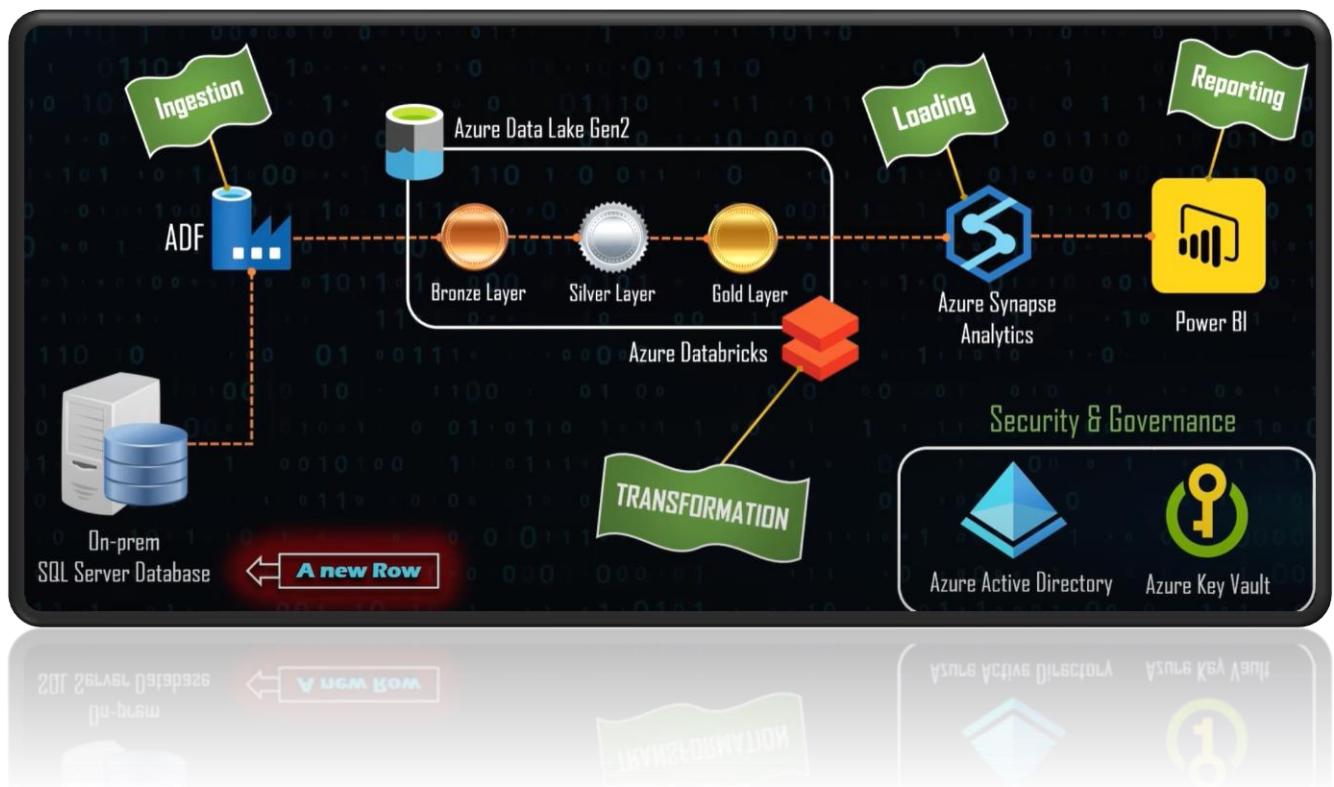


# *END To END ETL Implementation Project*



## Contents

<b><i>END To END ETL Implementation Project</i></b> .....	1
1. Set Up Environment .....	3
2. Data Ingestion using ADF .....	6
3. Data Transformation With Databricks.....	15
4. Creating Views Using Synapse.....	17
5. Publishing Data Analytics Reports – Power BI .....	22
6. Security and Governance .....	24
7. End to End Pipeline Testing .....	25
8. References.....	27

## Scenario :

Performing a migration from On-Premise PostgreSQL server to Azure cloud and doing a complete End to end implementation which includes Migration, Extraction, Transformation ,Loading and lastly Analytics .

### 1. Set Up Environment

1.1 For the set up environment I created a new resource group for the project and have the following resources up and running.

These Include : **Azure Databricks, Azure Data Factory, Azure Storage Account, Azure Synapse Analytics Workspace and Azure Key Vault**

The screenshot shows the Azure portal interface for a resource group named 'RG\_data\_eng\_project'. The left sidebar contains navigation links for Overview, Activity log, Access control (IAM), Tags, Resource visualizer, Events, Deployments, Security, Deployment stacks, Policies, and Properties. The main content area is titled 'Essentials' and displays a list of resources under the 'Resources' tab. The list includes:

Name	Type	Location
adb-data-eng-proj	Azure Databricks Service	East US
adf-data-eng-prj007	Data factory (V2)	East US
kvdataengproj	Key vault	East US
stgacdataengprj	Storage account	East US
syn-data-eng-proj	Synapse workspace	East US

Filtering options at the top include 'Type equals all' and 'Location equals all'. A tooltip on the right side of the table header says: 'Switch between a list view and a summary chart view counts.'

1.2 In my On-Premise Microsoft SQL Server/PostgreSQL, Tables which are being migrated :

The screenshot shows the pgAdmin 4 interface. At the top, there's a connection bar with the user 'postgres/postgres@PostgreSQL 16'. Below it is a toolbar with various icons for file operations, search, and navigation. The main area has tabs for 'Query' and 'Query History', with 'Query' selected. A query window contains the following SQL code:

```

1 select table_name from information_Schema.tables
2 where table_schema = 'public'

```

The results pane below shows a table with one column, 'table\_name', containing four rows: 'departments', 'regions', 'employees', and 'dupes'.

table_name
departments
regions
employees
dupes

### 1.3 Created Azure Key Vault secrets Which comprise secrets created for the SQL server user Login Id and Password

The screenshot shows the 'Create a secret' page in the Azure portal. The URL is 'Home > kvdataengproj | Secrets > Create a secret'. The form fields are as follows:

- Upload options:** Manual
- Name \***: SQL-server-username
- Secret value \***: ...
- Content type (optional)**: (empty)
- Set activation date**: (checkbox)
- Set expiration date**: (checkbox)
- Enabled**: Yes (selected)
- Tags**: 0 tags

Home > kvdataengproj

## kvdataengproj | Secrets

Key vault

Search Generate/Import Refresh Restore Backup View sample code Manage deleted secrets

The secret 'SQL-server-password' has been successfully created.

Name	Type	Status	Expiration date
SQL-server-password		✓ Enabled	
SQL-server-username		✓ Enabled	

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Access policies Events Objects Keys Secrets Certificates

Also, create a linked service for the key vault, as this needs to be linked whenever key vault is used.

New linked service

Azure Key Vault

Name \* AzureKeyVault1

Description Linked service for Azure Key vault

Azure key vault selection method  From Azure subscription  Enter manually

Azure subscription  Azure for Students [f4223b96-e74d-41b1-87de-96deceef0cf]

Azure key vault name \* kvdataengproj

Edit key vault

Authentication method System Assigned Managed Identity

Managed identity name: adf-data-eng-prj007  
Managed identity object ID: 3b95bc62-22b7-4fea-9c54-cf10aafc0ef1  
Grant Data Factory service managed identity access to your Azure Key Vault. Learn more

Test connection  To linked service  To secret

## 1.4 Storage account and containers

Home > stgacdataengprj

## stgacdataengprj | Containers

Storage account

Search Container Change access level Restore containers Refresh Delete Give feedback

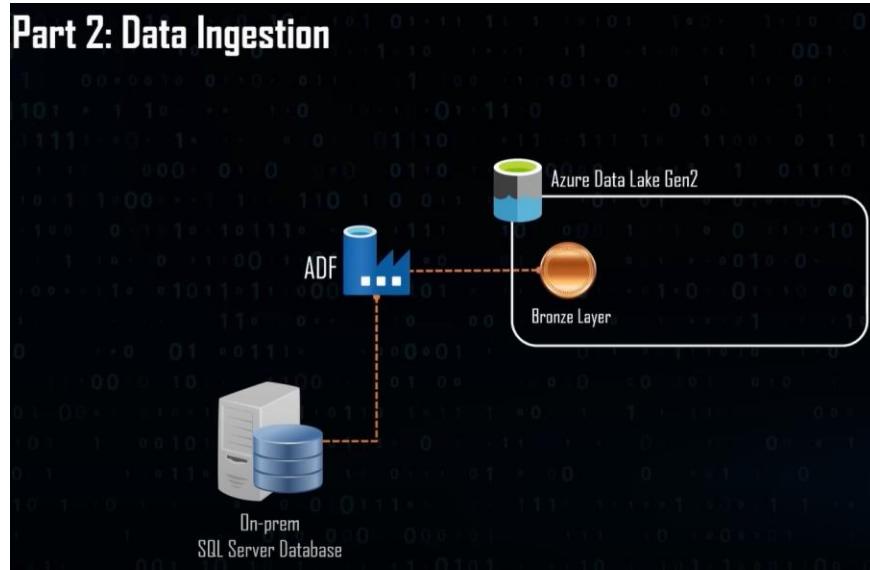
Search containers by prefix

Name	Last modified	Anonymous access level
\$logs	9/23/2023, 10:30:02 PM	Private
bronze	9/24/2023, 1:05:00 PM	Private
dbccontainer	9/24/2023, 12:02:35 AM	Private
gold	9/24/2023, 1:05:19 PM	Private
silver	9/24/2023, 1:05:14 PM	Private

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

## 2.Data Ingestion using ADF

### Part 2: Data Ingestion



#### 2.1 Creating a Self-Hosted Integration Runtime in Azure Data Factory as we are connecting to the on-premises server.

The screenshot shows the 'Integration runtime setup' page in the Azure Data Factory portal. The 'Settings' tab is selected. A new integration runtime named 'SelfHostedIR' is being created. The page provides options for Express setup or Manual setup, and includes fields for Name and Authentication key.

#### 2.2 Create a dataset and a linked service for the source of Copy data pipeline

Linked Service for Postgresql

The screenshot displays two side-by-side views of the Microsoft Azure Data Factory interface.

**Left View (Linked Services):**

- The title bar shows "Microsoft Azure | Data Factory > adf-data-eng-prj007".
- The left sidebar includes sections like "Connections", "Integration runtimes", "Microsoft Purview", "Source control", "Git configuration", "ARM template", "Author", "Triggers", "Global parameters", "Data flow libraries", "Security", "Credentials", "Customer managed key", "Outbound rules", and "Managed private endpoints".
- The main area is titled "Linked services" and shows a list of existing items:
 

Name	Type
AzureKeyVault1	Azure Key Vault
SqlServer1	SQL server

**Right View (Set properties for Copy data activity):**

- The title bar shows "Microsoft Azure | Data Factory > adf-data-eng-prj007".
- The left sidebar shows "Factory Resources" with "Pipelines" selected, displaying a list of pipelines including "Migrate\_data".
- The main area shows the "Activities" pane with "Copy data" selected.
- A modal dialog titled "Set properties" is open for the "Copy data1" activity:
 

Name	PostgreSQLTable1
Linked service	PostgreSQL1
Connect via integration runtime	integrationRuntime1
Table name	public.departments

### 2.3 Creating a pipeline in ADF with Lookup activity .

The screenshot shows the Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. The 'Activities' pane on the right shows a 'Lookup' activity selected. The 'Settings' tab is active, displaying the configuration for the 'Source dataset' (PostgreSQLTable2), 'First row only' (unchecked), 'Use query' (selected 'Query'), and the 'Query' text box containing the SQL script:

```
select table_name from information_Schema.tables  
where table_schema = 'public'
```

**Query :** select table\_name from information\_Schema.tables  
where table\_schema = 'public'

Validated the activity and debugged the activity to get the output as follows

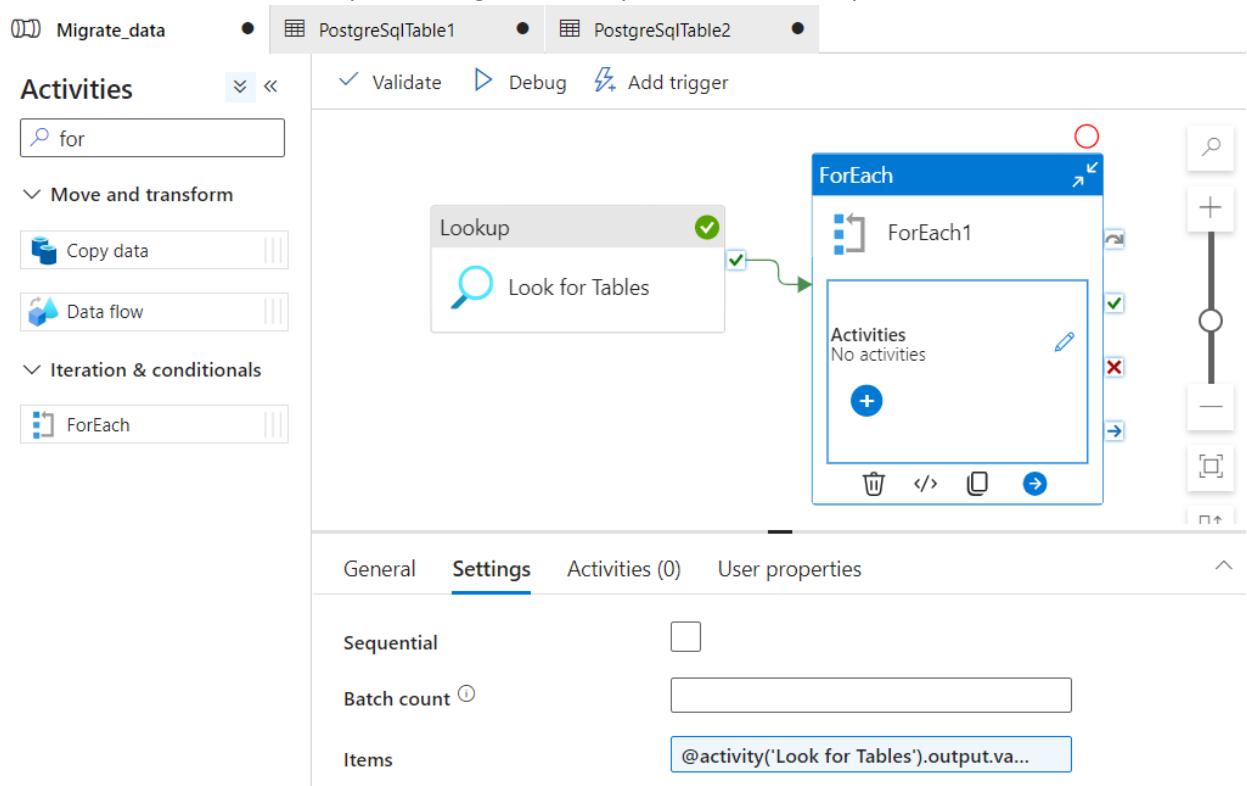
The screenshot shows the Azure Data Factory interface for a pipeline named "Migrate\_data". The pipeline consists of two stages: "PostgreSqlTable1" and "PostgreSqlTable2". The "PostgreSqlTable1" stage contains a "Lookup" activity named "Look for Tables". The output of this activity is displayed in JSON format:

```
{ "count": 4, "value": [ { "table_name": "departments" }, { "table_name": "regions" } ] }
```

The pipeline status bar at the bottom indicates the activity "Look for Tables" has succeeded. The overall run status is "Succeeded" and the run started on 9/24/2023, 12:00.

The JSON format output includes the list of tables in my PostgreSQL database.

2.4 Now, with the for each activity, we configure the output from the look up tables ;



In the items, we add dynamic content and take the output from lookup activity with expression

### Pipeline expression builder

Add dynamic content below using any combination of [expressions](#), [functions](#) and [system variables](#).

```
@activity('Look for Tables').output.value
```

[Clear contents](#)

2.5 Configure a **copy activity** as a sub activity in the for each activity

✓ Validate ✓ Validate copy runtime ⚡ Debug ⚡ Add trigger { }

Migration Pipeline > ForEach1

Copy data

Copy data1

General Source Sink Mapping Settings User properties

Source dataset \* PostgreSqlTable2 Open + New Preview data Learn more

Use query Table Query

Query @concat('Select \* FROM ', item().tab...)

Additional columns + New

Created a new dataset with the existing self hosted Integration runtime

Data Factory Validate all Publish all

Factory Resources

- Pipelines 1
  - Migrate\_data
- Datasets 3
  - Parquet1
  - PostgreSqlTable1
  - PostgreSqlTable2
- Data flows 0
- Power Query 0

Migrate\_data PostgreSqlTable1 PostgreSqlTable2 Parquet1

PostgreSQL PostgreSqlTable2

Connection Parameters

Linked service \* PostgreSql1 Test connection Edit + New Learn more

Integration runtime \* integrationRuntime1 Edit

Table Select... Refresh Preview data

A simple Concat query with the dynamic content from the previous activity

```
@{concat('Select * FROM ', item().table_name)}
```

Sink Dataset with parquet format , with table name being parameterised with dynamic content , as we have multiple tables being migrated.

The screenshot shows the 'Copy data' dialog within the Azure Data Factory interface. The 'Sink' tab is active, and the sink dataset is set to 'Parquet1'. In the 'Dataset properties' section, there is a mapping for 'tablename' with the value '@item().tablename'. The 'Copy behavior' dropdown is set to 'Select...'. The overall title bar indicates the pipeline name is 'Migrate\_data'.

Configured the dataset to store the dataset in parquet and store it in a hierarchical fashion.

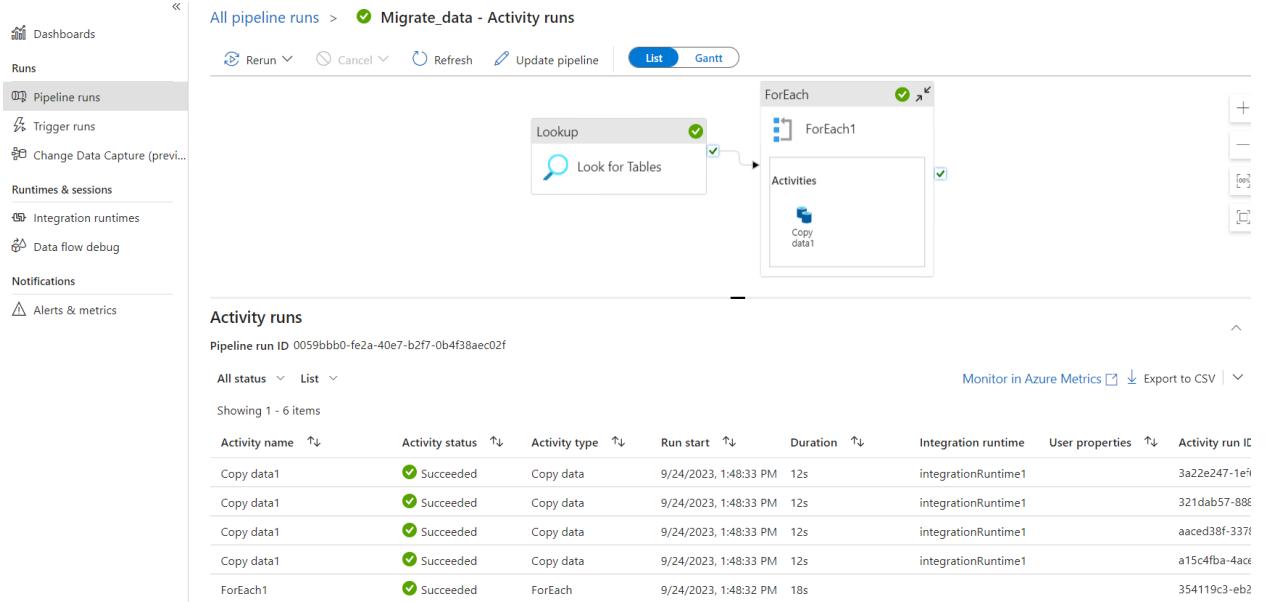
The screenshot shows the 'Parquet1' dataset configuration. The 'Connection' tab is selected, showing 'LS\_DataLake' as the linked service. The 'File path' is set to 'bronze' followed by a dynamic expression. The 'Compression type' is set to 'snappy'.

**Container\_name/OrganisationLT/Table\_name/Tablename.parquet**

**For Directory : @{{concat('OrganisationLT/ ', dataset().tablename)}}**

**For File name : @{{concat(dataset().tablename, '.parquet')}}**

## 2.6 Validated and published the pipeline and triggered the pipeline. Ran Successfully.



## 2.7 Validated for the files in the Storage account

The screenshot shows the Azure Storage Explorer interface for the 'bronze' container. The 'Overview' tab is selected. The container has an authentication method of 'Access key' and is located in the 'bronze / OrganisationLT' storage account. A search bar at the top right allows searching by blob prefix. Below the tabs, there is a list of blobs with their names, modified dates, access tiers, archive statuses, and blob types. The blobs listed are: [..], departments, dupes, employees, and regions.

Name	Modified	Access tier	Archive status	Blob
[..]				
departments				
dupes				
employees				
regions				

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

Home > stgacdataengprj | Containers >

**bronze** Container

Search Overview Diagnose and solve problems Access Control (IAM)

**Authentication method:** Access key ([Switch to Azure AD User Account](#))  
**Location:** bronze / OrganisationLT / departments

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

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[...]						...
departments.parquet	9/25/2023, 12:50:37 ...	Hot (Inferred)		Block blob	824 B	Available

Home > stgacdataengprj | Containers >

**bronze** Container

Search Overview Diagnose and solve problems Access Control (IAM)

**Authentication method:** Access key ([Switch to Azure AD User Account](#))  
**Location:** bronze / OrganisationLT / dupes

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

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[...]						...
dupes.parquet	9/25/2023, 12:50:35 ...	Hot (Inferred)		Block blob	460 B	Available

Home > stgacdataengprj | Containers >

**bronze** Container

Search Overview Diagnose and solve problems Access Control (IAM)

**Authentication method:** Access key ([Switch to Azure AD User Account](#))  
**Location:** bronze / OrganisationLT / employees

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

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[...]						...
employees.parquet	9/25/2023, 12:50:35 ...	Hot (Inferred)		Block blob	49.1 KiB	Available

Home > stgacdataengprj | Containers >

**bronze** Container

Search Overview Diagnose and solve problems Access Control (IAM)

**Authentication method:** Access key ([Switch to Azure AD User Account](#))  
**Location:** bronze / OrganisationLT / regions

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

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[...]						...
regions.parquet	9/25/2023, 12:50:35 ...	Hot (Inferred)		Block blob	696 B	Available

### 3.Data Transformation With Databricks

#### 3.1 Create a databricks single-node cluster while enabling credential passthrough for user-level data access

The screenshot shows the Azure Databricks interface for creating a cluster. The left sidebar lists various services like Workspace, Catalog, Workflows, Compute, SQL, and Data Engineering. The main panel is titled 'Data Project Cluster'. Under 'Compute', the 'Advanced options' section has a checked checkbox for 'Enable credential passthrough for user-level data access'. In the 'Spark config' section, there's a code block with three lines: 'spark.databricks.delta.preview.enabled true', 'spark.master local[\*], 4', and 'spark.databricks.cluster.profile singleNode'. Below that, in 'Environment variables', there's a line 'PYSPARK\_PYTHON=/databricks/python3/bin/python3'. At the bottom, there are 'Confirm and restart' and 'Cancel' buttons.

[Access Azure Data Lake Storage using Azure Active Directory credential passthrough \(legacy\) - Azure Databricks | Microsoft Learn](#)

I did mounting through credential passthrough method, which surprisingly worked well, unlike in my other projects where I couldn't make this work.

3.2 Created a databricks notebook activity for bronze to silver notebook, created linked service and tried utilizing Keyvault.

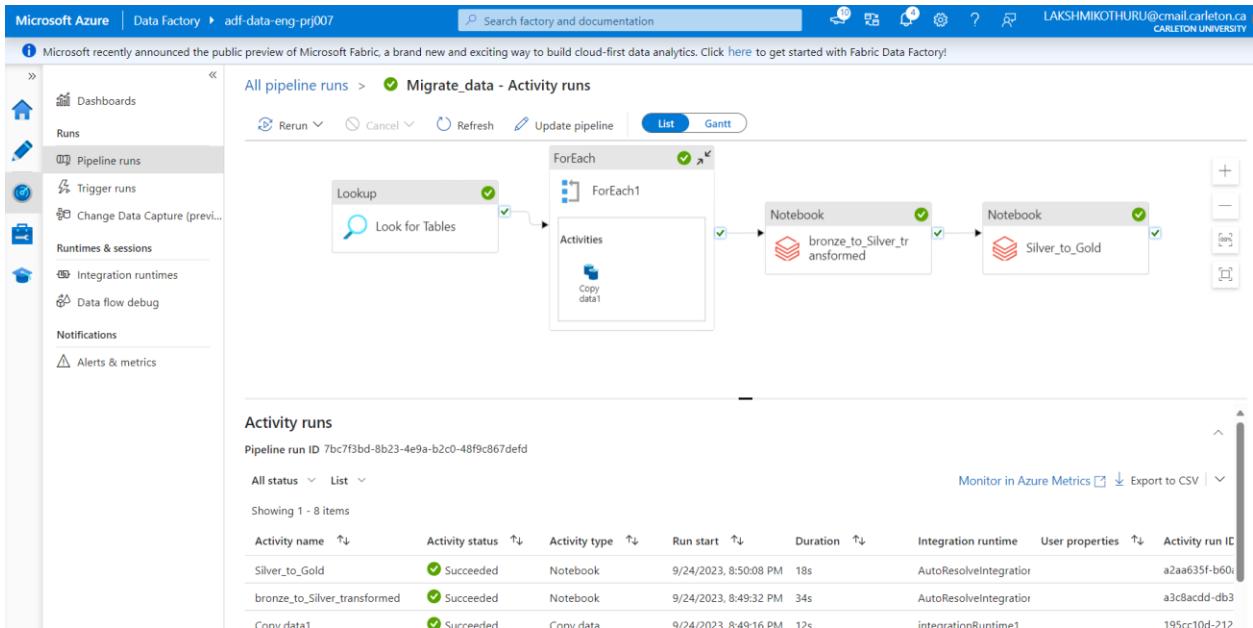
Since My account is not a premium account I'm unable to load the key vault secret name, even after applying role assignments.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. In the center, a pipeline named 'Migrate\_data' is being edited. A 'Lookup' activity is selected. On the right, a 'New linked service' dialog is open for 'Azure Databricks'. The 'Databricks workspace' dropdown is set to 'adb-data-eng-proj'. Under 'Select cluster', the 'Existing interactive cluster' radio button is selected. The 'Databrick Workspace URL' field contains 'https://adb-1358636769729125.azuredatabricks.net'. The 'Authentication type' dropdown is set to 'Access Token', with 'Access token' selected. The 'AKV linked service' dropdown is set to 'AzureKeyVault2'. The 'Secret name' dropdown is set to 'Loading failed', which has a red error icon. The 'Secret version' dropdown is set to 'Latest version'. At the bottom of the dialog are 'Create' and 'Cancel' buttons.

So I directly used the Access token , without Key Vault ,which is not a preferred approach

This screenshot is identical to the one above, but the 'Access token' radio button in the 'Authentication type' section is now checked, while 'Azure Key Vault' is unchecked. All other settings remain the same.

### 3.3 Pipeline Run Successfully



### 3.4 Verifying for data in the gold container

The screenshot shows the Azure Storage Explorer interface. On the left, a sidebar shows 'Containers' with 'gold' selected. The main area displays the contents of the 'gold' container, including blobs named 'departments', 'dupes', 'employees', and 'regions'. There are also entries for '[].' and '...'. At the top, there are buttons for 'Upload', 'Add Directory', 'Refresh', 'Rename', 'Delete', 'Change tier', 'Acquire lease', 'Break lease', and 'Give feedback'. A search bar at the top right allows searching by prefix. A table below lists the blobs with columns for Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state.

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[.]					-	***
departments					-	***
dupes					-	***
employees					-	***
regions					-	***

## 4.Creating Views Using Synapse

Created a Synapse Workspace,a database **gold\_db** and used the linked feature to look at the storage containers which also link the tables

The screenshot shows the Microsoft Synapse Analytics workspace interface. On the left, the Data pane is open, showing the workspace structure. Under 'Azure Data Lake Storage Gen2', there is a 'gold' folder. Inside the 'gold' folder are four sub-folders: 'departments', 'dupes', 'employees', and 'regions'. A message bar at the top right states: 'Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.' Below the file tree, a message says 'Showing 1 to 4 of 4 cached items'.

Created a procedure for parameterising and creating the views for all the tables in the gold container

```

1 USE gold_db
2 GO
3
4 CREATE OR ALTER PROC CreateSQLView_GOLD1 @ViewName NVARCHAR(100)
5 AS
6 BEGIN
7
8 DECLARE @statement VARCHAR(MAX)
9
10 SET @statement = N'CREATE OR ALTER VIEW ' + @ViewName + ' AS
11     SELECT *
12     FROM
13         OPENROWSET(
14             BULK ''https://stgacdataengprj.dfs.core.windows.net/gold/OrganisationLT/' + @ViewName + '/',
15             FORMAT = ''DELTA''
16         ) AS [result]
17
18 EXECUTE(@statement)
19
20 END
21 GO
22

```

Now I'm creating a pipeline for automating this task , creating a new linked service for the SQL database(serverless) which has the stored procedure.

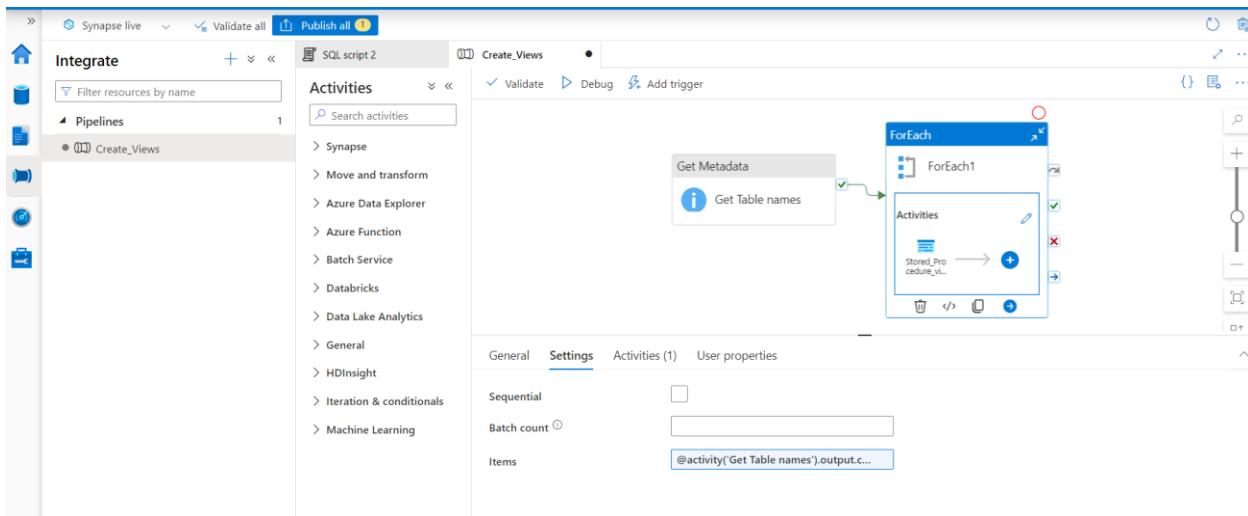
Entered Connection string details manually as we can't populate the database name .

Used the domain name which can be found under **syn-data-eng-proj | Properties| Serverless SQL endpoint**

Create a pipeline in the synapse studio and use the activity GET METADATA and create a dataset in binary format as we need only table names.

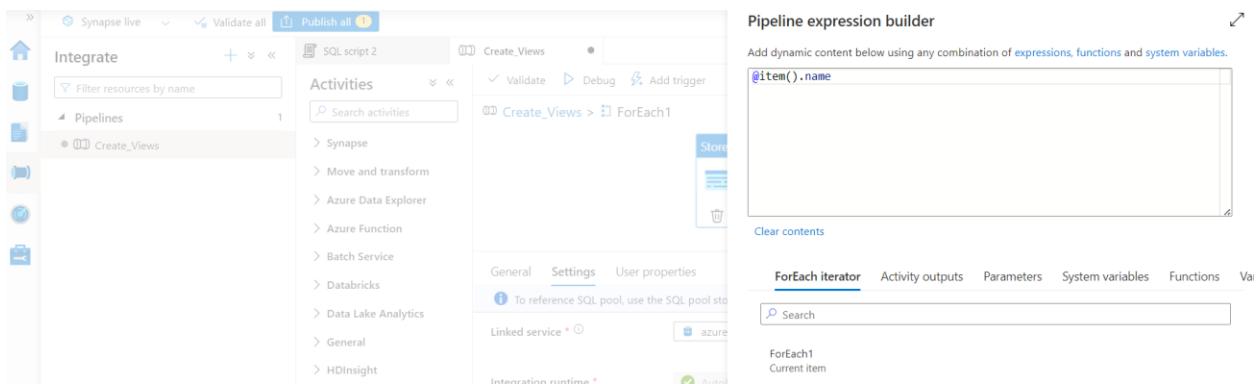
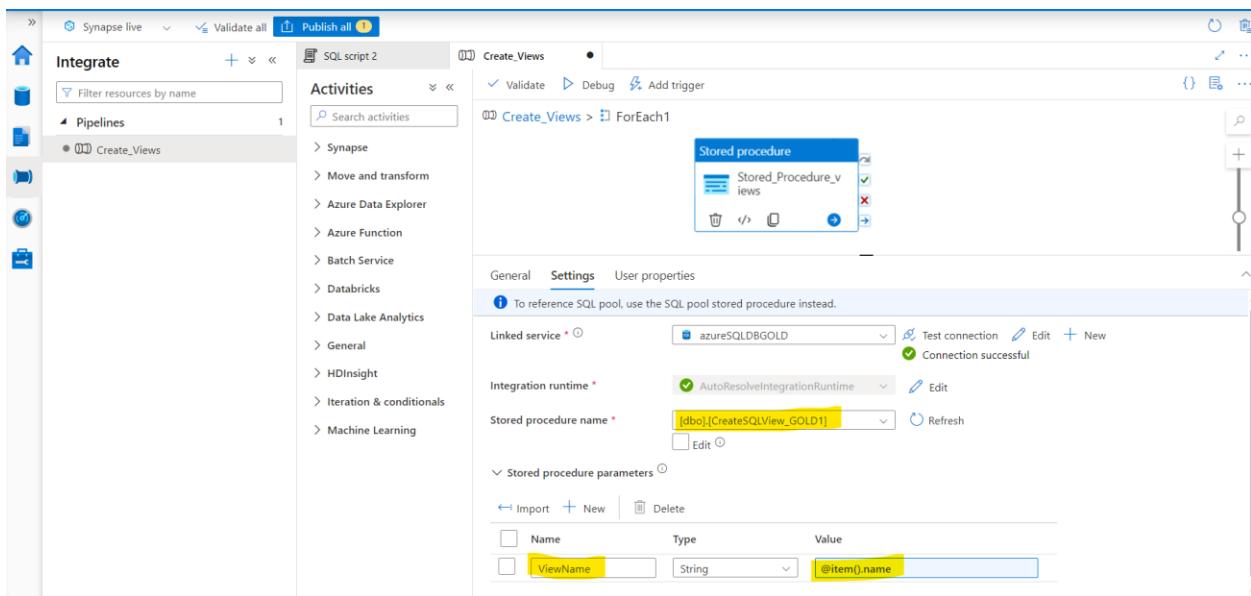
Pass the Field list argument as Child Items as we need table name .

Next, I'm using ForEach, passing the childnames to the Stored procedure I created earlier.

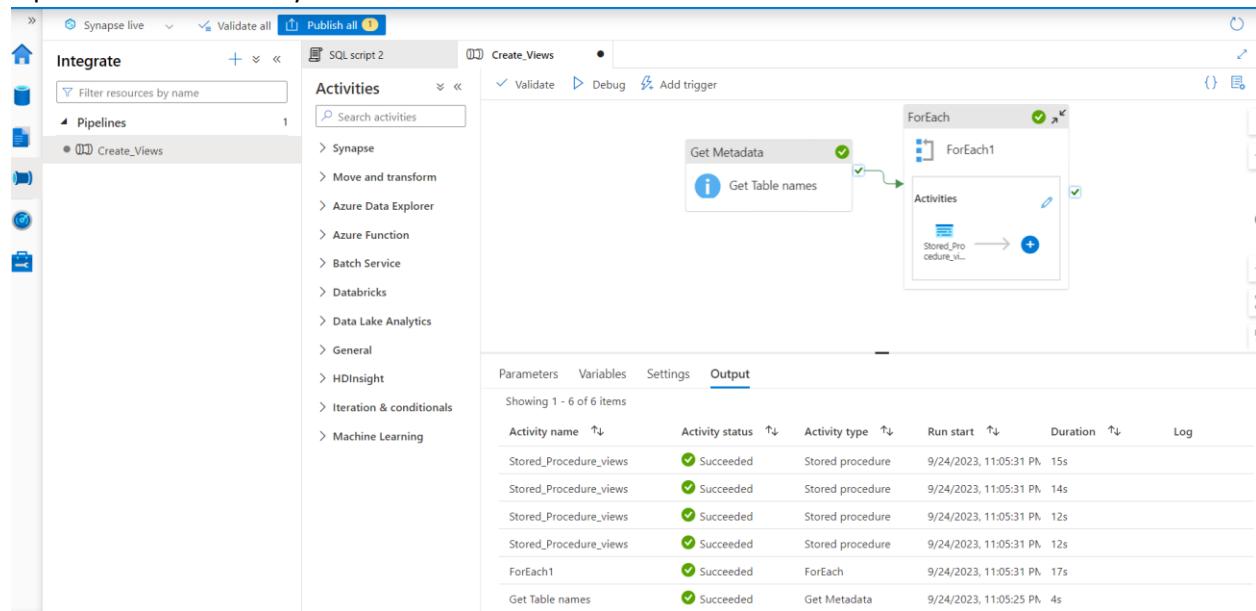


Expression to get table name from the GET METADATA activity

In the For each activity, I'm now calling a stored procedure activity and linking the stored procedure created in SQL Pool.



## Pipeline Ran successfully



## Validating the views in Serverless SQL POOL

The screenshot shows the Azure Synapse Studio interface. The left sidebar under 'Data' shows the 'gold\_db (SQL)' database, specifically the 'Views' section, which contains four views: 'dbo.departments', 'dbo.dupes', 'dbo.employees', and 'dbo.regions'. The right pane shows a query editor with the following SQL script:

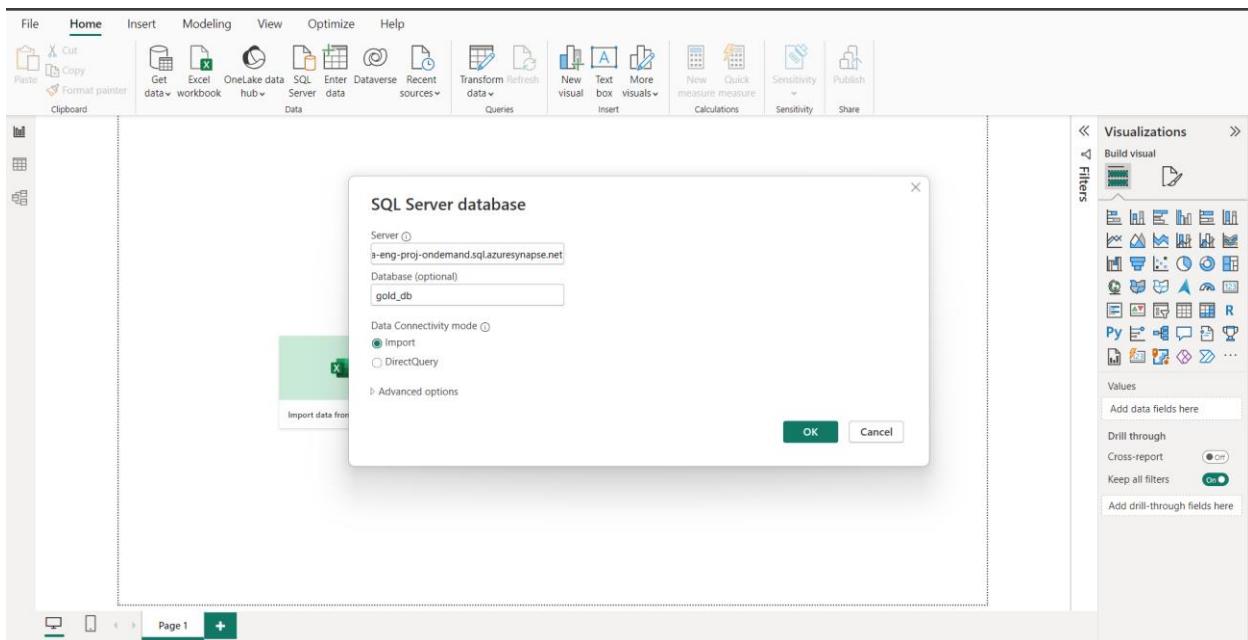
```
1 SELECT TOP (100) [employee_id]
2 ,[first_name]
3 ,[last_name]
4 ,[email]
5 ,[Hire_Date]
6 ,[department]
7 ,[gender]
8 ,[salary]
9 ,[region_id]
10 FROM [dbo].[employees]
```

The results pane displays the data from the 'dbo.employees' view, showing 6 rows of employee information.

employee_id	first_name	last_name	email	Hire_Date	department	gender	salary	region_id
1	Berrie	Manueau	bmaneau0@goo...	2006-04-20	Sports	F	154864	4
2	Aeriell	McNee	amcnee1@goo...	2009-01-26	Tools	F	56752	3
3	Sydney	Symonds	ssymonds2@h...	2010-05-17	Clothing	F	95313	4
4	Avrom	Rowantree	(NULL)	2014-08-02	Phones & Tablets	M	119674	7
5	Feliks	Morffew	fmorffew4@a8...	2003-01-14	Computers	M	55307	5
6	Bethena	Trow	btrow5@techn...	2003-06-08	Sports	F	134501	3

## 5. Publishing Data Analytics Reports – Power BI

Here I'm Importing data from Azure SQL Database. I've given the Serverless SQL server endpoint , which can be found **synapse-workspace | Properties| Serverless SQL endpoint**

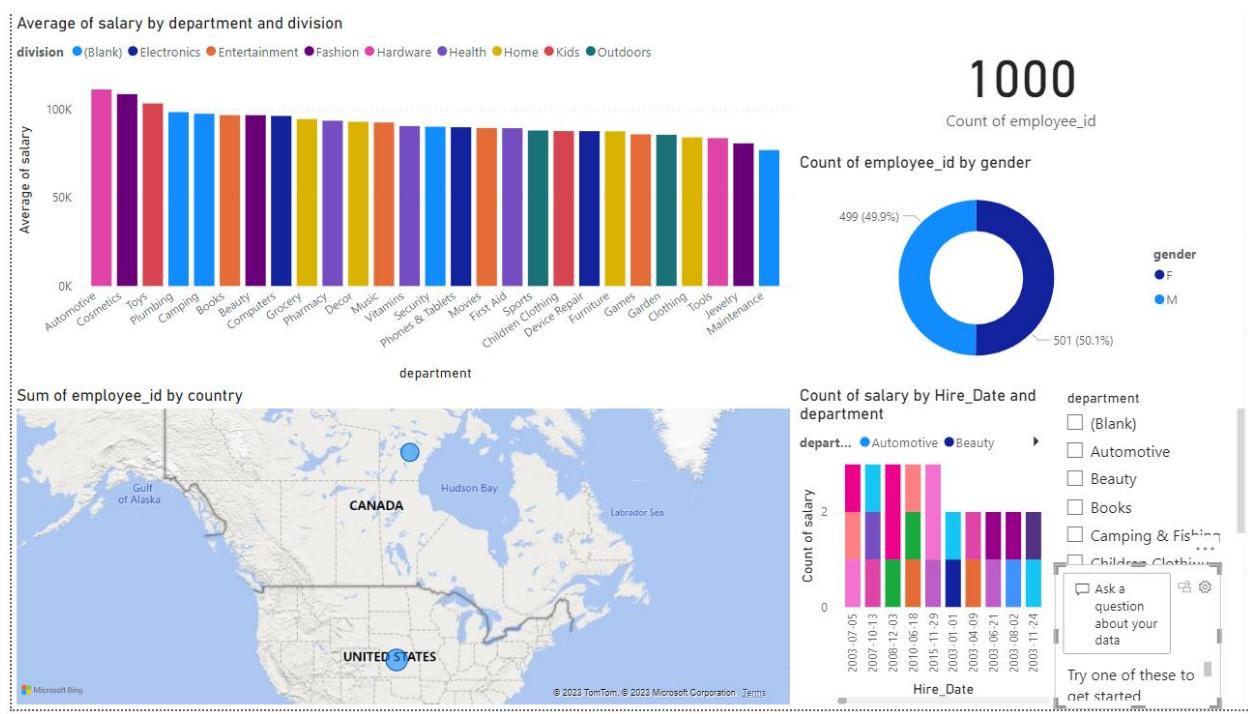


Signed-in with Microsoft account, now able to connect to synapse workspace and import data

employee_id	first_name	last_name	email
1	Berrie	Manueau	bmanueau0@dion.ne.jp
2	Aeriel	McNee	amcnee1@google.es
3	Sydney	Symonds	ssymonds2@hhs.gov
4	Avrom	Rowantree	
5	Feliks	Morffew	fmorffew4@a8.net
6	Bethena	Trow	btrow5@technorati.com
7	Ardeen	Curwood	acurwood6@lund1.de
8	Selline	Dubber	sdubber7@t-online.de
9	Dayle	Trail	dtrail8@tamu.edu
10	Redford	Roberti	
11	Nickey	Pointnta	npointnta@vistaprint.com
12	Leonora	Casaroli	lcasarolib@plala.or.jp
13	Anetta	Arnao	
14	Jodi	Hook	jhookd@booking.com
15	Alyson	Franzonello	
16	Merell	Yakovliv	myakovlivf@ucsd.edu
17	Annora	Bendelow	abendelowg@google.com.hk
18	Ronica	Armfield	
19	Bernardine	Hendricks	bhendricks@privacy.gov.au
20	Jessey	Colum	jcolumnj@pen.io
21	Bernardo	Davage	
22	Cayla	Duffer	
23	Vanda	Marwick	vmarwickm@upenn.edu

Relationship established automatically.

## Data Analysed and plotted as a report in Power BI



## 6. Security and Governance

Next I'll have to create a security group in Azure active directory/Microsoft Entra ID for easy access control in a real-time scenario.

Unfortunately, as mine is a student account I'm unable to access this page.

The screenshot shows a 'No access' error page. At the top, there's a 'Carleton University' header with a Microsoft Entra ID link. Below it is a large 'No access' icon (a cloud with a slash). The text 'No access' is centered below the icon. A summary card provides details: Session ID: 347b5ace82cc4197acb9c455de34ccaf; Resource ID: Not available; Extension: Microsoft\_AAD\_IAM; Content: ActiveDirectoryMenuBlade; Error code: 403.

## 7. End to End Pipeline Testing

This is a end to End Pipeline, meaning any update in the data in the database tables, should automatically be updated everywhere .

To do this we initiate a schedule trigger to run everyday at a specific time.

The screenshot shows the 'New trigger' configuration dialog in the Azure Data Factory interface. It includes fields for Start date (9/25/2023, 4:44:59 AM), Time zone (Eastern Time (US & Canada) (UTC-5)), Recurrence (Every 1 Day(s)), and Advanced recurrence options (Hours: 1, Minutes: 0). There are sections for Schedule execution times, Annotations, and Start trigger (with the 'Start trigger on creation' checkbox checked). Buttons for OK and Cancel are at the bottom.

Before this to see the updated the database with 2 records for testing end to end

Screenshot of pgAdmin showing two PostgreSQL connections and an Object Explorer.

**Object Explorer:**

- Extensions
- Foreign Data Wrappers
- Languages
- Publications
- Schemas (1)
  - public
  - Aggregates
  - Collations
  - Domains
  - FTS Configurations
  - FTS Dictionaries
  - FTS Parsers
  - FTS Templates
  - Foreign Tables
  - Functions
  - Materialized Views

**PostgreSQL 16 Connection 1:**

```

1 insert into employees values (1001, 'Srk', 'Ponnappalli', 'bmanueau@dion.ne.jp', '2023-04-20', 'Sports', 'M', 154864, 4);
2 insert into employees values (1002, 'Vamshi', 'Reddy', 'amcneel@google.es', '2023-01-26', 'Tools', 'M', 56752, 3);

```

**PostgreSQL 16 Connection 2:**

```

Data Output Messages Notifications
INSERT 0 1
Query returned successfully in 185 msec.

```

**PostgreSQL 16 Connection 3:**

```

-- insert into employees values (1001, 'Srk', 'Ponnappalli', 'bmanueau@dion.ne.jp', '2023-04-20', 'Sports', 'M', 154864,
-- insert into employees values (1002, 'Vamshi', 'Reddy', 'amcneel@google.es', '2023-01-26', 'Tools', 'M', 56752, 3);
3 Select * from Employees

```

employee_id	first_name	last_name	email	hire_date	department	gender
1	Berrie	Manueau	bmanueau@dion.ne.jp	2006-04-20	Sports	F
2	Aerill	McNee	amcneel@google.es	2009-01-26	Tools	F
3	Sydney	Symonds	ssymonds2@hhs.gov	2010-05-17	Clothing	F
4	Avrom	Rowantree	[null]	2014-08-02	Phones & Tablets	M
5	Feliks	Morfew	fmorphew4@a8.net	2003-01-14	Computers	M
6	Bethena	Trow	btrow5@technorati.com	2003-06-08	Sports	F
7	Ardene	Curwood	acurwood6@tund1.de	2006-02-19	Clothing	F
8	Seline	Dubber	sdupper7@t-online.de	2012-05-28	Phones & Tablets	F
9	Dayl	Trall	dtrall8@tamu.edu	2003-03-01	First Aid	F
10	Redford	Roberti	[null]	2008-07-21	Clothing	M
11	Nickey	Pointon	npointon@vistaprint.com	2006-12-30	Jewelry	M
12	Leonora	Casaroli	lcasaroli@plala.or.jp	2013-07-22	Beauty	F
13	Anetta	Armao	[null]	2009-05-23	Games	F
14	Jodi	Hook	jhookd@booking.com	2003-10-16	Tools	F
15	Alyson	Franzonello	[null]	2004-01-01	Furniture	F
16	Merell	Yakovliv	mvakovlivf@ucsd.edu	2008-08-16	Movies	M

Total rows: 1000 of 1002 Query complete 00:00:00.143 Ln 3, Col 24

The Trigger ran successfully, leading to both the pipelines in ADF in Synapse running successfully .

When the Power BI report is refreshed we get the following result.



## 8. References

1. [Azure Data Engineering Project – Mr K Talks](#)
2. [Access Azure Data Lake Storage using Azure Active Directory credential passthrough \(legacy\) - Azure Databricks | Microsoft Learn](#)
3. [Access Azure Data Lake Storage using Azure Active Directory credential passthrough \(legacy\) - Azure Databricks | Microsoft Learn](#)
4. [pyspark.sql.DataFrame.join — PySpark master documentation \(databricks.com\)](#)