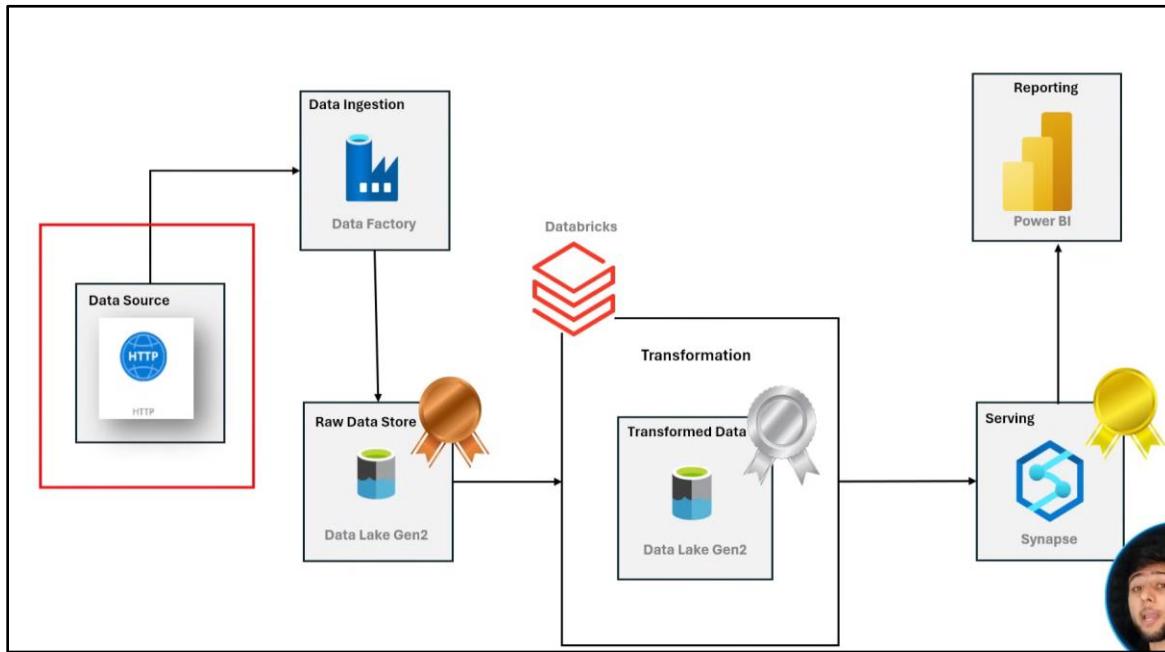


# Azure End-To-End Data Engineering Project (From Scratch!)

## Architecture:-



## Step 1 – Creating a Resource Group

## Step 2 – Creating Storage Account

### Create a storage account ...

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription \*

Azure subscription

Resource group \*

AzureProject1

[Create new](#)

#### Instance details

Storage account name \* ⓘ

azurestoragedatalake1

Region \* ⓘ

(US) East US

[Deploy to an Azure Extended Zone](#)

Primary service ⓘ

Azure Blob Storage or Azure Data Lake Storage Gen 2

Performance \* ⓘ

**Standard:** Recommended for most scenarios (general-purpose v2 account)

**Premium:** Recommended for scenarios that require low latency.

Redundancy \* ⓘ

Locally-redundant storage (LRS)

**Note:- This converts blob storage to data lake storage.**

**Hierarchical Namespace**

Hierarchical namespace, complemented by Data Lake Storage Gen2 endpoint, enables file and directory semantics, accelerates big data analytics workloads, and enables access control lists (ACLs) [Learn more](#)

Enable hierarchical namespace

## Step 3 – Creating Azure Data Factory

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

Home > Data factories >

### Create Data Factory ...

[Basics](#) [Git configuration](#) [Networking](#) [Advanced](#) [Tags](#) [Review + create](#)

One-click to create data factory with sample pipeline and datasets. [Try it](#)

**Project details**

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \*  Azure subscription 1

Resource group \*  AzureProject1 [Create new](#)

**Instance details**

Name \*  azureadoproject-rutuja

Region \*  East US

Version \*  V2

Home > **AzureProject1** [...](#)

Resource group

Search [Create](#) [Manage view](#) [Delete resource group](#) [Refresh](#) [Export to CSV](#) [Open query](#) [Assign tags](#) [Move](#) [Delete](#) [Export template](#) ... [JSON View](#)

**Overview** [Essentials](#)

**Resources** [Recommendations](#)

Filter for any field... Type equals all Location equals all Add filter

Showing 1 to 2 of 2 records.  Show hidden types

	Type ↑↓	Location ↑↓
<input type="checkbox"/> azstoragedatalake1rutu	Storage account	East US
<input type="checkbox"/> azureadoproject-rutuja	Data factory (V2)	East US

Home > AzureProject1 >  
**azureadfproject-rutuja** Data factory (V2)

Search Delete

**Overview**

Resource group (move) : [AzureProject1](#)  
Status : Succeeded  
Location : East US  
Subscription (move) : [Azure subscription 1](#)  
Subscription ID : d7f3756b-e1ae-4fd8-87b8-aa37971ee89d

Type : Data factory (V2)  
Getting started : [Quick start](#)

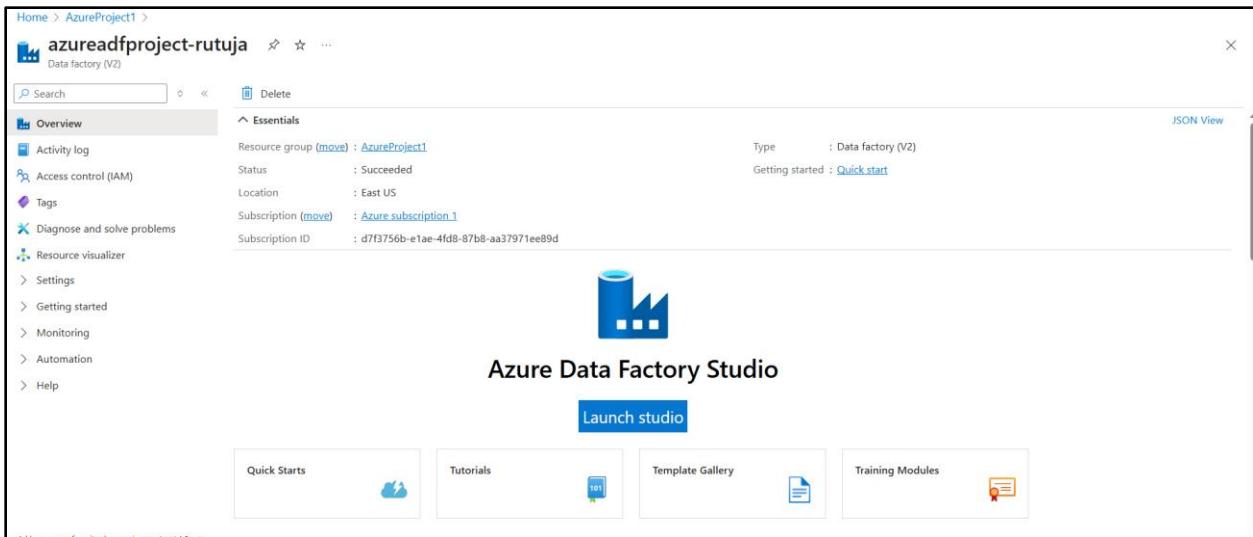
**Essentials**

**Azure Data Factory Studio**

**Launch studio**

**Quick Starts**   
**Tutorials**   
**Template Gallery**   
**Training Modules**

JSON View X



Microsoft Azure | Data Factory > [azureadfproject-rutuja](#) Search factory and documentation rutuja.lamkane.personal@gmail.com DEFAULT DIRECTORY X

Would you like to see Data Factory inside of Microsoft Fabric, Microsoft's newest cloud-first data analytics SaaS platform? Click [here](#) to get started with Fabric Data Factory!

» **Azure Data Factory** allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration. Learn more Set up code repository X

**Data factory**  
**azureadfproject-rutuja**

New ▾

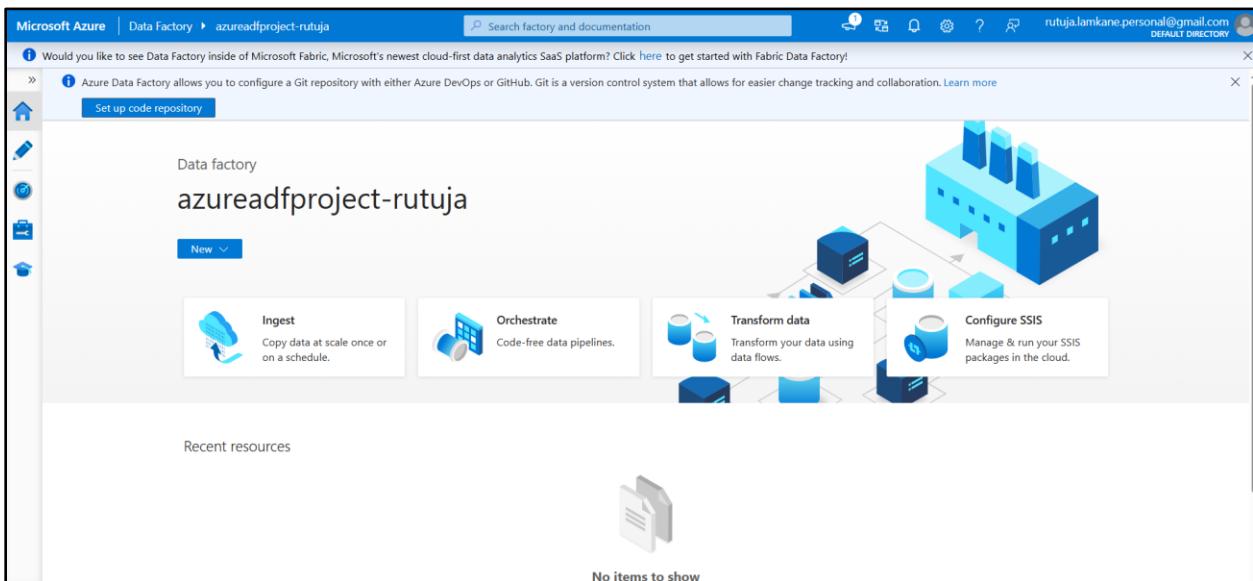
**Ingest** Copy data at scale once or on a schedule.

**Orchestrate** Code-free data pipelines.

**Transform data** Transform your data using data flows.

**Configure SSIS** Manage & run your SSIS packages in the cloud.

Recent resources   
No items to show



## Step 5 – Creating Containers

The screenshot shows the Azure Storage account properties for 'azstoragedatalake1rutu'. The 'Properties' tab is selected, displaying various storage settings. Under 'Data Lake Storage', it shows 'Hierarchical namespace' as Enabled, 'Default access tier' as Hot, and 'Blob anonymous access' as Disabled. Under 'Security', 'Require secure transfer for REST API operations' is Enabled. Under 'Networking', 'Allow access from' is set to 'All networks'. Other tabs like Monitoring, Capabilities, Recommendations, Tutorials, and Tools + SDKs are also visible.

Bronze(raw), silver(transformation), gold(serving) layer in storage account

The screenshot shows the 'Containers' page for the 'azstoragedatalake1rutu' storage account. It lists four containers: 'logs', 'bronze', 'gold', and 'silver'. All containers are private and available. The 'logs' container was modified on 10/07/2025, 20:09:37. The 'bronze', 'gold', and 'silver' containers were modified on 10/07/2025, 20:25:52, 20:26:21, and 20:26:05 respectively. The 'Container' section of the sidebar is highlighted.

Name	Last modified	Anonymous access level	Lease state
logs	10/07/2025, 20:09:37	Private	Available
bronze	10/07/2025, 20:25:52	Private	Available
gold	10/07/2025, 20:26:21	Private	Available
silver	10/07/2025, 20:26:05	Private	Available

## Step 6- Creating Copy Activity

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar is open, showing 'Pipelines' with one item named 'GitToRaw'. The main workspace displays an 'Activities' list with various options like 'Move and transform', 'Synapse', and 'Copy data'. A 'Copy data' activity is selected, shown in a preview window with the name 'CopyRawData'. Below the preview, the 'General' tab of the activity configuration pane is active, showing fields for 'Name' (set to 'CopyRawData'), 'Description', 'Activity state' (set to 'Activated'), and 'Timeout' (set to '0:12:00').

## Step 7 – Creating Linked service (here it is http) – dataset stored in GitHub in csv format

The screenshot shows the Microsoft Azure Data Factory 'Linked services' page. The left sidebar lists various settings and connections, with 'Linked services' currently selected. A 'New' button is visible on the right. The main area shows a 'No linked services' message and a 'Create linked service' button. On the right, a 'New linked service' configuration pane is open for an 'HTTP' type linked service. It includes fields for 'Name' (set to 'httplinkedservice'), 'Description', 'Connect via integration runtime' (set to 'AutoResolveIntegrationRuntime'), 'Base URL' (set to 'https://raw.githubusercontent.com/'), 'Server certificate validation' (set to 'Enable'), 'Authentication type' (set to 'Anonymous'), and 'Auth headers'. At the bottom, a 'Create' button is highlighted, and a success message 'Connection successful' is displayed.

The screenshot shows the Microsoft Azure Data Factory interface for the project 'azureadlproject-rutuja'. The left sidebar navigation includes General, Connections (selected), Source control, Author, and Security. The main content area is titled 'Linked services' and displays a table with one item: 'httplinkedservice' of type 'HTTP'. A search bar at the top right says 'Search factory and documentation'.

## Step 8 – Creating another link service (here it is data lake storage) – destination (raw/bronze layer)

The screenshot shows the 'New linked service' dialog in the Microsoft Azure Data Factory interface. The 'Name' field is set to 'storagedatalake'. The 'Connect via integration runtime' dropdown is set to 'AutoResolveIntegrationRuntime'. The 'Authentication type' dropdown is set to 'Account key'. Under 'Account selection method', the 'From Azure subscription' radio button is selected. The 'Storage account name' dropdown is set to 'azstoragedatalake1rutu'. At the bottom, there are 'Create', 'Back', 'Test connection', and 'Cancel' buttons.

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has 'Data Factory' selected under 'General'. The main area is titled 'Linked services' with the sub-instruction: 'Linked service defines the connection information to a data store or compute. Learn more'. It includes a search bar 'Annotations : Any' and a table listing two items:

Name	Type	Related	Annotations
httplinkedservice	HTTP	0	
storagedatalake	Azure Data Lake Storage Gen2	0	

## Step 9 - Adding source

The screenshot shows the Microsoft Azure Data Factory interface with 'Pipelines' selected in the left sidebar. A pipeline named 'GitToRaw' is selected. On the right, a 'Set properties' dialog is open for the 'Source' tab of an activity. The dialog fields include:

- Name: ds\_http
- Linked service: httplinkedservice
- Relative URL: e-Works-Data-Engineering-Project/refs/heads/main/Data/AdventureWorks\_Products.csv
- First row as header: checked
- Import schema: From connection/store (radio button selected)

The bottom of the dialog shows 'OK' and 'Back' buttons.

## Step 10 - Adding sink(destination)

**Set properties**

Name: ds\_raw

Linked service: storagedatalake

File path: bronze/products/products.csv

First row as header:

Import schema:  From connection/store  From sample file  None

Advanced

Schema import failed: ADLS Gen2 operation failed for: Operation returned error 'Not found'. Account: 'awstoragedatalake'. FileSystem: 'bronze'. Path: 'b011-00b-597z-1eb363000000'. ErrorCode: 'PathNotFound'. Message: 'The specified path does not exist.' TimeStamp: Mon, 14 Oct 2024 19:40:00 UTC. Status code: '404'. Status message: 'The specified path does not exist.' Activity ID: '30e7f457949d560734'.

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Datasets' (2), and 'Data flows' (0). The main workspace displays a 'Copy data' activity named 'CopyRawData'. The 'Sink' tab is selected, showing the 'Sink dataset' dropdown set to 'ds\_raw'. Other settings include 'Copy behavior' (Select...), 'Max concurrent connections' (1), 'Block size (MB)' (1), and 'Metadata' (New). The top navigation bar includes 'Validate all', 'Publish all', and 'Preview experience' toggle.

## Step 11 - Run the pipeline ( here it is static pipeline)

The screenshot shows the Microsoft Azure Data Factory pipeline status page. The pipeline 'CopyRawData' is listed under 'Output' with a status of 'Succeeded'. The table below details the run information:

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime	User prop...
CopyRawData	Succeeded	Copy data	7/11/2025, 8:38:54 PM	14s	AutoResolveIntegrationRuntime (East US)	

Successful

The screenshot shows the Microsoft Azure Storage Explorer interface. The left sidebar displays the navigation path: Home > AzureProject1 > azstoragedatalake1rtrutu | Containers > bronze. The main area shows a list of blobs in the 'bronze' container. There is one item listed: 'products.csv'. The details for this file are as follows:

Name	Last modified	Access tier	Blob type	Size	Lease state
products.csv	11/07/2025, 20:39:05	Hot (Inferred)	Block blob	56.76 KiB	Available

## Step 12 -- Dynamic Pipeline

Creating three parameters for copy activity

One for relative path

One for folder

One for file name

## Step 13 – Adding source for dynamic copy activity

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. In the center, a pipeline named 'DynamicGitToRaw' is displayed. A 'Copy data' activity is selected in the 'Activities' list. The right panel shows the 'Set properties' configuration for this activity. The 'Source' tab is selected, and the 'Source dataset' dropdown is set to 'ds\_git\_dynamic'. Other visible settings include 'Linked service': 'httplinkedservice', 'Relative URL': (empty), 'First row as header': checked, 'Import schema': 'None', and 'Advanced' section with a note about opening the dataset for parameterization.

## Step 14 – Creating a parameter for relative path in the source

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 'GitToRaw' is selected. A dataset named 'ds\_git\_dynamic' is shown under the 'Datasets' section. The 'Connection' tab is active, displaying connection details: Linked service is 'httplinkedservice', Base URL is 'https://raw.githubusercontent.com', Relative URL is empty, and other parameters like Compression type, Column delimiter, Row delimiter, and Encoding are set. A preview video in the bottom right corner shows a person speaking.

This screenshot is similar to the previous one, but the 'Parameters' tab is now active in the Pipeline expression builder. The 'Functions' tab is also visible. A search bar is present at the top of the builder area. The preview video continues to play in the bottom right corner.

Microsoft Azure | Data Factory > adf-aw-project-ansh

Search factory and documentation

Validate all Publish all

Factory Resources

- Pipelines
- GitToRaw
- DynamicGitToRaw
- ds.git\_dynamic

Datasets

- ds.git\_dynamic
- ds.http
- ds.raw

Data flows

Power Query

DelimitedText ds.git.dynamic

New parameter

Name: p\_rel\_url

Type: String

Default value: I

Connection Schema Parameters

Linked service: httplinkedservice

Base URL: https://raw.githubusercontent.com

Relative URL: Add dynamic content [Alt+Shift+D]

Compression type: Select...

Column delimiter: Comma (,)

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

Encoding: Default(UTF-8)

Quote character: Double quote (")

Save Cancel

16 / 4:39:16 • Real-Time Scenarios with ADF >

Microsoft Azure | Data Factory > adf-aw-project-ansh

Search factory and documentation

Validate all Publish all

Factory Resources

- Pipelines
- GitToRaw
- DynamicGitToRaw
- ds.git\_dynamic

Datasets

- ds.git\_dynamic
- ds.http
- ds.raw

Data flows

Power Query

DelimitedText ds.git.dynamic

Pipeline expression builder

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

@dataset().p\_rel\_url

Clear contents

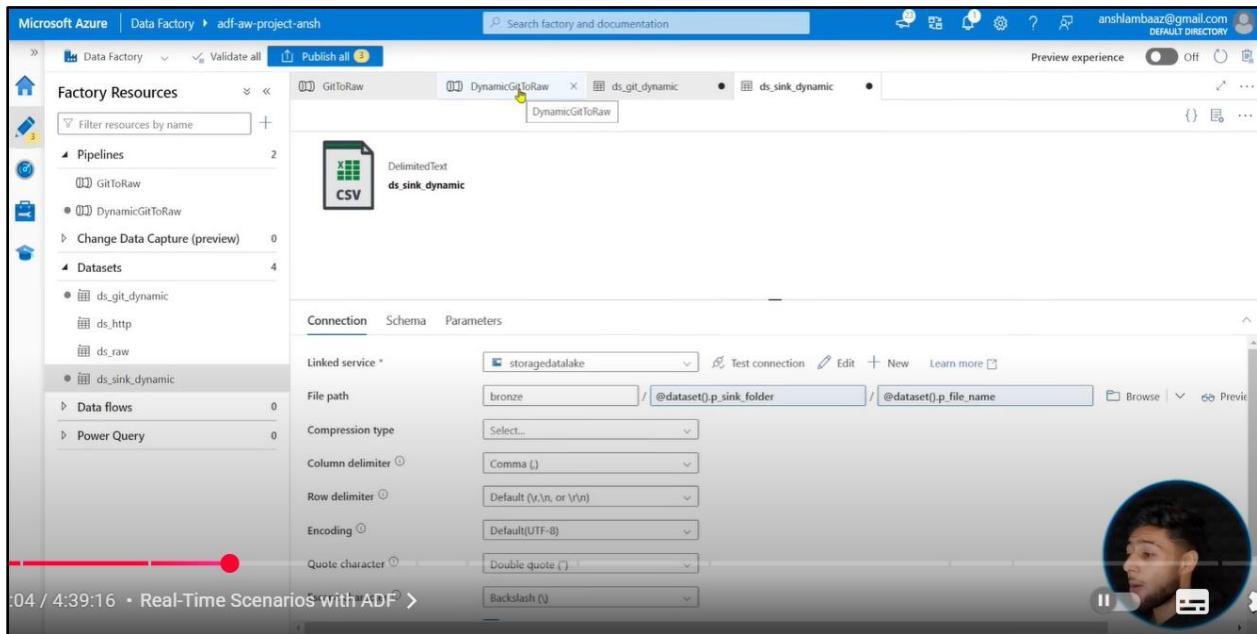
Parameters Functions

p\_rel\_url Pipeline parameter

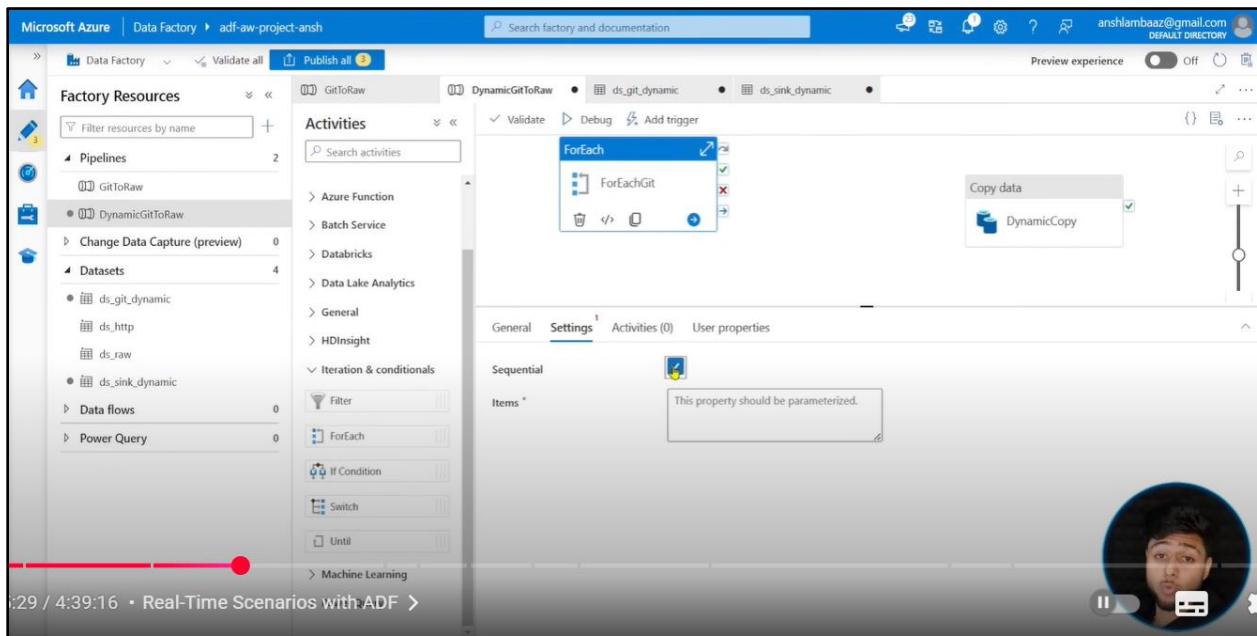
OK Cancel

142 / 4:39:16 • Real-Time Scenarios with ADF >

## Step 15 – Similarly create two more parameters for sink

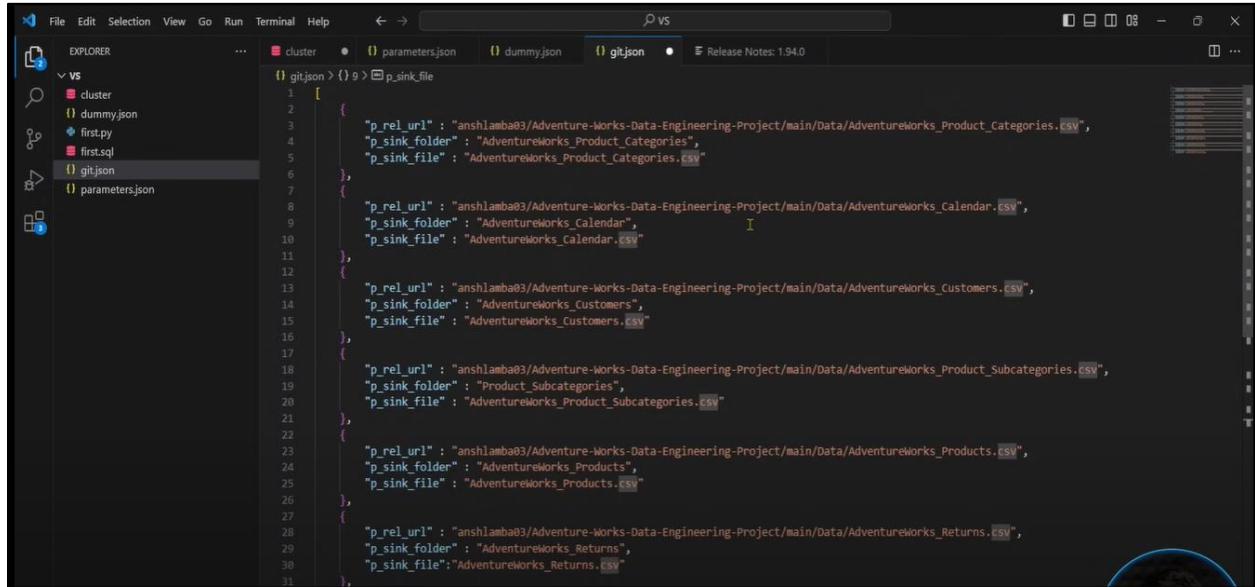


## Step 16 – Creating ForEach Activity



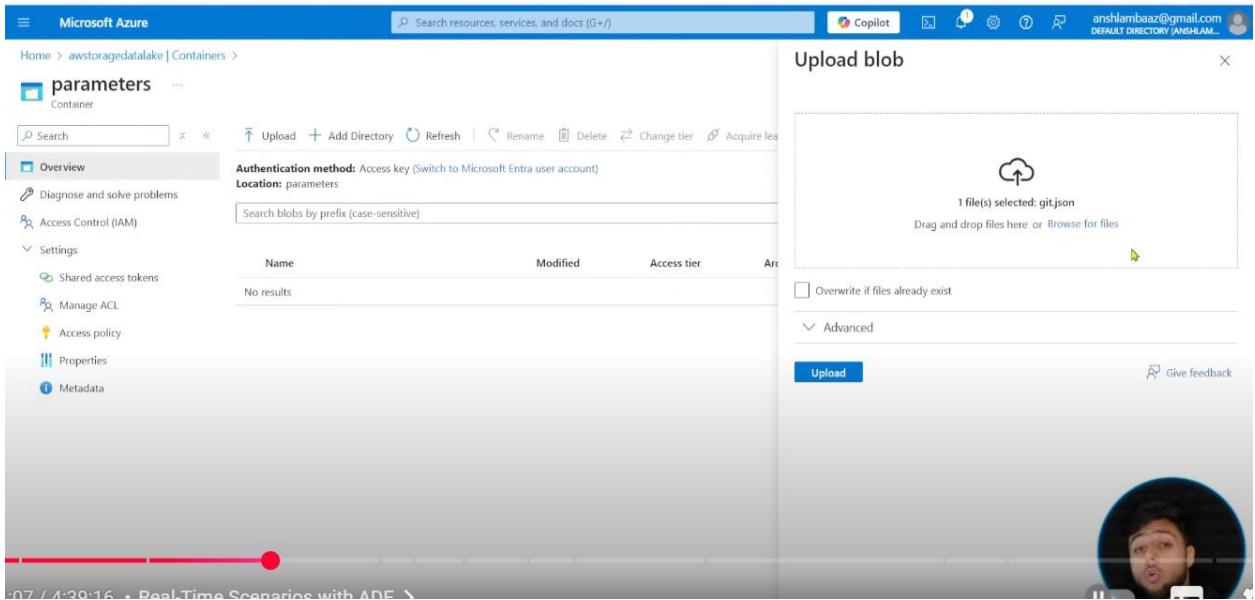
Note – Turn on sequential

## Step 17 – Creating Json file for three parameters values



```
[{"p_rel_url": "anshlamba03/Adventure-Works-Data-Engineering-Project/main/Data/Adventureworks_Product_Categories.csv", "p_sink_folder": "Adventureworks_Product_Categories", "p_sink_file": "Adventureworks_Product_Categories.csv"}, {"p_rel_url": "anshlamba03/Adventure-Works-Data-Engineering-Project/main/Data/Adventureworks_Calendar.csv", "p_sink_folder": "Adventureworks_Calendar", "p_sink_file": "Adventureworks_Calendar.csv"}, {"p_rel_url": "anshlamba03/Adventure-Works-Data-Engineering-Project/main/Data/Adventureworks_Customers.csv", "p_sink_folder": "Adventureworks_Customers", "p_sink_file": "Adventureworks_Customers.csv"}, {"p_rel_url": "anshlamba03/Adventure-Works-Data-Engineering-Project/main/Data/Adventureworks_Product_Subcategories.csv", "p_sink_folder": "Product Subcategories", "p_sink_file": "Adventureworks_Product_Subcategories.csv"}, {"p_rel_url": "anshlamba03/Adventure-Works-Data-Engineering-Project/main/Data/Adventureworks_Products.csv", "p_sink_folder": "Adventureworks_Products", "p_sink_file": "Adventureworks_Products.csv"}, {"p_rel_url": "anshlamba03/Adventure-Works-Data-Engineering-Project/main/Data/Adventureworks_Returns.csv", "p_sink_folder": "Adventureworks_Returns", "p_sink_file": "Adventureworks_Returns.csv"}]
```

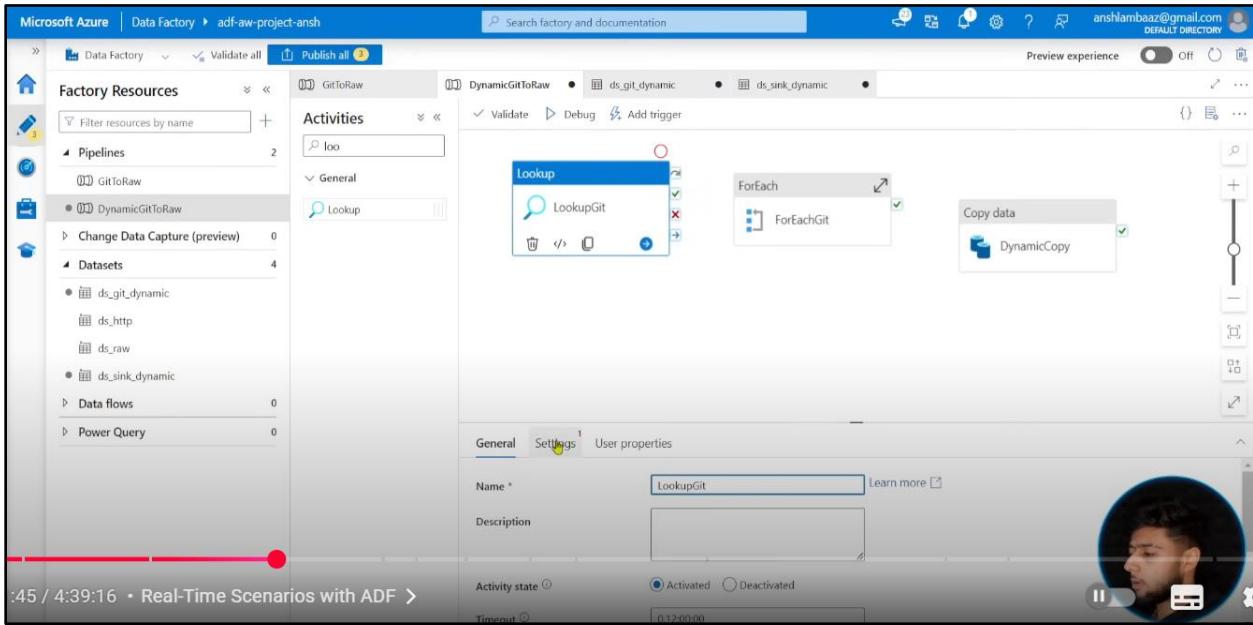
## Step 18 – Adding the Json file to storage account container (create a new container “parameters” and then upload Json file there)



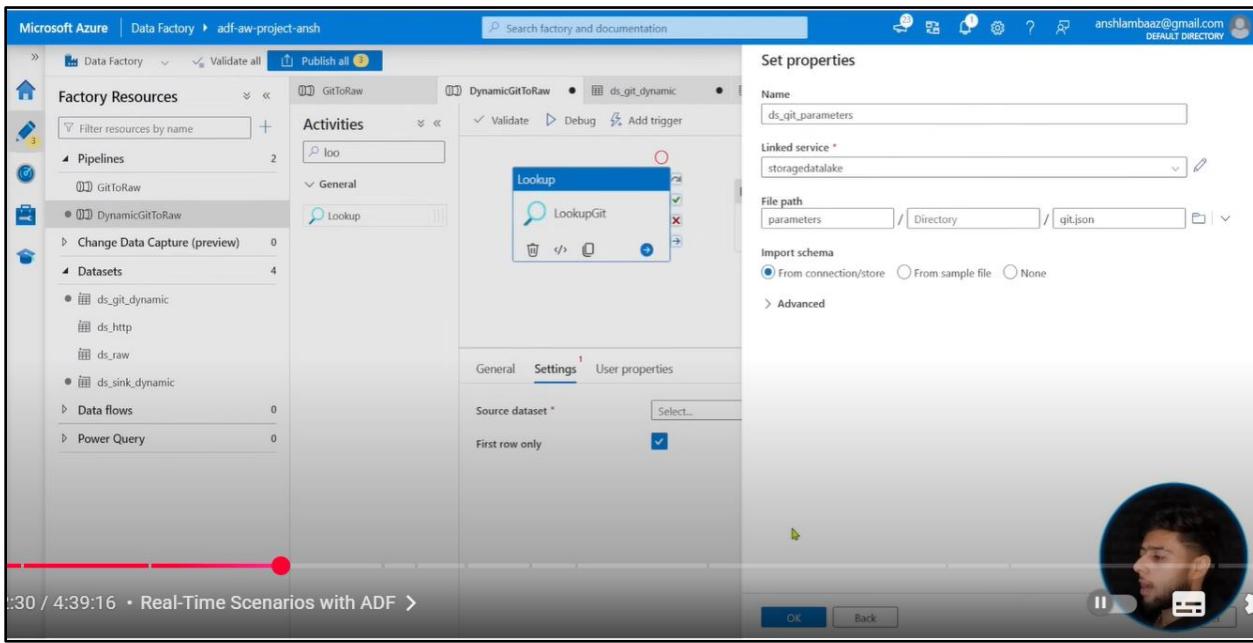
The screenshot shows the Microsoft Azure Storage Explorer interface. On the left, the 'parameters' container is selected under 'Containers'. The 'Overview' tab is active, showing the container's properties: 'Authentication method: Access key (Switch to Microsoft Entra user account)' and 'Location: parameters'. A search bar at the top says 'Search resources, services, and docs (G +/)'.

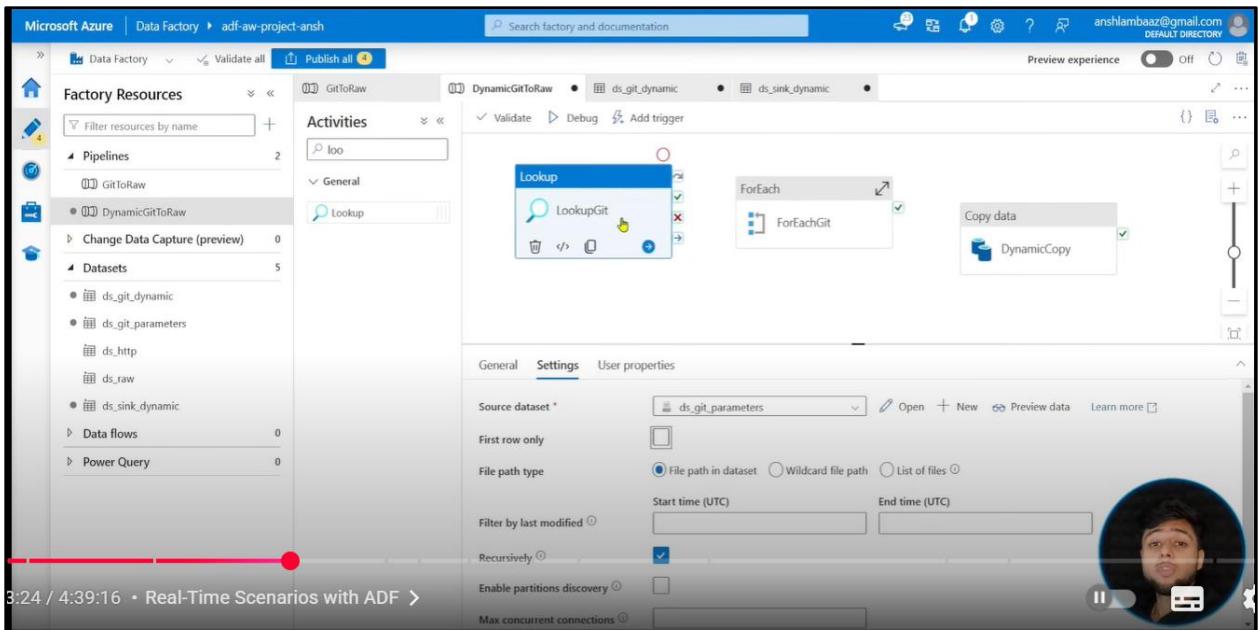
On the right, an 'Upload blob' dialog is open. It displays a cloud icon and the message '1 file(s) selected: git.json'. Below it, there's a note 'Drag and drop files here or browse for files'. A checkbox 'Overwrite if files already exist' is present. At the bottom, there's an 'Upload' button and a 'Give feedback' link.

## Step 19 – Creating Lookup activity



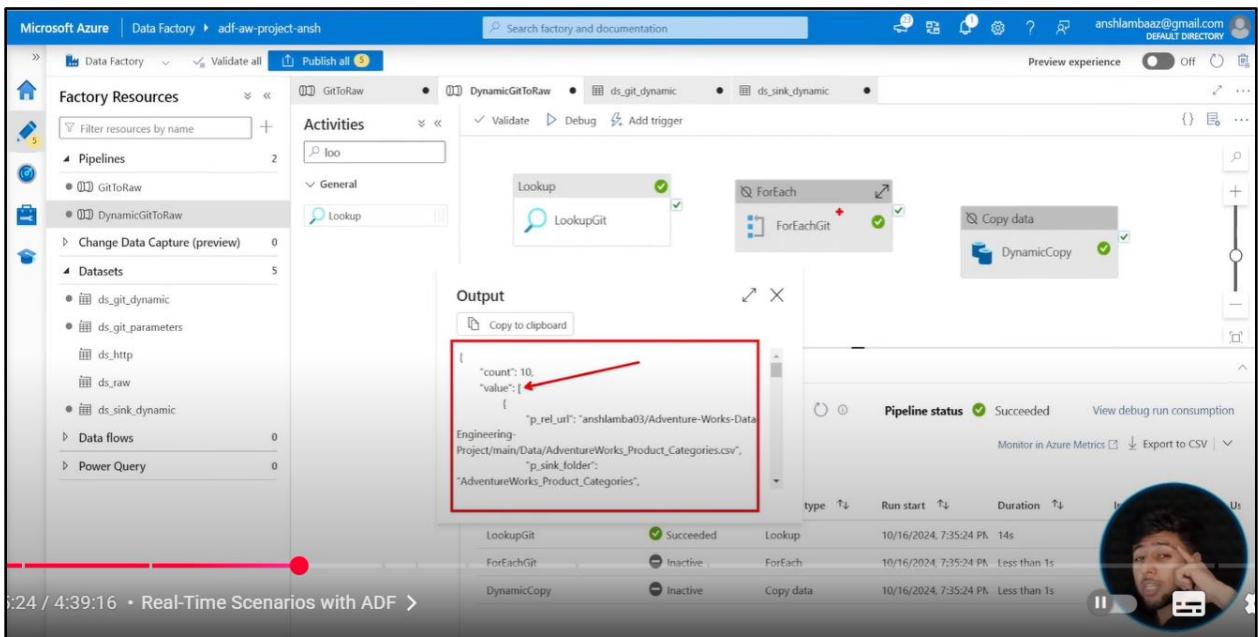
## Step 20 – Adding Json file to lookup activity



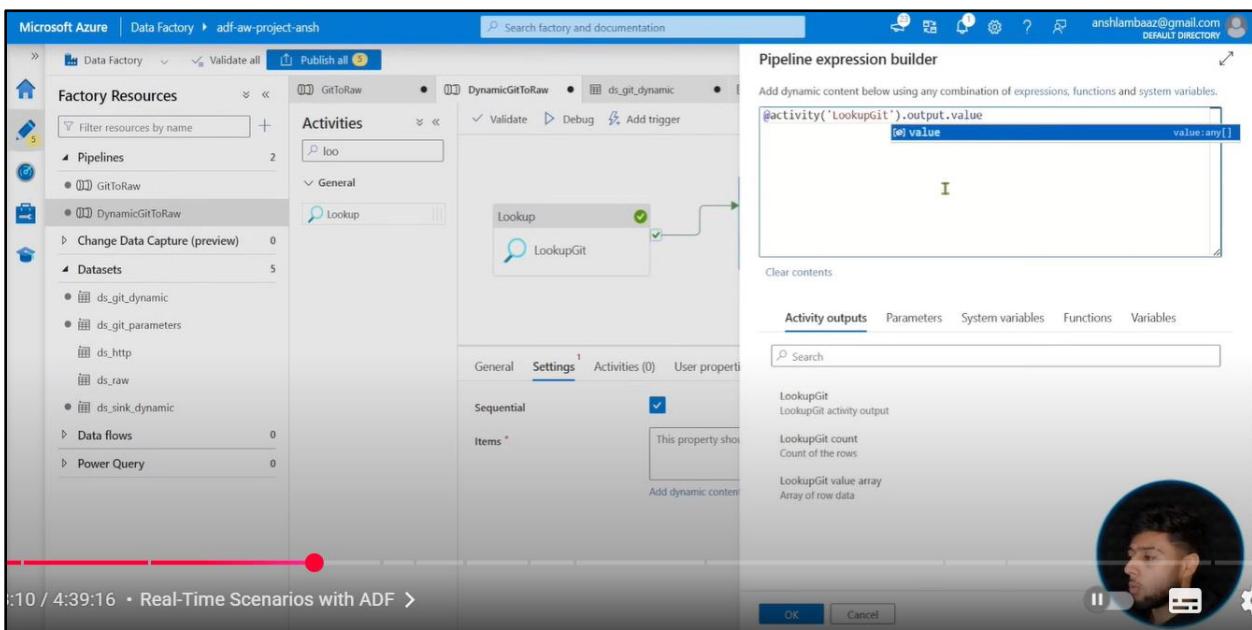
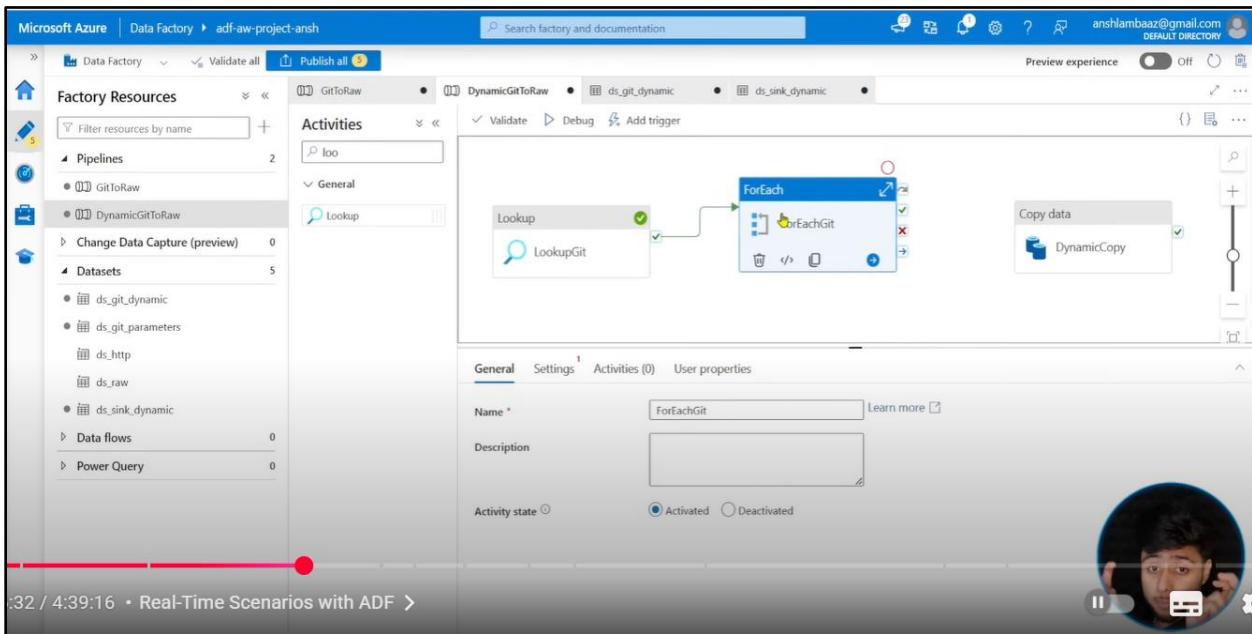


Note – Uncheck First row only box as we want to iterate all the values

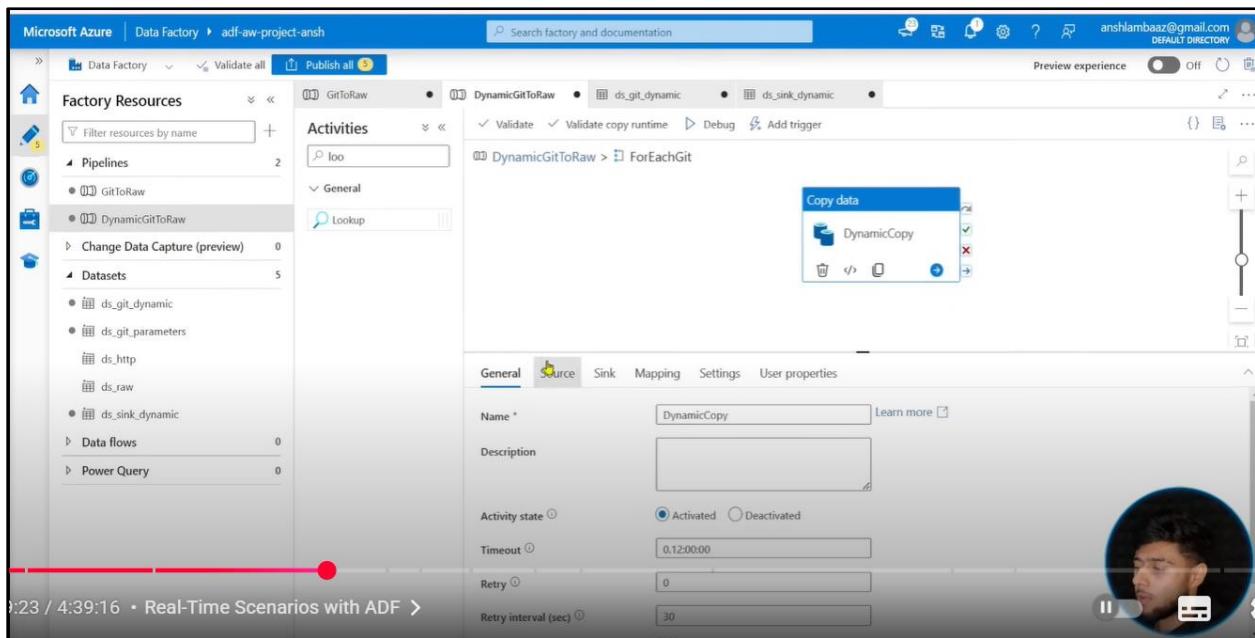
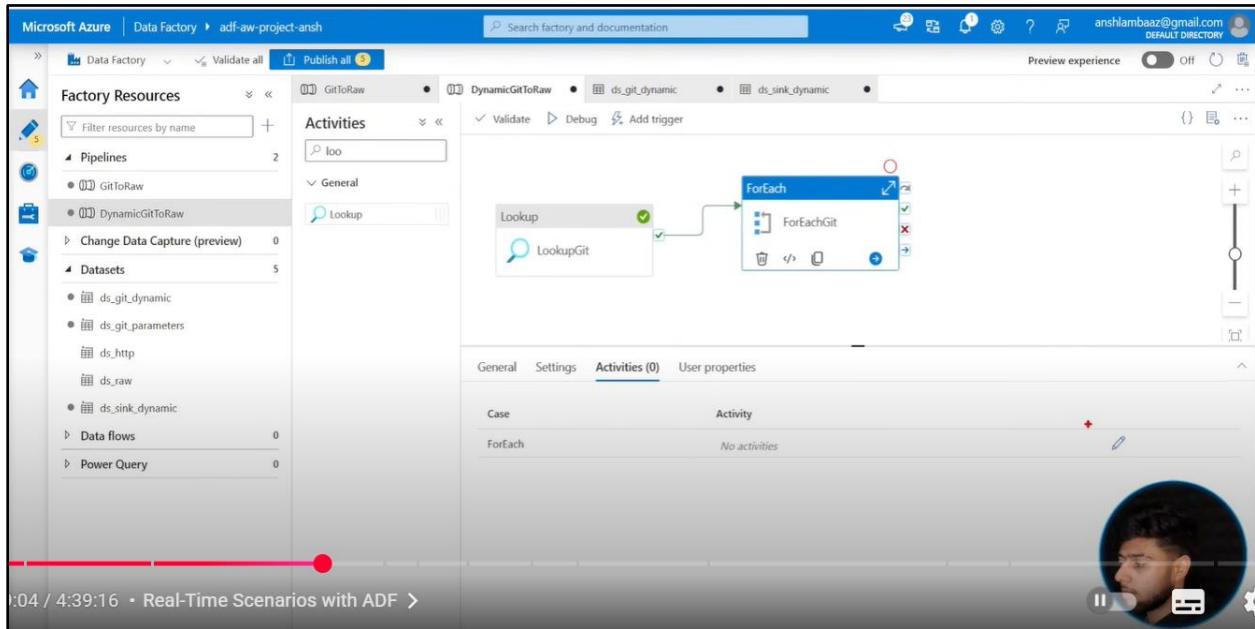
## Step 21 – Check if Lookup activity is working properly



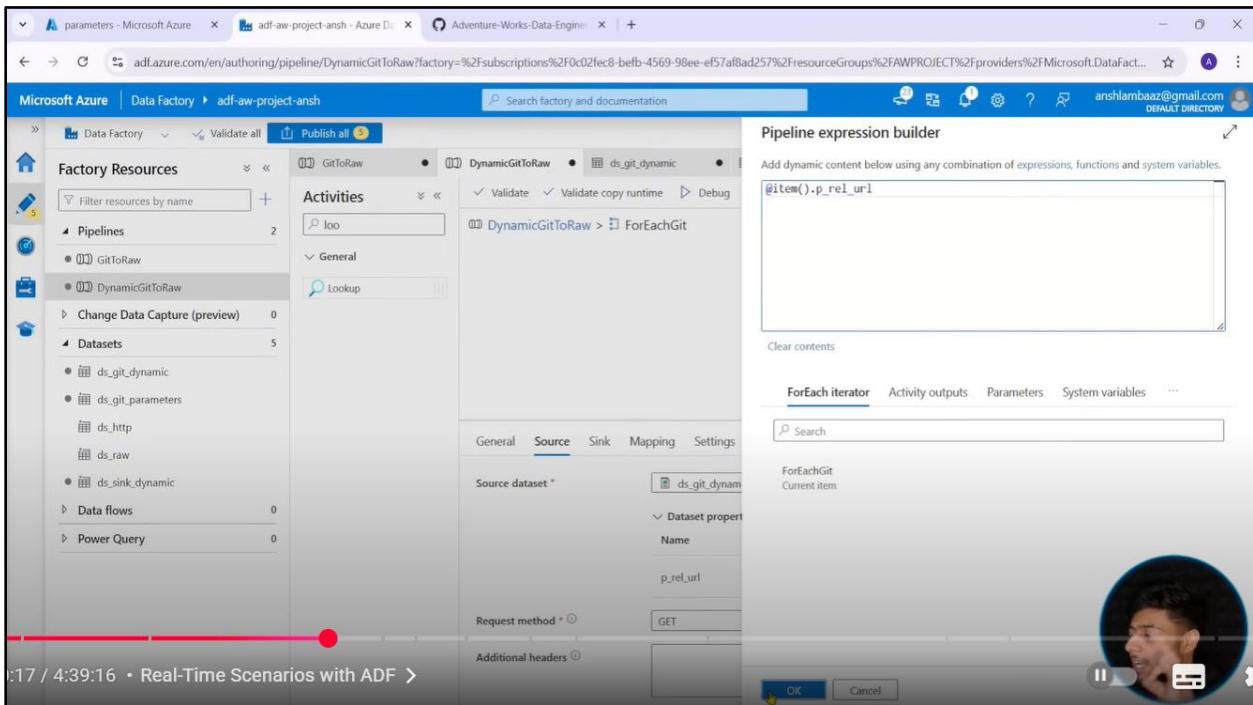
## Step 22 – Connect lookup activity with foreach activity



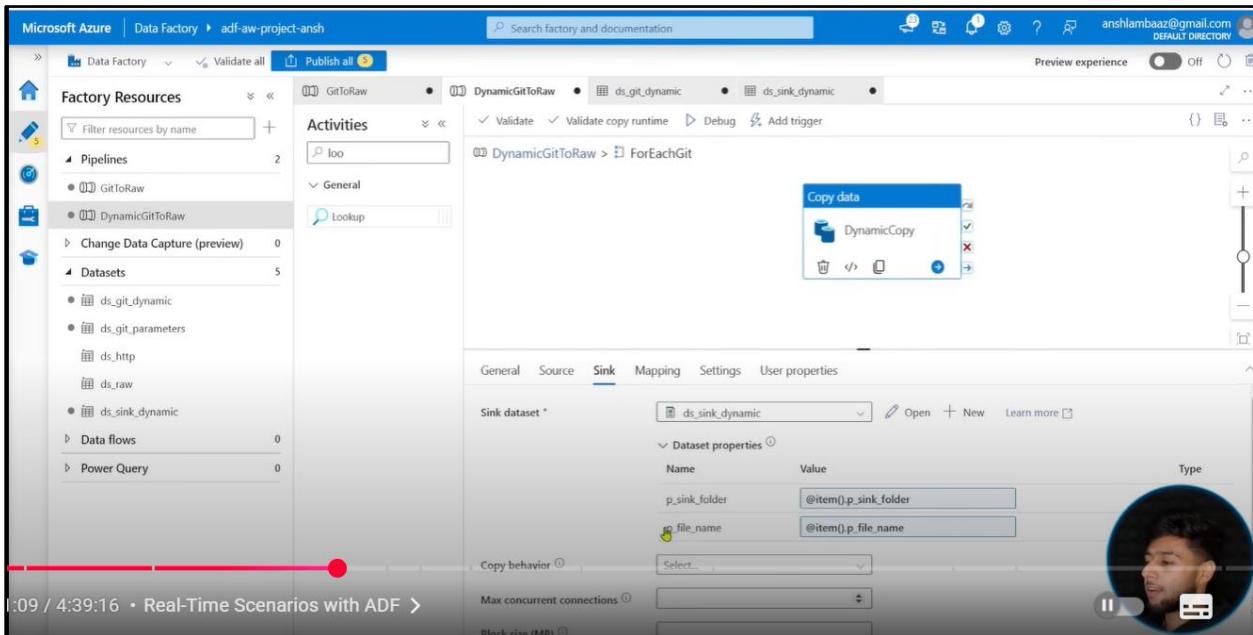
## Step 23 – Cut the copy activity and paste in activity tab of foreach activity



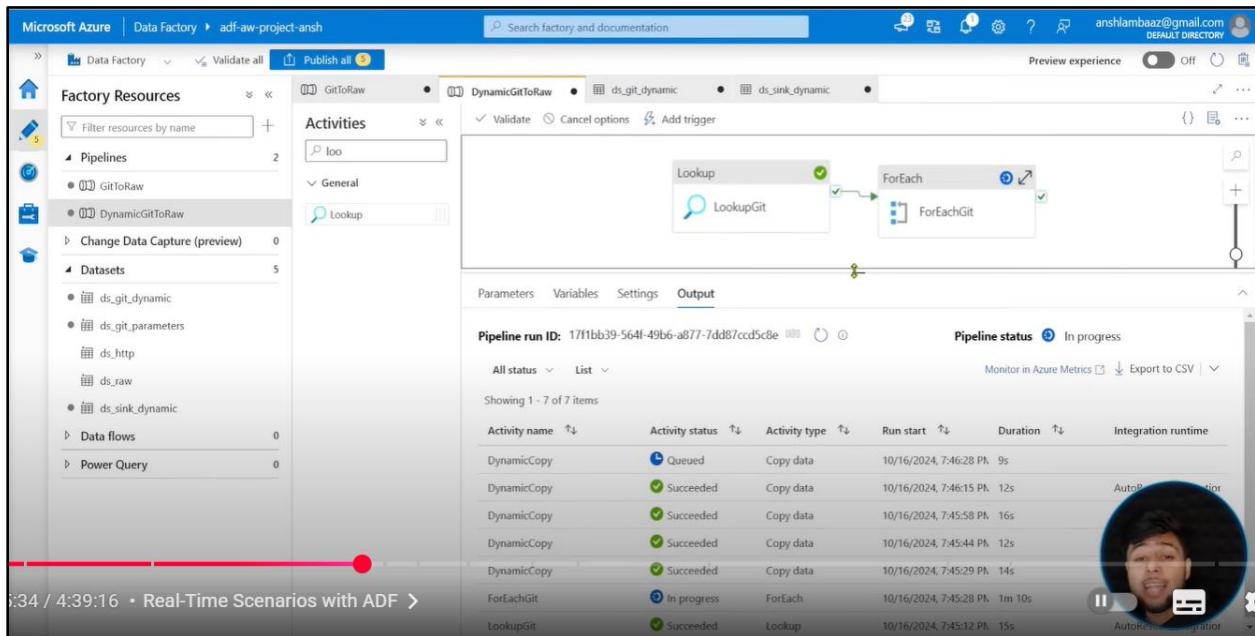
## Step 24 – Add source value



## Step 25 – Add sink value

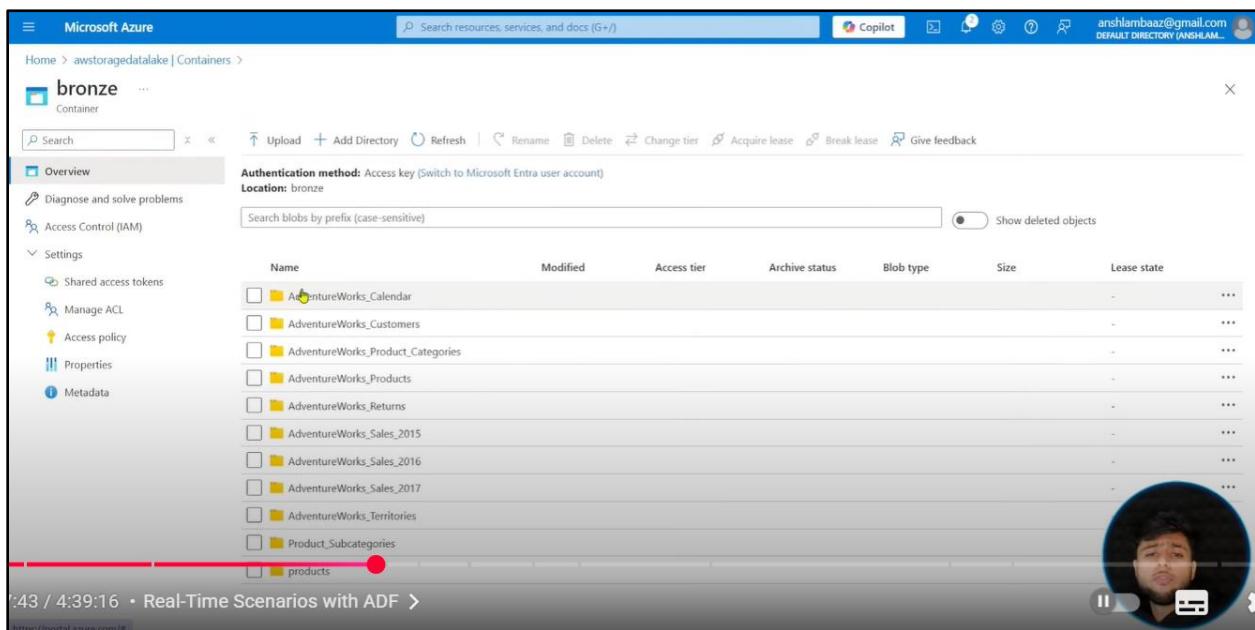


## Step 26 – Run the pipeline



The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. The main workspace displays a pipeline named 'DynamicGitToRaw'. The pipeline consists of a 'Lookup' activity followed by two 'ForEach' activities. The 'Output' tab shows the pipeline run ID: 17f1bb39-564f-49b6-a877-7dd87cc5c8e. The 'Pipeline status' is 'In progress'. Below the status, a table lists the activity details:

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime
DynamicCopy	Queued	Copy data	10/16/2024, 7:46:28 PM	9s	AutoReserve
DynamicCopy	Succeeded	Copy data	10/16/2024, 7:46:15 PM	12s	AutoReserve
DynamicCopy	Succeeded	Copy data	10/16/2024, 7:45:58 PM	16s	AutoReserve
DynamicCopy	Succeeded	Copy data	10/16/2024, 7:45:44 PM	12s	AutoReserve
DynamicCopy	Succeeded	Copy data	10/16/2024, 7:45:29 PM	14s	AutoReserve
ForEachGit	In progress	ForEach	10/16/2024, 7:45:28 PM	1m 10s	AutoReserve
LookupGit	Succeeded	Lookup	10/16/2024, 7:45:12 PM	15s	AutoReserve



The screenshot shows the Microsoft Azure Storage Container overview for 'bronze'. The left sidebar includes options like Overview, Diagnose and solve problems, Access Control (IAM), Settings, Shared access tokens, Manage ACL, Access policy, Properties, and Metadata. The main area displays a list of blobs with their names, modified dates, access tiers, archive statuses, blob types, sizes, and lease states. A search bar at the top allows filtering by prefix.

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
AdventureWorks_Calendar					-	...
AdventureWorks_Customers					-	...
AdventureWorks_Product_Categories					-	...
AdventureWorks_Products					-	...
AdventureWorks_Returns					-	...
AdventureWorks_Sales_2015					-	...
AdventureWorks_Sales_2016					-	...
AdventureWorks_Sales_2017					-	...
AdventureWorks_Territories					-	...
Product_Subcategories					-	...
products					-	...

## Step 27 – Creating azure databricks

The screenshot shows the 'Create an Azure Databricks workspace' wizard. The 'Basics' tab is selected. In the 'Project Details' section, the 'Subscription' is set to 'Azure subscription 1' and the 'Resource group' is 'AzureProject1'. In the 'Instance Details' section, the 'Workspace name' is 'adb-azproject1-rutu', 'Region' is 'East US', 'Pricing Tier' is 'Trial (Premium - 14-Days Free DBUs)', and the 'Managed Resource Group name' is 'managed-adb-azproject1'. At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next : Networking >'.

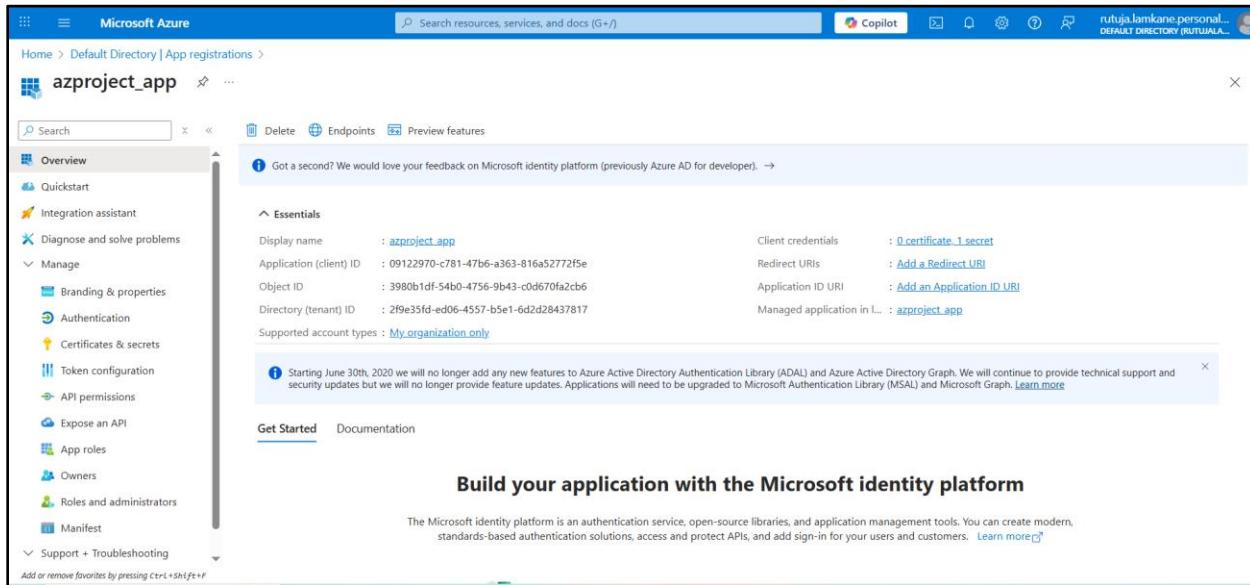
## Step 28 – Creating databricks cluster

The screenshot shows the 'Rutuja Lamkane's Cluster' configuration page. The 'Compute' tab is selected. Under 'Configuration', the 'Policy' is 'Unrestricted', 'Access mode' is 'No isolation shared', and the 'Databricks Runtime Version' is '15.4 LTS (includes Apache Spark 3.5.0, Scala 2.12)'. The 'Node type' is 'Standard\_D4ds\_v5' with '16 GB Memory, 4 Cores'. A checkbox for 'Terminate after 15 minutes of inactivity' is checked. On the right, a 'Summary' panel shows 1 Driver, 16 GB Memory, 4 Cores, Runtime 15.4.x-scala2.12, and Standard\_D4ds\_v5 1 DBU/h.

## Step 29 -- Got to Microsoft Entra ID

### App registration

#### Copy the app id and tenant id



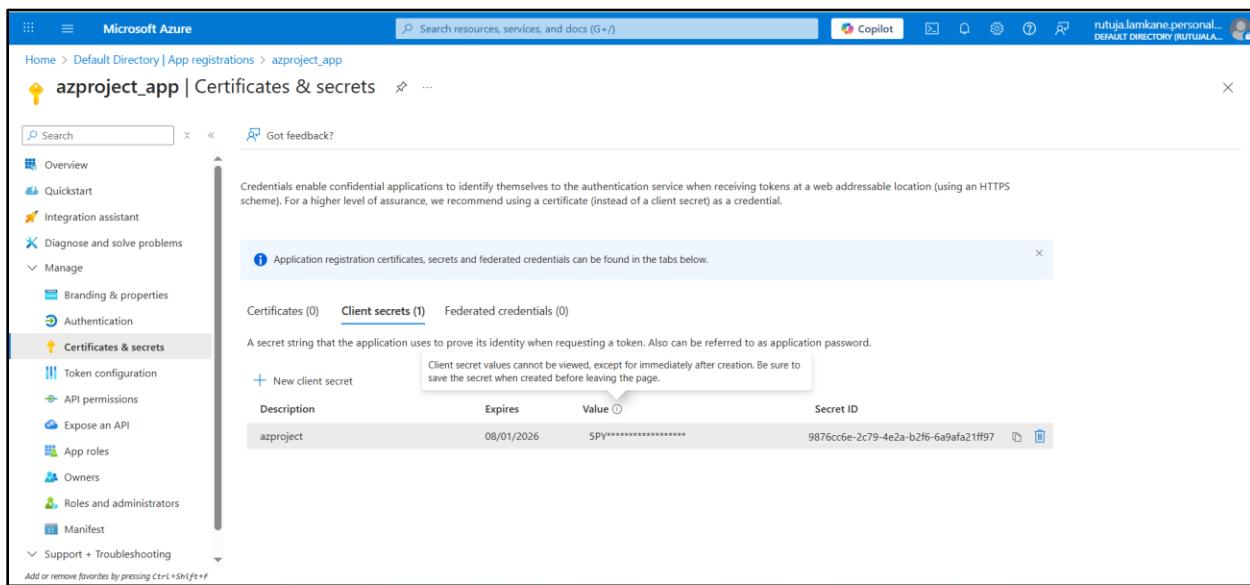
The screenshot shows the Microsoft Azure portal with the URL [https://portal.azure.com/#blade/Microsoft\\_AAD\\_IAM/AppRegistrationsBlade/Overview/~/resourceId/09122970-c781-47b6-a363-816a52772f5e](#). The page title is "azproject\_app". The left sidebar shows the navigation menu for app registration, with "Certificates & secrets" selected. The main content area displays the app's details under the "Essentials" section:

Display name	: azproject_app	Client credentials	: 0_certificate_1_secret
Application (client) ID	: 09122970-c781-47b6-a363-816a52772f5e	Redirect URIs	: Add a Redirect URI
Object ID	: 3980b1df-54b0-4756-9b43-c0d670fa2cb6	Application ID URI	: Add an Application ID URI
Directory (tenant) ID	: 2f9e35fd-ed06-4557-b5e1-6d2d28437817	Managed application in ...	: azproject_app
Supported account types	: My organization only		

A note at the bottom states: "Starting June 30th, 2020 we will no longer add any new features to Azure Active Directory Authentication Library (ADAL) and Azure Active Directory Graph. We will continue to provide technical support and security updates but we will no longer provide feature updates. Applications will need to be upgraded to Microsoft Authentication Library (MSAL) and Microsoft Graph. [Learn more](#)".

## Step 30 -- Then go to certificates and secret

### Copy the secret value



The screenshot shows the Microsoft Azure portal with the URL [https://portal.azure.com/#blade/Microsoft\\_AAD\\_IAM/AppRegistrationsBlade/CertificatesAndSecretsBlade/~/resourceId/09122970-c781-47b6-a363-816a52772f5e](#). The page title is "azproject\_app | Certificates & secrets". The left sidebar shows the navigation menu for app registration, with "Certificates & secrets" selected. The main content area displays the "Client secrets" tab:

Credentials enable confidential applications to identify themselves to the authentication service when receiving tokens at a web addressable location (using an HTTPS scheme). For a higher level of assurance, we recommend using a certificate (instead of a client secret) as a credential.

Application registration certificates, secrets and federated credentials can be found in the tabs below.

Description	Expires	Value	Secret ID
azproject	08/01/2026	5PY*****	9876cc6e-2c79-4e2a-b2f6-6a9afa21ff97

A tooltip for the "Value" column states: "Client secret values cannot be viewed, except for immediately after creation. Be sure to save the secret when created before leaving the page."

## Step 31 -- The go to storage account

### Give the role to it – access control

The screenshot shows the Microsoft Azure Access Control (IAM) page for the storage account 'azstoragedatalake1runtu'. The left sidebar includes options like Overview, Activity log, Tags, Diagnose and solve problems, and Access Control (IAM). The main content area has tabs for Check access, Role assignments, Roles, Deny assignments, and Classic administrators. Under 'My access', there's a 'View my access' button. Under 'Check access', there's a 'Check access' button. Below these are three cards: 'Grant access to this resource' (with 'Add role assignment' button), 'View access to this resource' (with 'View' button), and 'View deny assignments' (with 'View' button).

The screenshot shows the 'Add role assignment' dialog. It has tabs for Role, Members\*, Conditions, and Review + assign. The Role tab is selected, showing a search bar for 'storage blob' and filters for Type: All and Category: All. A table lists six built-in roles under 'Job function roles': Defender CSPM Storage Data Scanner, Defender for Storage Data Scanner, Storage Blob Data Contributor, Storage Blob Data Owner, Storage Blob Data Reader, and Storage Blob Delegator. Each row includes columns for Name, Description, Type, Category, and Details. At the bottom, there are buttons for 'Review + assign', 'Previous', and 'Next'.

Name	Description	Type	Category	Details
Defender CSPM Storage Data Scanner	Grants access to read blobs and files. This role is used by the data scanner of Dfender CSM.	BuiltinRole	None	View
Defender for Storage Data Scanner	Grants access to read blobs and update index tags. This role is used by the data scanner of Defender for Storage.	BuiltinRole	None	View
Storage Blob Data Contributor	Allows for read, write and delete access to Azure Storage blob containers and data	BuiltinRole	Storage	View
Storage Blob Data Owner	Allows for full access to Azure Storage blob containers and data, including assigning POSIX access control.	BuiltinRole	Storage	View
Storage Blob Data Reader	Allows for read access to Azure Storage blob containers and data	BuiltinRole	Storage	View
Storage Blob Delegator	Allows for generation of a user delegation key which can be used to sign SAS tokens	BuiltinRole	Storage	View

Microsoft Azure

Home > AzureProject1 > azstoragedatalake1ruru | Access Control (IAM) >

### Add role assignment

Role Members\* Conditions Review + assign

Selected role Storage Blob Data Contributor

Assign access to  User, group, or service principal  Managed identity

Members + Select members

Name	Object ID	Type
No members selected		

Description Optional

Review + assign Previous Next

## Step 32 – Select the app that you created as a member

Microsoft Azure

Home > AzureProject1 > azstoragedatalake1ruru | Access Control (IAM) >

### Add role assignment

Role Members\* Conditions Review + assign

Selected role Storage Blob Data Contributor

Assign access to  User, group, or service principal  Managed identity

Members + Select members

Name	Object ID	Type
No members selected		

Description Optional

Review + assign Previous Next

Select members

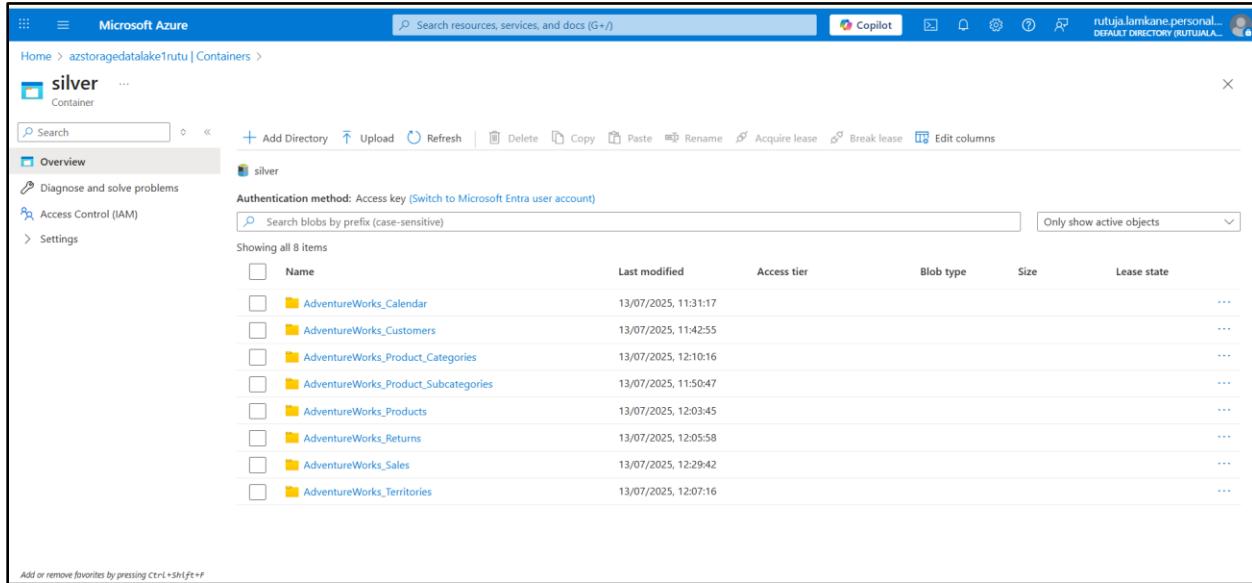
- azproject\_app Application

Selected members:

- azproject\_app Application

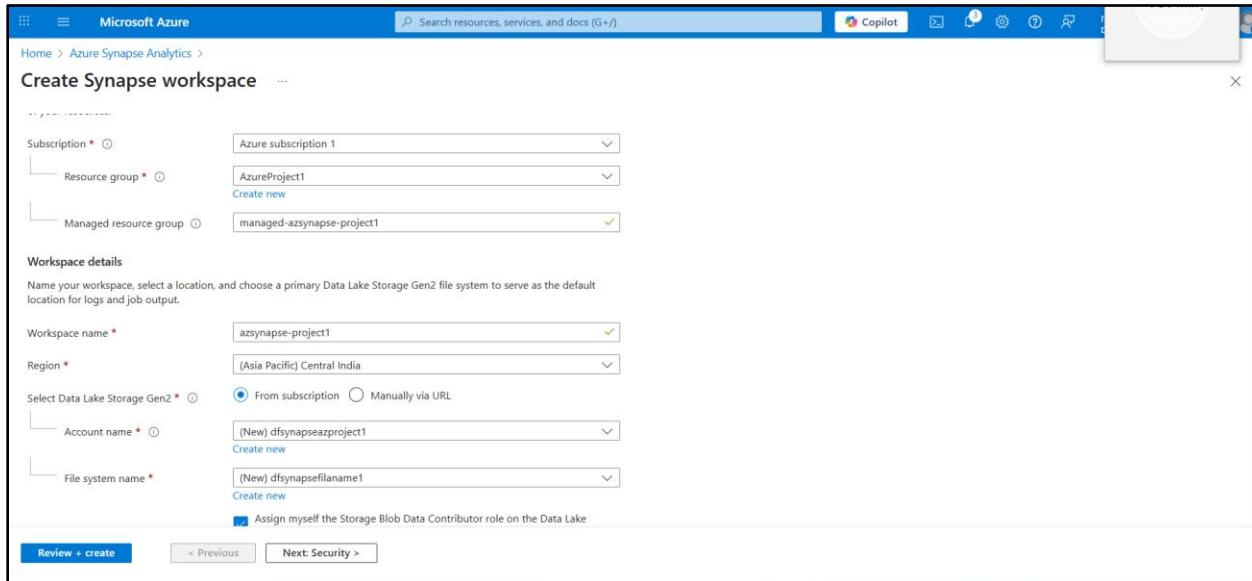
Select Close

**Step 33 – Access is provided. Now the databricks can access the data from bronze layer. Perform some transformations in databricks notebook and load the data to silver layer.**



The screenshot shows the Microsoft Azure Storage Explorer interface. The left sidebar shows a tree structure with 'silver' selected under 'Containers'. The main area is titled 'Overview' and displays a list of blobs. The table includes columns for Name, Last modified, Access tier, Blob type, Size, and Lease state. The blobs listed are: AdventureWorks\_Calendar, AdventureWorks\_Customers, AdventureWorks\_Product\_Categories, AdventureWorks\_Product\_Subcategories, AdventureWorks\_Products, AdventureWorks\_Returns, AdventureWorks\_Sales, and AdventureWorks\_Territories. All blobs were last modified on 13/07/2025 at different times between 11:31:17 and 12:07:16.

## Step 34 -- Creating Azure Synapse



The screenshot shows the 'Create Synapse workspace' wizard in the Microsoft Azure portal. The first step, 'Set workspace details', is displayed. It requires selecting a subscription (Azure subscription 1), resource group (AzureProject1), and managed resource group (managed-azsynapse-project1). The 'Workspace details' section allows naming the workspace (azsynapse-project1), choosing a region (Asia Pacific) Central India, and selecting a Data Lake Storage Gen2 account (New dfsynapseazproject1) and file system (New dfsynapsefilename1). A checkbox option is available to assign the Storage Blob Data Contributor role on the Data Lake. Navigation buttons at the bottom include 'Review + create' (highlighted in blue), '< Previous', and 'Next: Security >'.

**Microsoft Azure** Microsoft.Azure.SynapseAnalytics-20250713142114 | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

**Overview**

Inputs Outputs Template

Deployment is in progress

Deployment name : Microsoft.Azure.SynapseAnalytics-20250713142114 Start time : 13/07/2025, 14:29:17  
Subscription : Azure subscription 1 Correlation ID : f1f9c8d9-f254-4ad1-9abb-28f55f78e109  
Resource group : AzureProject1

Deployment details

Resource	Type	Status	Operation details
dfsynapseazproject1/default/dfsynapsefilenam	Microsoft.Storage/storageAccount	Created	Operation details
dfsynapseazproject1	Storage account	OK	Operation details

Give feedback Tell us about your experience with deployment

Add or remove favorites by pressing **Ctrl+Shift+F**

**Microsoft Defender for Cloud**  
Secure your apps and infrastructure  
[Go to Microsoft Defender for Cloud >](#)

**Free Microsoft tutorials**  
[Start learning today >](#)

**Work with an expert**  
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.  
[Find an Azure expert >](#)

**Microsoft Azure** Microsoft.Azure.SynapseAnalytics-20250713143516 | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

**Overview**

Inputs Outputs Template

Your deployment is complete

Deployment name : Microsoft.Azure.SynapseAnalytics-20250713143516 Start time : 13/07/2025, 14:38:01  
Subscription : Azure subscription 1 Correlation ID : d3bddf02-a47e-4631-9f01-b666ee180611  
Resource group : AzureProject1

Deployment details

Next steps

**Go to resource group**

Give feedback Tell us about your experience with deployment

Add or remove favorites by pressing **Ctrl+Shift+F**

**Cost management**  
Get notified to stay within your budget and prevent unexpected charges on your bill.  
[Set up cost alerts >](#)

**Microsoft Defender for Cloud**  
Secure your apps and infrastructure  
[Go to Microsoft Defender for Cloud >](#)

**Free Microsoft tutorials**  
[Start learning today >](#)

**Work with an expert**  
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.  
[Find an Azure expert >](#)

Microsoft Azure

Home > Microsoft.Azure.SynapseAnalytics-20250713143516 | Overview >

AzureProject1

Resource group

Overview

Essentials

Resources Recommendations (1)

Filter for any field... Type equals all Location equals all Add filter

Showing 1 to 5 of 5 records. Show hidden types

Name	Type	Location
azdb-project1	Azure Databricks Service	Central India
azstoragedatalake1rutu	Storage account	East US
azsynapse-project1	Synapse workspace	Central India
azureadfproject-rutuja	Data factory (V2)	East US
dfsynapseazproject1	Storage account	Central India

< Previous Page 1 of 1 Next >

Give feedback

## Step 35 -- Adding role to azure synapse just like we did before for databricks

Only difference is we now select managed identity instead of service principle because we do not need third party application to give the access

Microsoft Azure

Home > AzureProject1 > azstoragedatalake1rutu | Access Control (IAM) >

Add role assignment

Role Members\* Conditions Review + assign

Selected role Storage Blob Data Contributor

Assign access to

User, group, or service principal (radio button)

Managed identity (radio button, selected)

Members + Select members

Name Object ID Type

No members selected

Description

Optional

Select managed identities

Some results might be hidden due to your ABAC condition.

Subscription \* Azure subscription 1

Managed identity Synapse workspace (1)

Select

Search by name

Selected members: azsynapse-project1 /subscriptions/d7f3756b-e1ae-4fd8-87b8-aa37971ee89d/resourceGroups/...

Review + assign Previous Next Select Close Feedback

## Step 36 -- Adding role to yourself --- the user

The screenshot shows the Microsoft Azure Access Control (IAM) interface for adding a role assignment. The 'Members' tab is selected. Under 'Selected role', 'Storage Blob Data Contributor' is chosen. Under 'Assign access to', 'User, group, or service principal' is selected. In the 'Members' section, a search bar shows 'rutuja'. Below it, a list shows 'Rutuja Lamkane(Guest)' with the email 'rutuja.lamkane.personal\_gmail.com#EXT#@rutujalamkanepersonalgmail.on...'. A 'Select' button is at the bottom right of the modal.

## Step 37 – Create Views and External tables and load in Gold(serving) layer

The screenshot shows the Microsoft Azure Synapse Analytics workspace. On the left, the 'Data' menu is open, showing 'Workspace' and 'SQL database'. A 'Linked' section lists 'azdatabase1 (SQL)'. The main area is a query editor with a SQL script. The script reads data from a blob storage location using a BULK command. The results pane shows a table with three rows:

Date	Month	Year
2015-01-01	1	2015
2015-01-02	1	2015

A message at the bottom says '00:00:10 Query executed successfully.'

Microsoft Azure

Home > AzureProject1 > azstoragedatalake1runtu | Containers >

gold Container

Overview

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

Showing all 8 items

Name	Last modified	Access tier	Blob type	Size	Lease state
extcalendar	13/07/2025, 19:28:36				...
excUSTOMERS	13/07/2025, 19:30:18				...
extproduct	13/07/2025, 19:35:10				...
extProduct_Categories	13/07/2025, 19:32:43				...
extProduct_SubCategories	13/07/2025, 19:34:10				...
extReturns	13/07/2025, 19:36:05				...
extSales	13/07/2025, 19:25:59				...
extTerritories	13/07/2025, 19:40:01				...

Add or remove favorites by pressing Ctrl+Shift+F

## Step 38 -- To connect with power BI

### Copy the serverless sql endpoint on right side

Microsoft Azure

Home > AzureProject1 >

azsynapse-project1 Synapse workspace

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Resource visualizer

Settings

Analytics pools

Security

Monitoring

Automation

Help

Essentials

Resource group (move)	: AzureProject1
Status	: Succeeded
Location	: Central India
Subscription (move)	: Azure subscription 1
Subscription ID	: d7f3756b-e1ae-4fdb-87b8-aa37971ee89d
Managed virtual network	: No
Managed identity object ...	: 96425b0f-da78-4d0a-a66c-acd6280e01fc
Workspace web URL	: <a href="https://web.azuresynapse.net?workspace=%2bsubscriptions%2fd7...">https://web.azuresynapse.net?workspace=%2bsubscriptions%2fd7...</a>
Tags (edit)	: Add tags

Networking : Show firewall settings

Primary ADLS Gen2 account : <https://dfsynapseappproject1.dfs.core.windows.net>

Primary ADLS Gen2 file s... : dfsynapsefilename1

SQL admin username : adminruntu

SQL Microsoft Entra admin : <live.com#rutuja.lamkane.personal@gmail.com>

Dedicated SQL endpoint : <azsynapse-project1.sql.azuresynapse.net>

Serverless SQL endpoint : <azsynapse-project1-ondemand.sql.azuresynapse.net>

Development endpoint : <https://azsynapse-project1.dev.azuresynapse.net>

Getting started

Open Synapse Studio

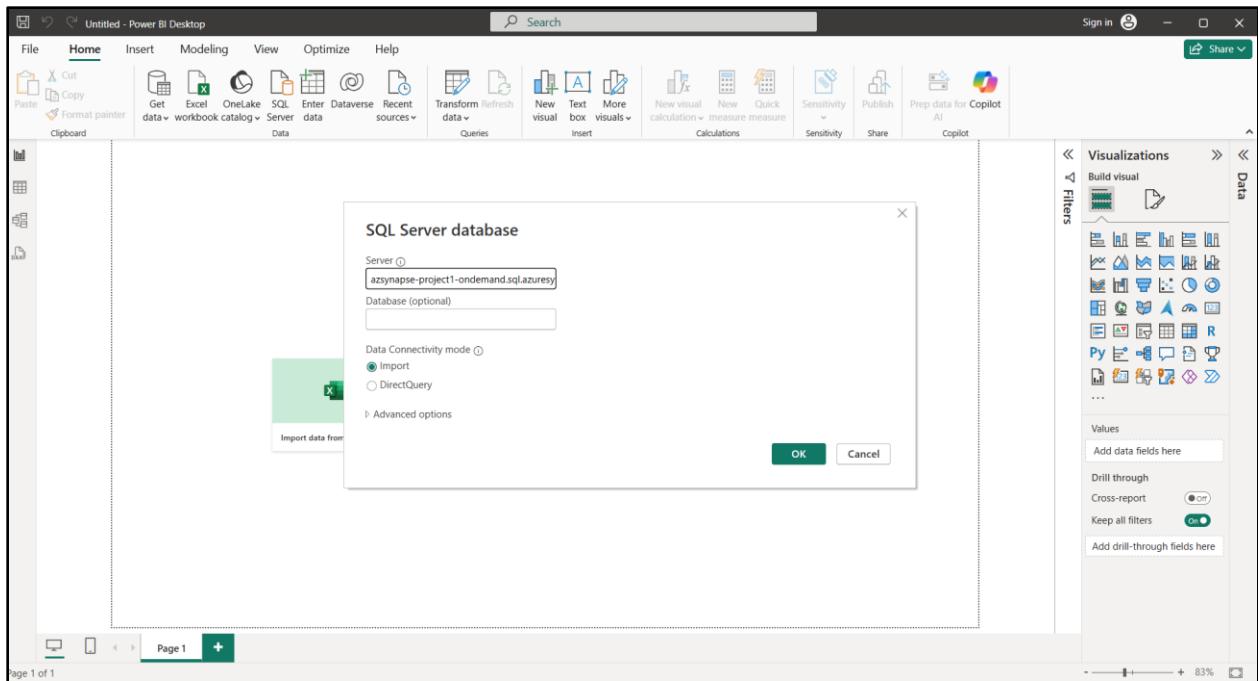
Start building your fully-integrated analytics solution and unlock new insights.

Open

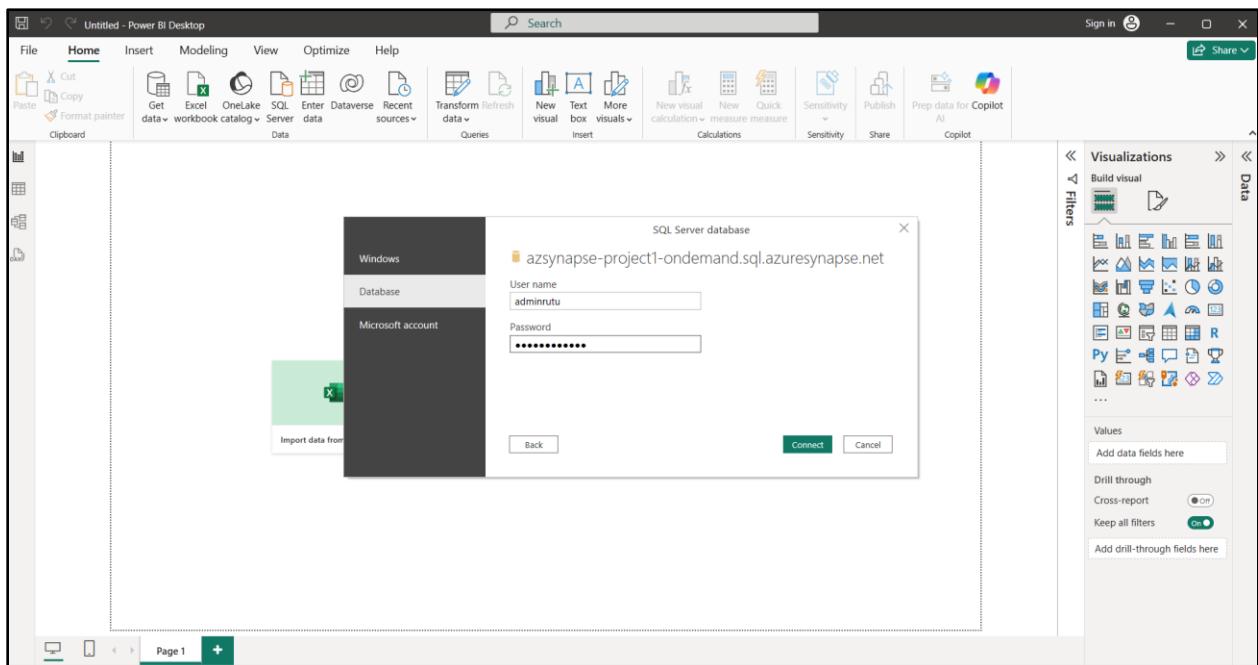
Read documentation

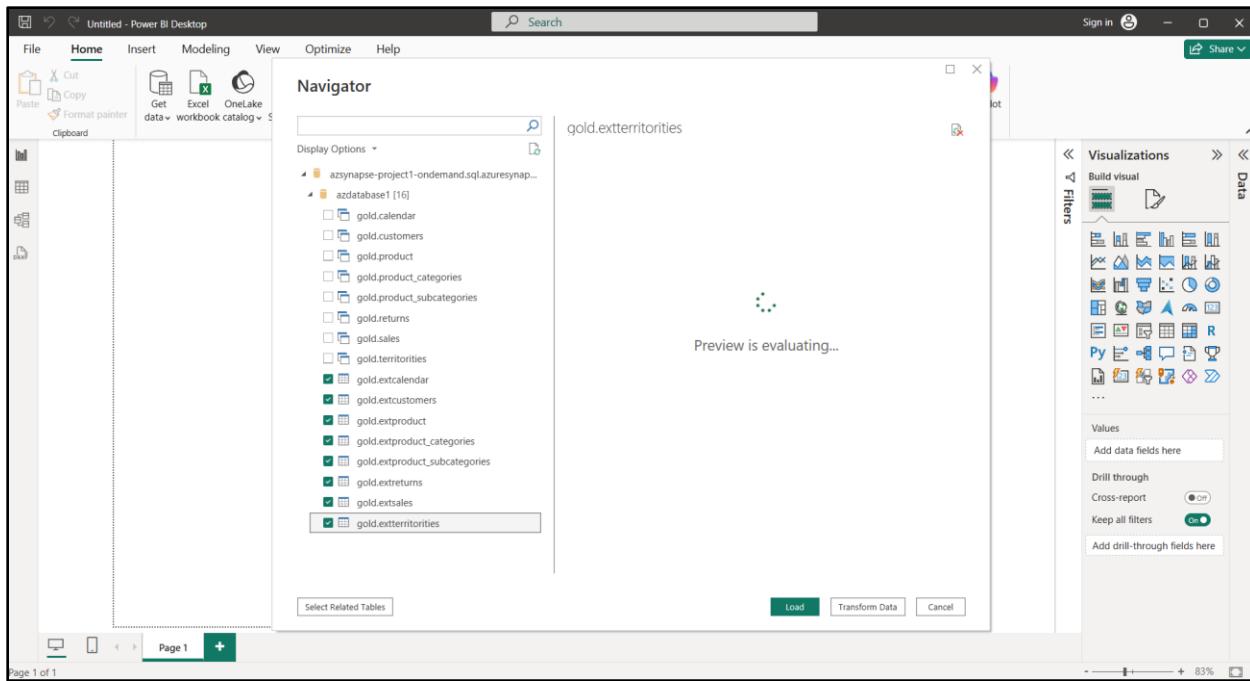
Learn how to be productive quickly. Explore concepts, tutorials, and samples.

Learn more



## Step 39 – Add the credentials that we created while creating azure synapse resource





## Step 40 – Final Output

