ETL - extract, transform, load

New runtime to perform transformations

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

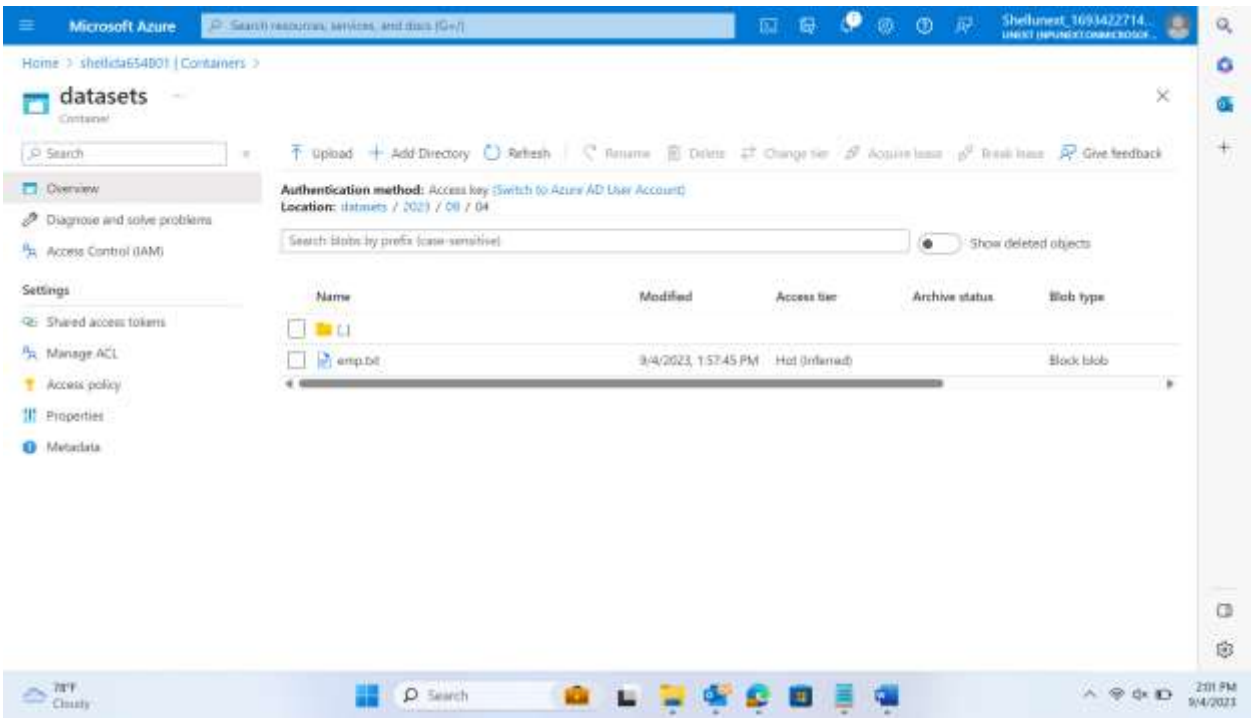
****

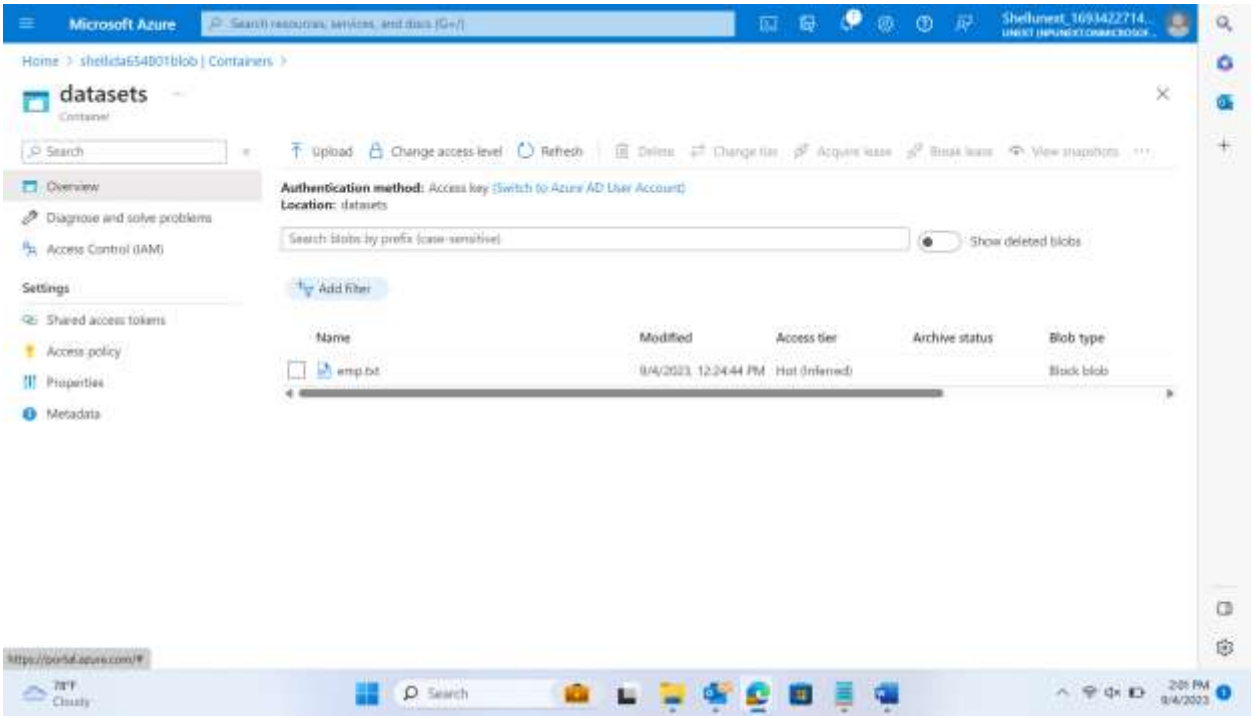
resource group

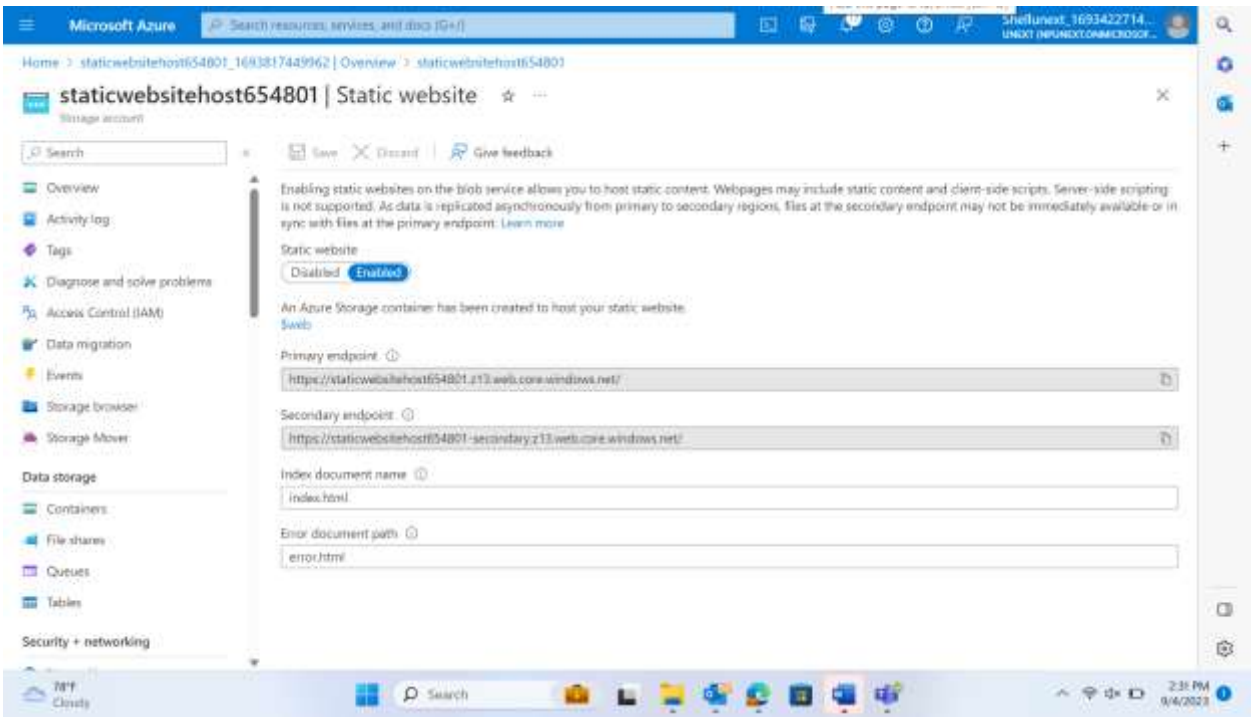
blob storage

key vault --> secrets --> blob key

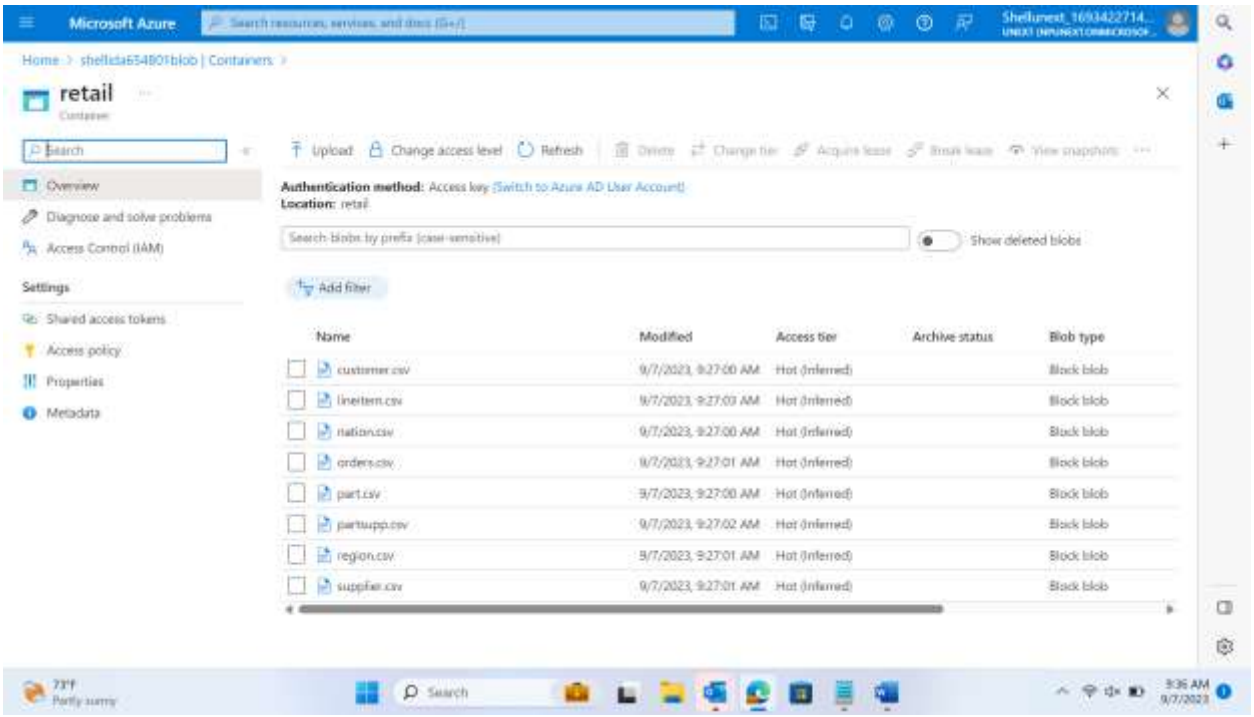
adls storage



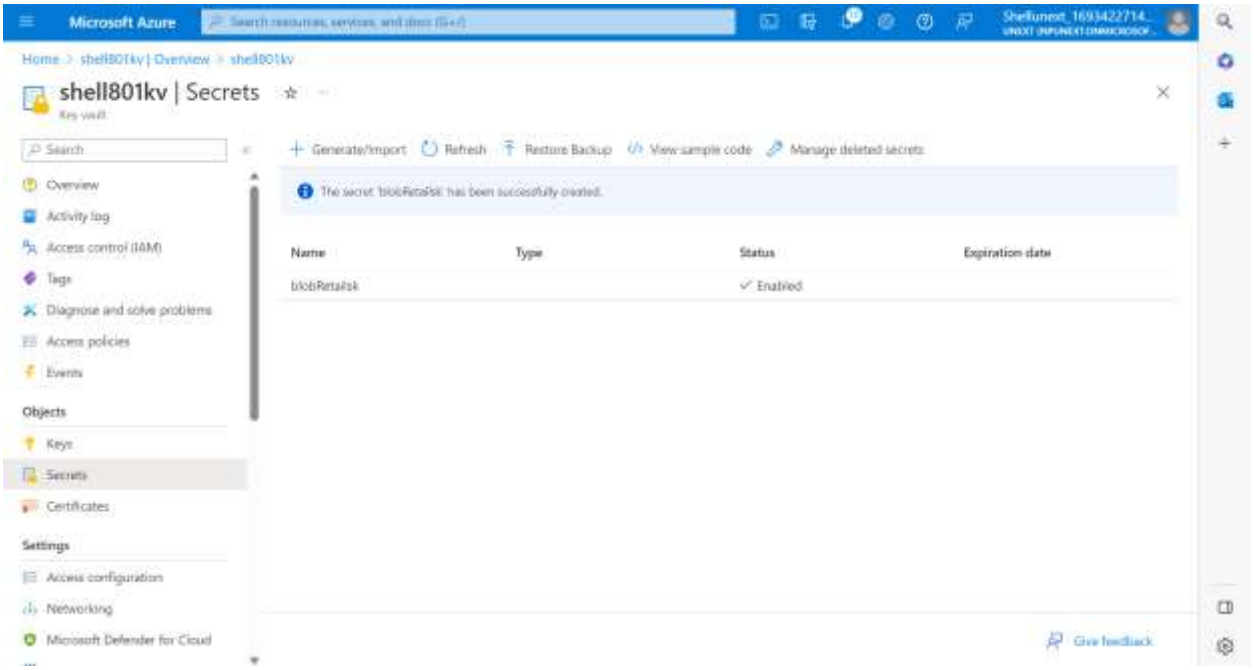

Microsoft Azure
Home > sheilda654801blob | Containers >
datasets
Container
Search
Upload
Change access level
Refresh
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Access policy
Properties
Metadata
Authentication method: Access key (Switch to Azure AD User Account)
Location: datasets
Search blobs by prefix (case-sensitive)
Show deleted blobs
Add Filter
Name
Modified
Access tier
Archive status
Blob type
emp.txt
9/4/2023, 12:24:44 PM
Hot (Inferred)
Block blob


Microsoft Azure
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.
$web
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

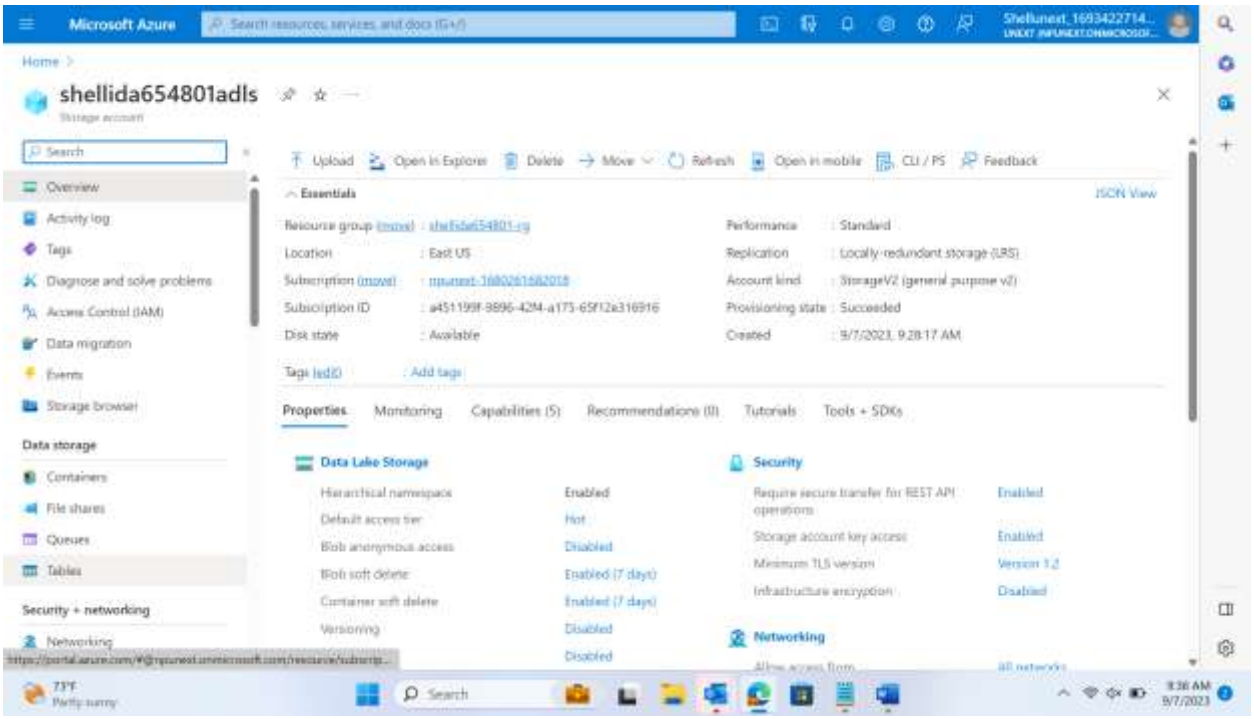
Blob storage



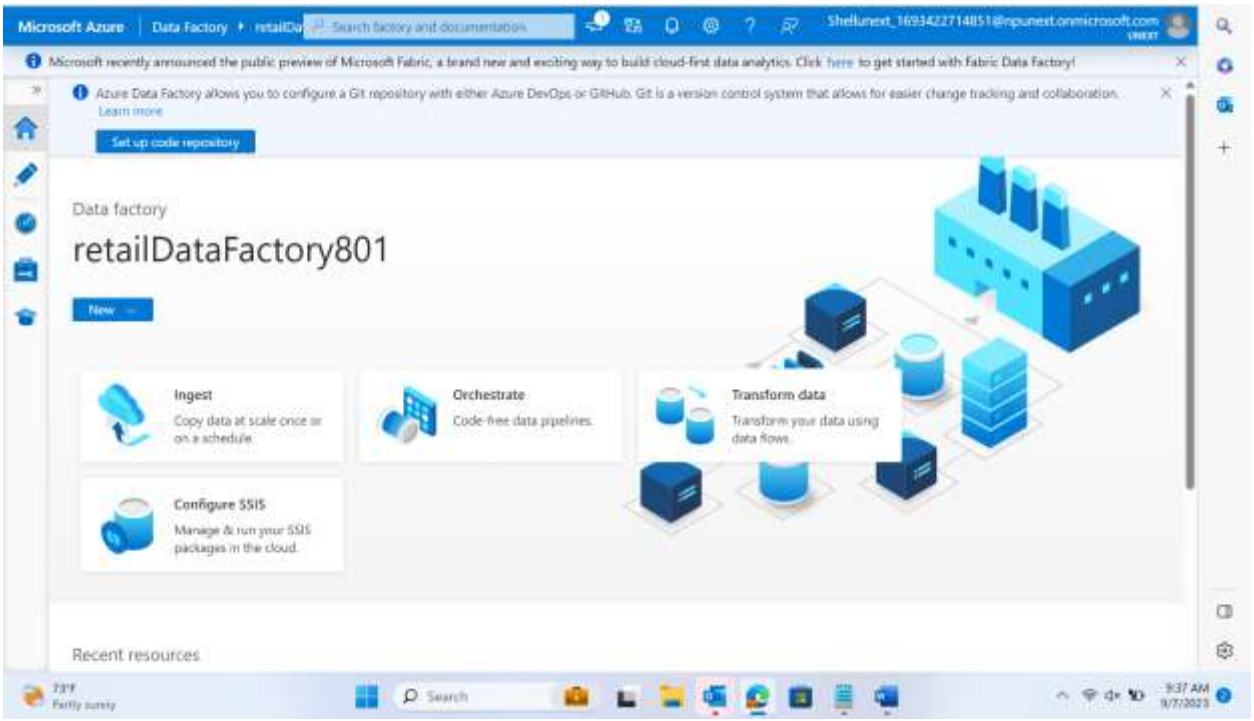
Key vault



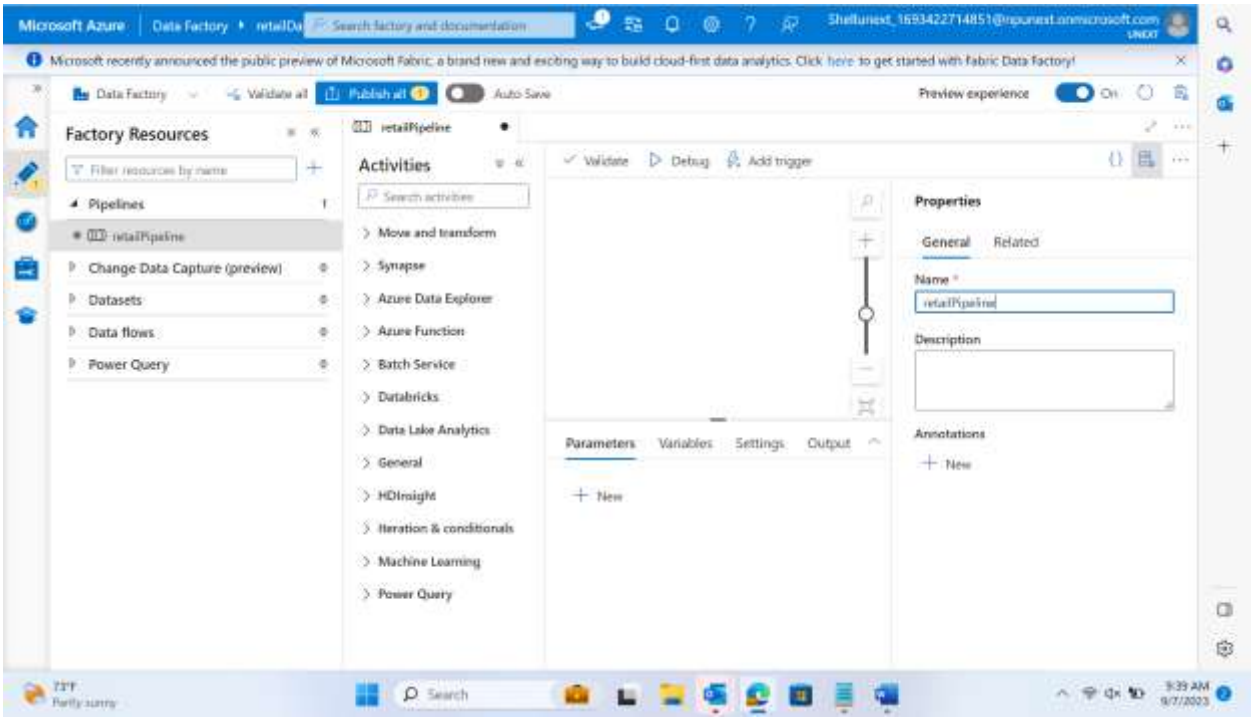
Adls storage



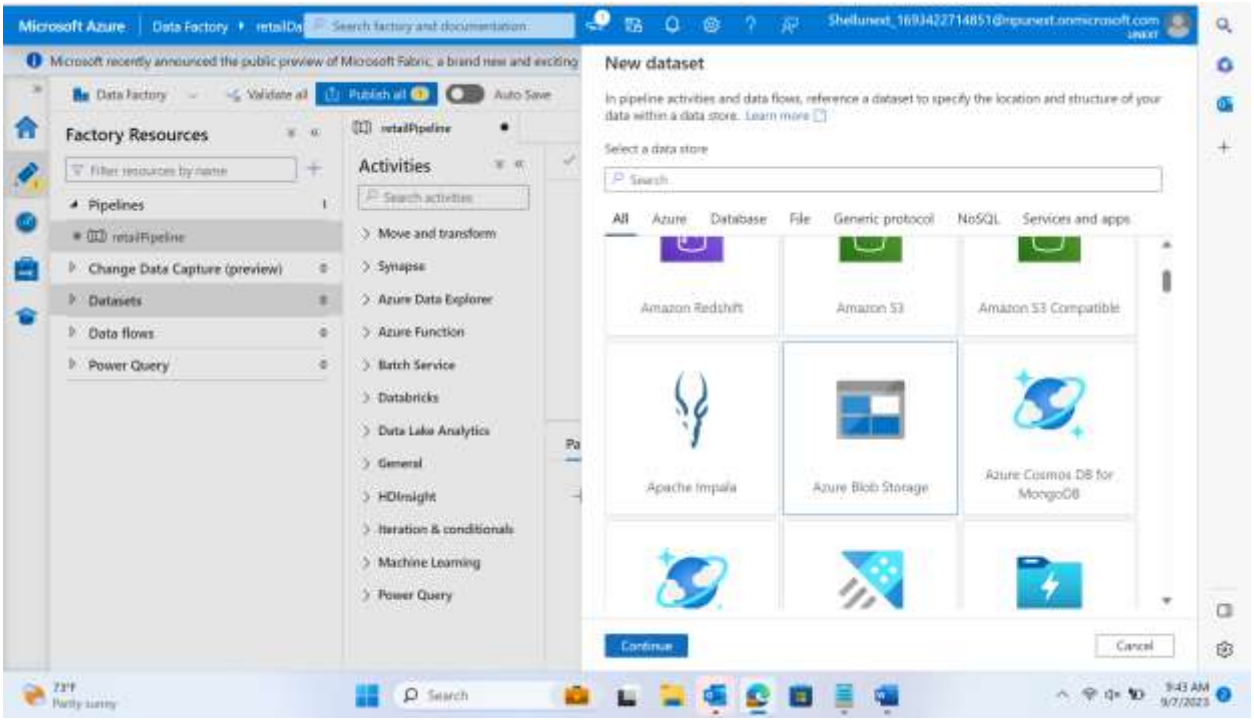
Data factory



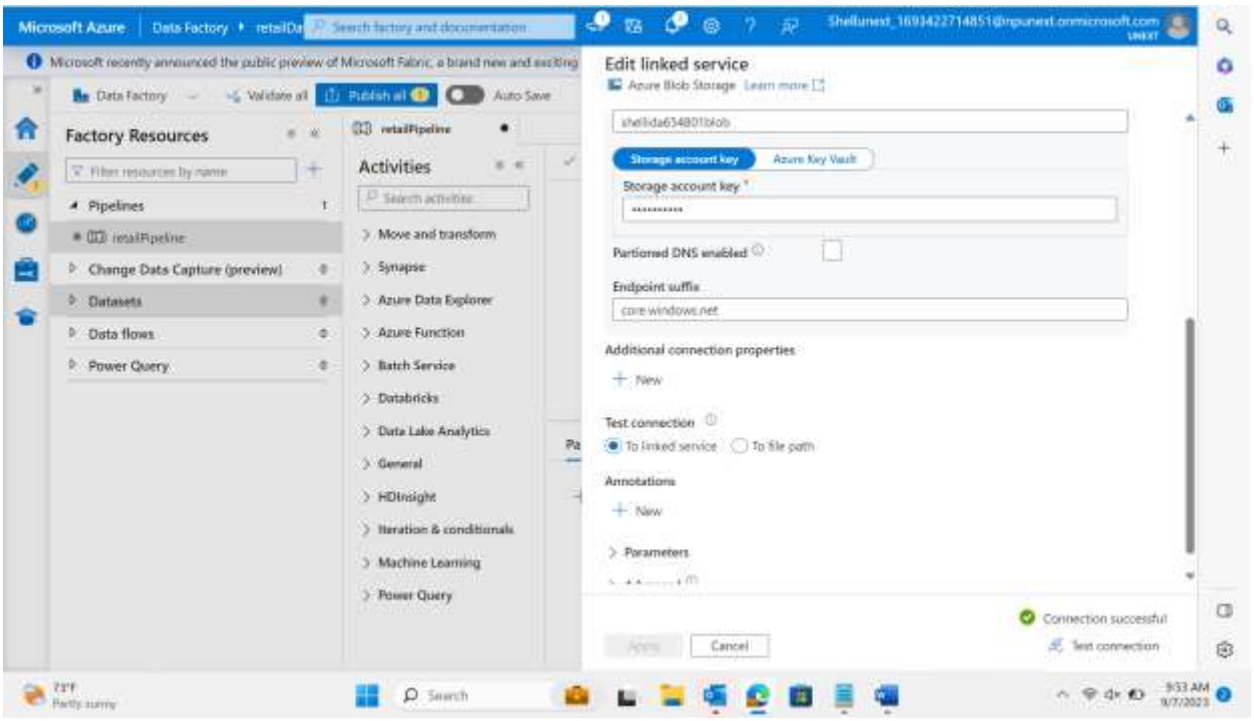
Pipeline

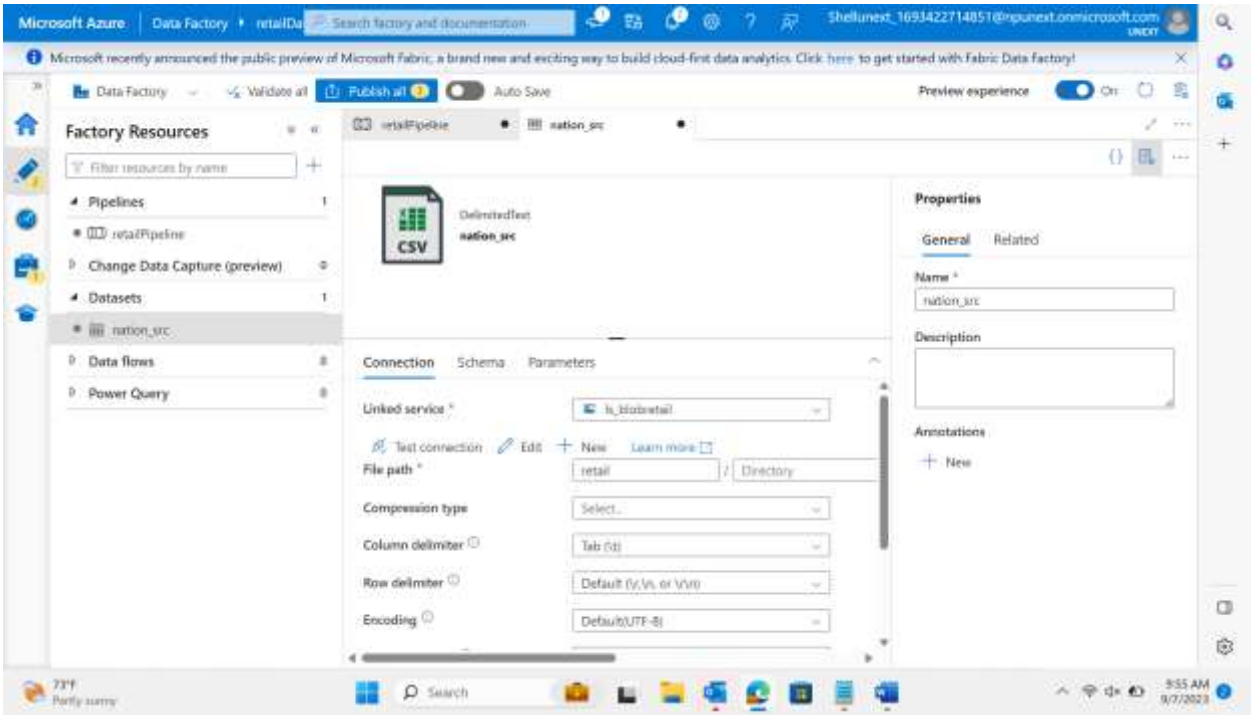


Dataset


Microsoft Azure
Data Factory
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 Redshift
Amazon S3
Amazon S3 Compatible
Apache Impala
Azure Blob Storage
Azure Cosmos DB for MongoDB
Continue
Cancel
Factory Resources
Pipelines
retailPipeline
Change Data Capture (preview)
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
Power Query

Ls


Microsoft Azure
Data Factory
Edit linked service
Azure Blob Storage
Storage account key
Azure Key Vault
Storage account key
Partioned DNS enabled
Endpoint suffix
core.windows.net
Additional connection properties
New
Test connection
To linked service
To file path
Annotations
New
Parameters
Connection successful
Test connection
Apply
Cancel
Factory Resources
Pipelines
retailPipeline
Change Data Capture (preview)
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
Power Query

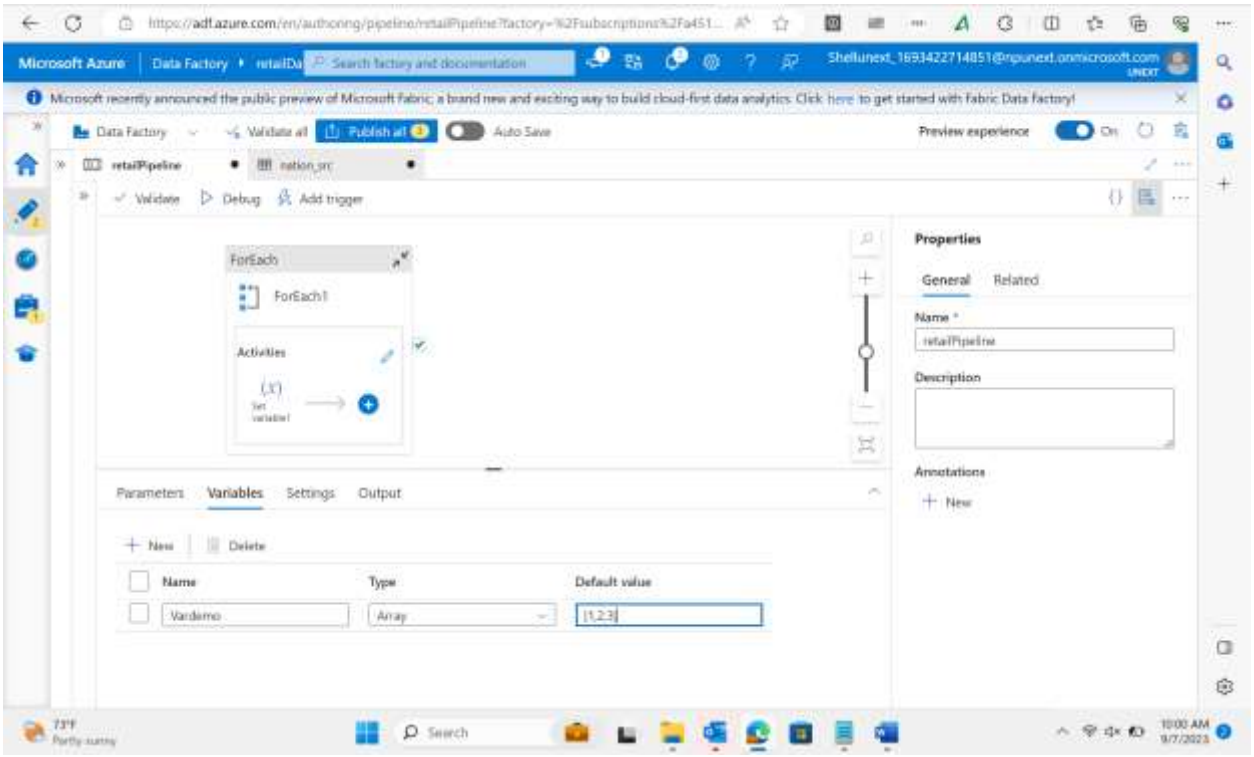


For each activity

Activity + set variable

Click outside

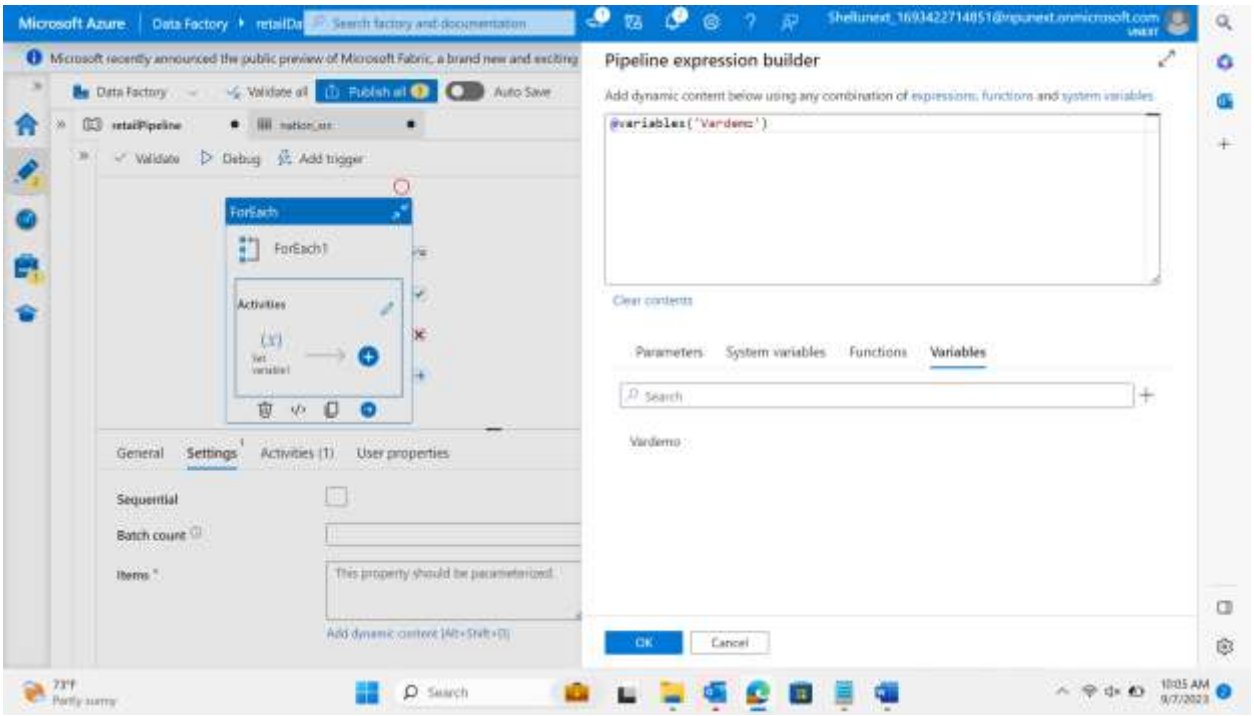
Variables → Vardemo, Array, [1,2,3]

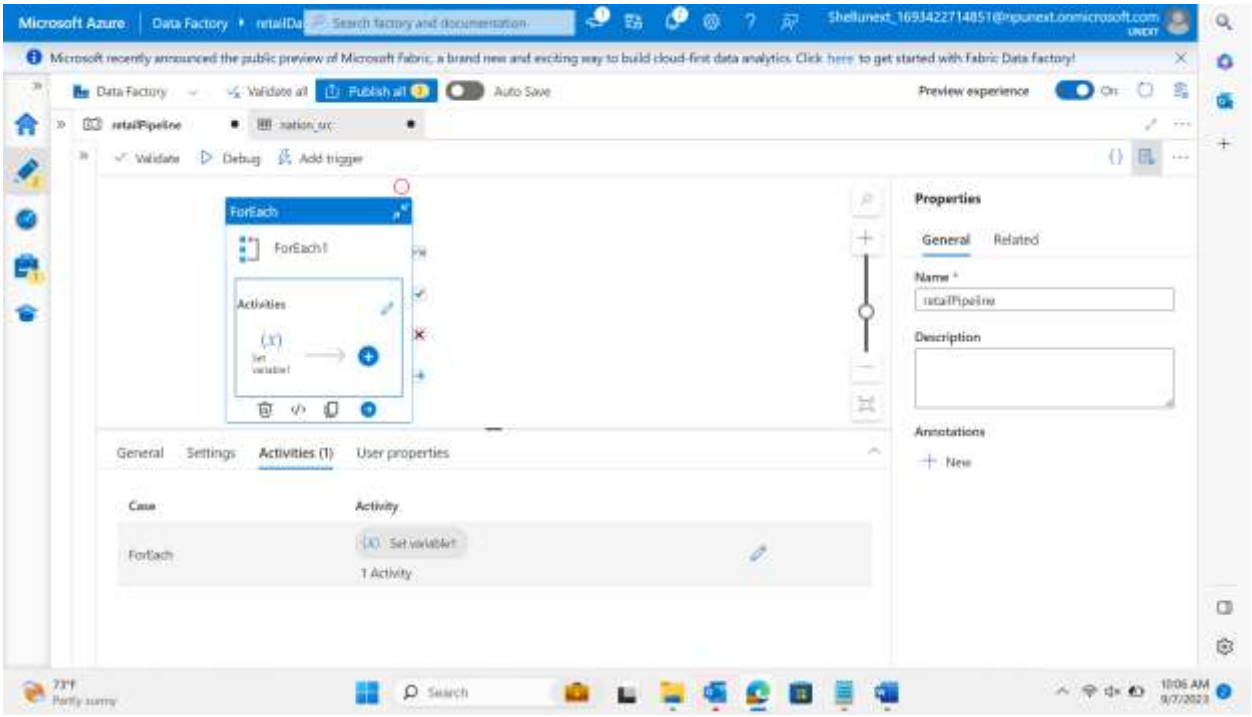


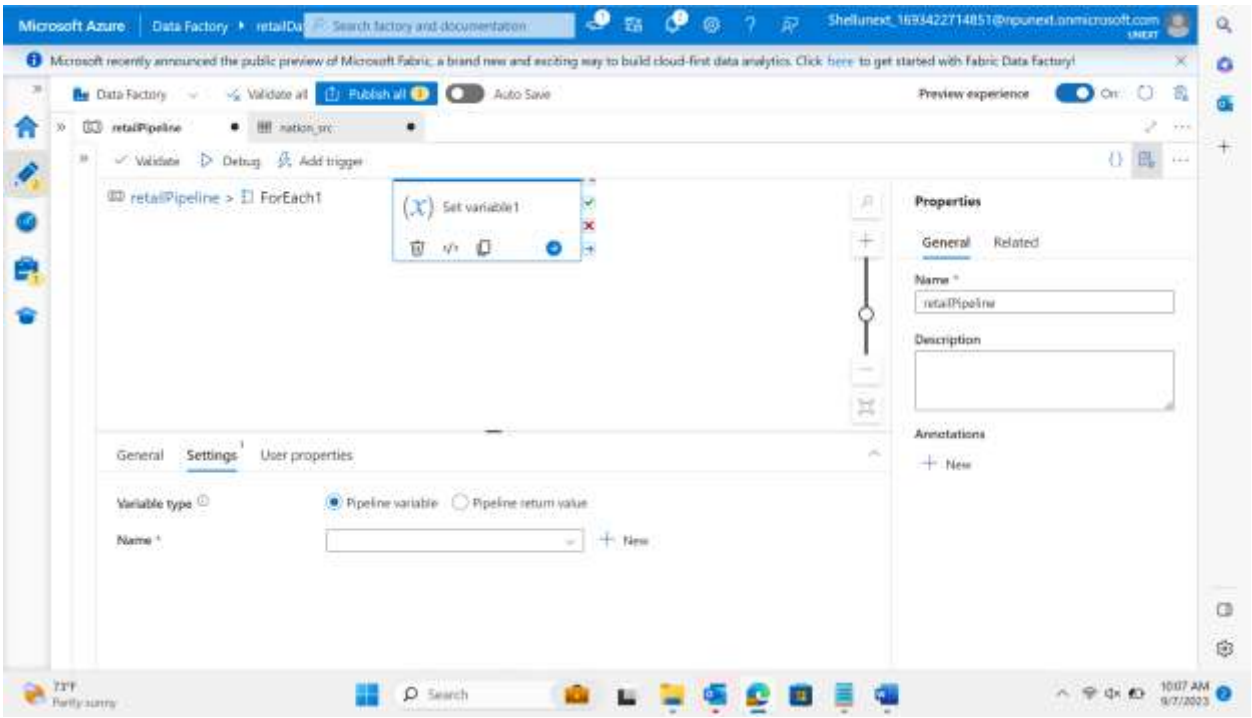
Parameters → dynamic content -value -runtime

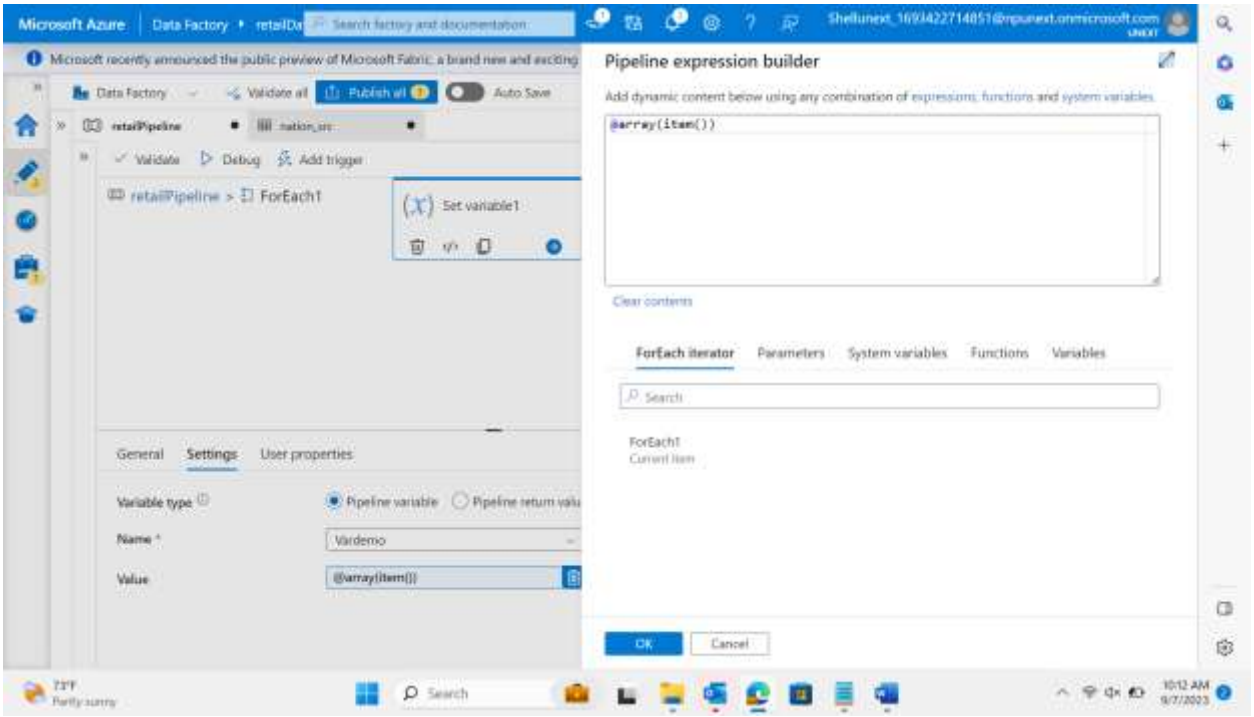
Variables → pipeline → append variable

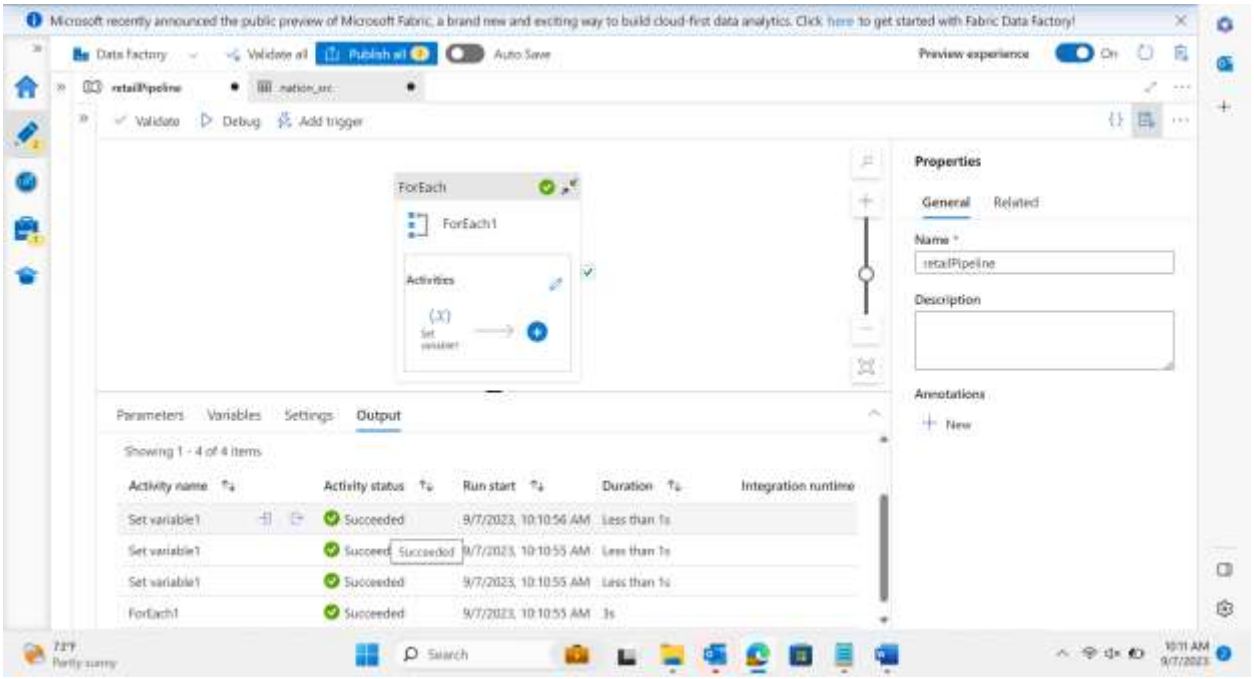
Double click on foreach → settings → items → add dynamic content

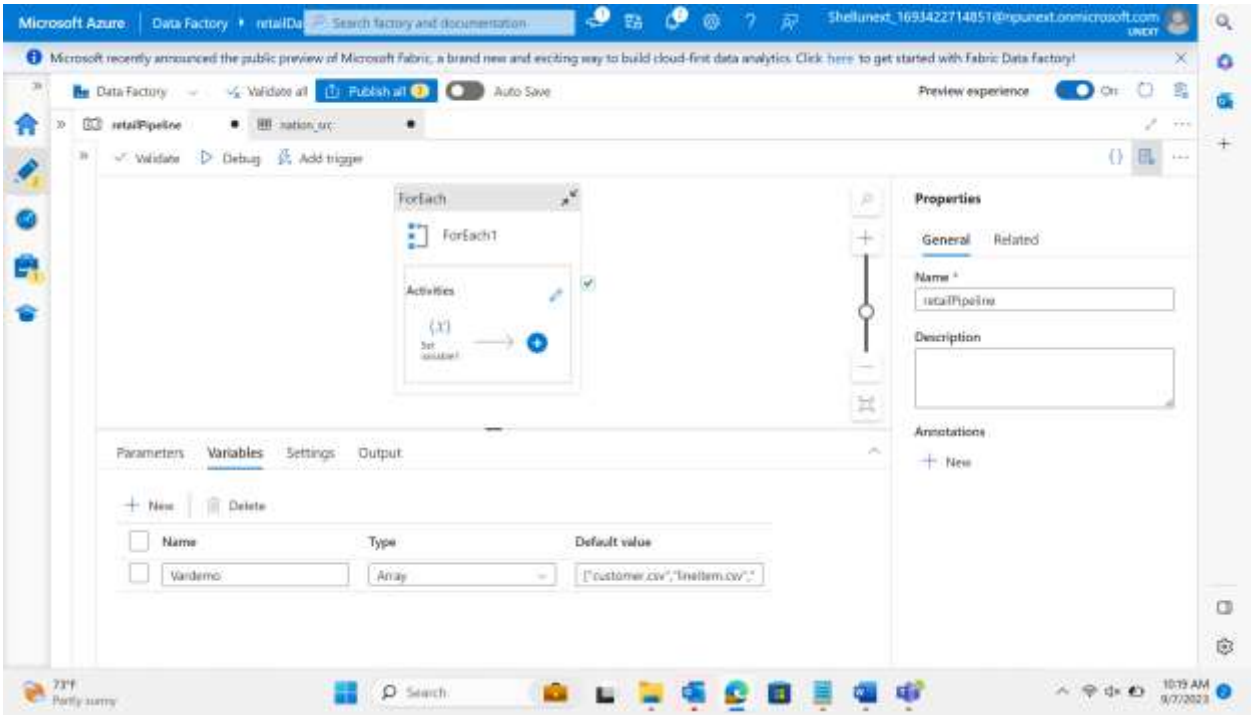


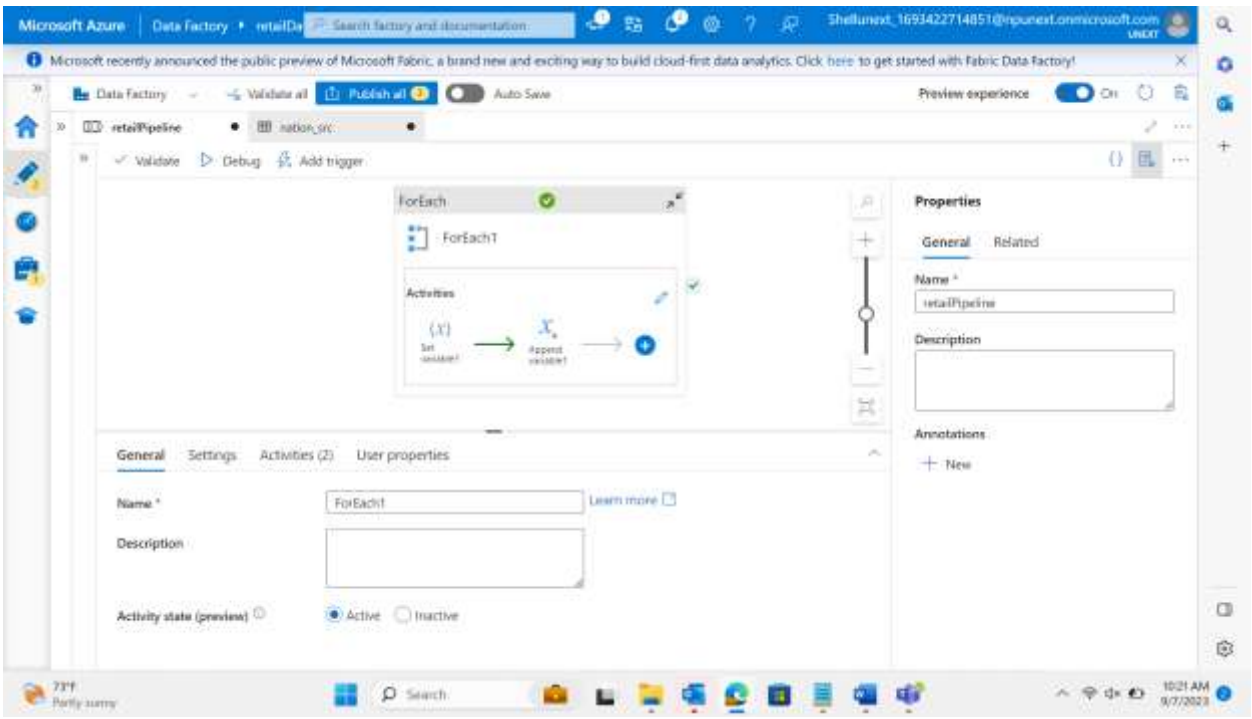

Microsoft Azure
Data Factory
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Validate all
Publish all
Auto Save
Preview experience
retailPipeline
nation_src
Validate
Debug
Add trigger
ForEach
ForEach1
Activities
Set variable1
Properties
General
Related
Name *
retailPipeline
Description
Annotations
New
General
Settings
Activities (1)
User properties
Case
Activity
ForEach
Set variable1
1 Activity


Microsoft Azure
Data Factory
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!
Validate all
Publish all
Auto Save
Preview experience
retailPipeline
nation_src
Validate
Debug
Add trigger
retailPipeline > ForEach1
Set variable1
Properties
General
Related
Name *
retailPipeline
Description
Annotations
New
General
Settings
User properties
Variable type
Pipeline variable
Pipeline return value
Name *
New


Microsoft Azure
Data Factory
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting
Validate all
Publish all
Auto Save
Validate
Debug
Add trigger
retailPipeline > ForEach1
Set variable1
General
Settings
User properties
Variable type
Pipeline variable
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

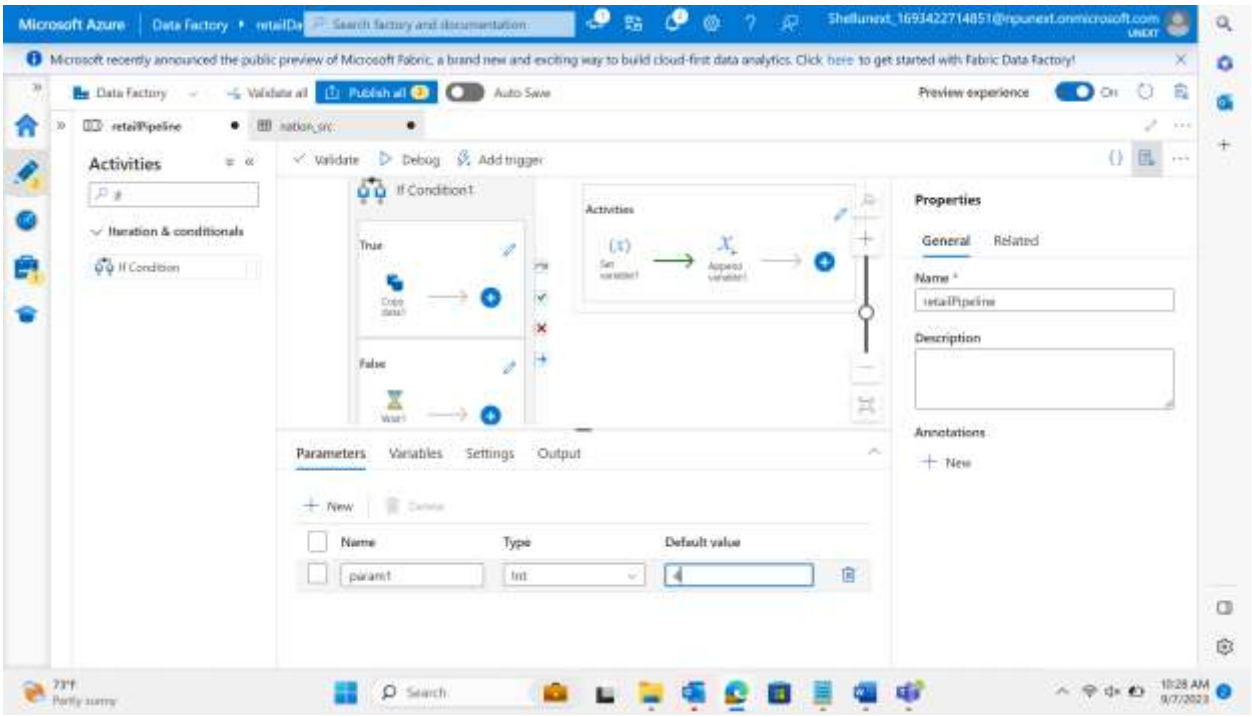

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Data Factory
Validate all
Publish all
Auto Save
Preview experience
On
Validate
Debug
Add trigger
ForEach
ForEach1
Activities
Properties
General
Related
Name *
retailPipeline
Description
Annotations
New
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





If condition

Pipeline parameter



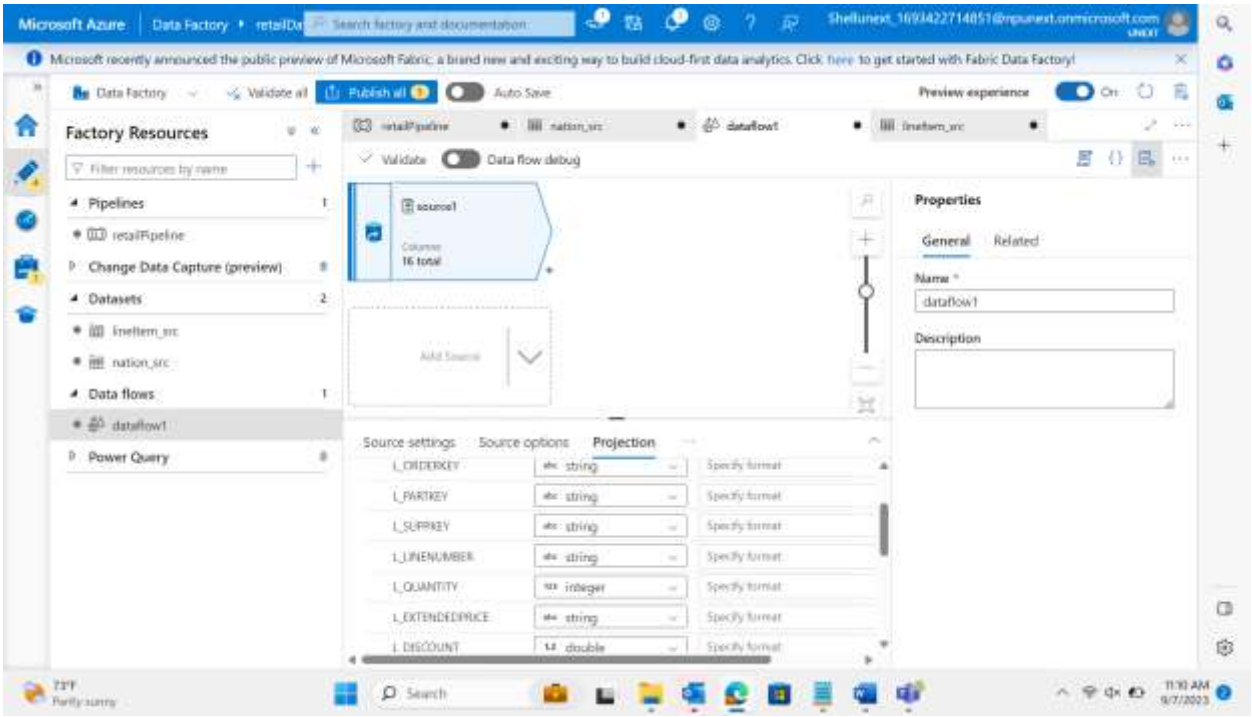
Lubricant retail sales data :

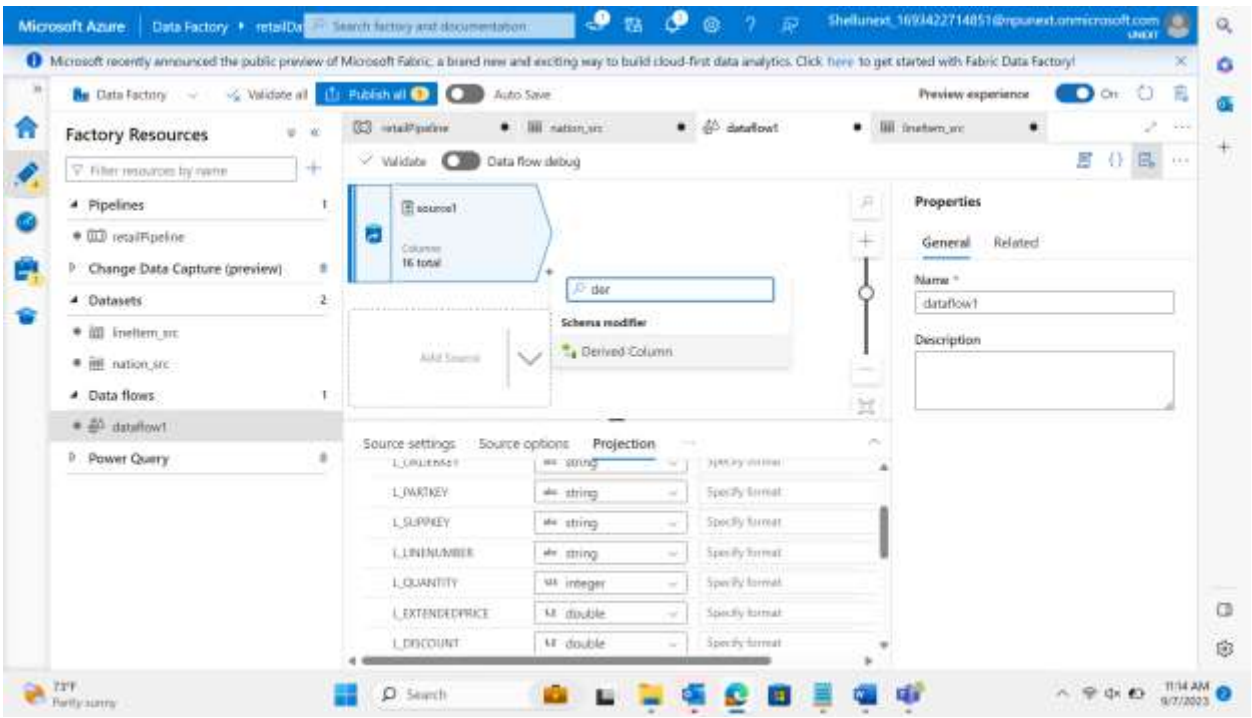
Data Flow -> transformation

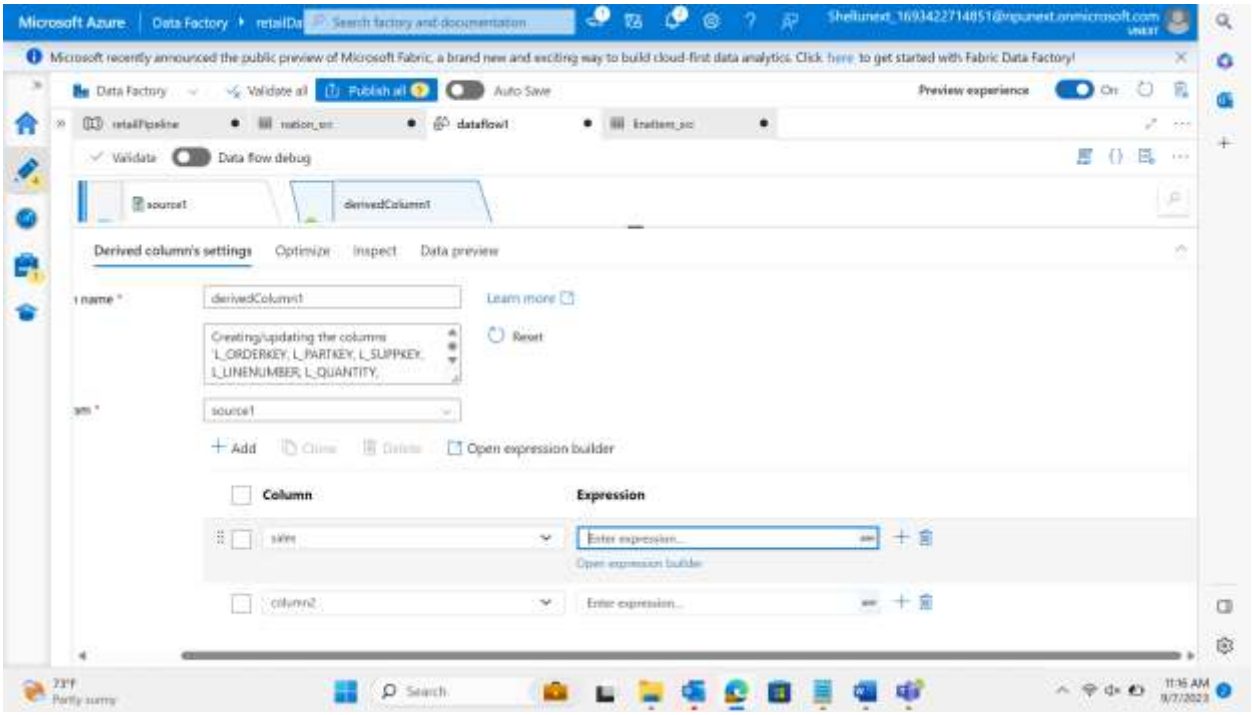
Line item --> column →

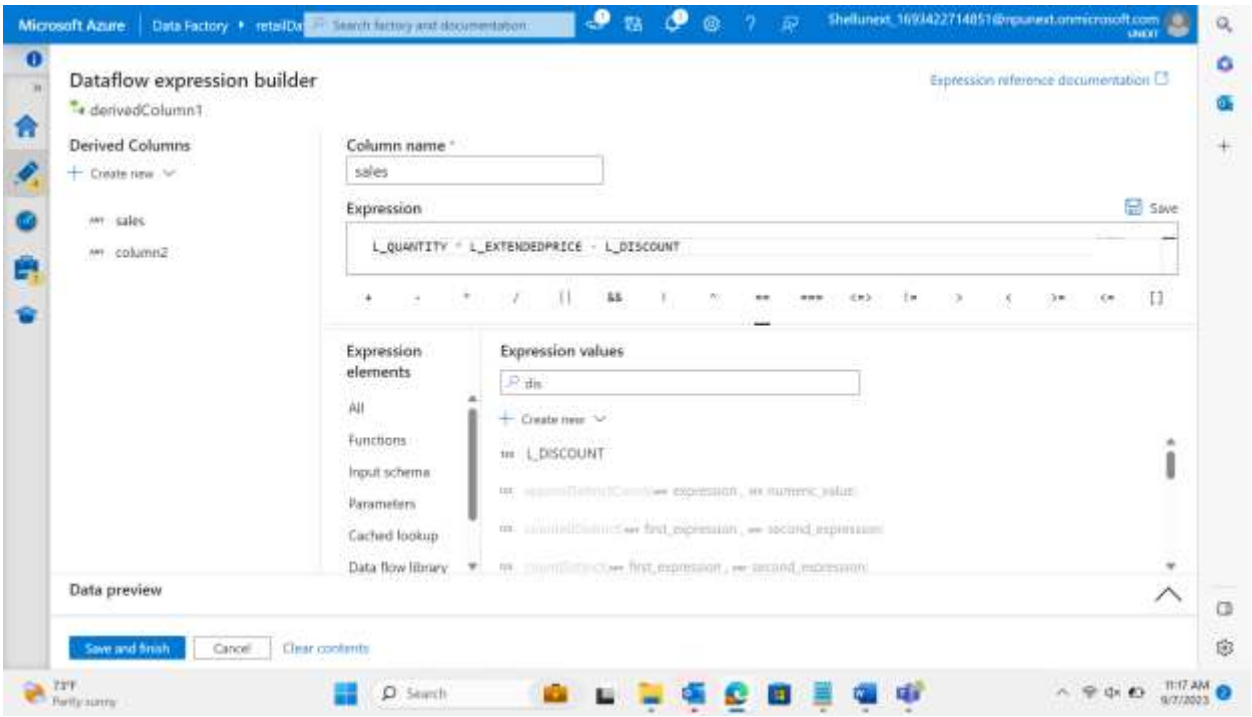
Df1 →

Calculate something which is not present in source table

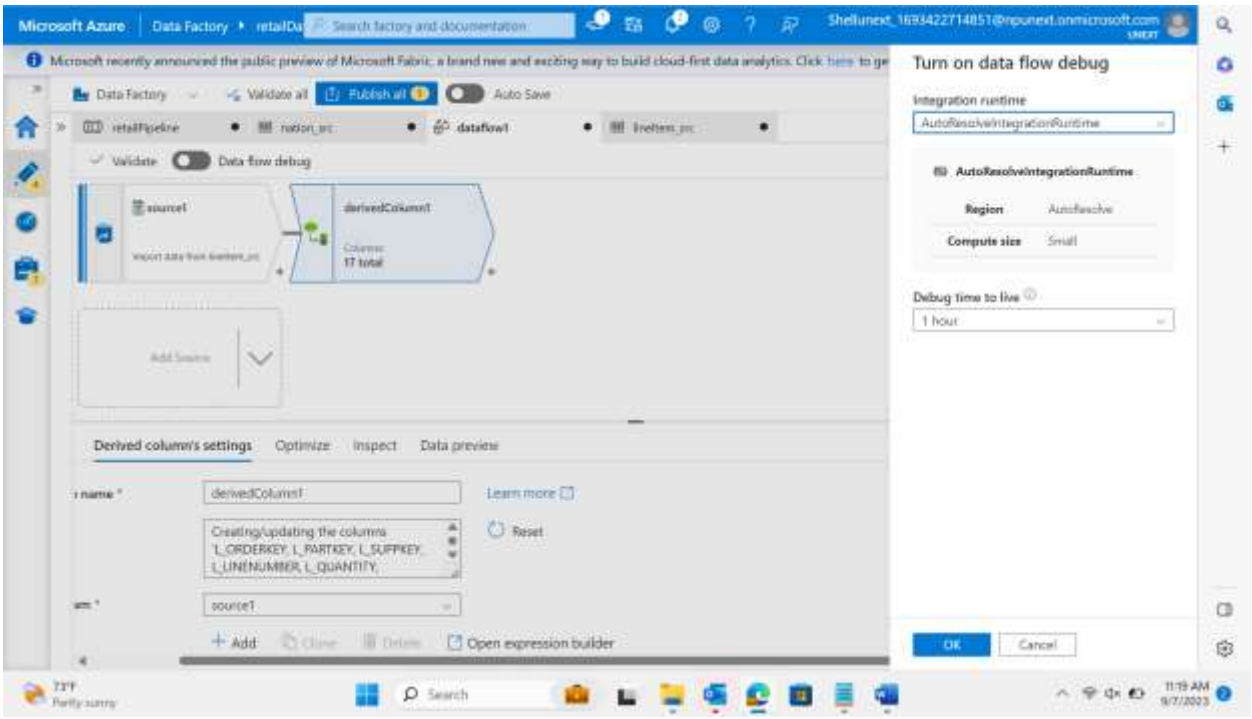

Microsoft Azure
Data Factory
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Validate all
Publish all
Auto Save
Preview experience
Factory Resources
Pipelines
retailPipeline
Change Data Capture (preview)
Datasets
lineitem_src
nation_src
Data flows
dataflow1
Power Query
Validate
Data flow debug
source1
16 total
Properties
General
Related
Name *
dataflow1
Description
Source settings
Source options
Projection
L_ORDERKEY
L_PARTKEY
L_SUPPKEY
L_LINENUMBER
L_QUANTITY
L_EXTENDEDPRICE
L_DISCOUNT
string
integer
double

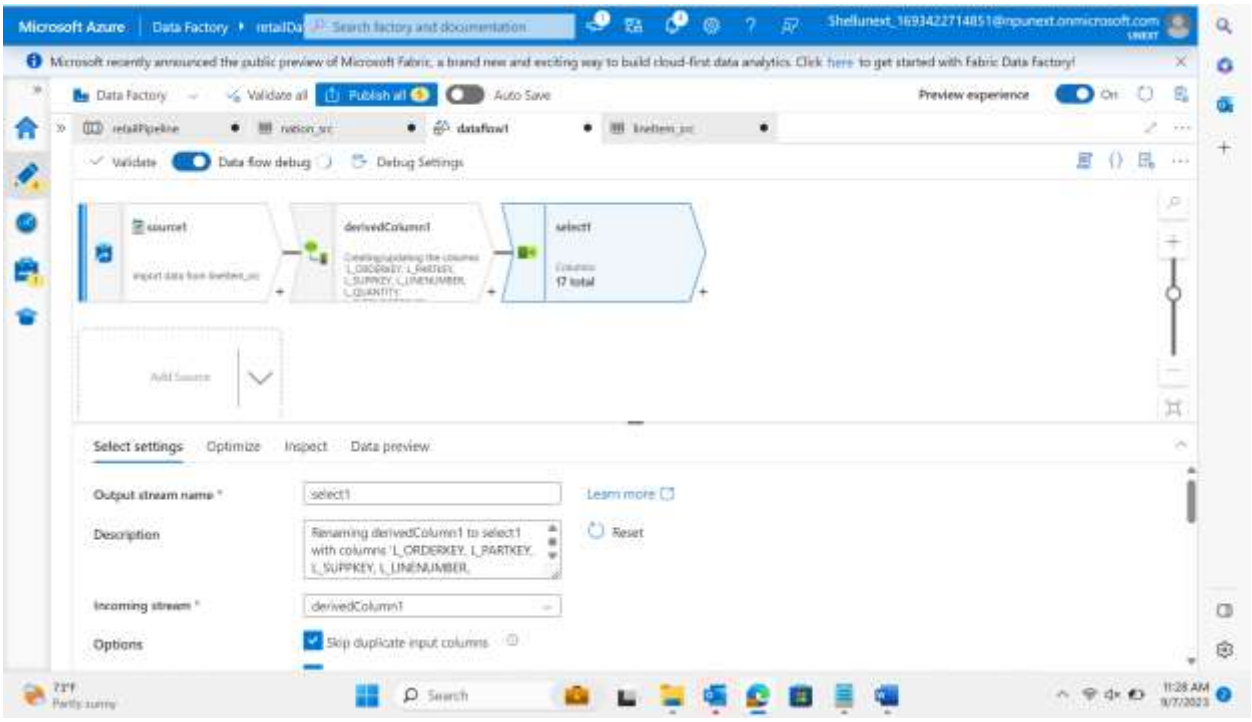

Microsoft Azure
Data Factory
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Validate all
Publish all
Auto Save
Preview experience
Factory Resources
Pipelines
retailPipeline
Change Data Capture (preview)
Datasets
lineitem_src
nation_src
Data flows
dataflow1
Power Query
Validate
Data flow debug
source1
16 total
der
Schema modifier
Derived Column
Properties
General
Related
Name *
dataflow1
Description
Source settings
Source options
Projection
L_PARTKEY
L_SUPPKEY
L_LINENUMBER
L_QUANTITY
L_EXTENDEDPRICE
L_DISCOUNT
string
integer
double



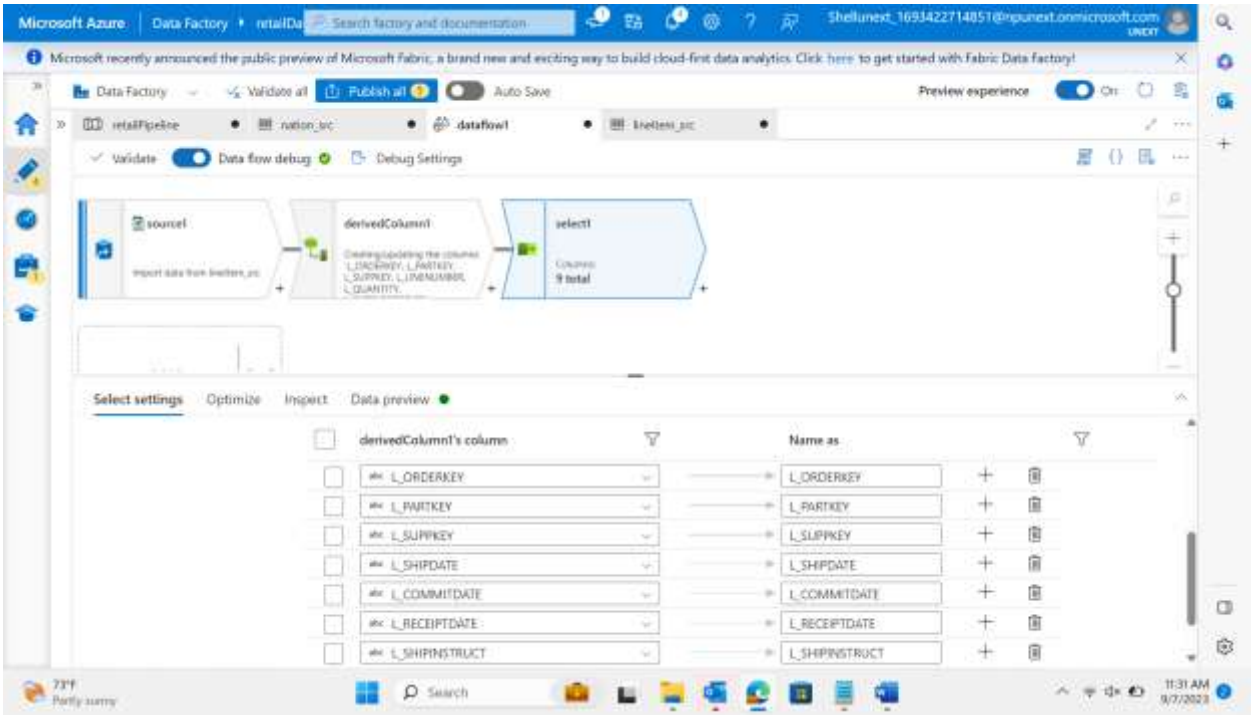


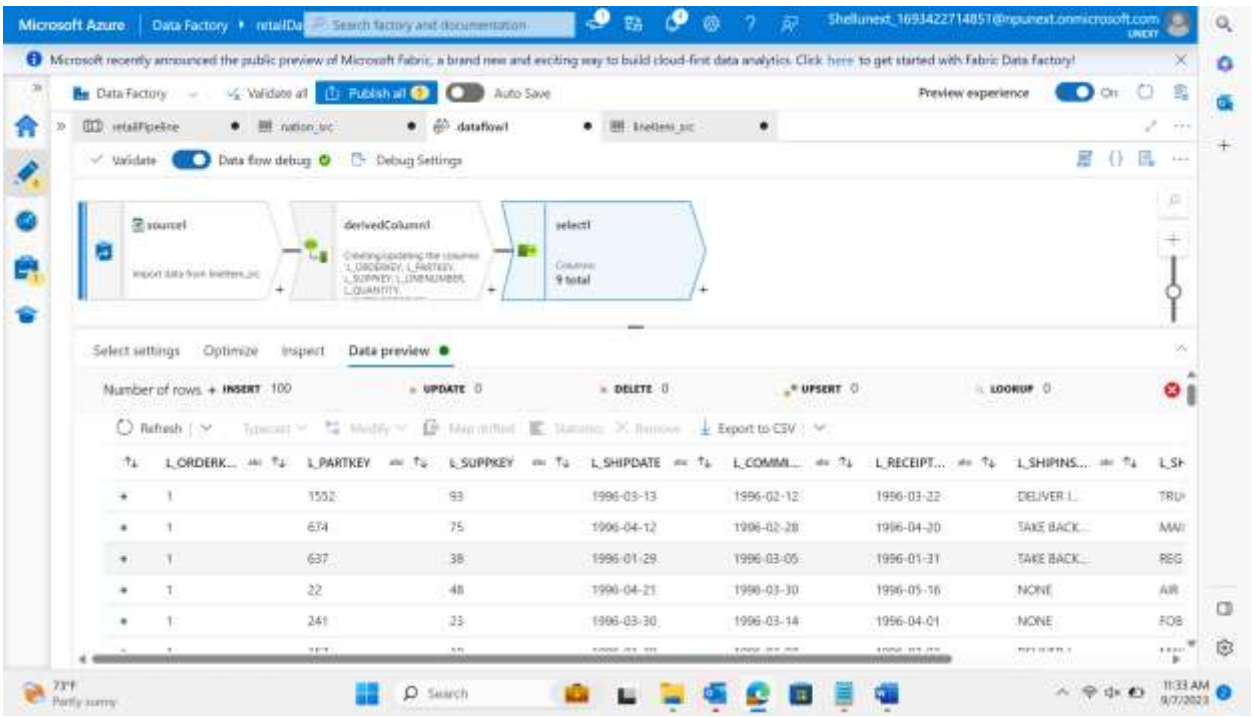
Enable data flow debug



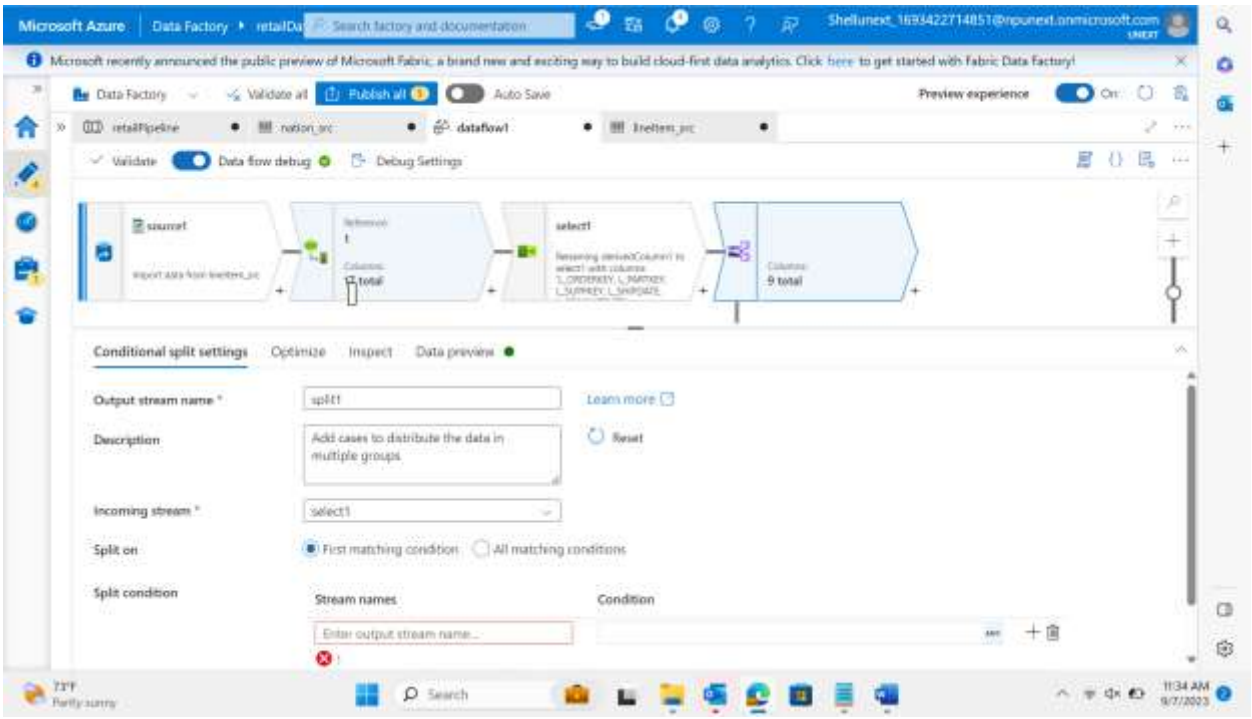


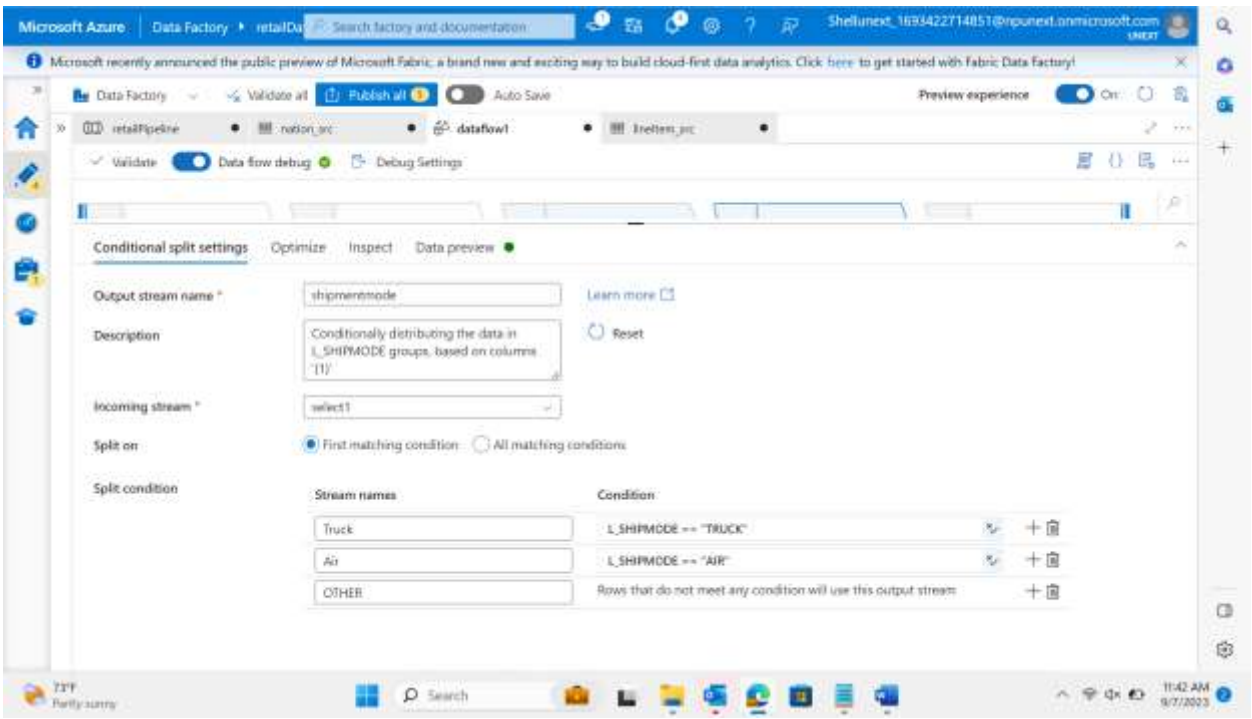
Selected only the required future columns

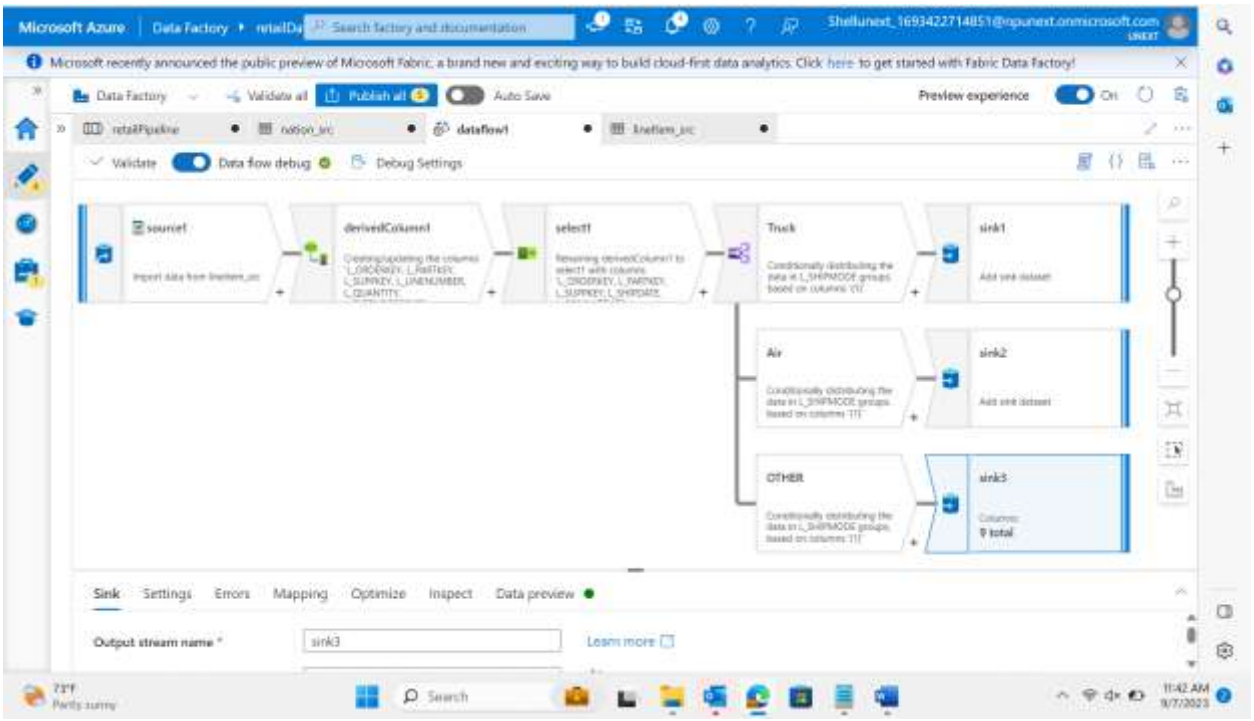




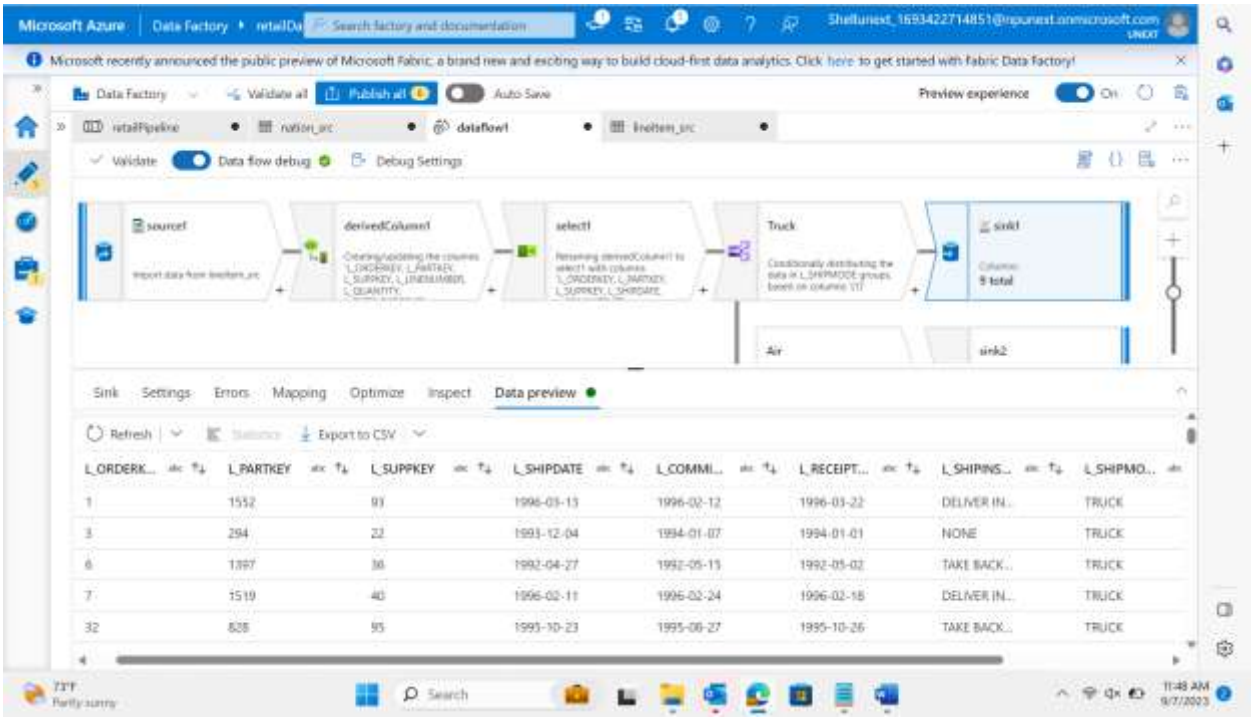
Conditional split – route your data to multiple destinations/sinks

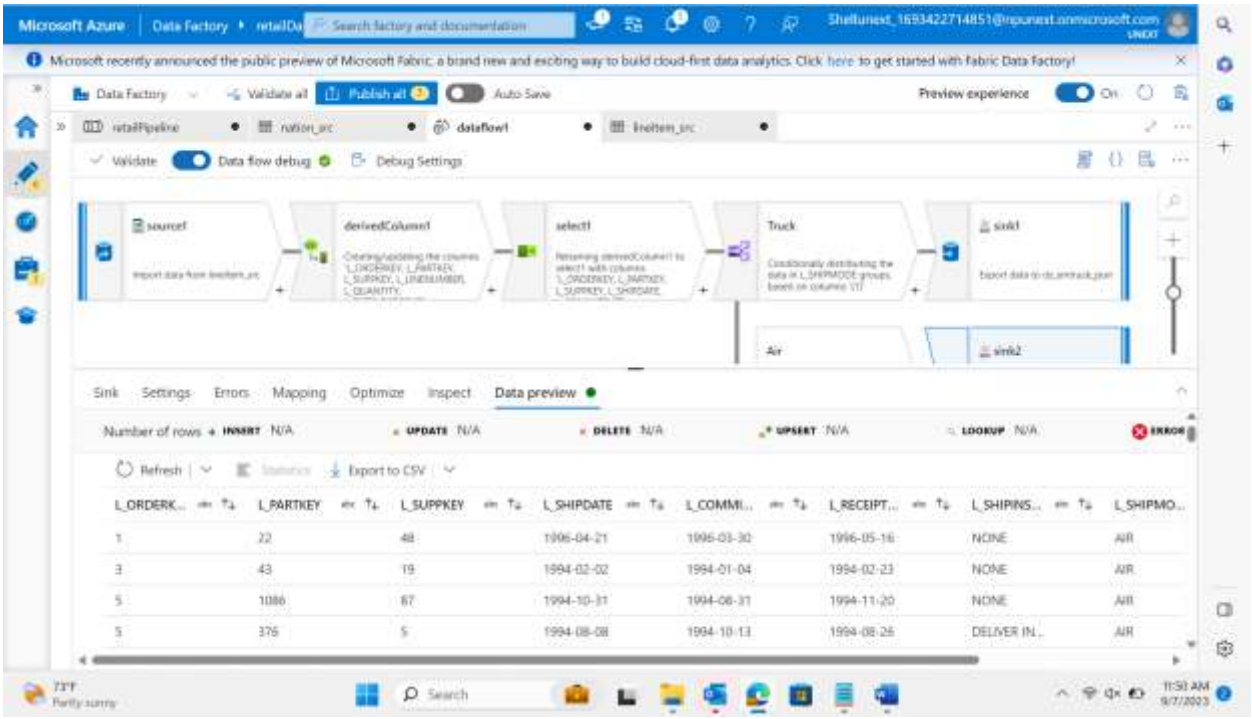

Microsoft Azure
Data Factory
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Validate all
Publish all
Auto Save
Preview experience
Validate
Data flow debug
Debug Settings
source1
select1
9 total
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...

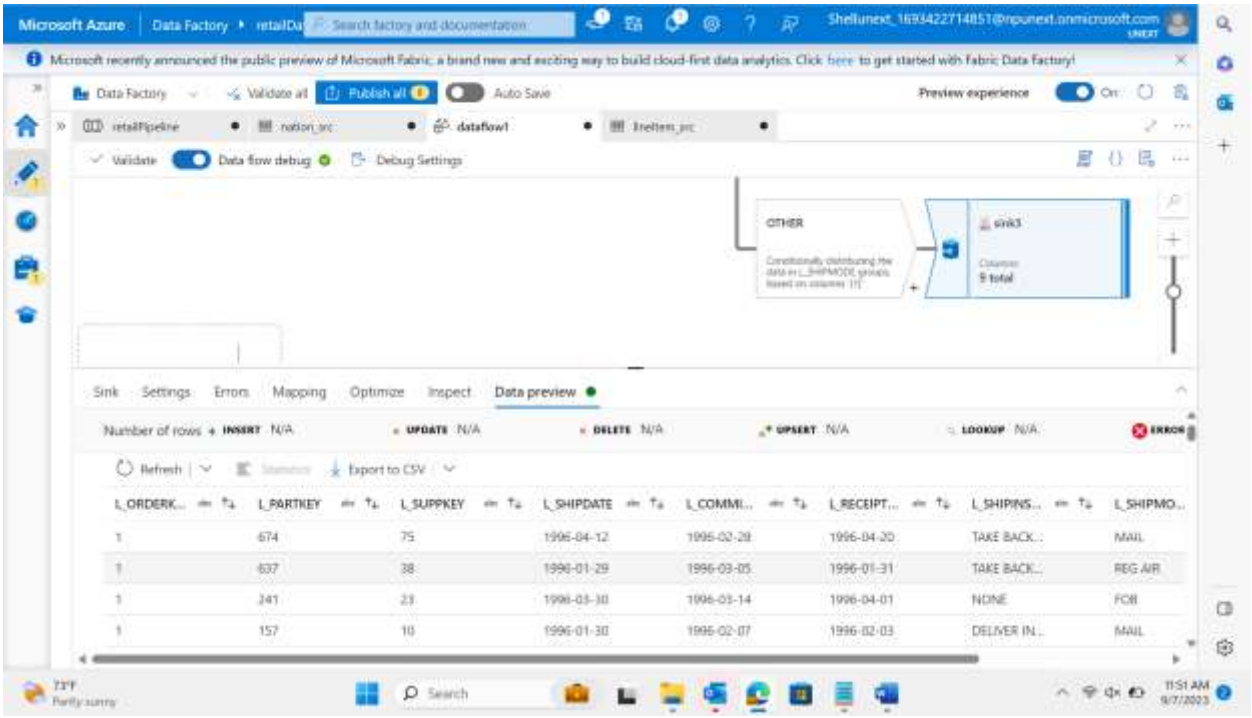

Microsoft Azure
Data Factory
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Validate all
Publish all
Auto Save
Preview experience
Validate
Data flow debug
Debug 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 columns '(1)'
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

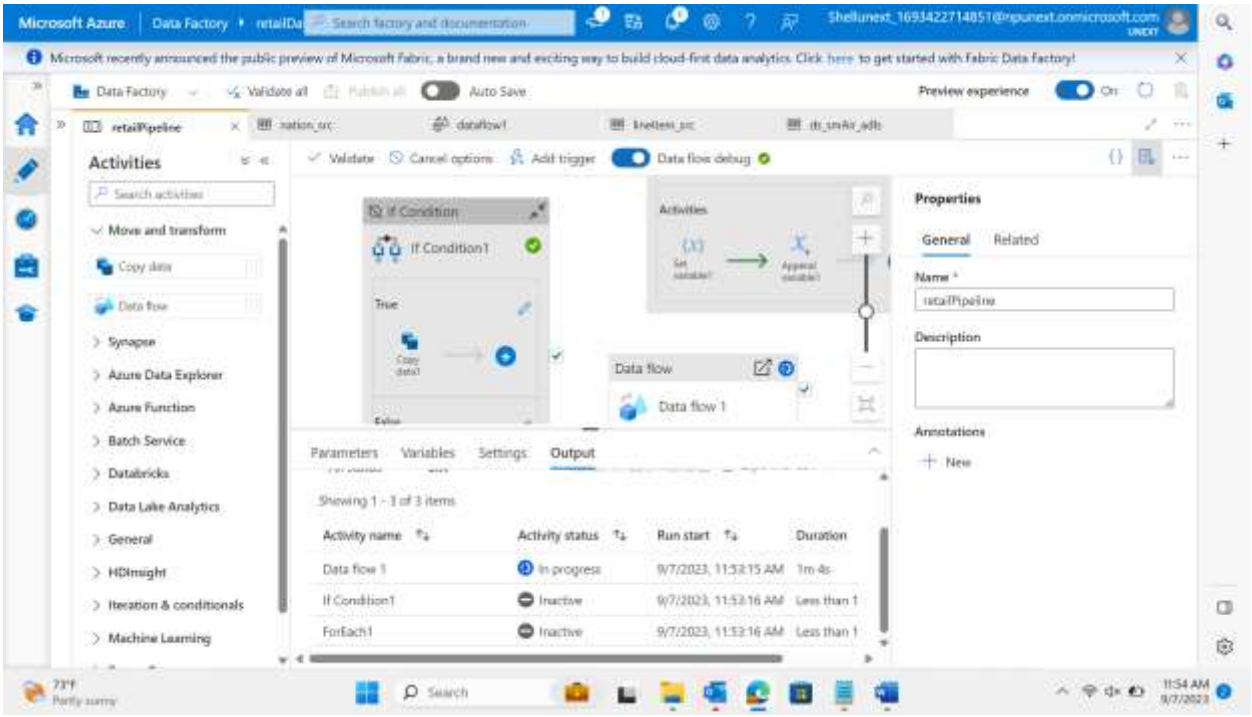

Microsoft Azure
Data Factory
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Validate all
Publish all
Auto Save
Preview experience
Validate
Data flow debug
Debug Settings
source1
derivedColumn1
select1
Truck
sink1
Air
sink2
OTHER
sink3
9 total
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Output stream name *
sink3
Learn more

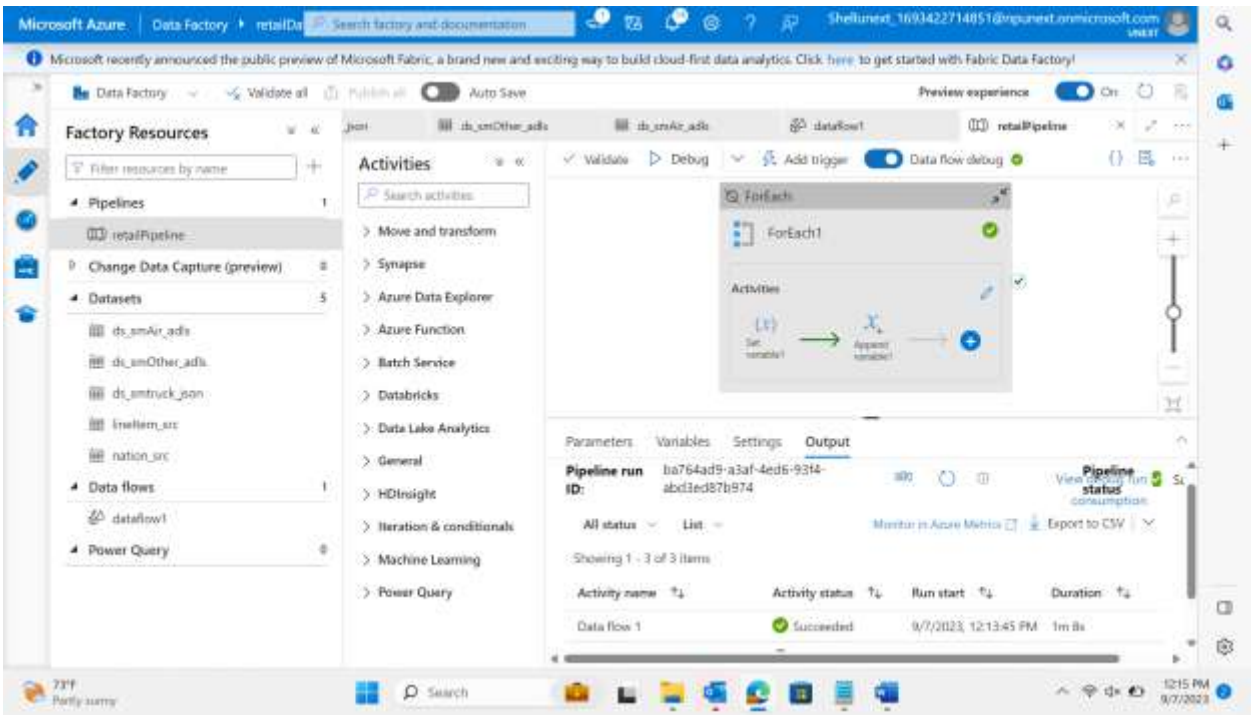
ADLS2 → json


Microsoft Azure
Data Factory
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!
Validate all
Publish all
Auto Save
Preview experience
Validate
Data flow debug
Debug Settings
source1
derivedColumn1
select1
Truck
sink1
9 total
Air
sink2
Sink
Settings
Errors
Mapping
Optimize
Inspect
Data preview
Refresh
Export to CSV
L_ORDERK...
L_PARTKEY
L_SUPPKEY
L_SHIPDATE
L_COMMI...
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 1995-08-27 1995-10-26 TAKE BACK... TRUCK

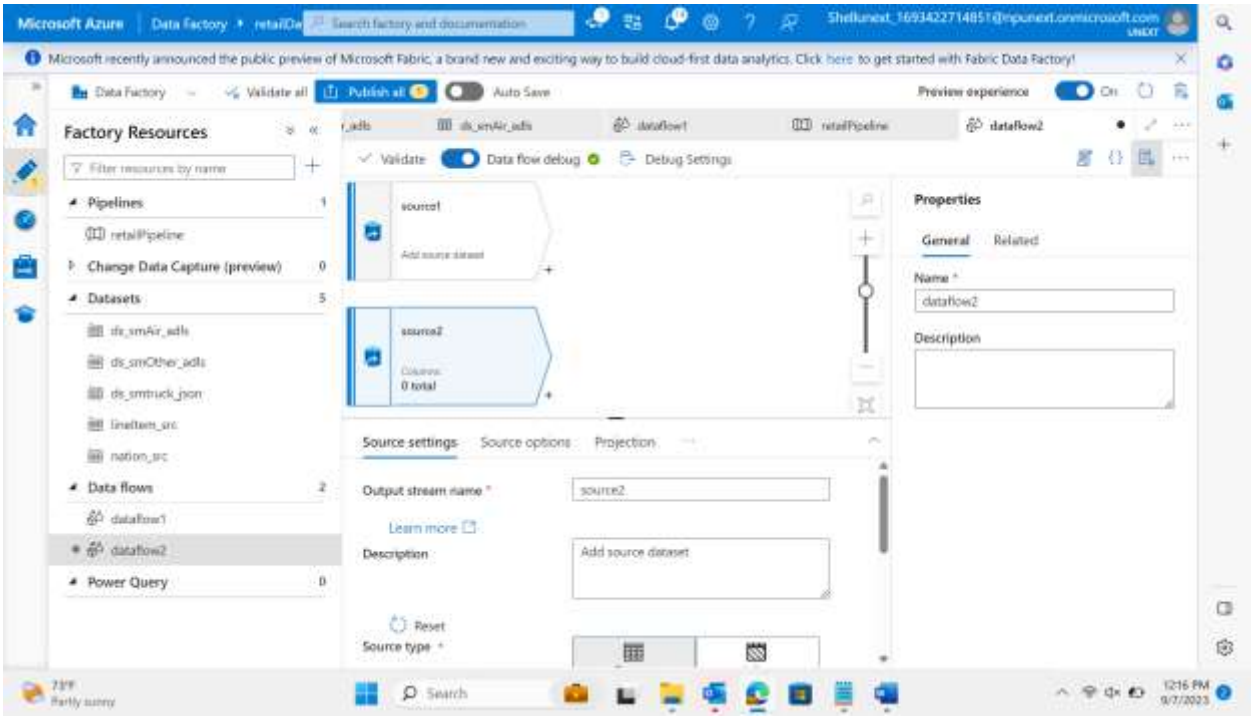

Microsoft Azure
Data Factory
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Publish all
Auto Save
Preview experience
Validate
Data flow debug
Debug Settings
source1
derivedColumn1
select1
Truck
sink1
Air
sink2
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
Export to CSV
L_ORDERK...
L_PARTKEY
L_SUPPKEY
L_SHIPDATE
L_COMMI...
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 87 1994-10-31 1994-08-31 1994-11-20 NONE AIR
5 376 5 1994-08-08 1994-10-13 1994-08-26 DELIVER IN... AIR


Microsoft Azure
Data Factory
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!
Publish all
Auto Save
Preview experience
Validate
Data flow debug
Debug Settings
OTHER
sink3
9 total
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
Export to CSV
L_ORDERK...
L_PARTKEY
L_SUPPKEY
L_SHIPDATE
L_COMMI...
L_RECEIPT...
L_SHIPINS...
L_SHIPMO...
1 674 75 1996-04-12 1996-02-28 1996-04-20 TAKE BACK... MAIL
1 637 38 1996-01-29 1996-03-05 1996-01-31 TAKE BACK... REG AIR
1 241 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



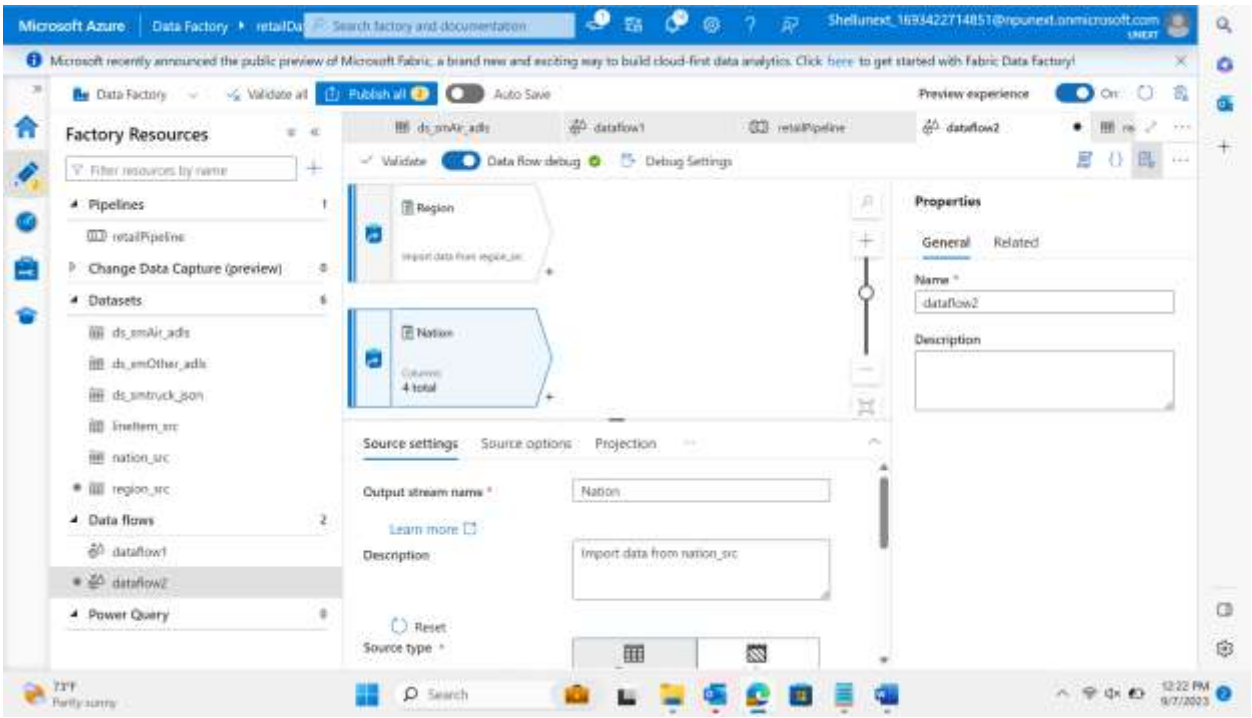


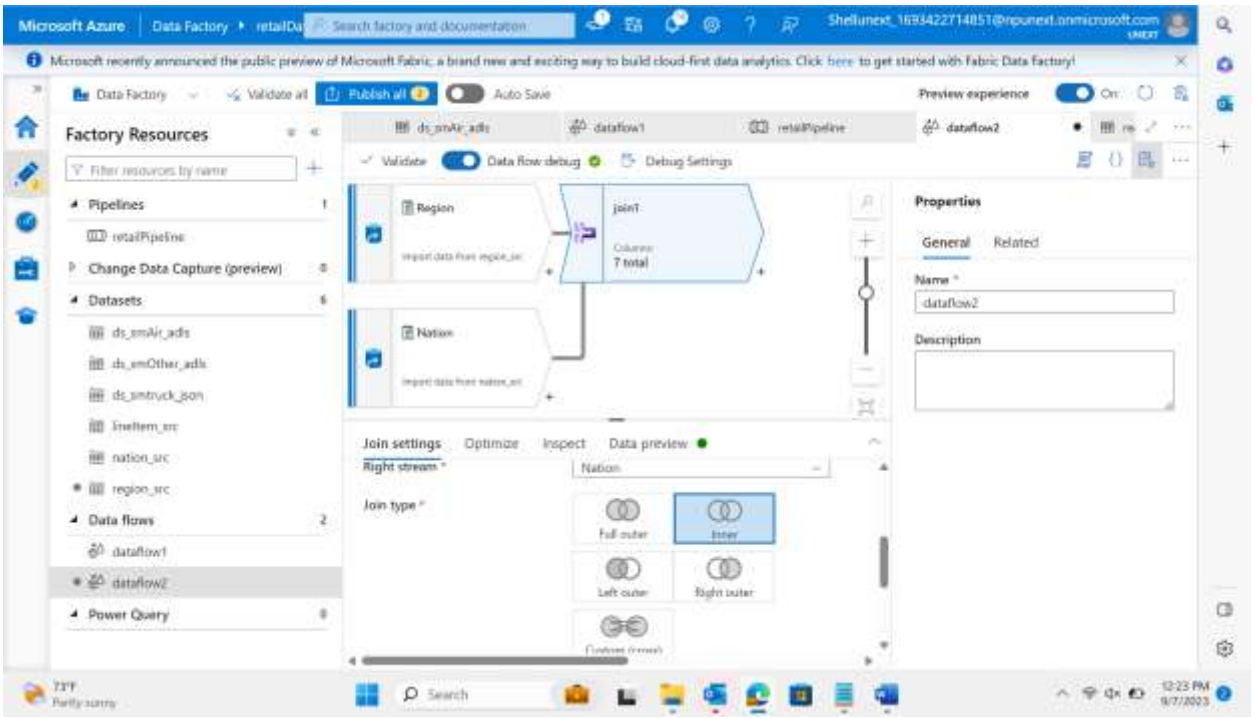
Another Data flow task

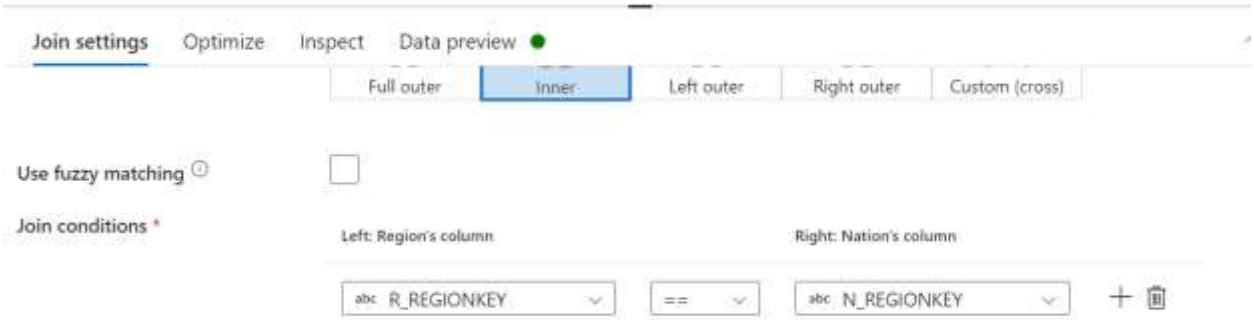


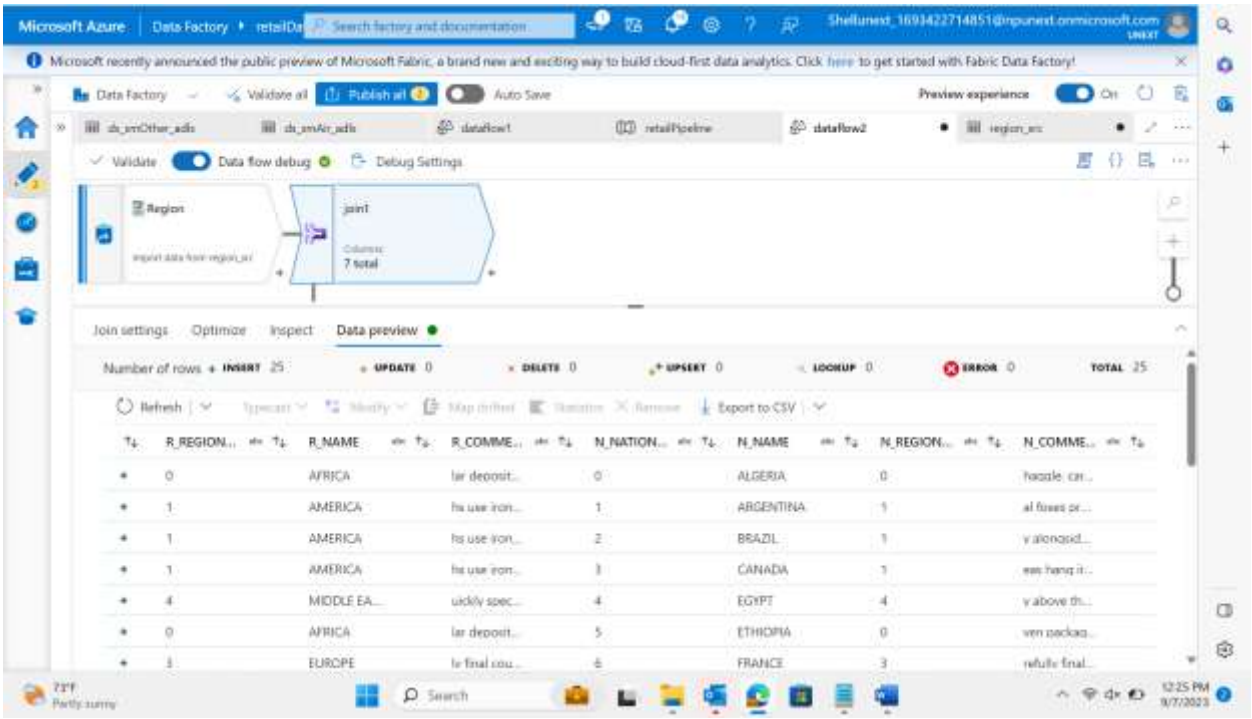
Region source

Nation second source



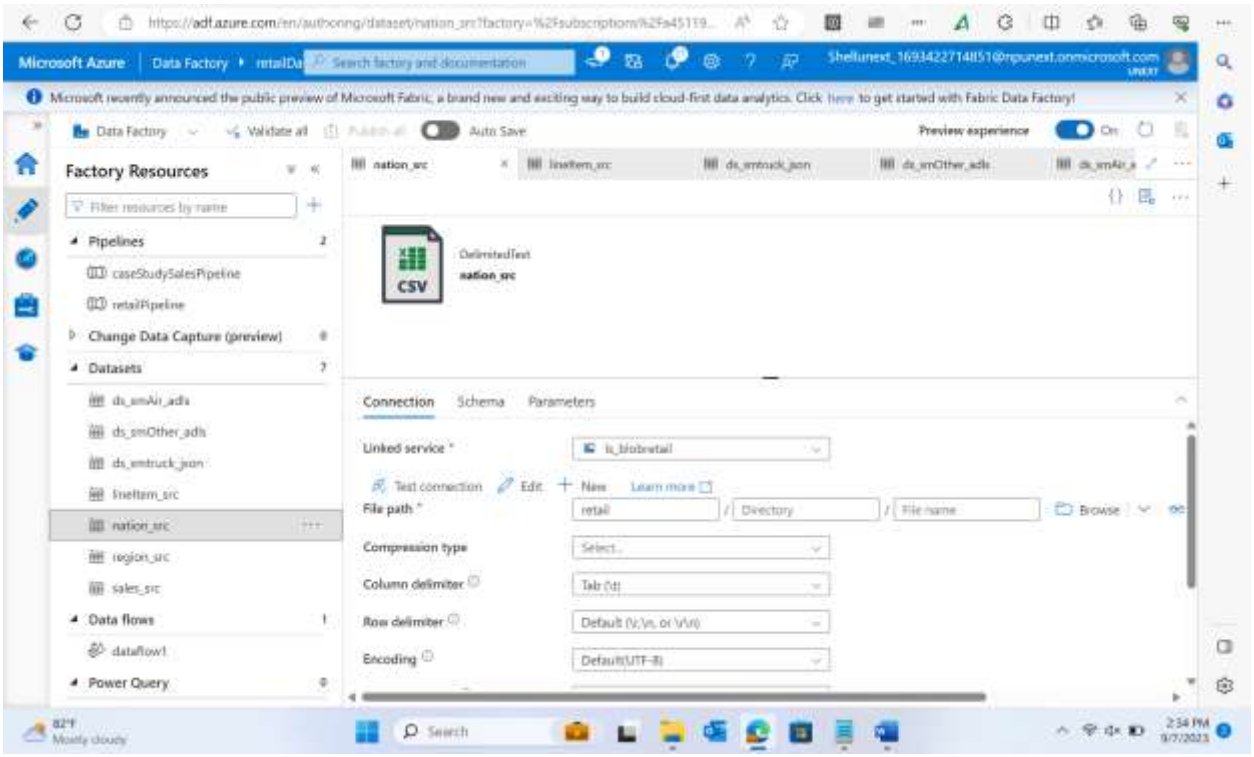

Microsoft Azure
Data Factory
Factory Resources
Pipelines
retailPipeline
Change Data Capture (preview)
Datasets
ds_smAir_adls
ds_smOther_adls
ds_smtruck_json
lineitem_src
nation_src
region_src
Data flows
dataflow1
dataflow2
Power Query
Validate all
Publish all
Auto Save
Preview experience
Validate
Data flow debug
Debug Settings
Region
Nation
join1
7 total
Properties
General
Related
Name *
dataflow2
Description
Join settings
Optimize
Inspect
Data preview
Right stream *
Join type *
Full outer
Inner
Left outer
Right outer

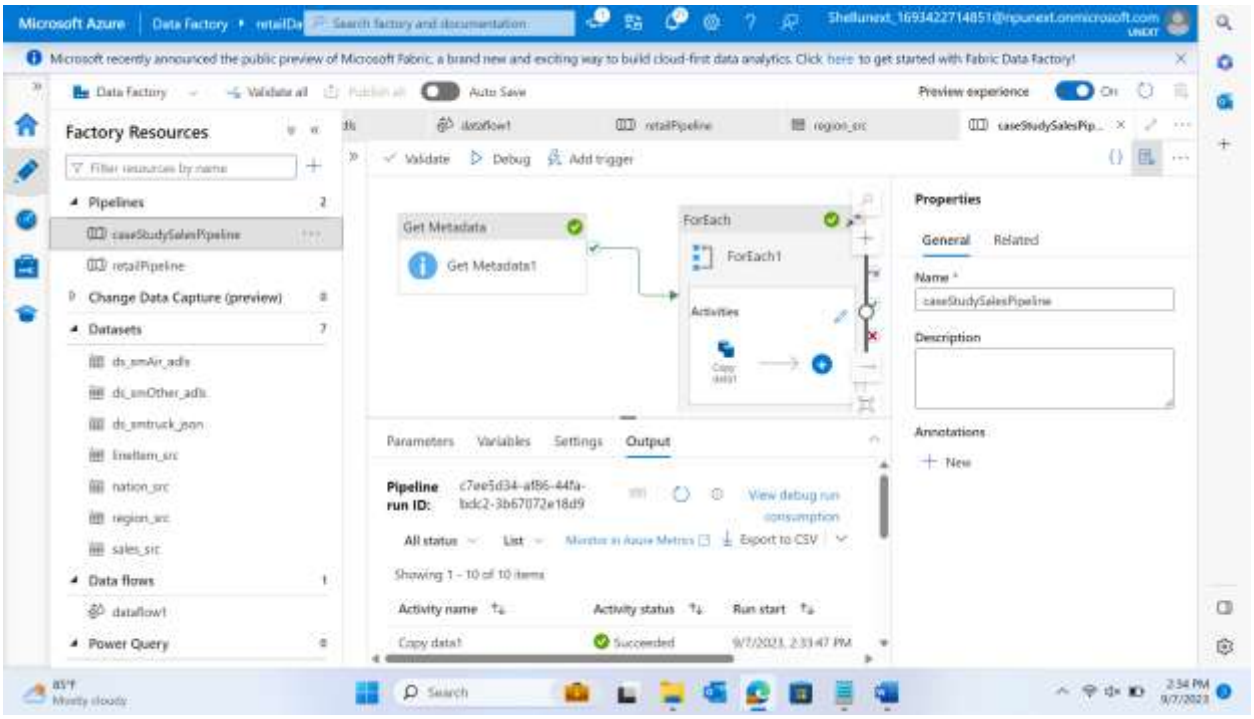

Join settings
Optimize
Inspect
Data preview
Full outer
Inner
Left outer
Right outer
Custom (cross)
Use fuzzy matching
Join conditions *
Left: Region's column
Right: Nation's column
R_REGIONKEY
==
N_REGIONKEY

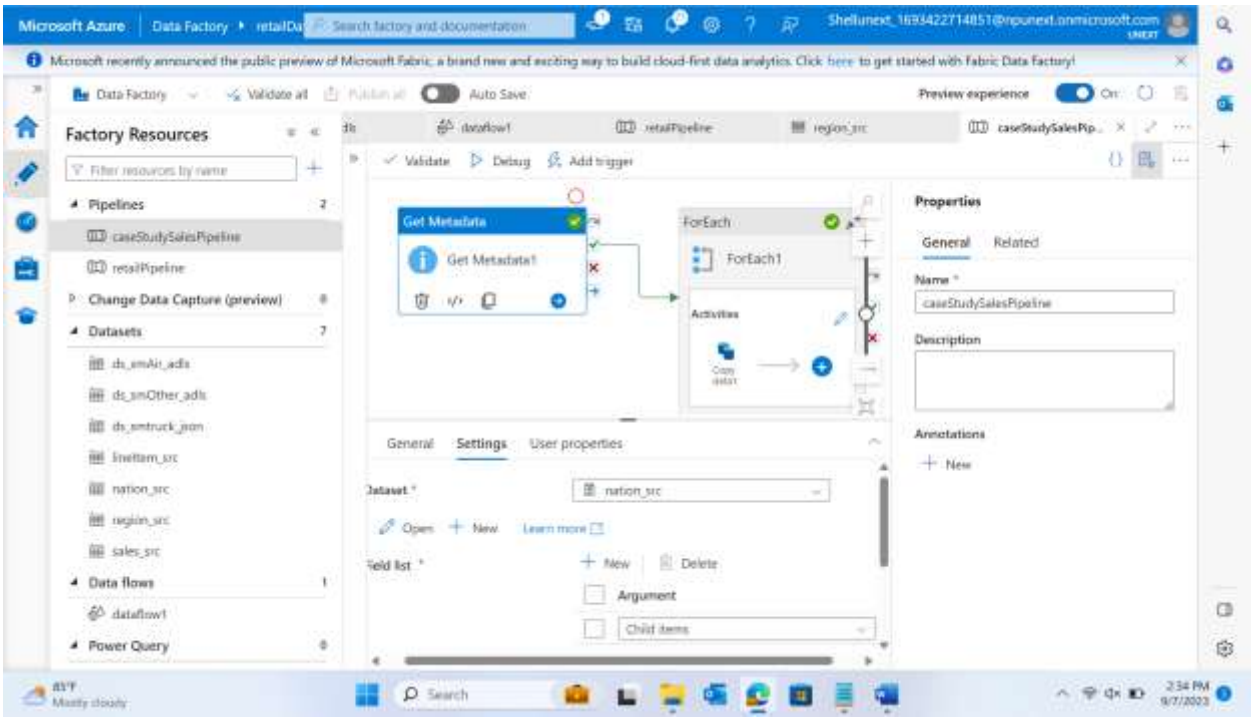


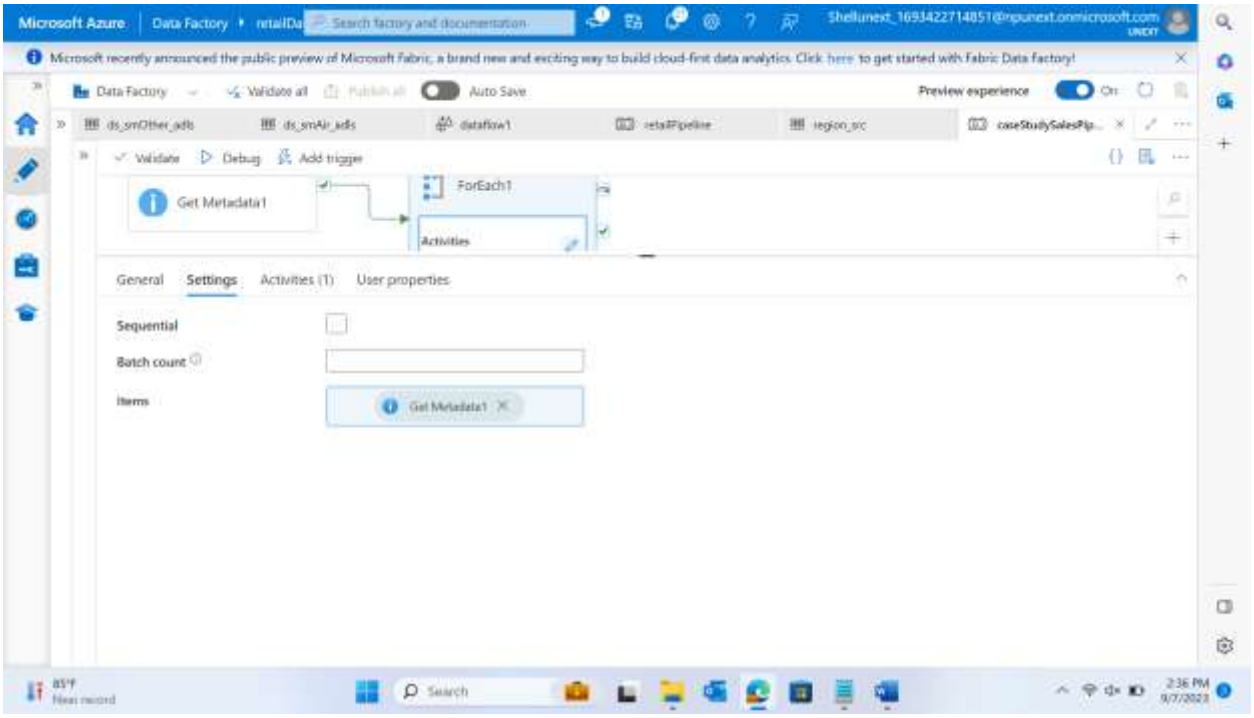
Copy Data: Ingest the raw sales data from various sources (e.g., CSV files, databases) into Azure Data Lake Storage.

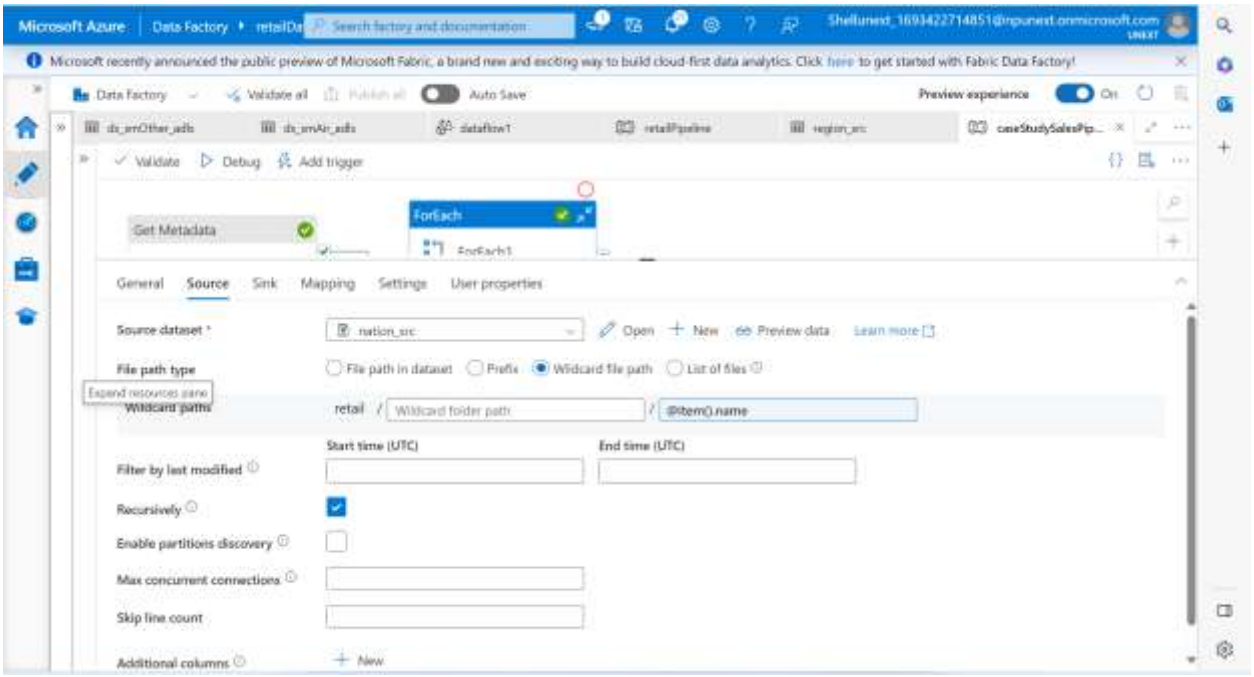
****Ingesting multiple files all at once**

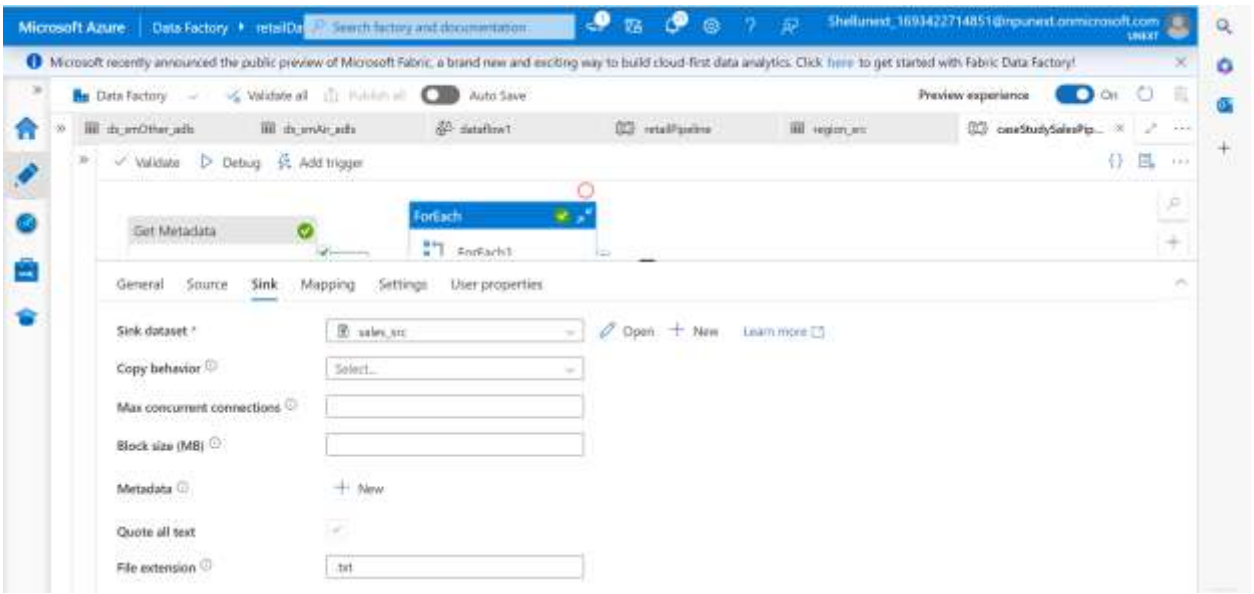

Microsoft Azure
Data Factory
Factory Resources
Pipelines
caseStudySalesPipeline
retailPipeline
Change Data Capture (preview)
Datasets
ds_smAir_adls
ds_smOther_adls
ds_smtruck_json
lineItem_src
nation_src
region_src
sales_src
Data flows
dataflow1
Power Query
Auto Save
Preview experience
Connection
Schema
Parameters
Linked service *
Test connection
Edit
New
File path *
Compression type
Column delimiter
Row delimiter
Encoding
Browse

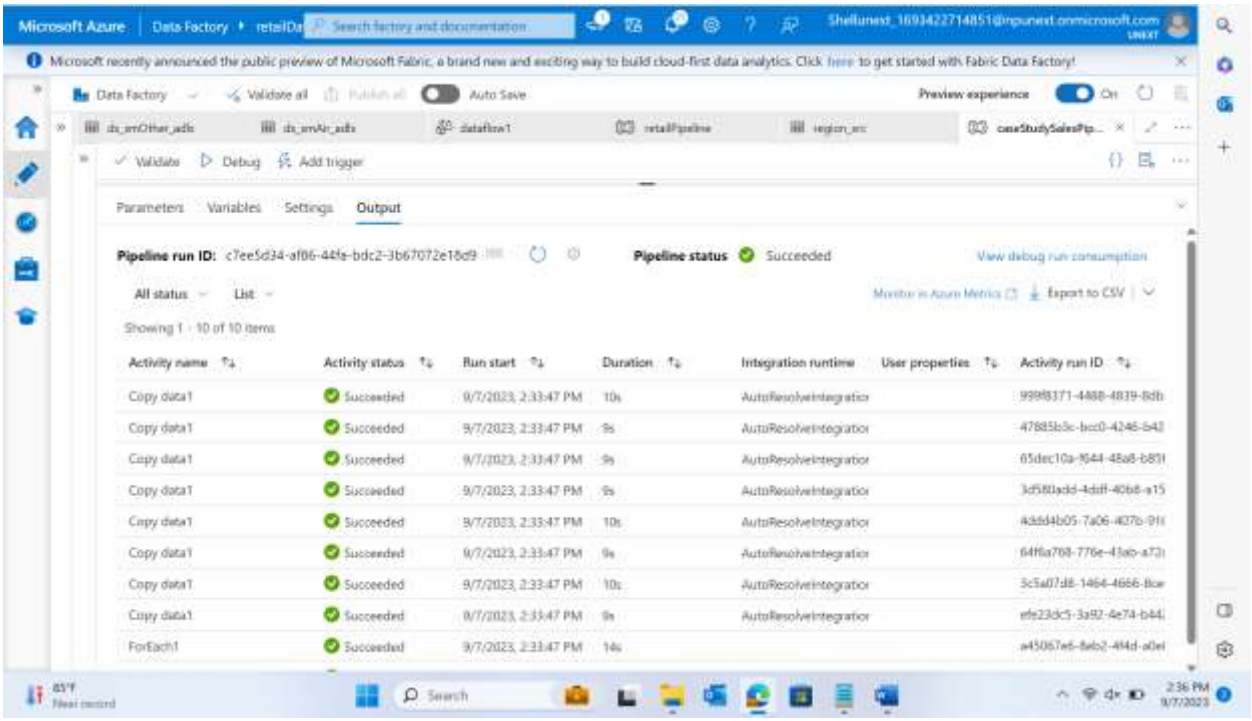

Microsoft Azure
Data Factory
Factory Resources
Pipelines
caseStudySalesPipeline
retailPipeline
Change Data Capture (preview)
Datasets
Data flows
dataflow1
Power Query
Validate
Debug
Add trigger
Get Metadata
Get Metadata1
ForEach
ForEach1
Activities
Properties
General
Related
Name *
Description
Annotations
New
Parameters
Variables
Settings
Output
Pipeline run ID:
c7ee5d34-af86-44fa-bdc2-3b67072e18d9
View debug run consumption
All status
List
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


Microsoft Azure
Data Factory
Factory Resources
Pipelines
caseStudySalesPipeline
retailPipeline
Change Data Capture (preview)
Datasets
Data flows
dataflow1
Power Query
Get Metadata
Get Metadata1
ForEach
ForEach1
Activities
Properties
General
Related
Name *
caseStudySalesPipeline
Description
Annotations
New
General
Settings
User properties
Dataset *
nation_src
Open
New
Learn more
Field list *
Delete
Argument
Child items
Validate
Debug
Add trigger
Auto Save
Preview experience


Microsoft Azure
Data Factory
Get Metadata1
ForEach1
Activities
General
Settings
Activities (1)
User properties
Sequential
Batch count
Items
Get Metadata1
Validate
Debug
Add trigger
Auto Save
Preview experience

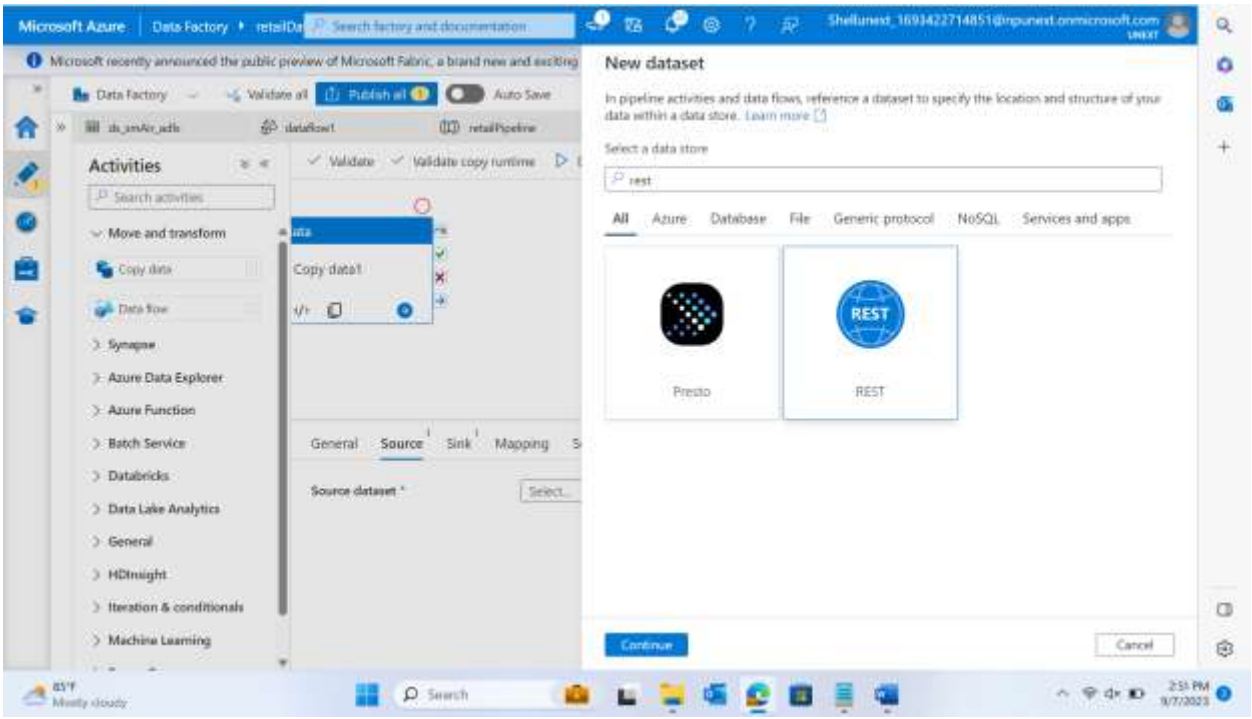

Microsoft Azure
Data Factory
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Validate
Debug
Add trigger
Get Metadata
ForEach
General
Source
Sink
Mapping
Settings
User properties
Source dataset *
Open
New
Preview data
Learn more
File path type
File path in dataset
Prefix
Wildcard file path
List of files
Wildcard paths
retail
Wildcard folder path
@item().name
Start time (UTC)
End time (UTC)
Filter by last modified
Recursively
Enable partitions discovery
Max concurrent connections
Skip line count
Additional columns
New

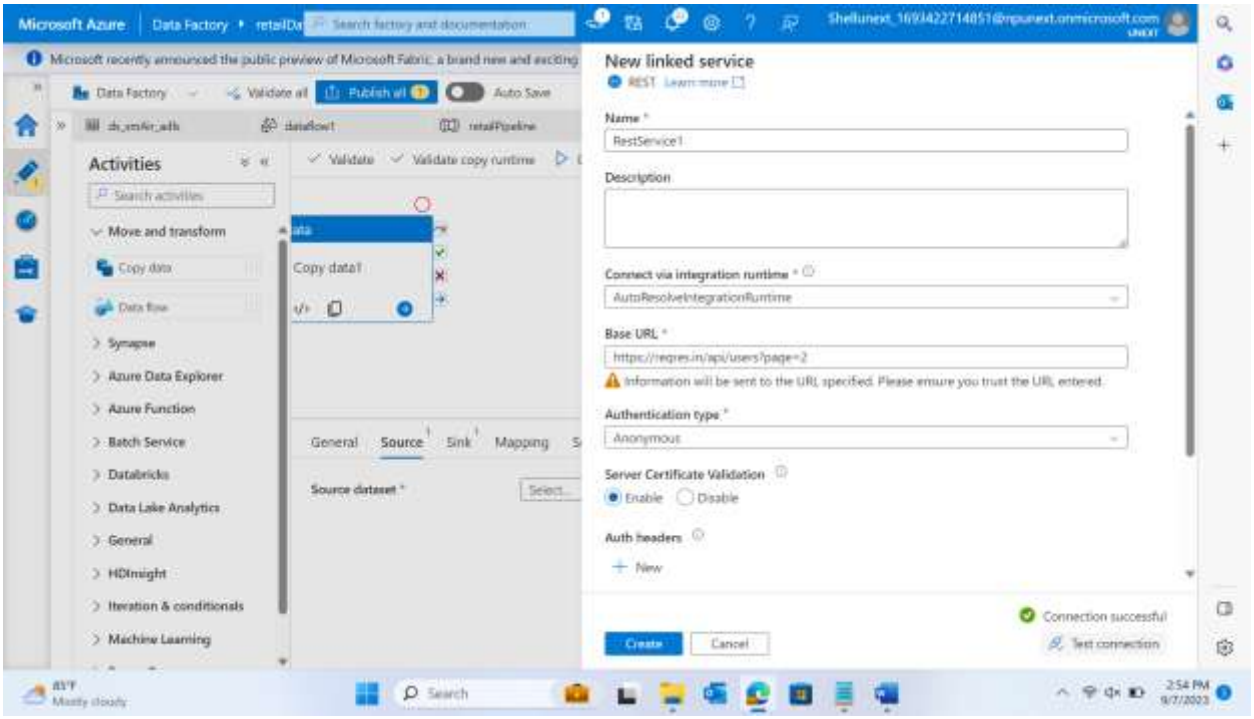

Microsoft Azure
Data Factory
Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!
Validate
Debug
Add trigger
Get Metadata
ForEach
General
Source
Sink
Mapping
Settings
User properties
Sink dataset *
Open
New
Learn more
Copy behavior
Select...
Max concurrent connections
Block size (MB)
Metadata
New
Quote all text
File extension
txt



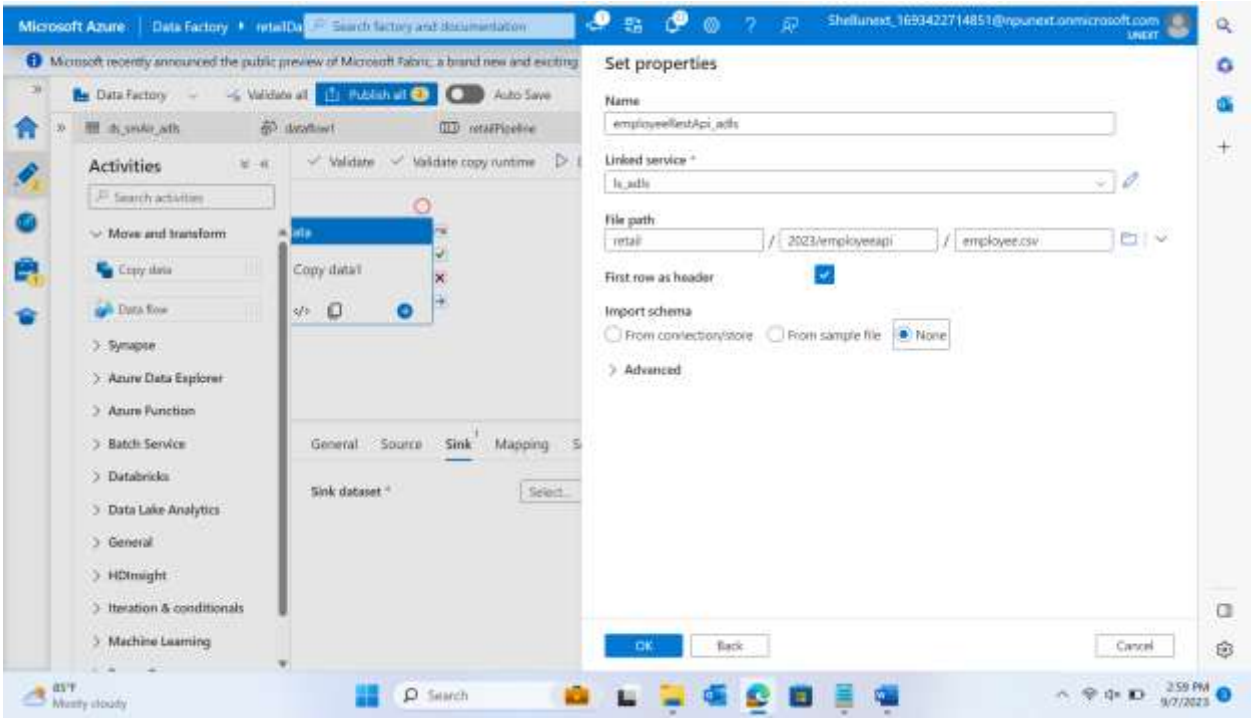
API-based integration

reqres.in/api/users?page=2




Microsoft Azure
Data Factory
New linked service
REST
Name *
RestService1
Description
Connect via integration runtime *
AutoResolveIntegrationRuntime
Base URL *
https://reqres.in/api/users?page=2
Information will be sent to the URL specified. Please ensure you trust the URL entered.
Authentication type *
Anonymous
Server Certificate Validation
Enable
Disable
Auth headers
New
Connection successful
Test connection
Create
Cancel
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 *

Microsoft Azure
Data Factory
Preview data
Linked service: RestService1
Object:
[
{
"page": 2,
"per_page": 6,
"total": 12,
"total_pages": 2,
"data": [
{
"id": 7,
"email": "michael.lawson@reqres.in",
"first_name": "Michael",
"last_name": "Lawson",
"avatar": "https://reqres.in/img/faces/7-image.jpg"
},
{
"id": 8,
"email": "lindsay.ferguson@reqres.in",
"first_name": "Lindsay",
"last_name": "Ferguson",
"avatar": "https://reqres.in/img/faces/8-image.jpg"
},
{
"id": 9,
"email": "tobias.funke@reqres.in",
"first_name": "Tobias",
"last_name": "Funke",
"avatar": "https://reqres.in/img/faces/9-image.jpg"
}
Additional headers
New
Activities
Move and transfo
Copy data
Data flow
Synapse
Azure Function
Batch Service
Databricks
General
HDInsight
Machine Learning



General | Source | Sink | **Mapping** | Settings | User properties

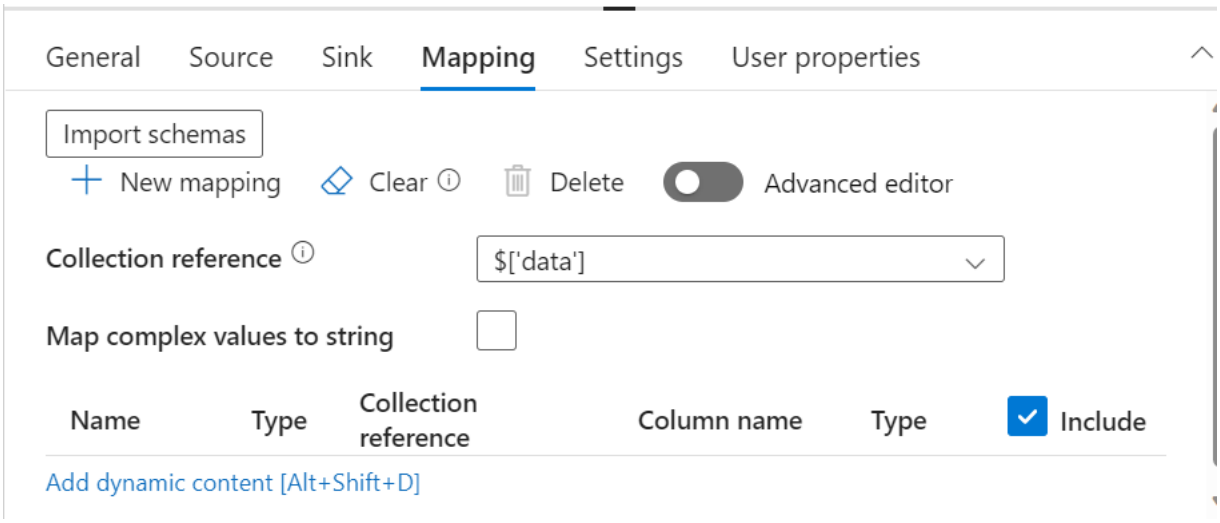
Import schemas

+ New mapping | Clear | Delete | Advanced editor

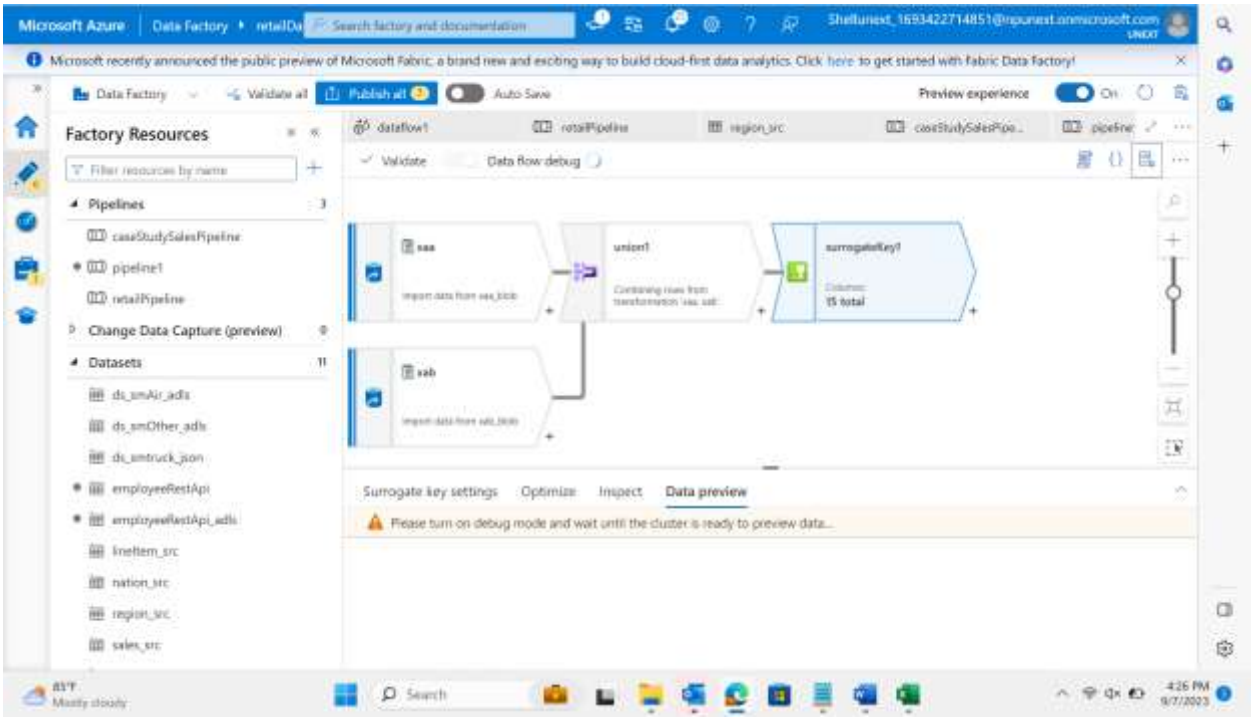
Collection reference

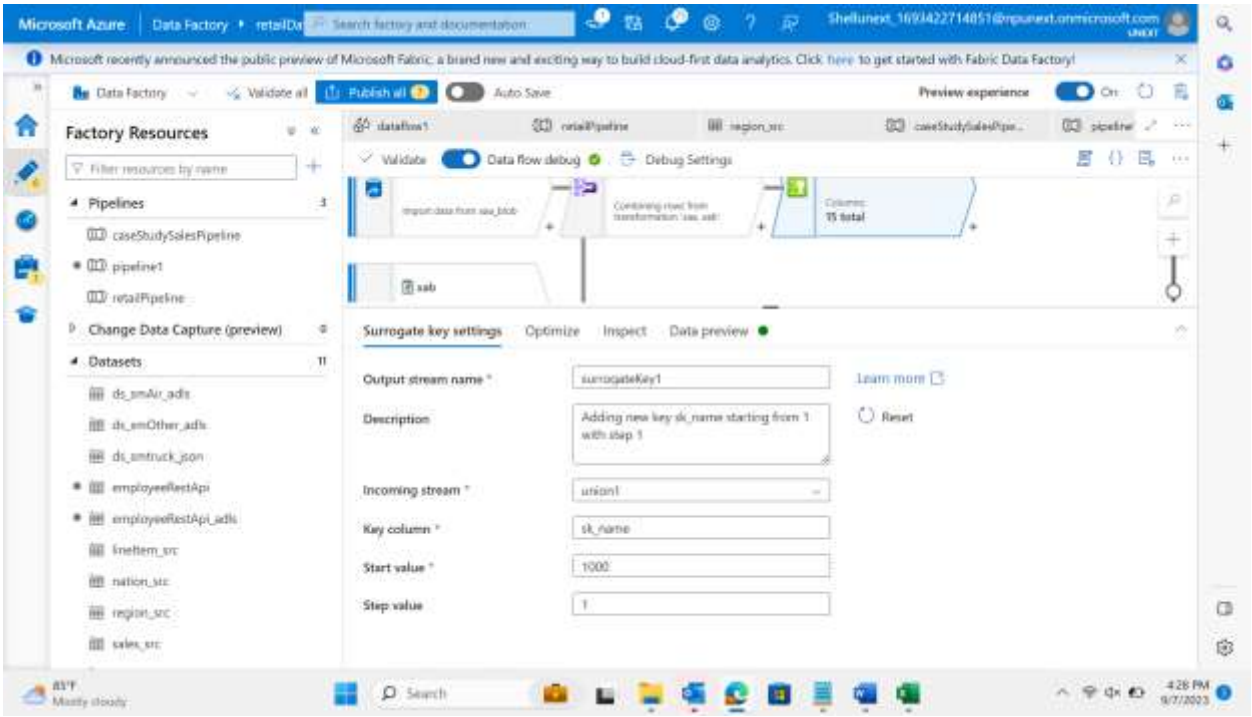
Map complex values to string

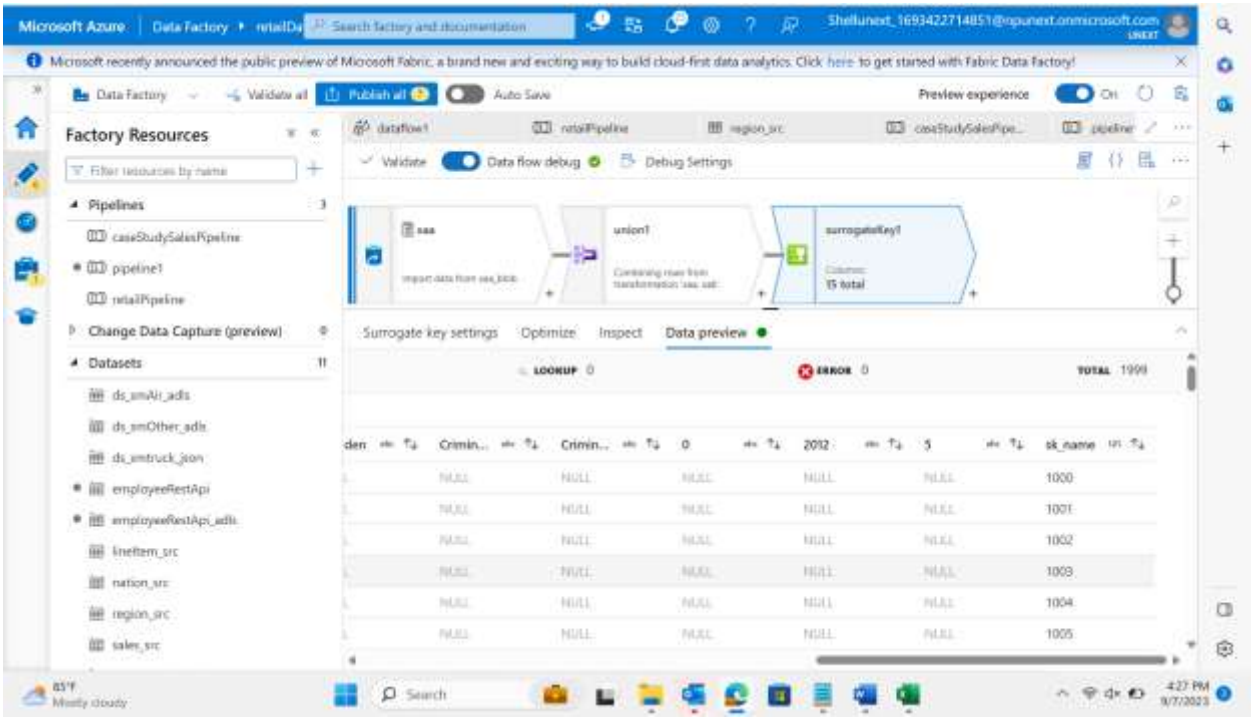
Name	Type	Collection reference	Column name
page	any, any		



Import schemas




Microsoft Azure
Data Factory
Factory Resources
Pipelines
caseStudySalesPipeline
pipeline1
retailPipeline
Change Data Capture (preview)
Datasets
ds_smAir_adls
ds_smOther_adls
ds_smtruck_json
employeeRestApi
employeeRestApi_adls
lineitem_src
nation_src
region_src
sales_src
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

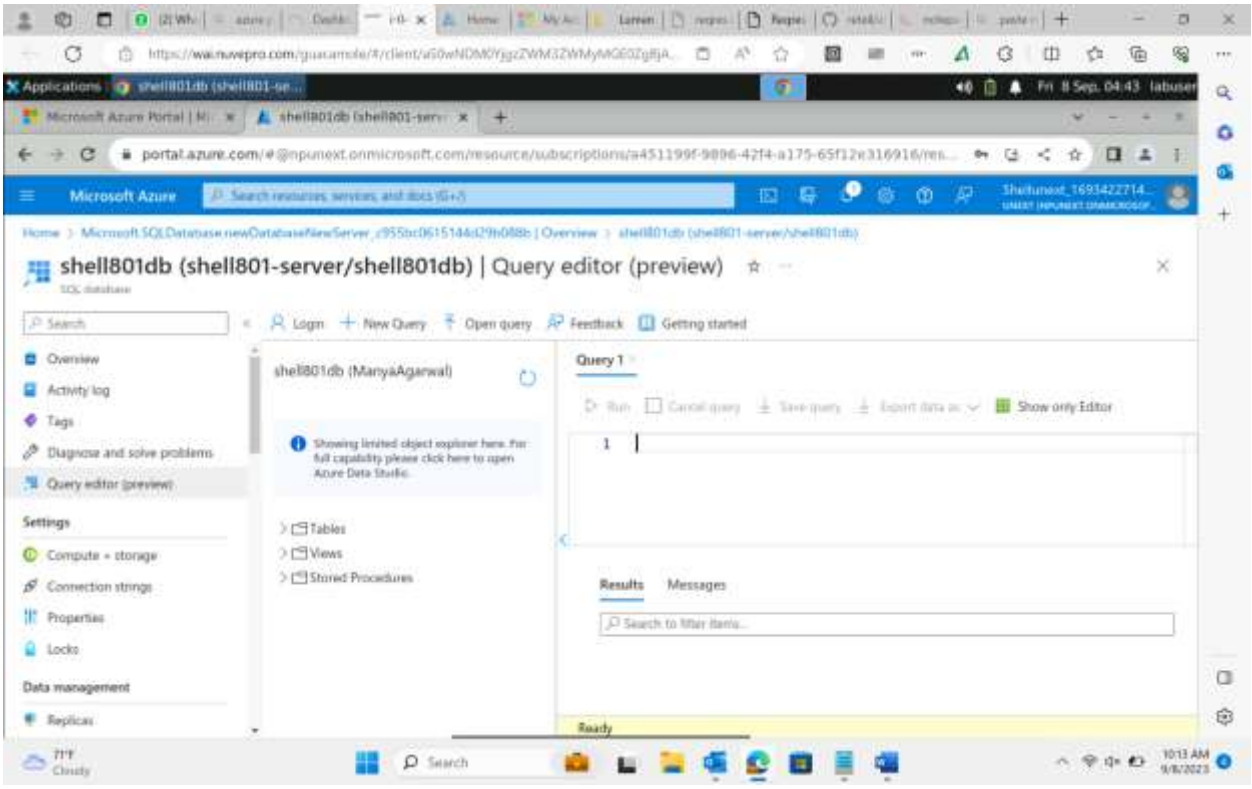

Microsoft Azure
Data Factory
Factory Resources
Pipelines
caseStudySalesPipeline
pipeline1
retailPipeline
Change Data Capture (preview)
Datasets
ds_smAir_adls
ds_smOther_adls
ds_smtruck_json
employeeRestApi
employeeRestApi_adls
lineitem_src
nation_src
region_src
sales_src
saa
union1
surrogateKey1
Surrogate key settings
Optimize
Inspect
Data preview
LOOKUP 0
ERROR 0
TOTAL 1999
sk_name
1000
1001
1002
1003
1004
1005

Friday – 8/9/23

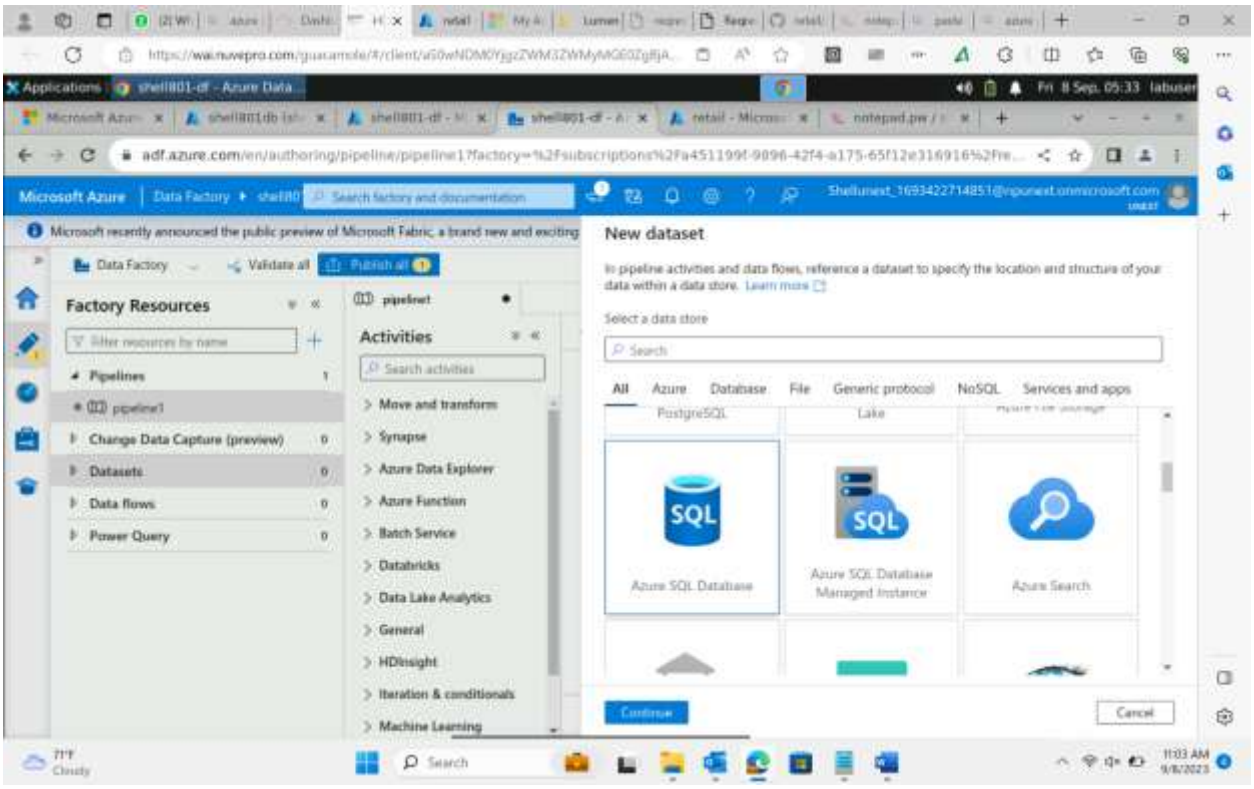
To calculate month wise data from date → derived column

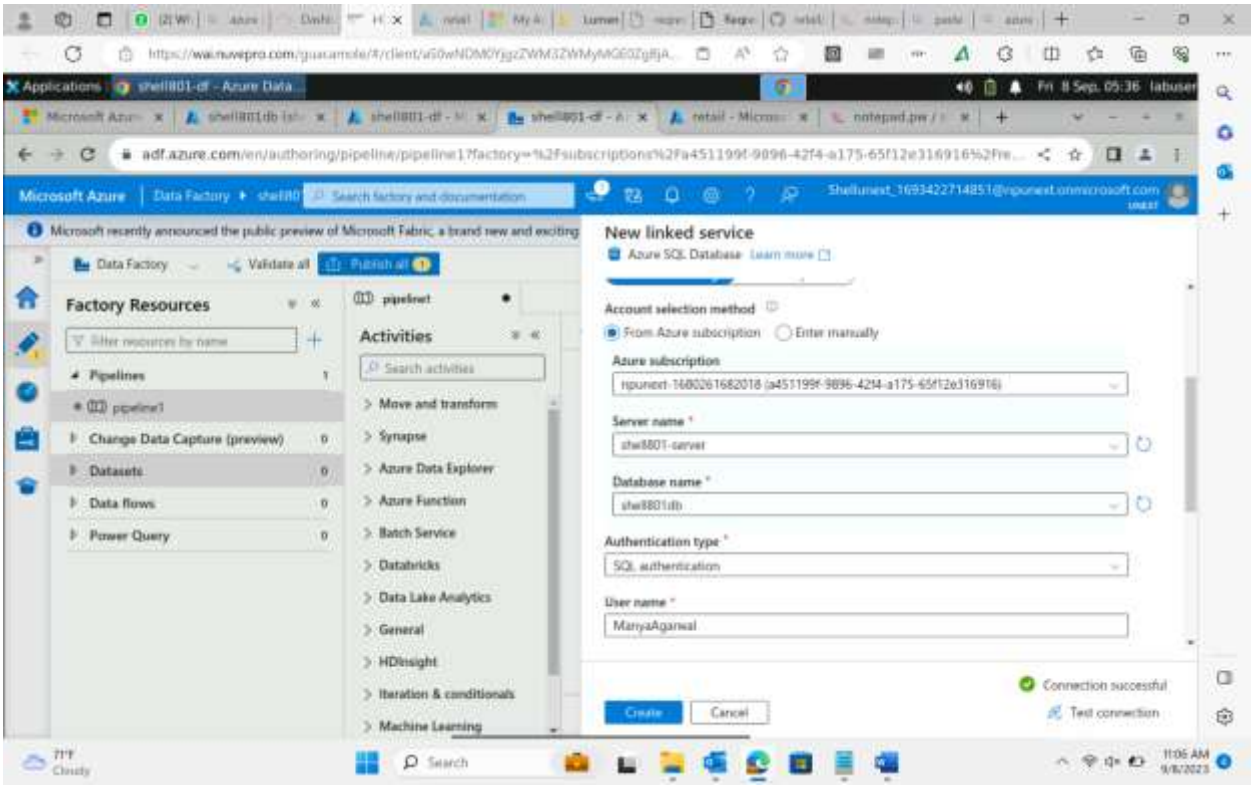
Today's task : Data migration – from on premises to cloud DW (file sources to azure sql)

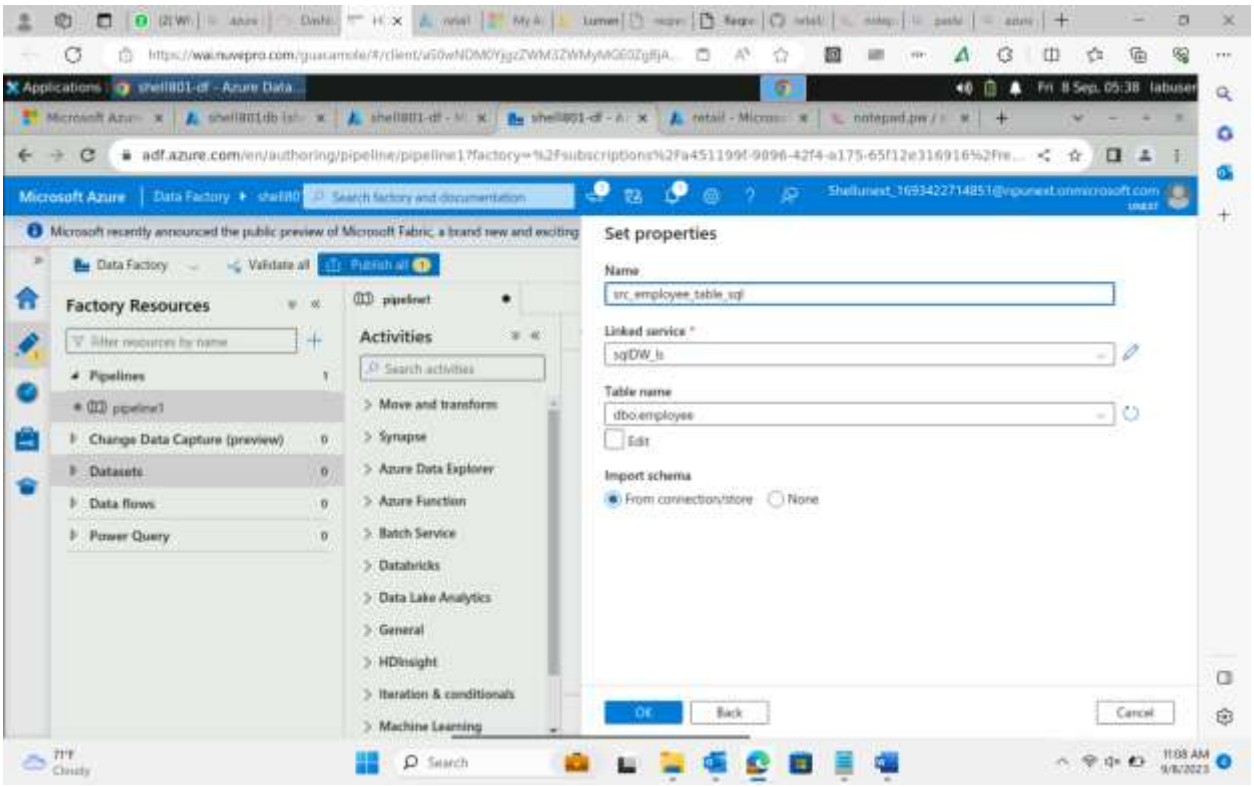
1. **Create azure sql**
2. **Sql databases – single database**
3. **Create server**
   ManyaAgarwal
   ManyaAgarwa4
   Configure – basic
   Development
   Local redundant
   Network → public endpoint, yes, yes
   Query editor

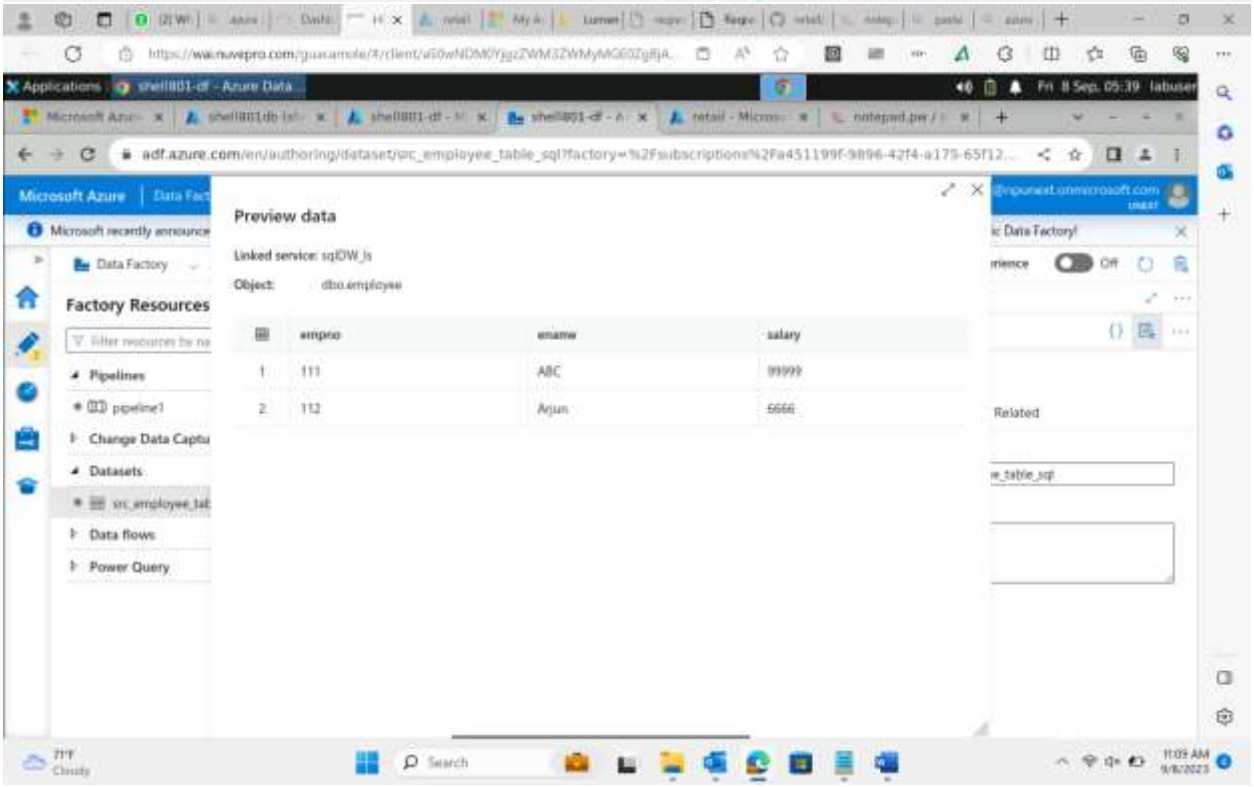


Create data factory
Azure adls storage account
Retail/2023/sep/08/superstore.csv

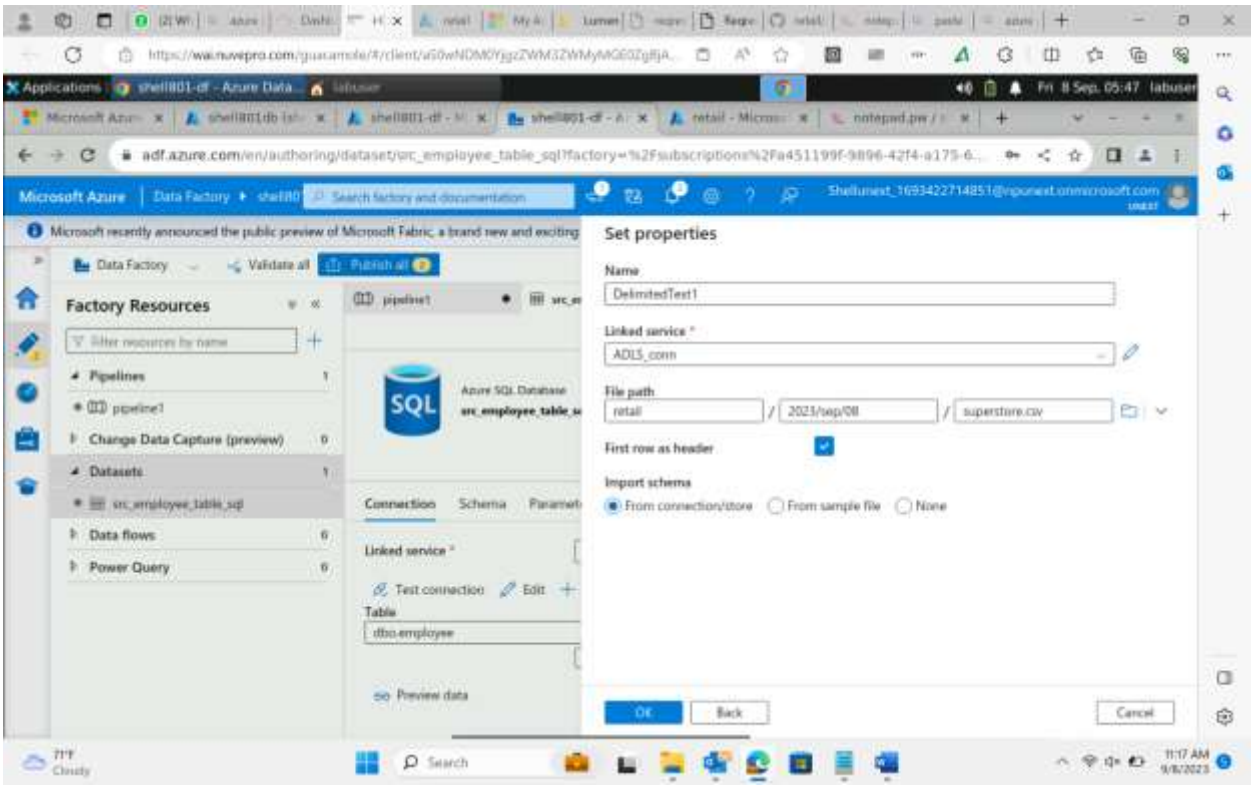

New dataset
Select a data store
Azure SQL Database
Azure SQL Database Managed Instance
Azure Search
Continue
Cancel

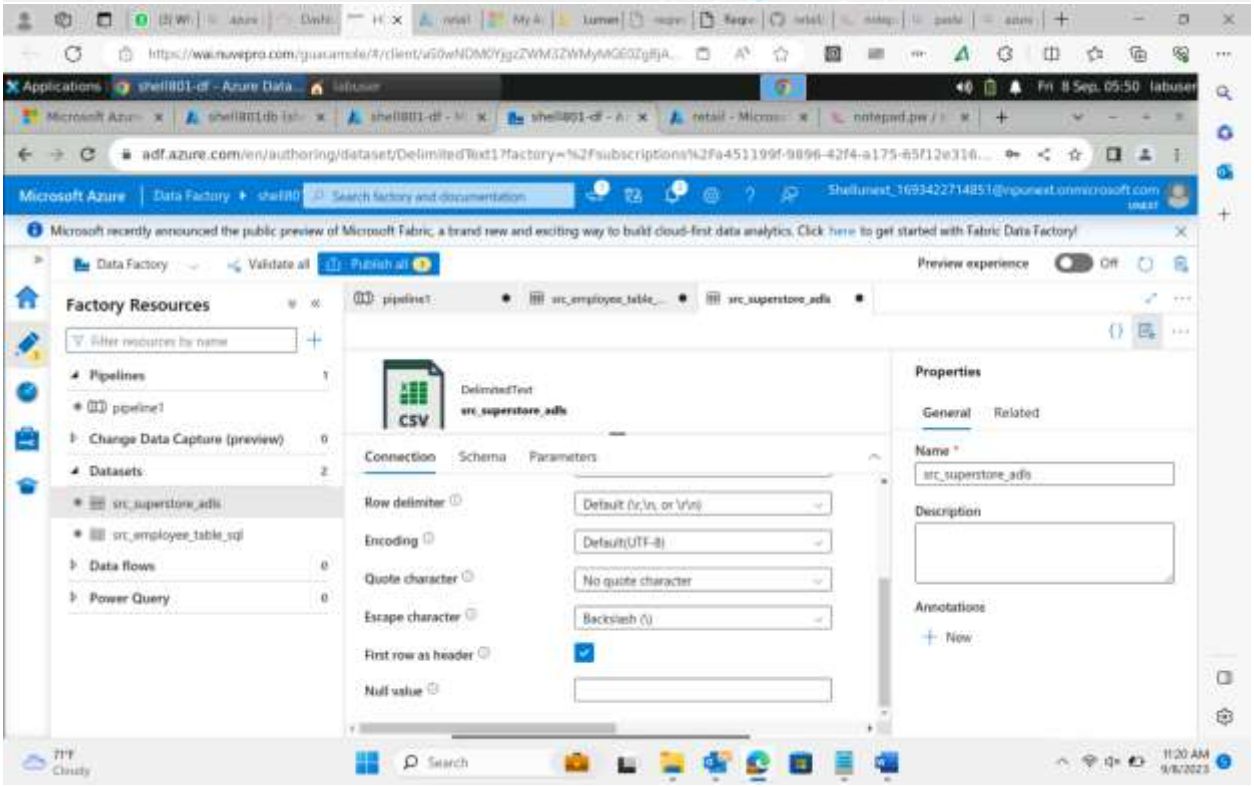

New linked service
Azure SQL Database
Account selection method
From Azure subscription
Enter manually
Azure subscription
Server name *
sha8801-server
Database name *
sha8801db
Authentication type *
SQL authentication
User name *
ManyaAgarwal
Connection successful
Test connection
Create
Cancel





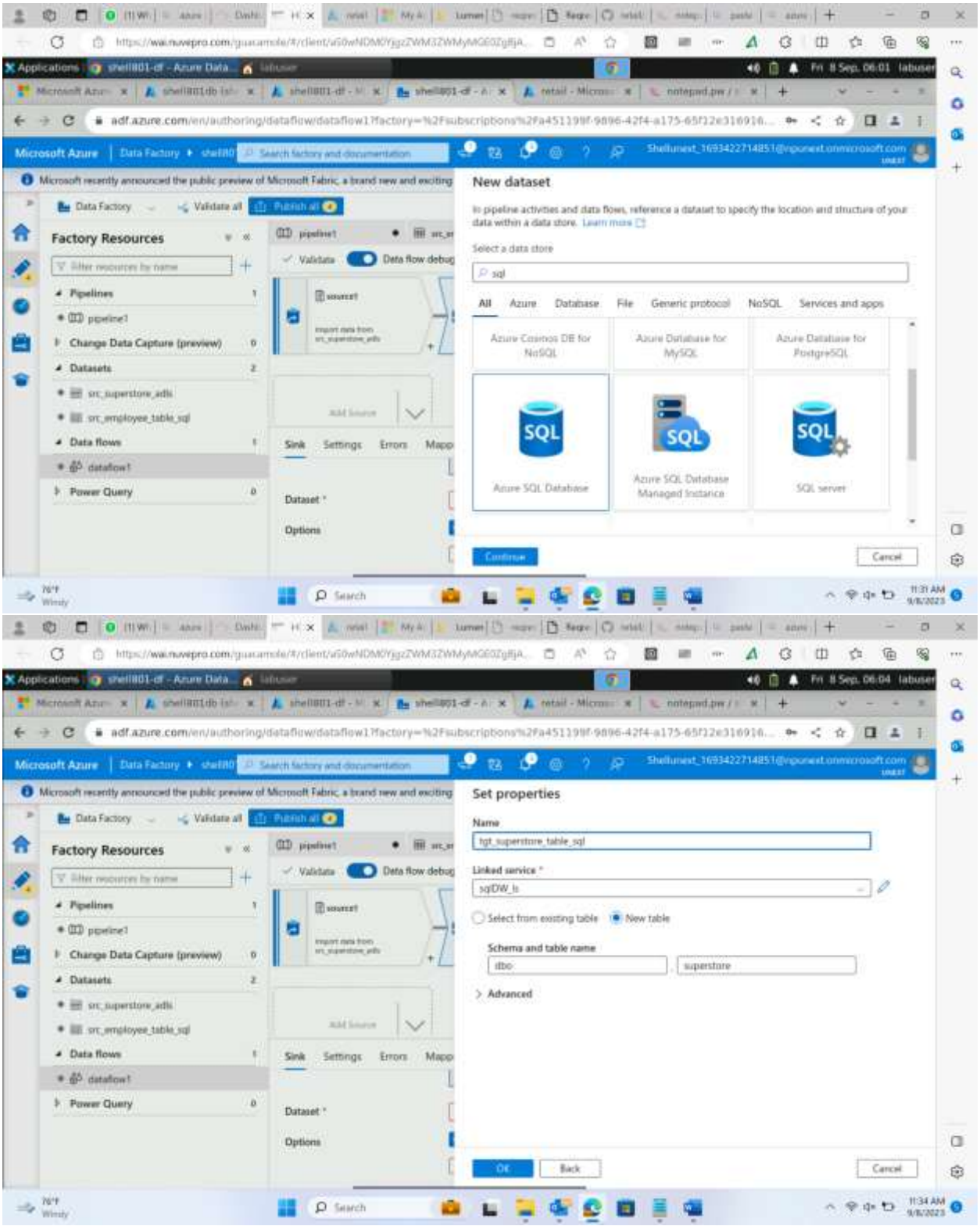
New dataset
Adls2



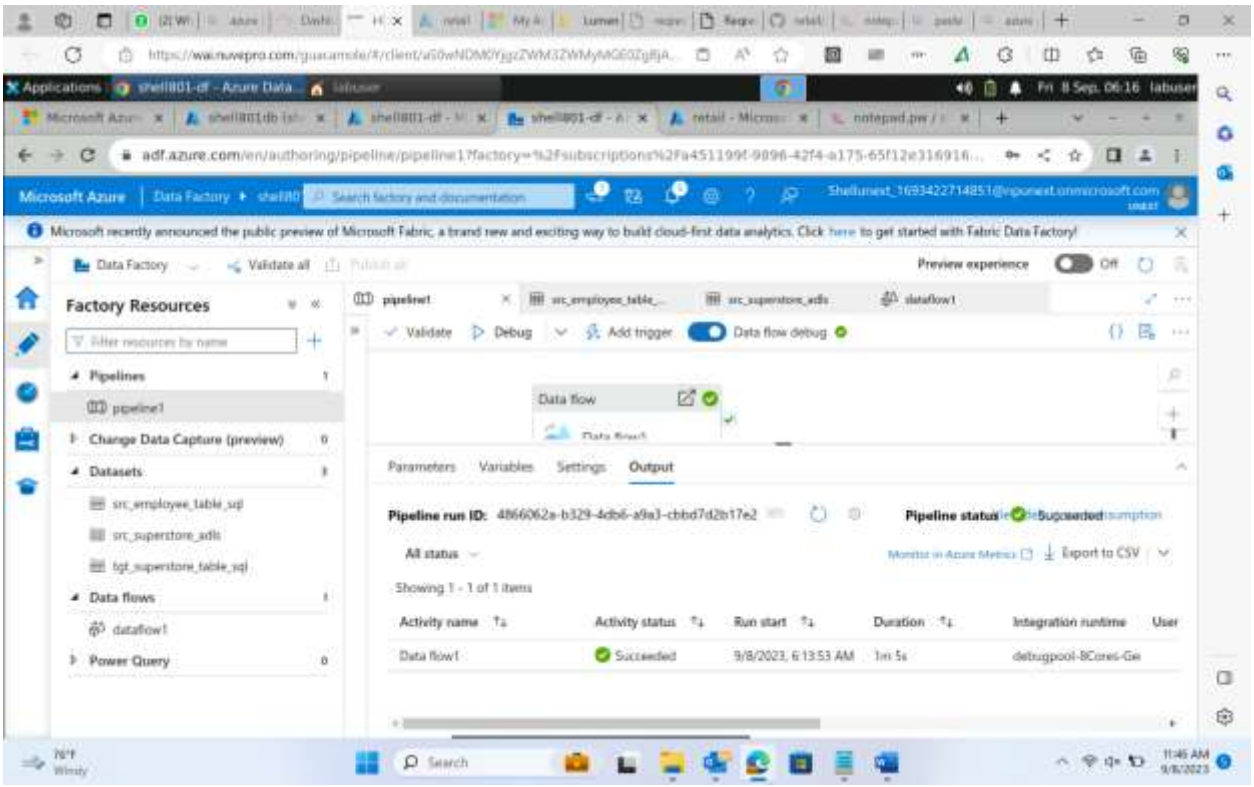


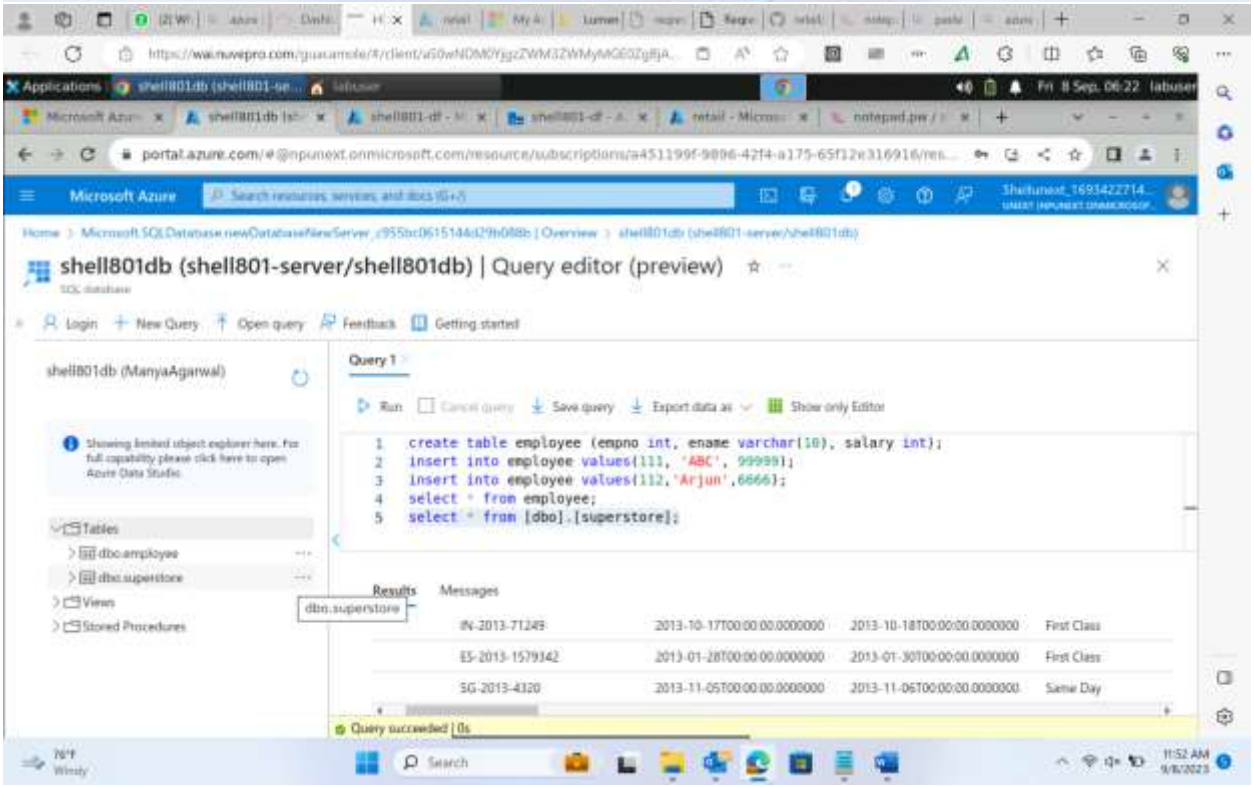
Data flow
Superstore_adls dataset
Projection – id, date, sales, qty, discount

sink

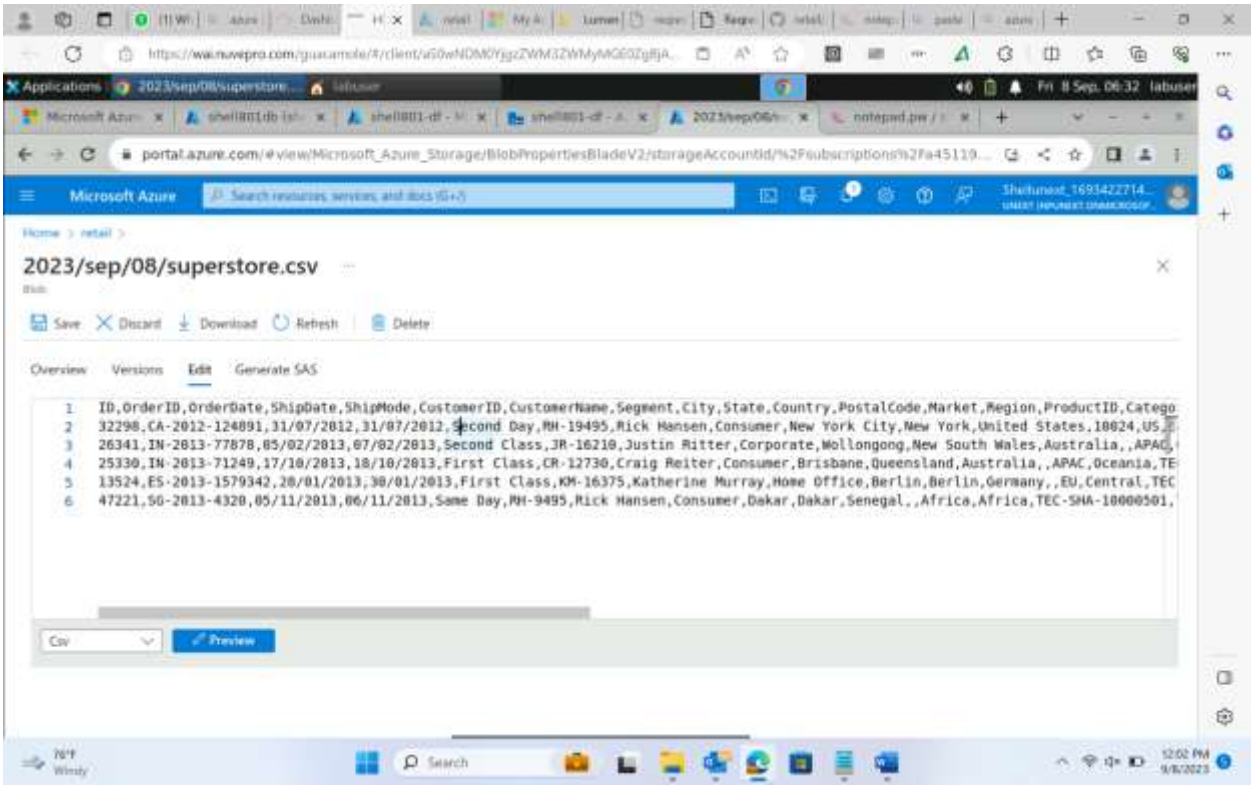


Pipeline → activity → dataflow

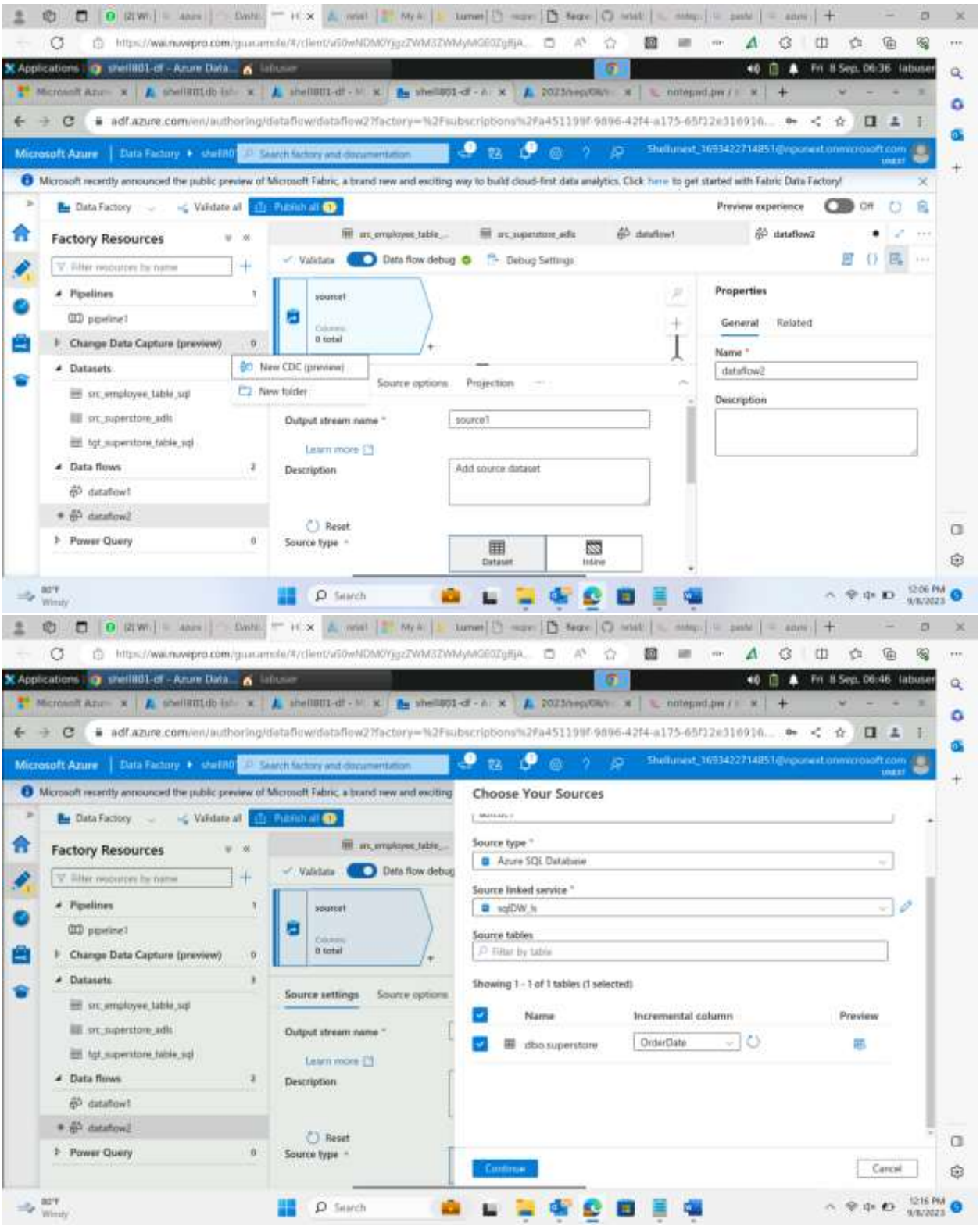

Microsoft Azure
Data Factory
Factory Resources
Pipelines
pipeline1
Change Data Capture (preview)
Datasets
src_employee_table_sql
src_superstore_adls
tgt_superstore_table_sql
Data flows
dataflow1
Power Query
Validate
Debug
Add trigger
Data flow debug
Data flow
Parameters
Variables
Settings
Output
Pipeline run ID: 4866062a-b329-4db6-a9a3-cbbd7d2b17e2
Pipeline status Succeeded
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
shell801db (shell801-server/shell801db) | Query editor (preview)
Login
New Query
Open query
Feedback
Getting started
shell801db (ManyaAgarwal)
Query 1
Run
Save query
Export data as
Show only Editor
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];
Tables
dbo.employee
dbo.superstore
Views
Stored Procedures
Results
Messages
dbo.superstore
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
Query succeeded | 0s

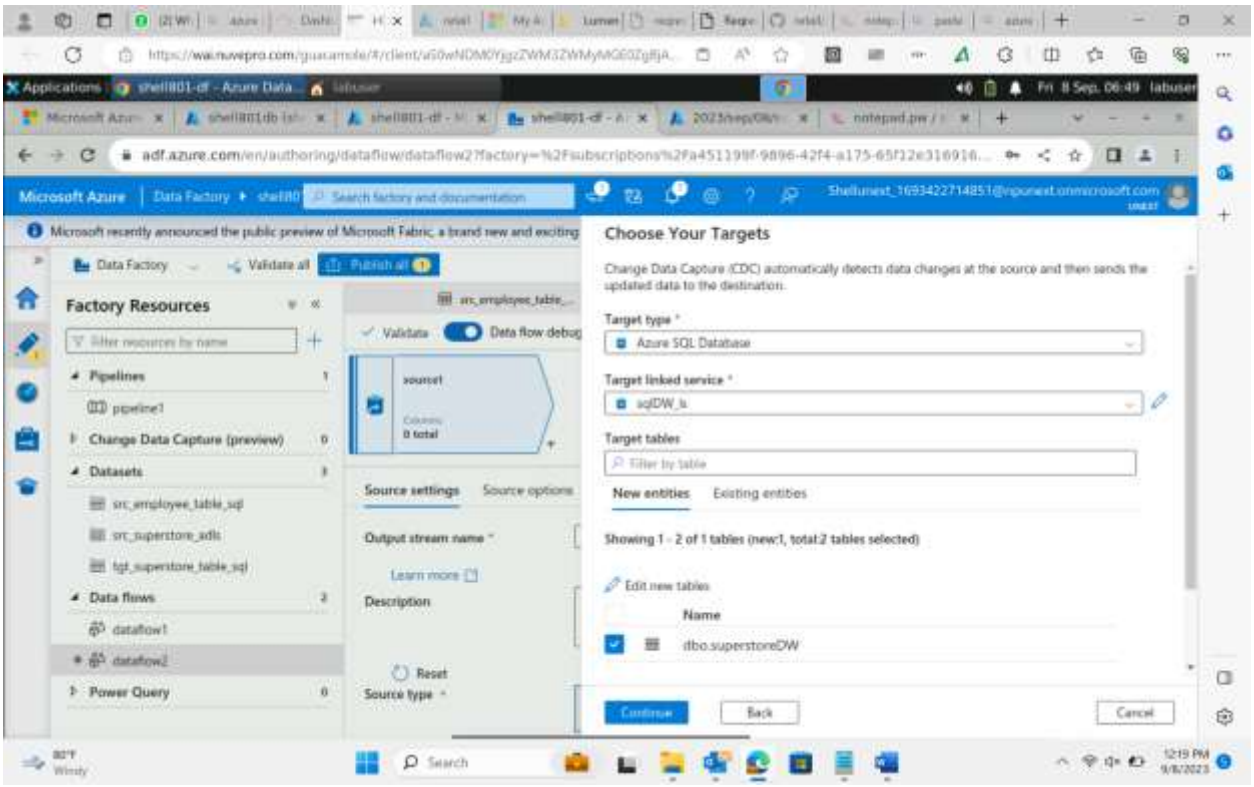
Full import – first time load
Incremental load/delta import – changes we make

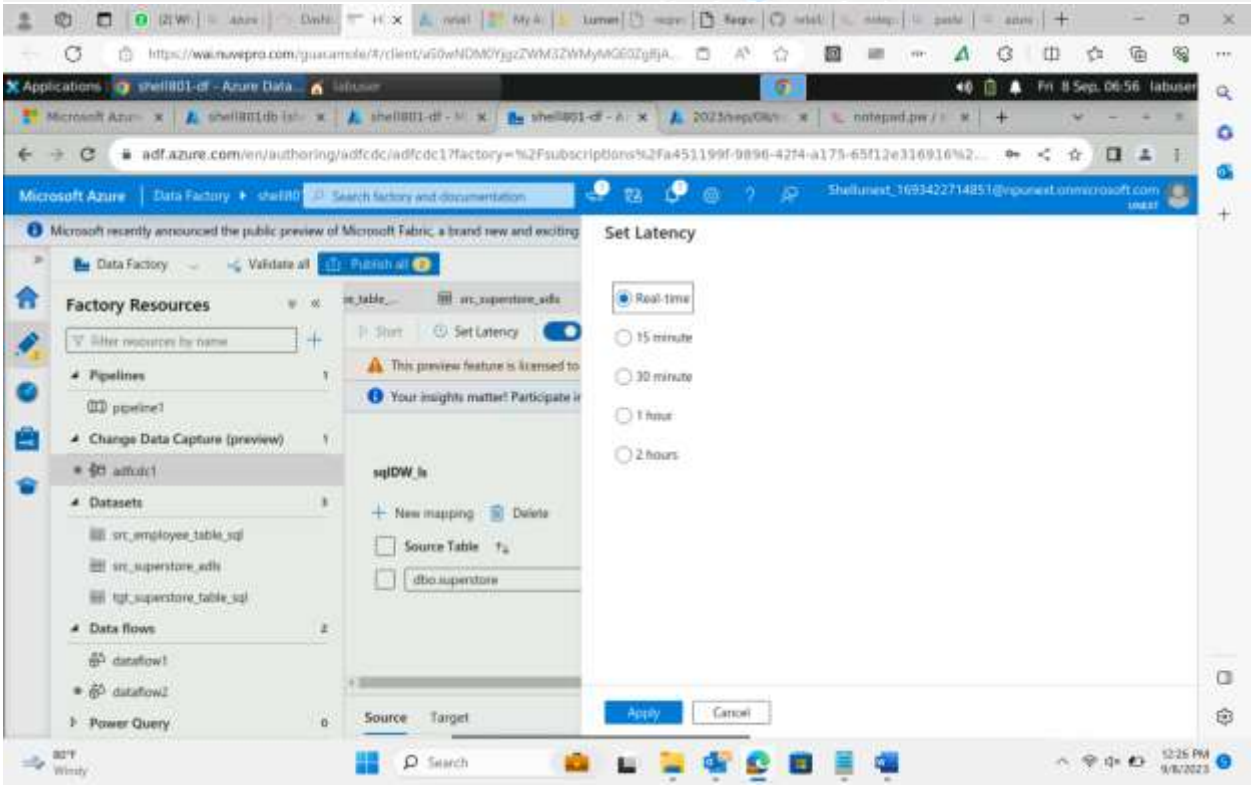


CDC -

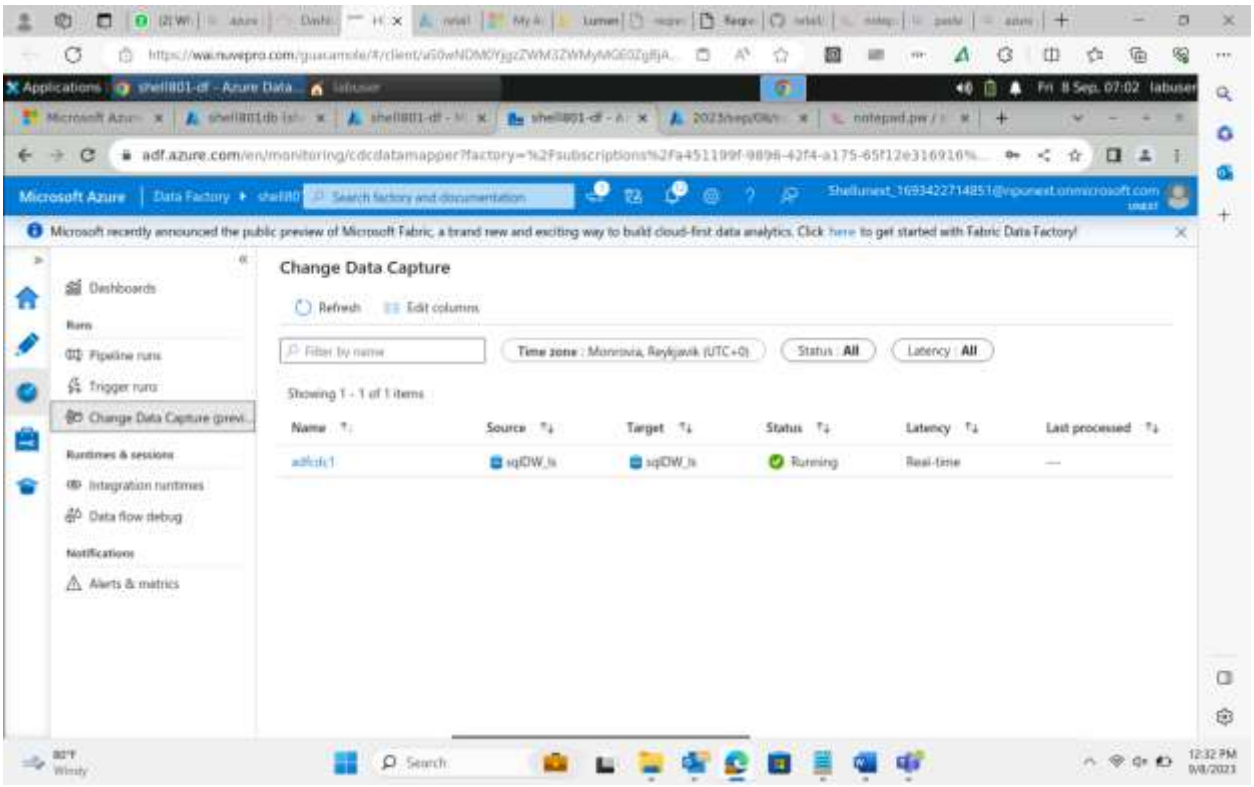

Factory Resources
Pipelines
pipeline1
Change Data Capture (preview)
Datasets
src_employee_table_sql
src_superstore_adls
tgt_superstore_table_sql
Data flows
dataflow1
dataflow2
Power Query
New CDC (preview)
New folder
Source options
Projection
Output stream name *
source1
Description
Add source dataset
Source type *
Dataset
Inline
Properties
General
Related
Name *
dataflow2
Description
Choose Your Sources
Source type *
Azure SQL Database
Source linked service *
Source tables
Showing 1 - 1 of 1 tables (1 selected)
Name
Incremental column
Preview
dbo.superstore
OrderDate
Continue
Cancel

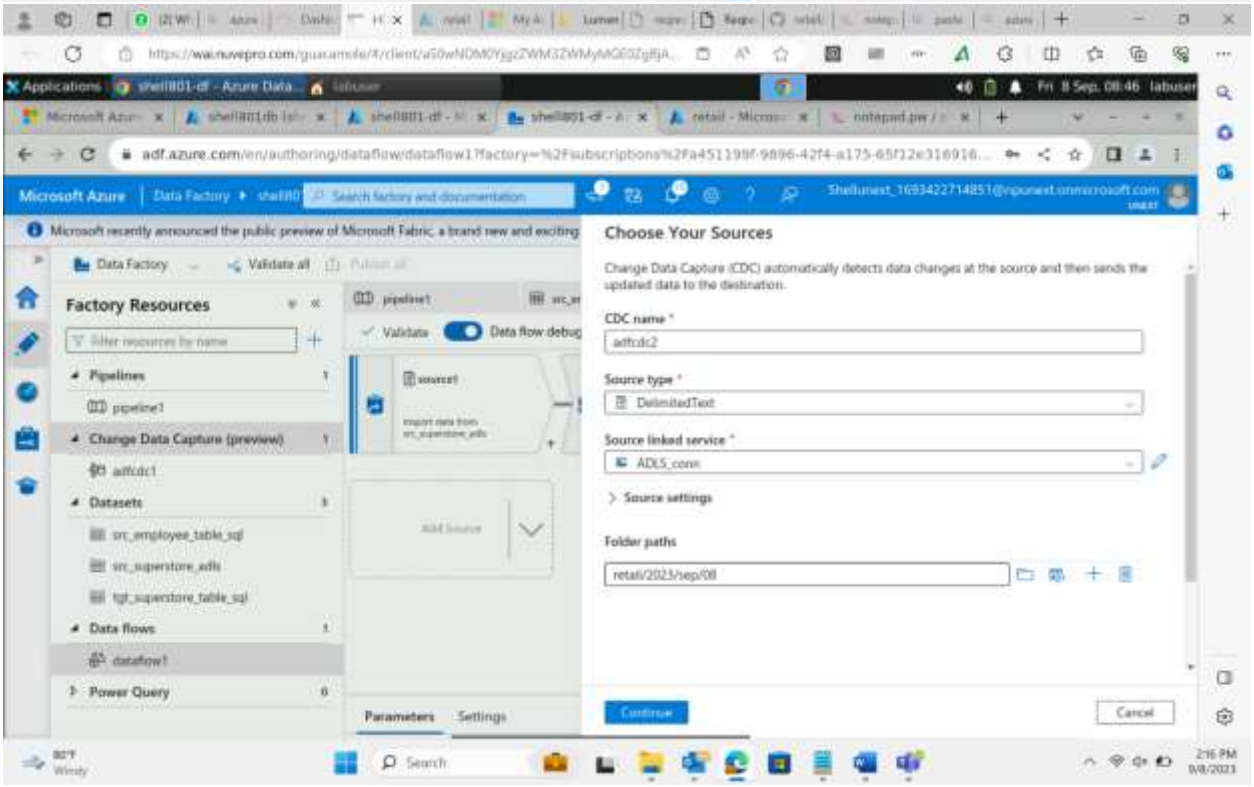
New entity

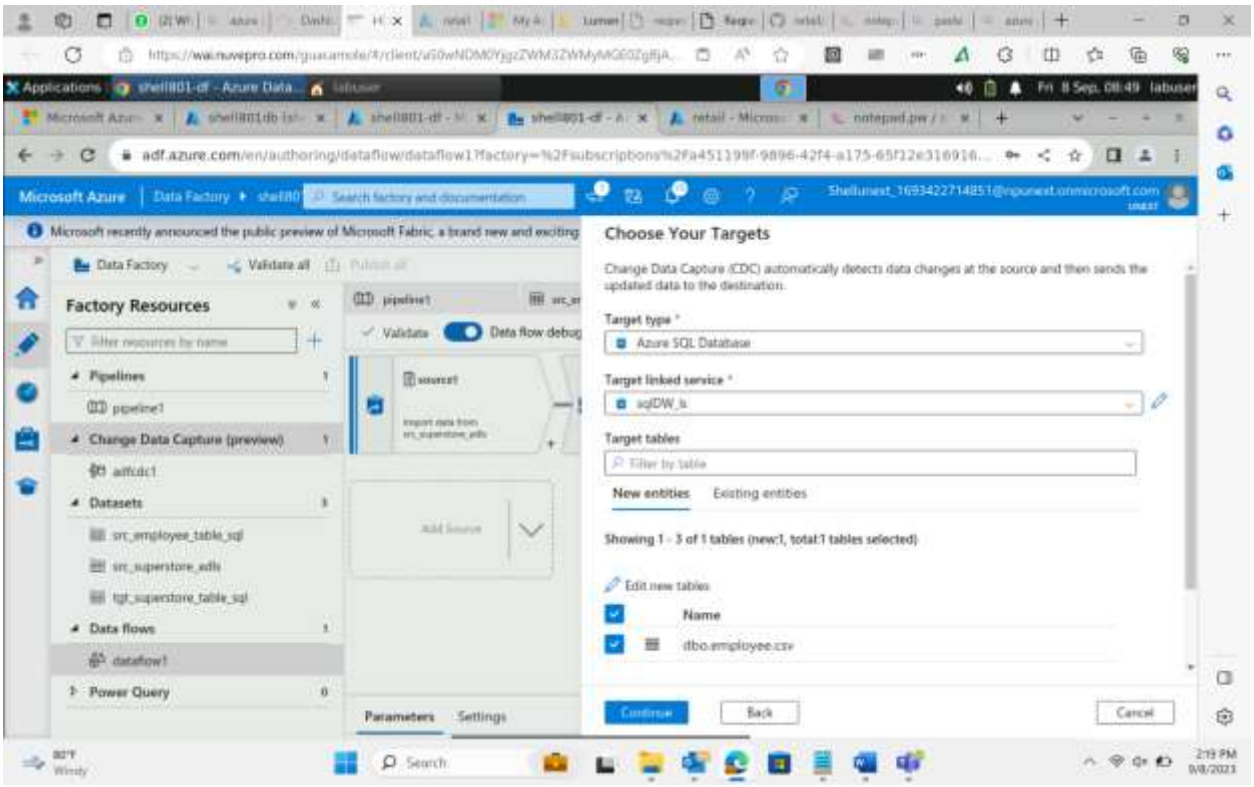


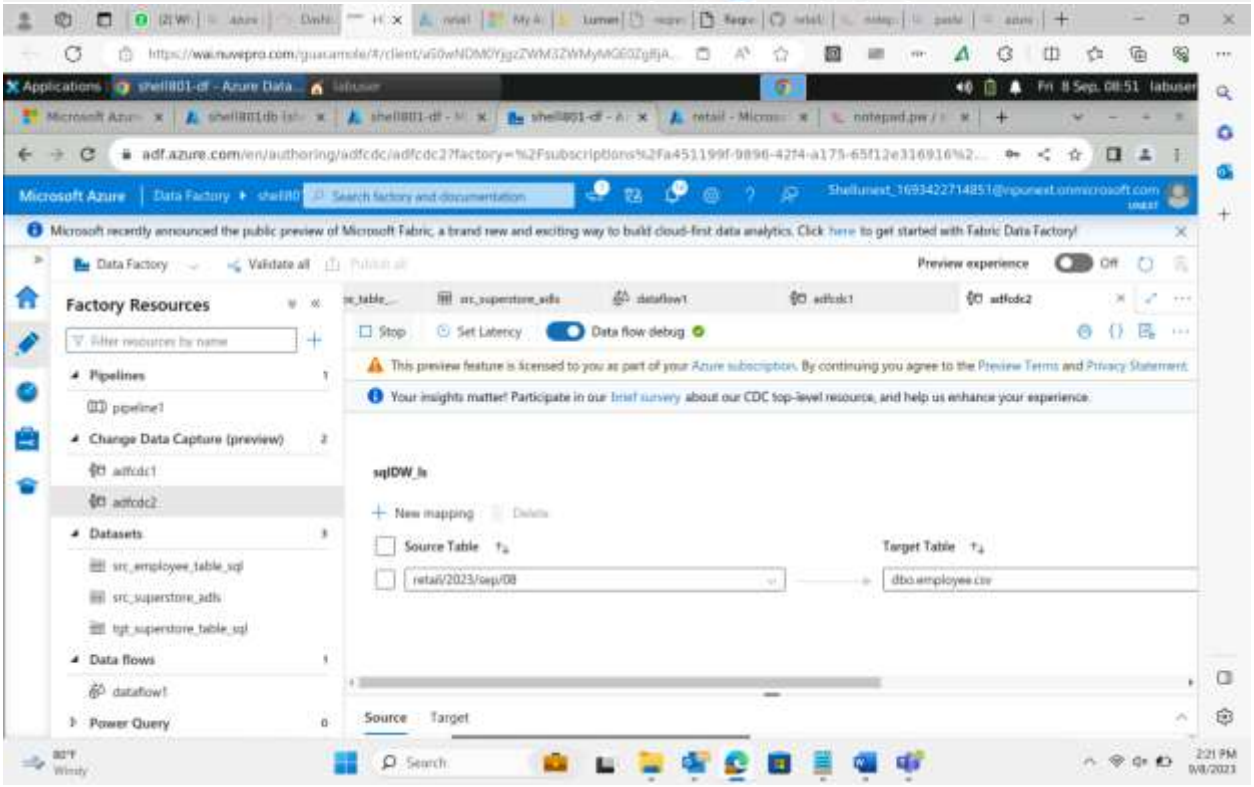


Publish
Start


Change Data Capture
Refresh
Edit columns
Dashboards
Pipeline runs
Trigger runs
Change Data Capture (previ...
Integration runtimes
Data flow debug
Alerts & metrics
Showing 1 - 1 of 1 items
Name
Source
Target
Status
Latency
Last processed
adfcdc1
Running
Real-time

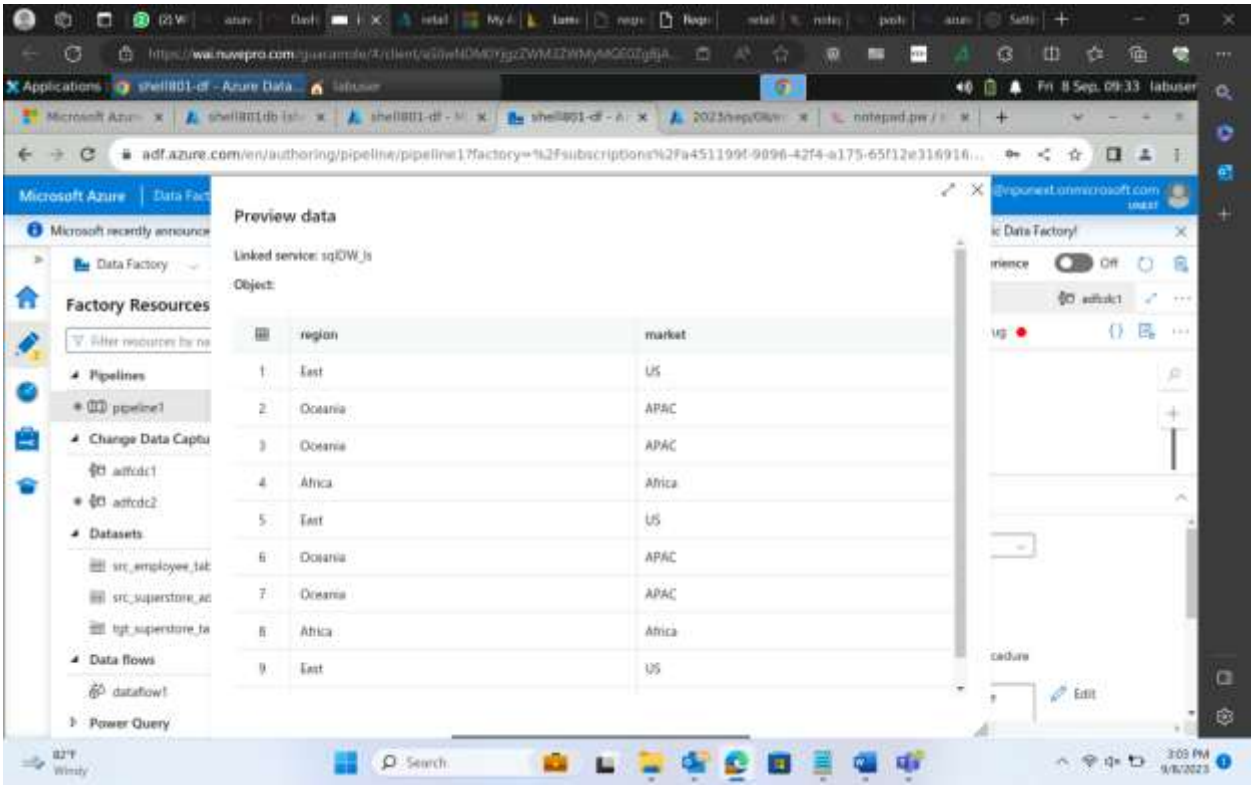

Factory Resources
Pipelines
pipeline1
Change Data Capture (preview)
adfcdc1
Datasets
src_employee_table_sql
src_superstore_adls
tgt_superstore_table_sql
Data flows
dataflow1
Power Query
Choose Your Sources
Change Data Capture (CDC) automatically detects data changes at the source and then sends the updated data to the destination.
CDC name *
adfcdc2
Source type *
DelimitedText
Source linked service *
ADLS_conn
Source settings
Folder paths
retail/2023/sep/08
Parameters
Settings
Continue
Cancel

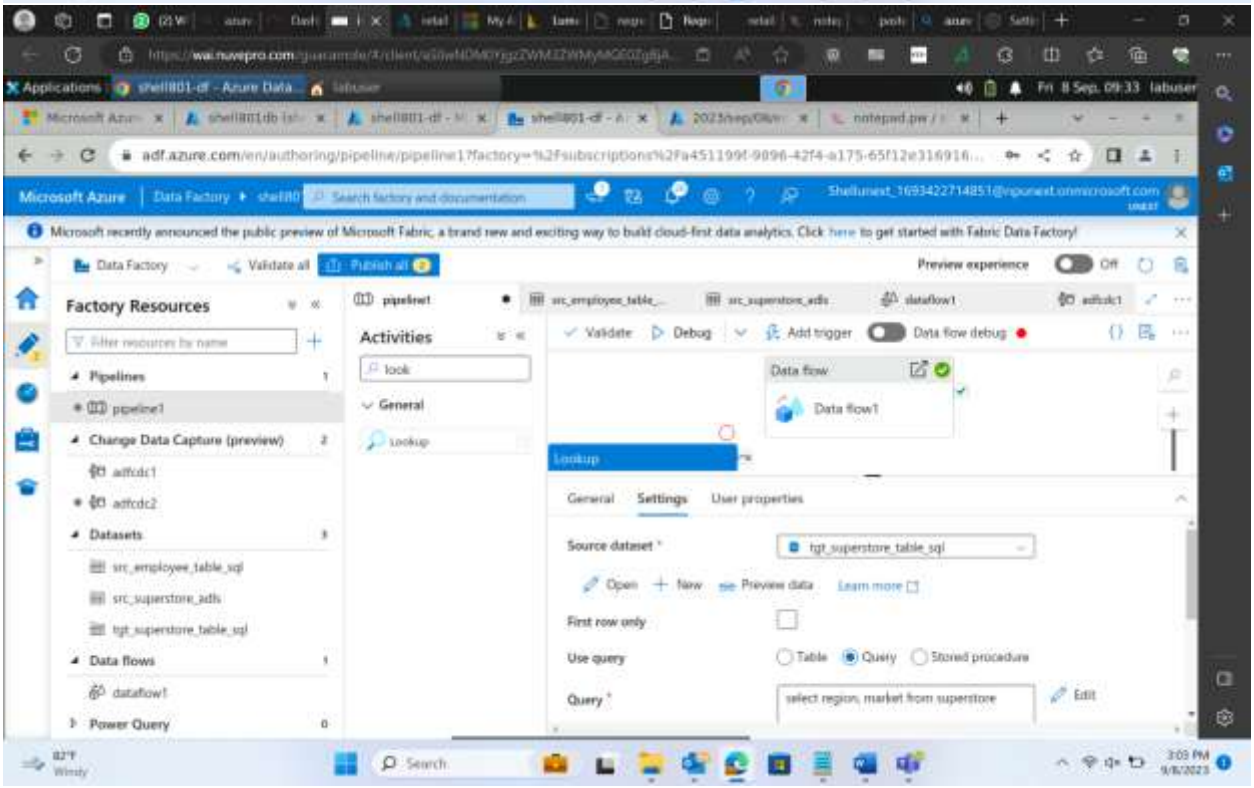




Edit employee.csv in storage
Delete one row and update something

Lookup activity



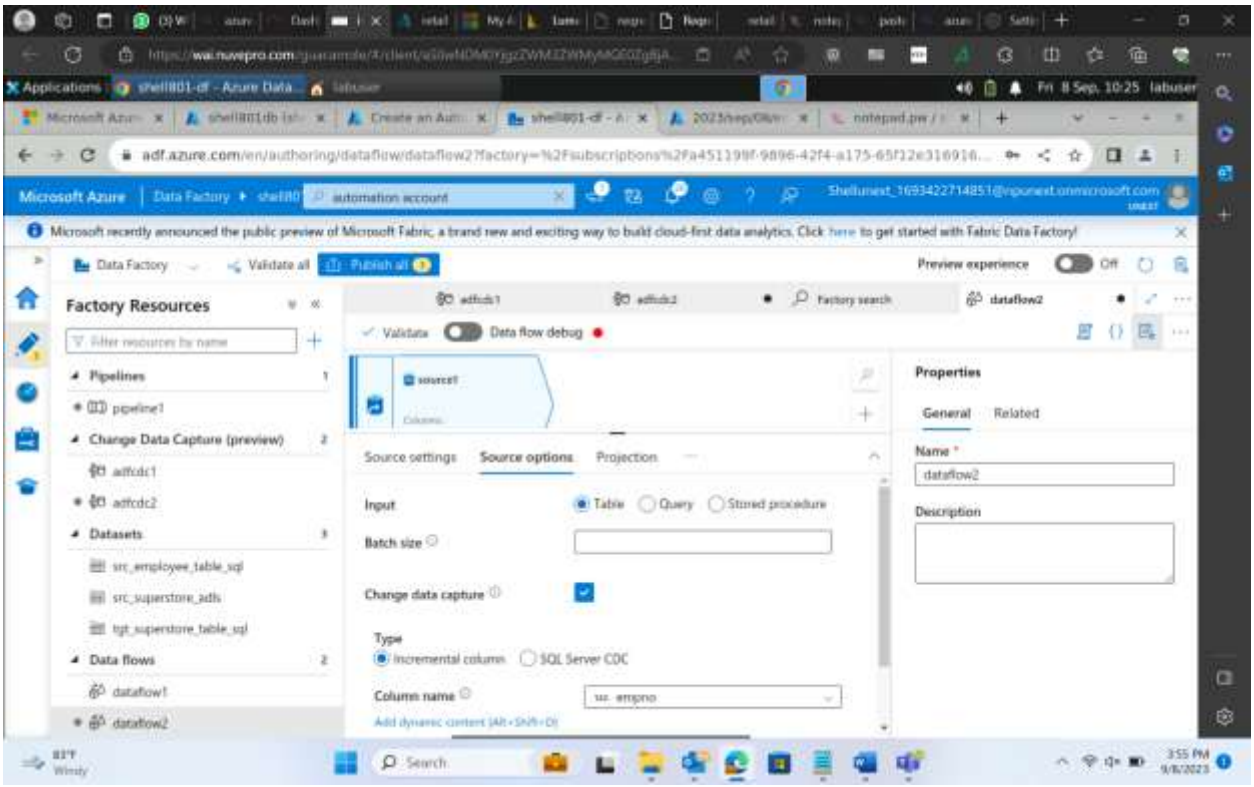


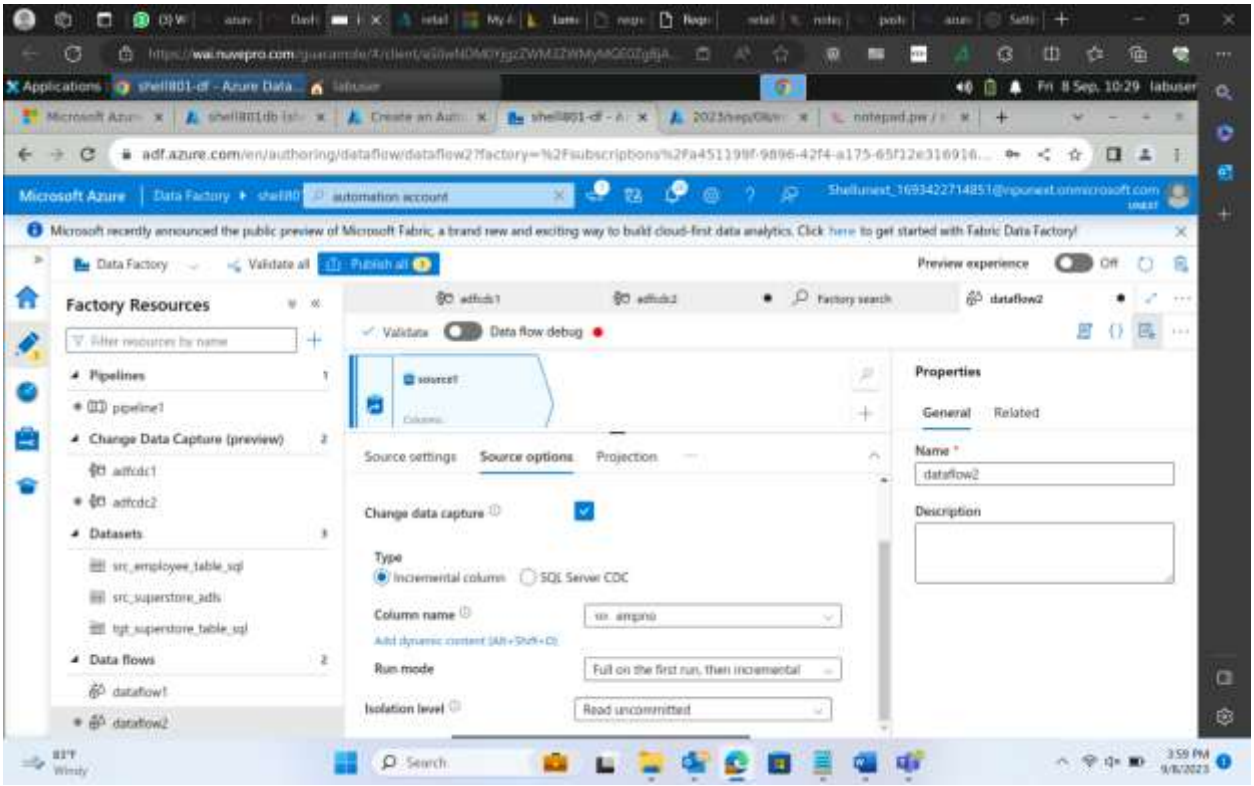
ADF – ETL tool

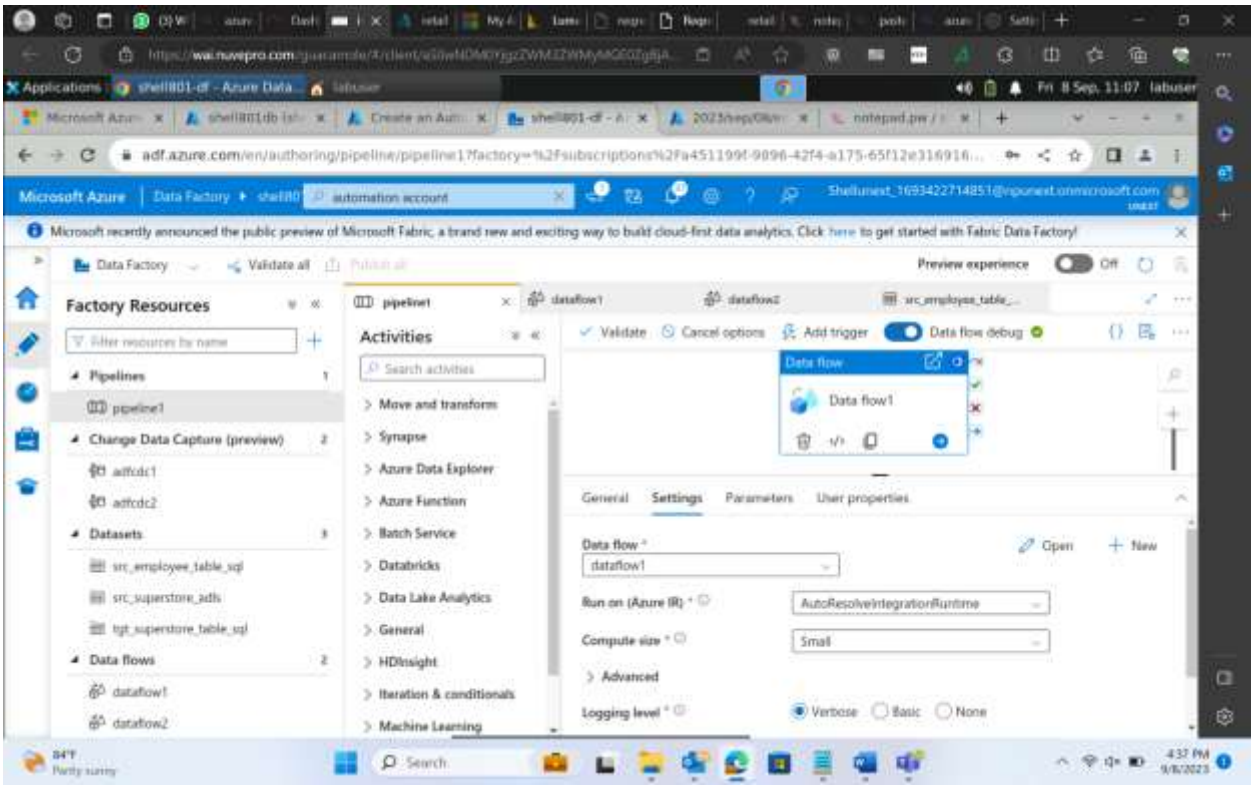
Orchestrating the data

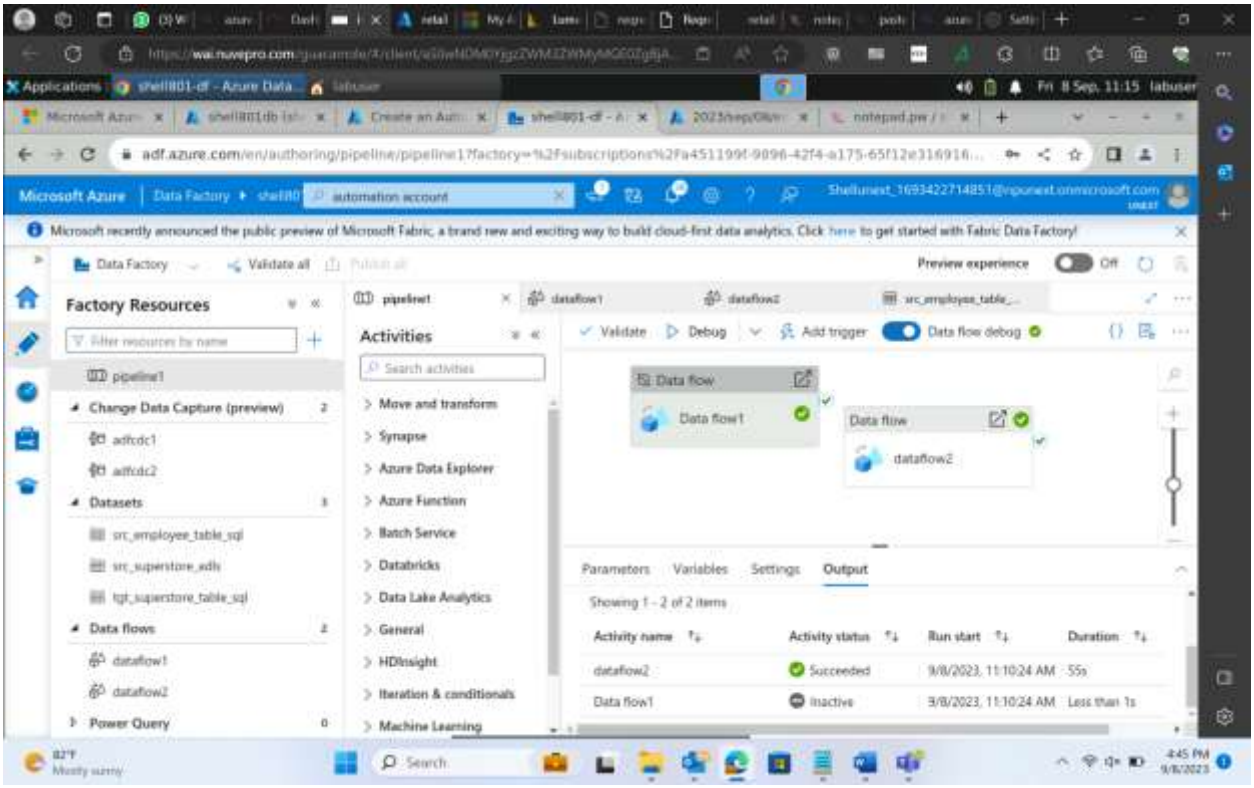
1. **Pipeline**
2. **Activity**
3. **Control flow**
4. **Get metadata – get information from file (childitems, file size, etc)**
5. **Lookup – get information for table**
6. **Copy data – copying bulk data (source and sink)**
7. **Linked service and datasets are reusable.**
8. **Foreach – iteration**
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


Microsoft Azure
Data Factory
Factory Resources
Pipelines
pipeline1
Change Data Capture (preview)
adfcdc1
adfcdc2
Datasets
src_employee_table_sql
src_superstore_adls
tgt_superstore_table_sql
Data flows
dataflow1
dataflow2
Source settings
Source options
Projection
Input
Table
Query
Stored procedure
Batch size
Change data capture
Type
Incremental column
SQL Server CDC
Column name
Properties
General
Related
Name
Description


Microsoft Azure
Data Factory
Factory Resources
Pipelines
pipeline1
Change Data Capture (preview)
adfcdc1
adfcdc2
Datasets
src_employee_table_sql
src_superstore_adls
tgt_superstore_table_sql
Data flows
dataflow1
dataflow2
Source settings
Source options
Projection
Change data capture
Type
Incremental column
SQL Server CDC
Column name
Run mode
Full on the first run, then incremental
Isolation level
Read uncommitted
Properties
General
Related
Name
Description





Logging level :L none if data flow gives error