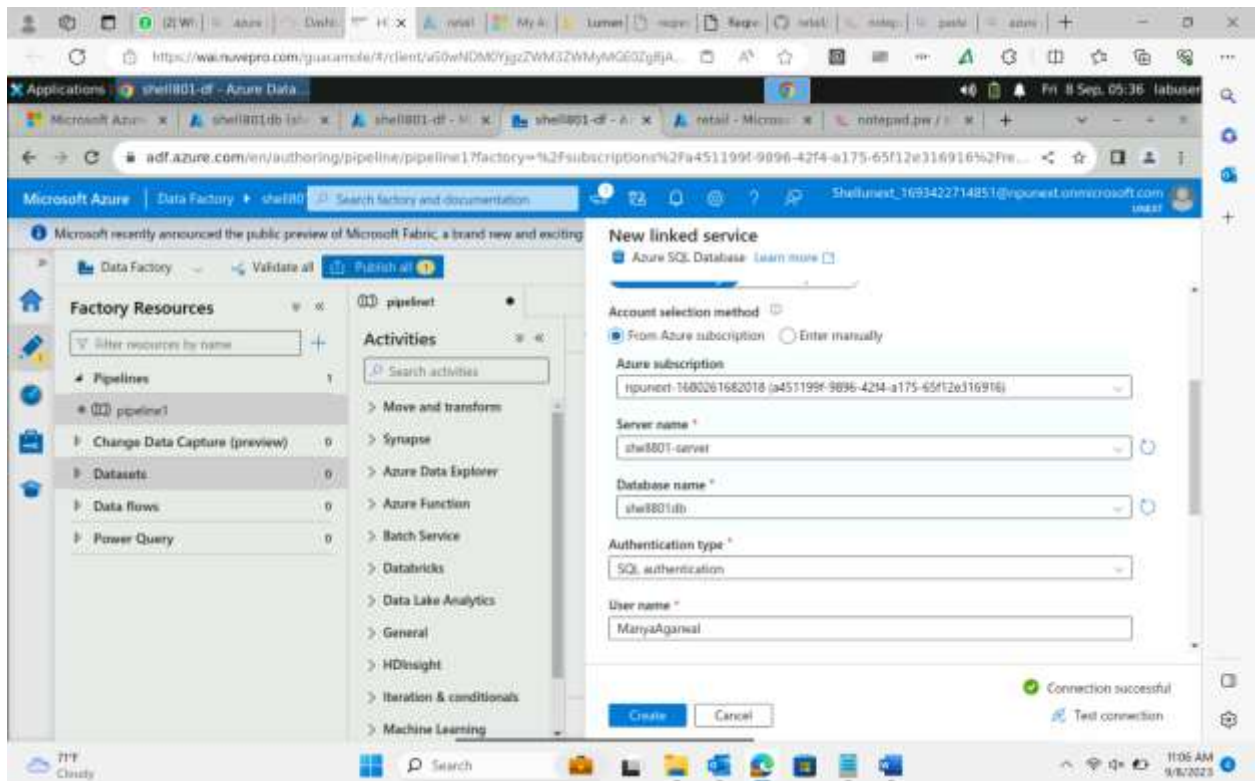
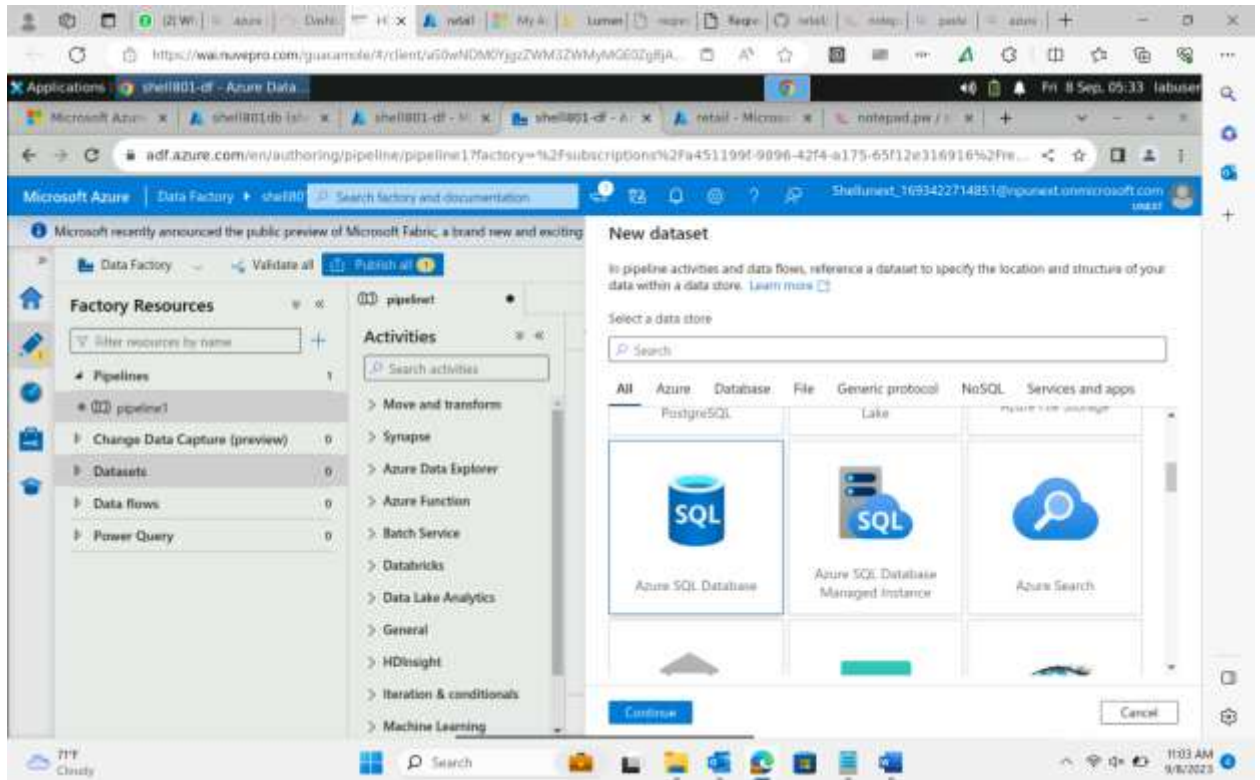
To calculate month wise data from date → derived column

Today's task : Data migration – from on premises to cloud DW (file sources to azure sql)

1. Create azure sql
2. Sql databases – single database
3. Create server
ManyarAgarwal
ManyarAgarwa4
Configure – basic
Development
Local redundant
Network → public endpoint, yes, yes
Query editor



Create data factory
Azure adls storage account
Retail/2023/sep/08/superstore.csv



Microsoft Azure | Data Factory | shell801-df - Azure Data

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting

Factory Resources

- Pipelines
- Datasets
- Data flows
- Power Query

Activities

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

Set properties

Name: src_employee_table_sql

Linked service: sqdW_ls

Table name: dbo.employee

Import schema: ☒ From connection/store ☐ None

Preview data

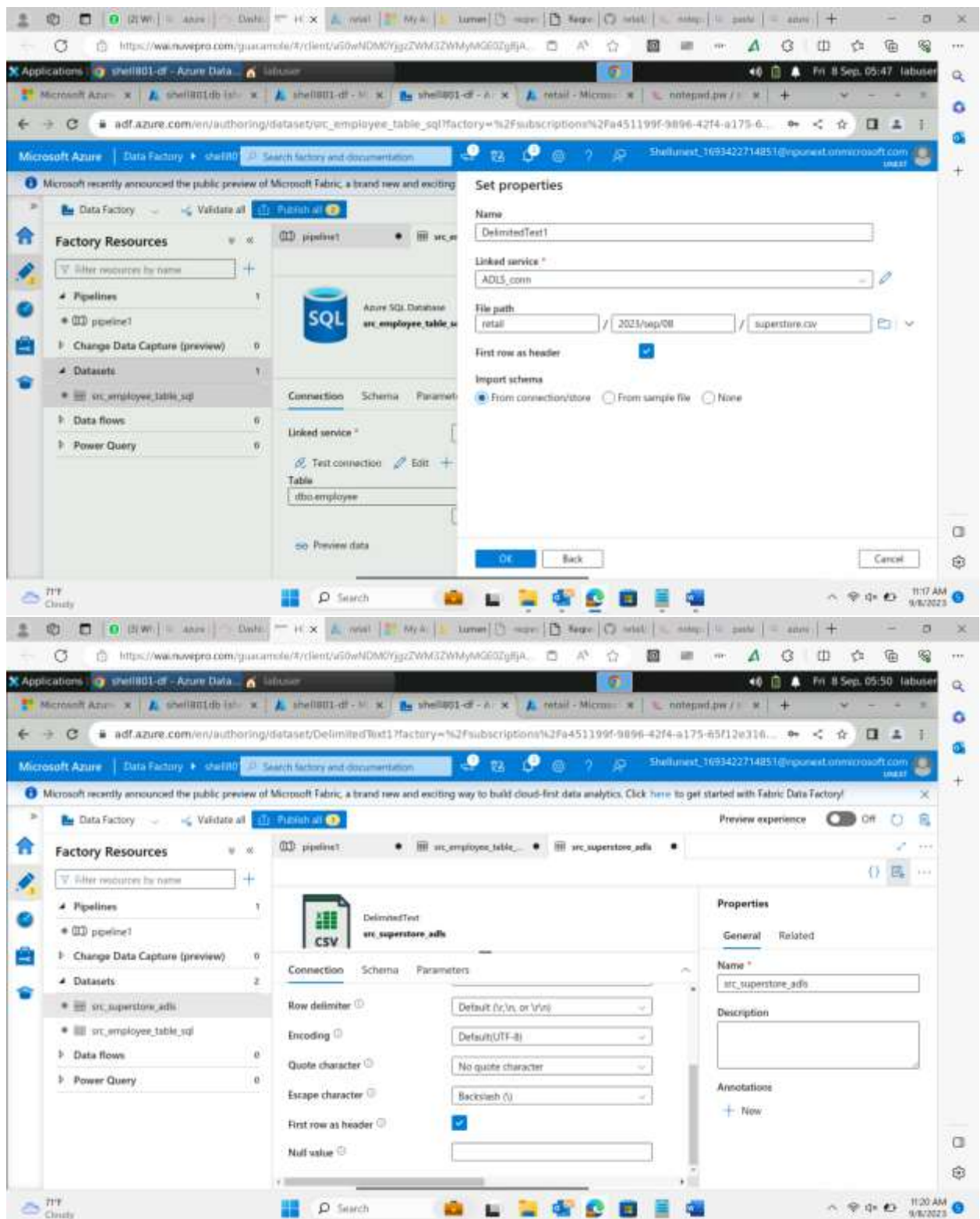
Linked service: sqdW_ls

Object: dbo.employee

	empno	ename	salary
1.	111	ABC	9999
2.	112	Arun	5555

src_employee_table_sql

New dataset
Adls2

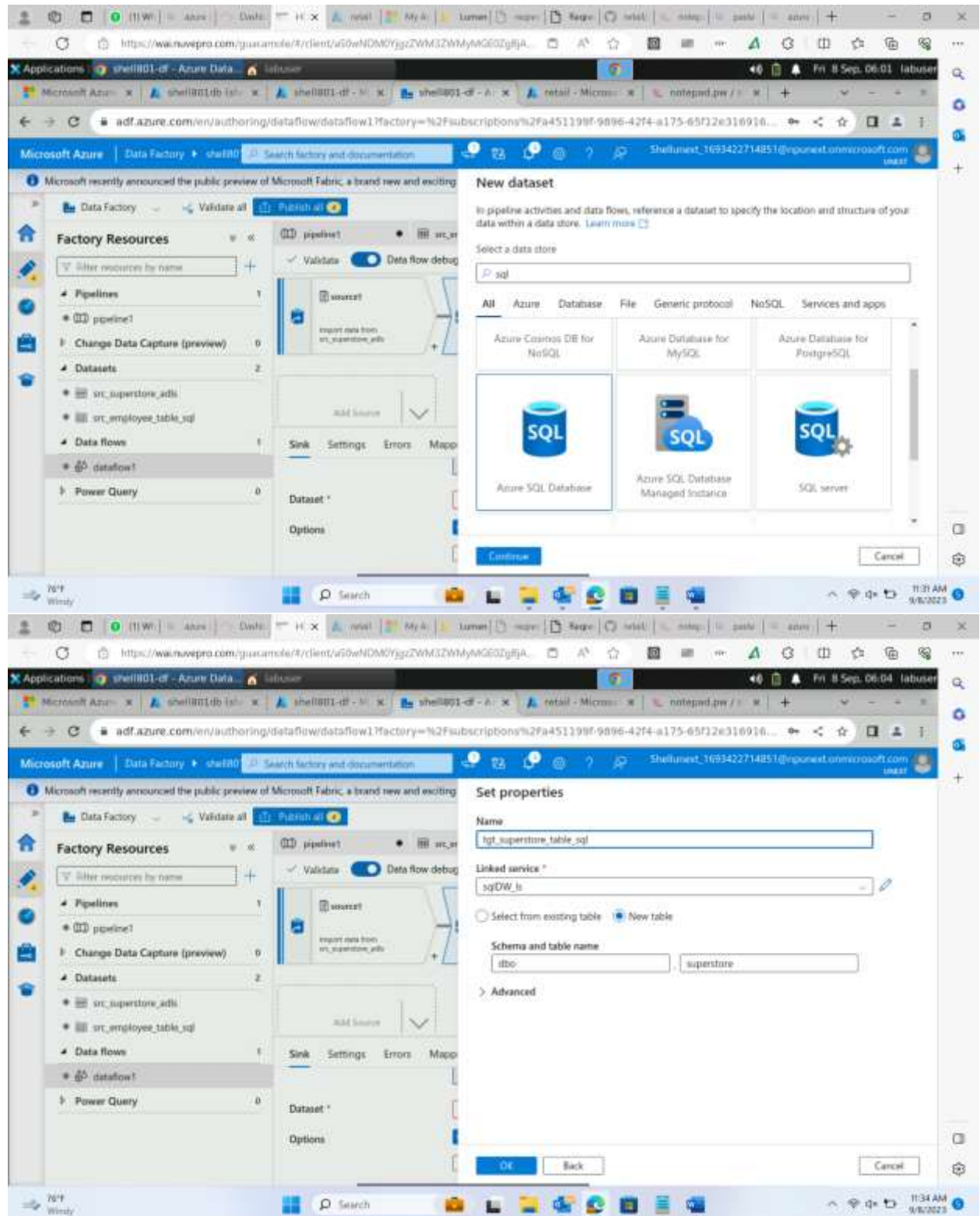


Data flow

Superstore_adls dataset

Projection – id, date, sales, qty, discount

sink



Pipeline → activity → dataflow

Microsoft Azure | Data Factory | shell801 | Search factory and documentation

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

Factory Resources

- Pipelines
 - pipeline1
- Change Data Capture (preview) 0
- Datasets
 - src_employee_table_sq
 - src_superstore_ads
 - tgt_superstore_table_sq
- Data flows
 - dataflow1
- Power Query 0

pipeline1

Validate Debug Add trigger Data flow debug

Data flow

Parameters Variables Settings Output

Pipeline run ID: 4866062a-b329-4db6-a9a3-cbbd7d2b17e2 Pipeline status: Succeeded

All status

Showing 1 - 1 of 1 items

Activity name	Activity status	Run start	Duration	Integration runtime	User
Data flow1	Succeeded	9/8/2023, 6:13:53 AM	1m 5s	debugpool-8Cores-Ge	

Microsoft Azure | Search resources, services, and docs (G+5)

Home > Microsoft SQL Database newDatabaseNewServer_c955bc0615144c29b088b | Overview > shell801db (shell801-server/shell801db)

shell801db (shell801-server/shell801db) | Query editor (preview)

SQL database

Login New Query Open query Feedback Getting started

shell801db (ManyarAgarwal)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- dbo.employee
- dbo.superstore

Views

Stored Procedures

Query 1

```
1 create table employee (empno int, ename varchar(10), salary int);
2 insert into employee values(111, 'ABC', 99999);
3 insert into employee values(112, 'Arjun', 6666);
4 select * from employee;
5 select * from [dbo].[superstore];
```

Results

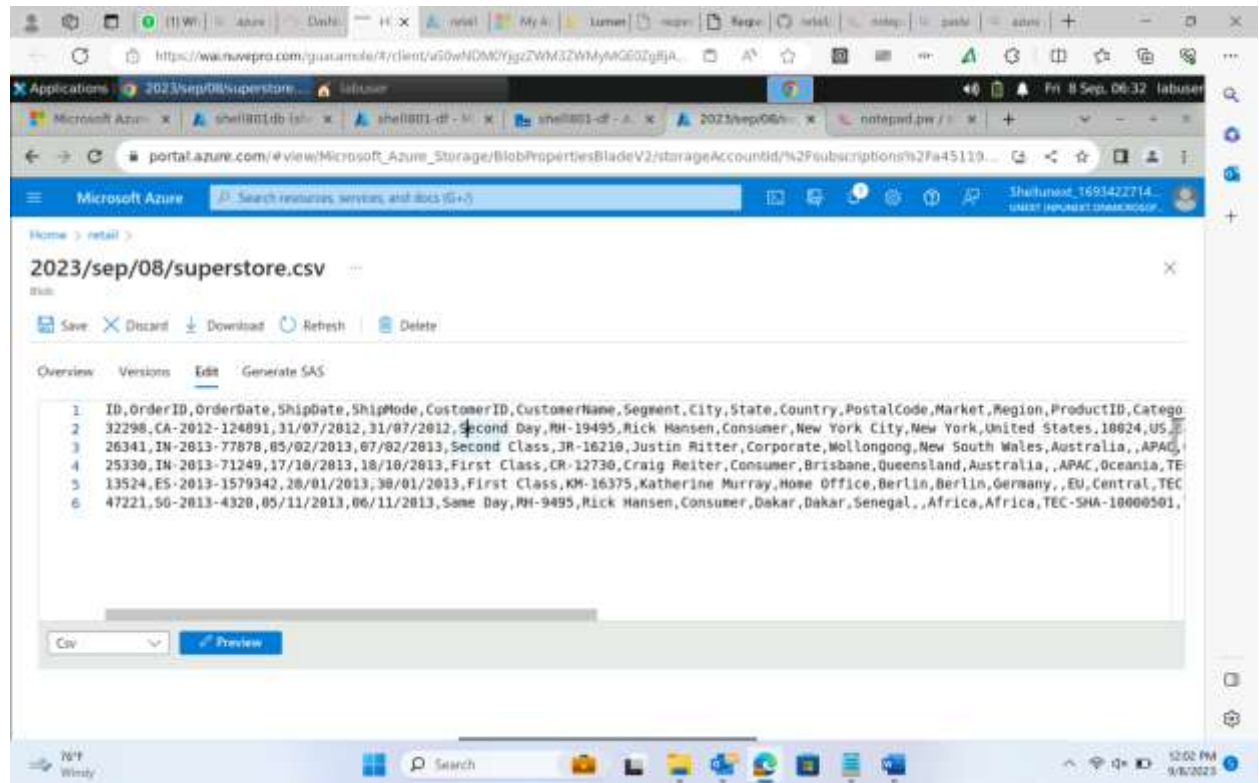
id	date	time	class
IN-2013-71249	2013-10-17T00:00:00.0000000	2013-10-18T00:00:00.0000000	First Class
ES-2013-1579342	2013-01-28T00:00:00.0000000	2013-01-30T00:00:00.0000000	First Class
SG-2013-4320	2013-11-05T00:00:00.0000000	2013-11-06T00:00:00.0000000	Same Day

Messages

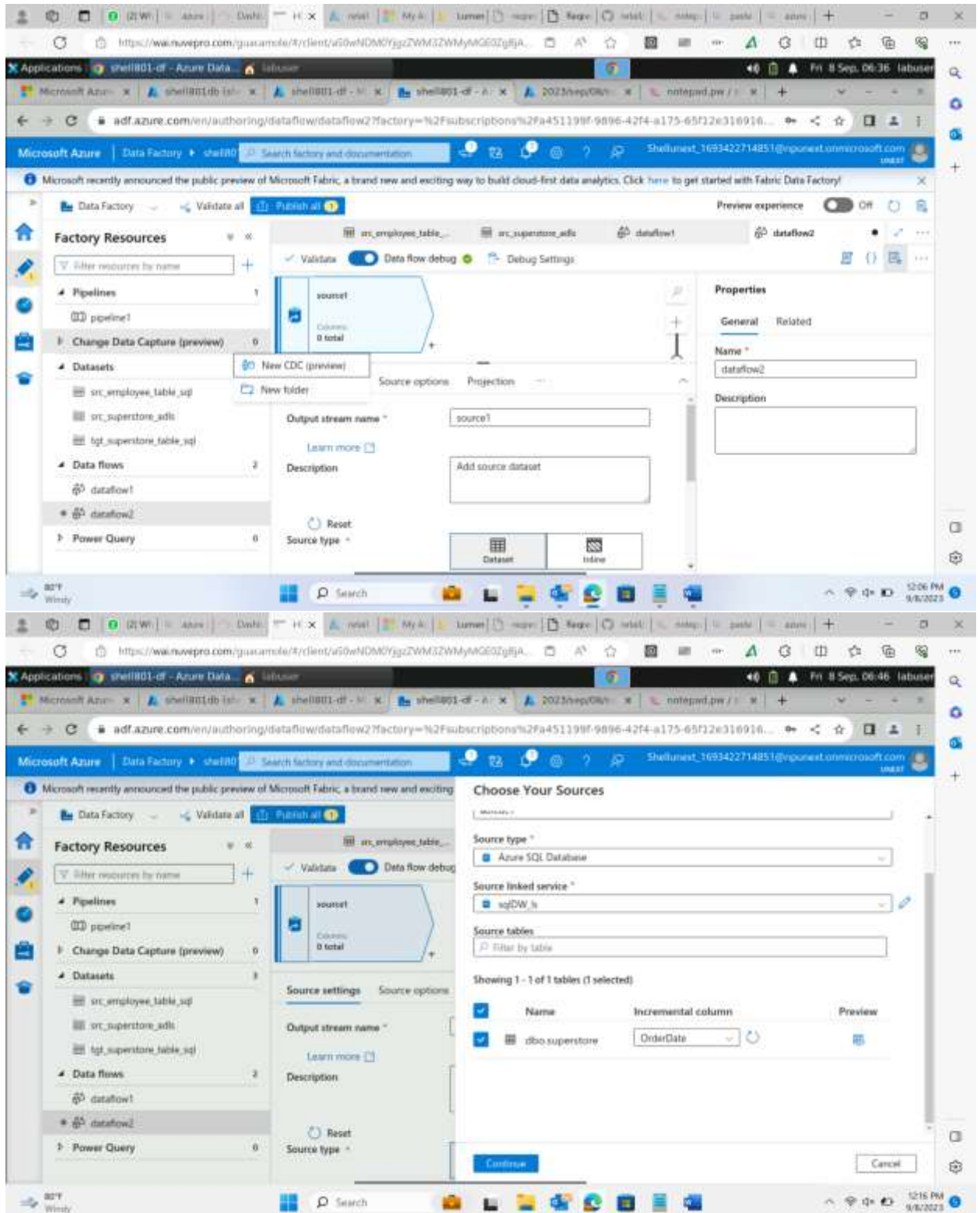
Query succeeded | 0s

Full import – first time load

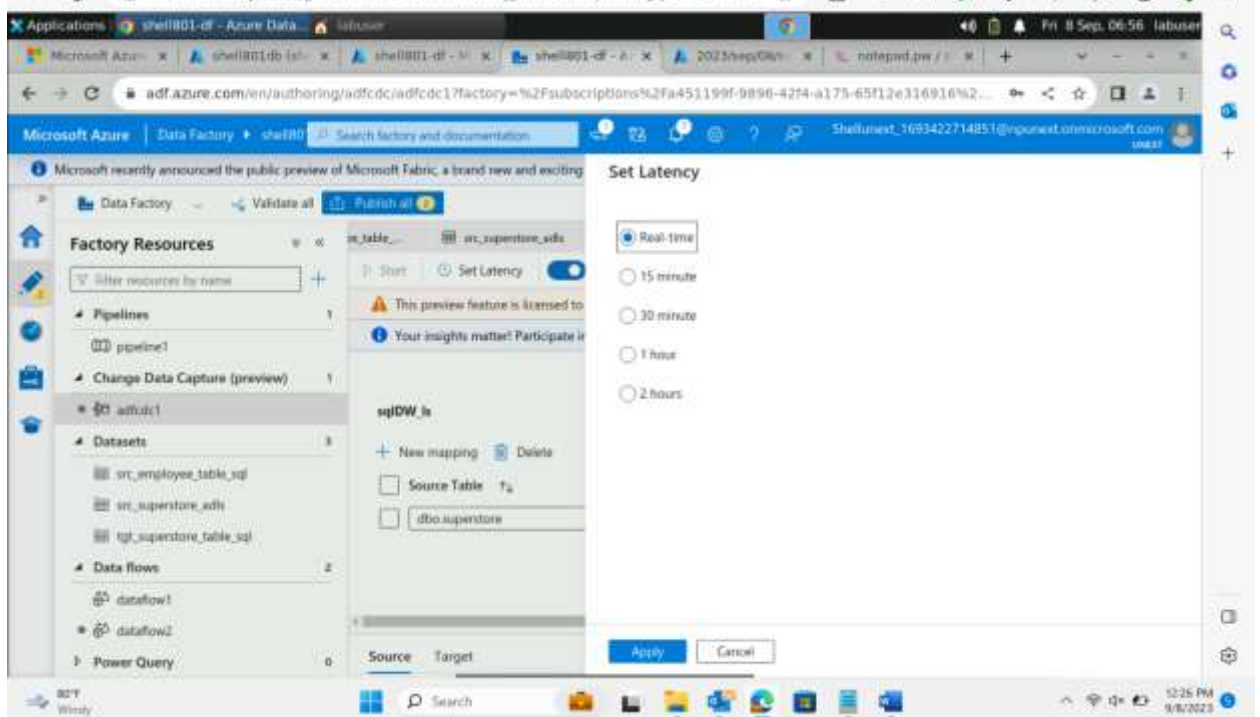
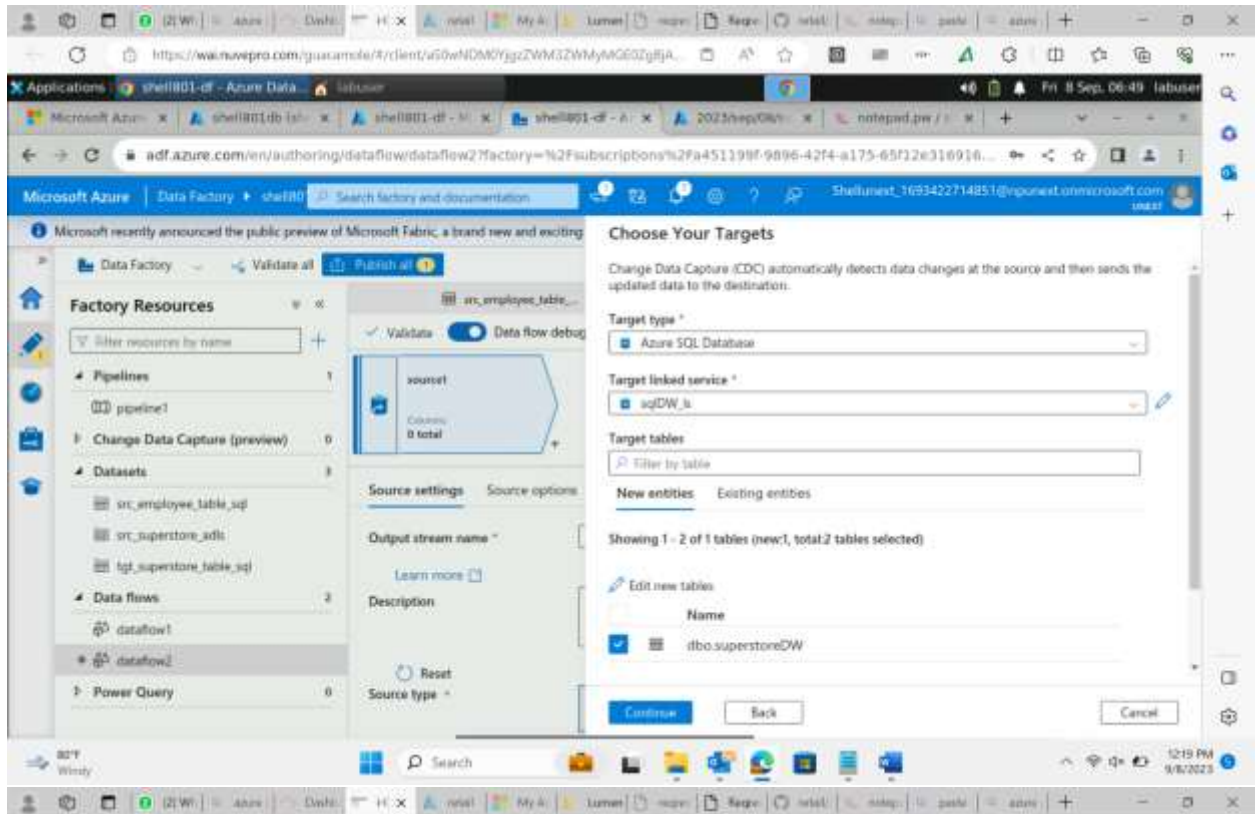
Incremental load/delta import – changes we make



CDC -



New entity



Publish
Start

Microsoft Azure | Data Factory | shell801-df | Search factory and documentation

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

Change Data Capture

Refresh Edit columns

Filter by name Time zone: Monrovia, Reykjavik (UTC+0) Status: All Latency: All

Showing 1 - 1 of 1 items

Name	Source	Target	Status	Latency	Last processed
adfdc1	sqdW_h	sqdW_h	Running	Real-time	—

80°F Windy 12:32 PM 9/8/2023

Microsoft Azure | Data Factory | shell801-df | Search factory and documentation

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

Choose Your Sources

Change Data Capture (CDC) automatically detects data changes at the source and then sends the updated data to the destination.

CDC name * adfdc2

Source type * DelimitedText

Source linked service * ADLS_conn

Source settings

Folder paths retail/2023/sep/08

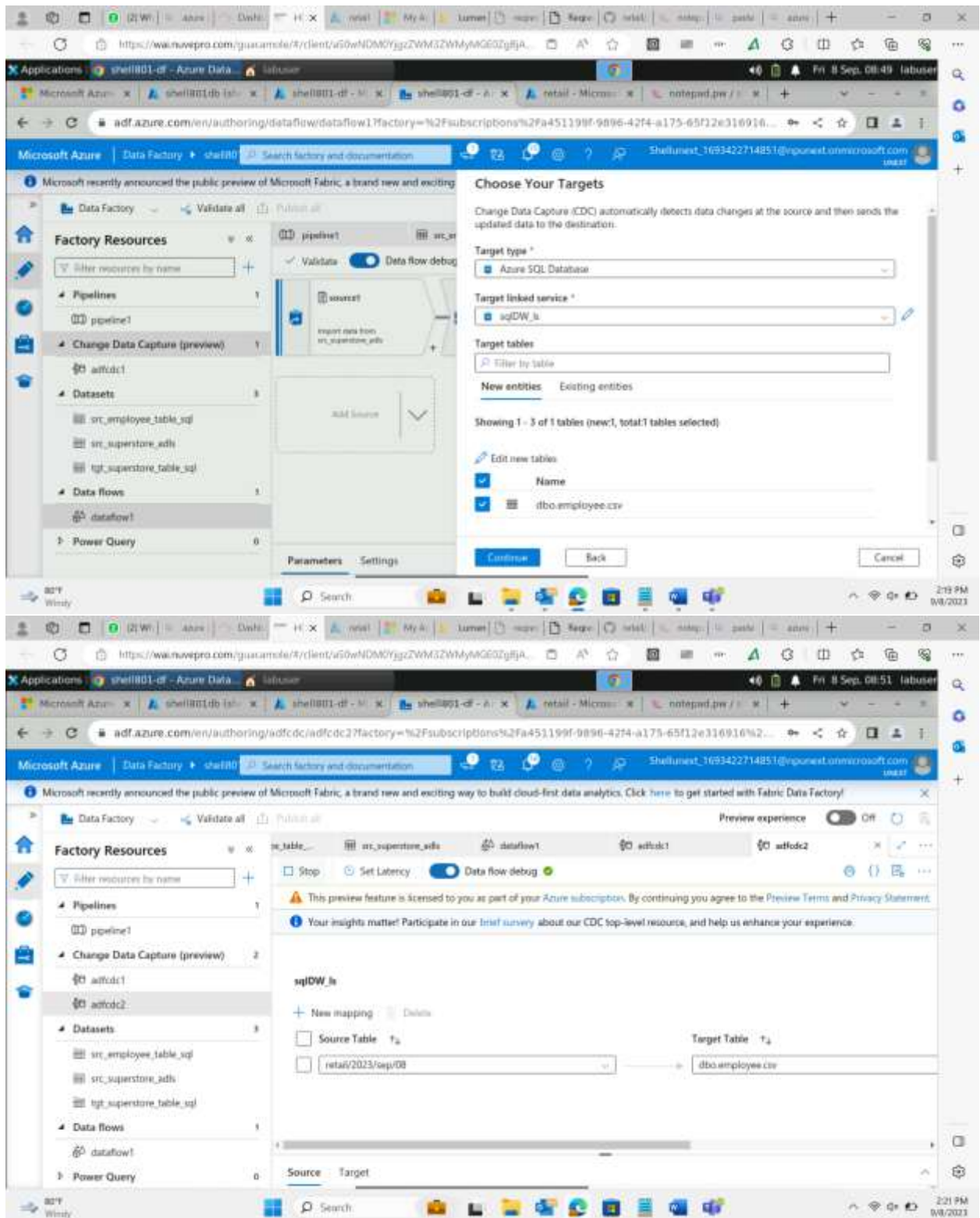
Continue Cancel

Factory Resources

- Pipelines 1
 - pipeline1
- Change Data Capture (preview) 1
 - adfdc1
- Datasets 3
 - src_employee_table_sql
 - src_superstore_adls
 - tgt_superstore_table_sql
- Data flows 1
 - dataflow1
- Power Query 0

Parameters Settings

80°F Windy 2:16 PM 9/8/2023



Edit employee.csv in storage
Delete one row and update something

Lookup activity

The screenshot displays the Microsoft Azure Data Factory (ADF) interface. The left sidebar shows the 'Factory Resources' tree with 'Pipelines' expanded, showing 'pipeline1'. The main canvas displays the 'Preview data' view for a 'Lookup' activity. The 'Linked service' is 'sqlDW_js' and the 'Object' is 'tgt_superstore_table_sql'. The data preview shows a table with two columns: 'region' and 'market'.

	region	market
1	East	US
2	Oceania	APAC
3	Oceania	APAC
4	Africa	Africa
5	East	US
6	Oceania	APAC
7	Oceania	APAC
8	Africa	Africa
9	East	US

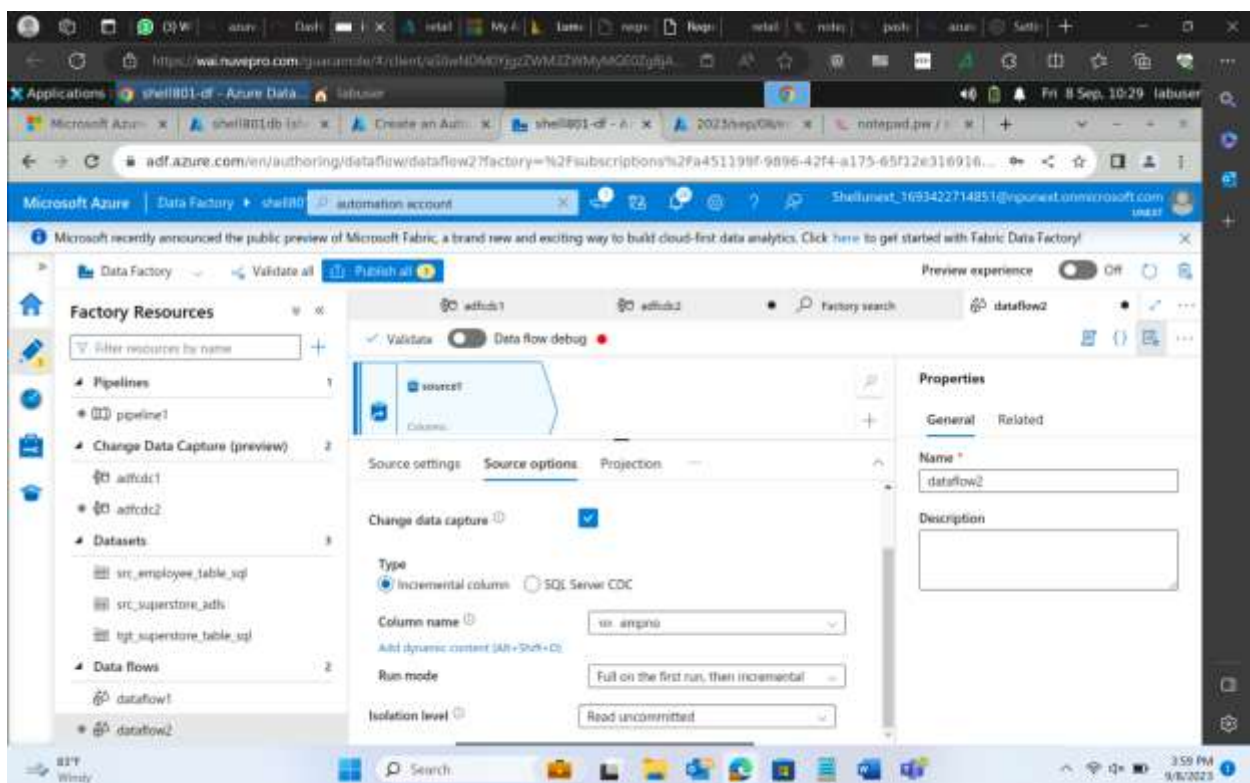
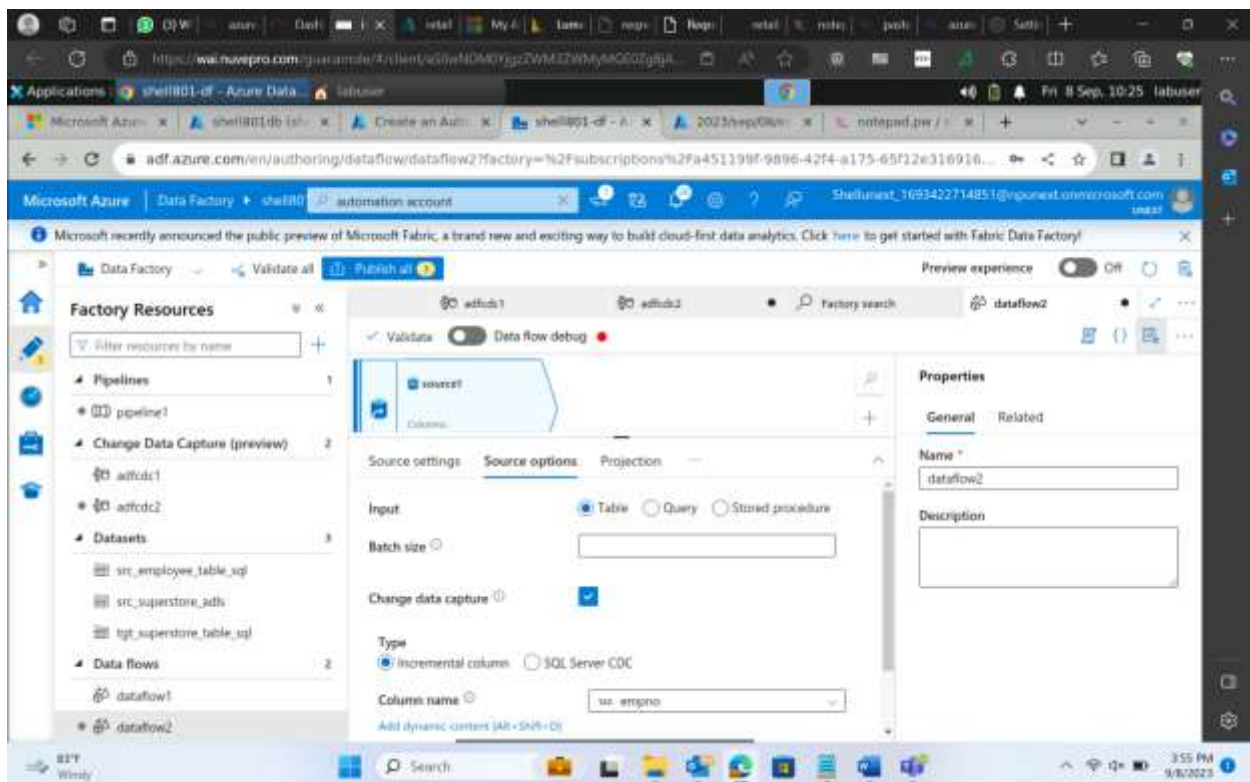
The right sidebar shows the 'Lookup' activity settings. The 'Source dataset' is 'tgt_superstore_table_sql'. The 'First row only' checkbox is unchecked. The 'Use query' radio button is selected, and the query is 'select region, market from superstore'. The 'Query' field is highlighted with a blue selection bar.

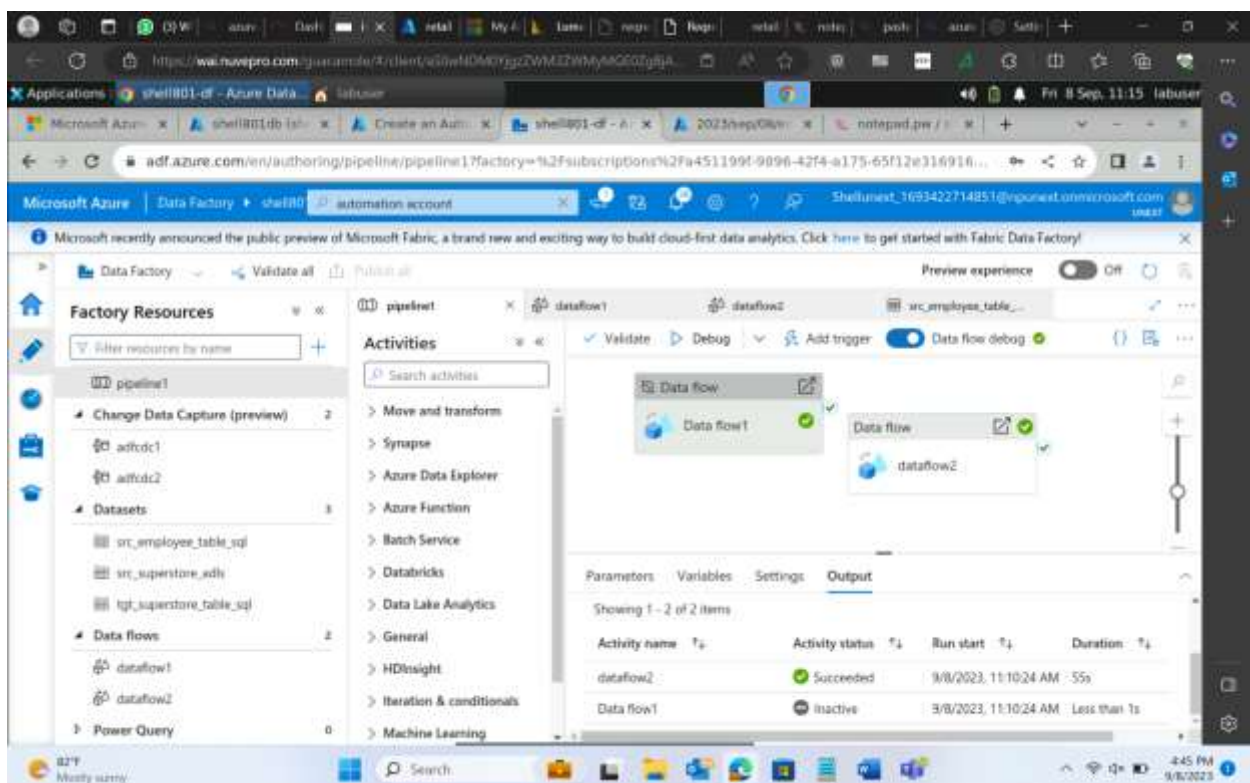
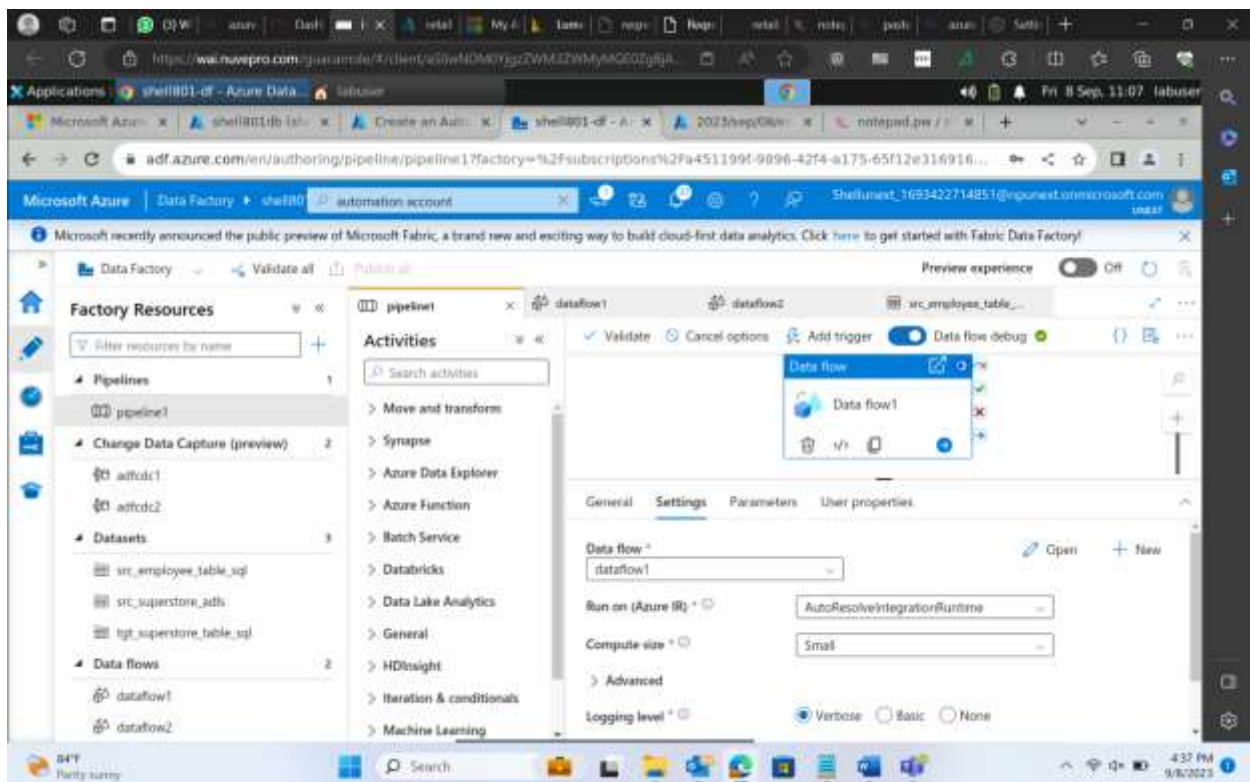
ADF – ETL tool

Orchestrating the data

- 1. Pipeline**
- 2. Activity**
- 3. Control flow**
- 4. Get metadata – get information from file (childitems, file size, etc)**
- 5. Lookup – get information for table**
- 6. Copy data – copying bulk data (source and sink)**
- 7. Linked service and datasets are reusable.**
- 8. Foreach – iteration**
- 9. If**
- 10. Change data capture (CDC) – in source if your data updates in real time.**
- 11. Full import and incremental load**
- 12. SCD story changing dimension – surrogate key – flag, version, begin-date end-date**
- 13. Data flow – particular data transformation/processing**
- 14. Source, derived column → total sales, concat, extract, joins, conditional split, sink, union→**
- 15. Join __> joining tables on condition based**
- 16. Union → join tables**
- 17. Videos, large datasets are stored in separate containers.**
- 18. Dataset → have info about datain**

Automation accounts





Logging level :L none if data flow gives error