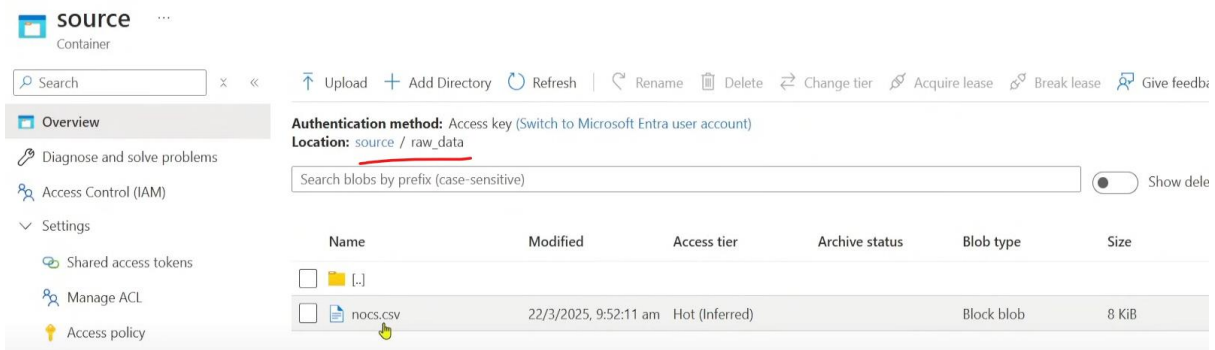


For Silver notebooks, will use Dynamic Notebooks using Workflows

GitHub repository view for 'anshlambagit / AzureProjectWithCICD'. The commit history shows a series of 'Add files via upload' commits for 'athletes.csv', 'coaches.csv', 'events.csv', and 'nocs.csv'. A red handwritten note 'OLYMPIC dataset' is written over the commit history.

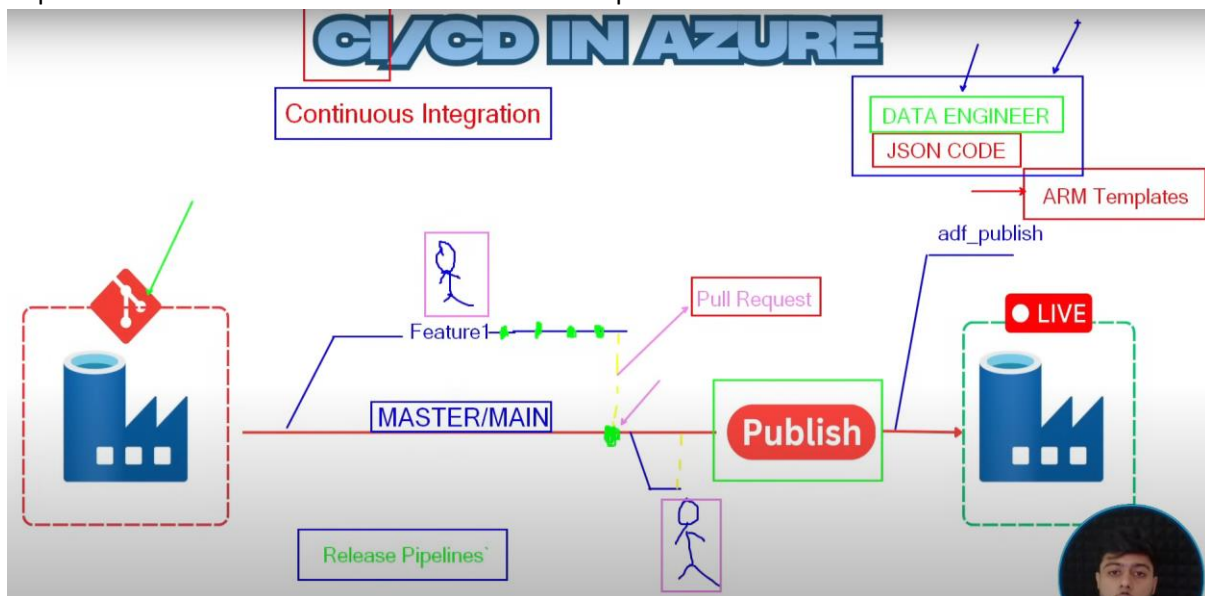
Azure portal view for the 'containersprojdl' storage account. The 'Containers' section lists five containers: '\$logs', 'bronze', 'gold', 'silver', and 'source'. A red handwritten note '4 containers' is written over the list, indicating that the first container '\$logs' is excluded.

Name	Last modified	Anonymous access level
<input type="checkbox"/> \$logs	22/3/2025, 9:43:25 am	Private
<input type="checkbox"/> bronze	22/3/2025, 9:46:50 am	Private
<input type="checkbox"/> gold	22/3/2025, 9:46:57 am	Private
<input type="checkbox"/> silver	22/3/2025, 9:46:54 am	Private
<input type="checkbox"/> source	22/3/2025, 9:46:44 am	Private



Upload nocs data in source ADLS, rest will do directly from GitHub

Popular version controls → GitHub and Azure DevOps



We always configure ADF development environment with Git, master/main branch will be created automatically once we integrate both. Master branch → Main branch/collaboration branch/source of truth.

Git/ADO integrated ADF instance is different from ADF Live mode. Publish to make it available.

ADF Live mode → actually this is the one which runs live.

Once you publish from main to Live mode, 1)all the changes will be available in Live mode 2)there will be another branch `adf_publish` which will carry the ARM templates (BTS JSON codes) which will help us to deploy the code to higher environments using Release pipelines.

To create Azure DevOps account, best way is to create from Azure portal

Search Azure DevOps in Azure → Click on My Azure DevOps Organisations → Choose Default Directory → Create new organization



Ansh Lamba

Edit profile

anshlambaed@outlook.com

Microsoft account

Default Directory

Microsoft account

✉ anshlambaed@outlook.com

### Visual Studio Dev Essentials

Get everything you need to build and deploy your app on any platform.

Use your benefits



## Get started with Azure DevOps

Plan better, code together, ship faster with Azure DevOps

Create new organization

portal.azure.com/#view/AzureIlsExtension/OrganizationsTemplateBlade

Microsoft Azure Upgrade Search resources, services, and docs (G+/) Copilot anshlambaed@outlook... DEFAULT DIRECTORY

Home > Azure DevOps

Name: Ansh Lamba  
Email: anshlambaed@outlook.com  
Directory: Default Directory (6d9d0ac5-2bac-45da-afe2-475d4b7b52fd)  
Domain: anshlambaedoutlook.onmicrosoft.com

We've made it easier to manage Azure DevOps billing and subscriptions. You can [set up billing](#), [change your subscription](#) or [pay for more users and resources](#) within Azure DevOps. [Learn more](#).

## Azure DevOps

Plan smarter, collaborate better, and ship faster with a set of modern dev



It should be Default Directory not Microsoft account

Azure DevOps

anshlambaed

New organization

Home page

### Create a project to get started

Project name \*

Visibility

Public  
Anyone on the internet can view the project. Certain features like TFVC are not supported.

Private  
Only people you give access to will be able to view this project.

Public projects are disabled for your organization. You can turn on public visibility with [organization policies](#).

+ Create project

Organization settings

Azure DevOps anshlambaed / Settings / Microsoft Entra

## Organization Settings

anshlambaed

Search Settings

General

- Overview
- Projects
- Users
- Billing
- Global notifications
- Usage
- Extensions
- Microsoft Entra

Security

- Security overview

### Microsoft Entra

Your organization is connected to the **Default Directory** directory.

**Default Directory**  
anshlambaedoutlook.onmicrosoft.com  
Tenant Id: 6d9d0ac5-2bac-45da-afe2-475d4b7b52fd

Check out [other frequently asked questions](#).

**Disconnect directory**

Download Azure DevOps organizations connected to **Default Directory** directory.

**Download**

**Organization Settings**  
anshlambaed  
Search Settings

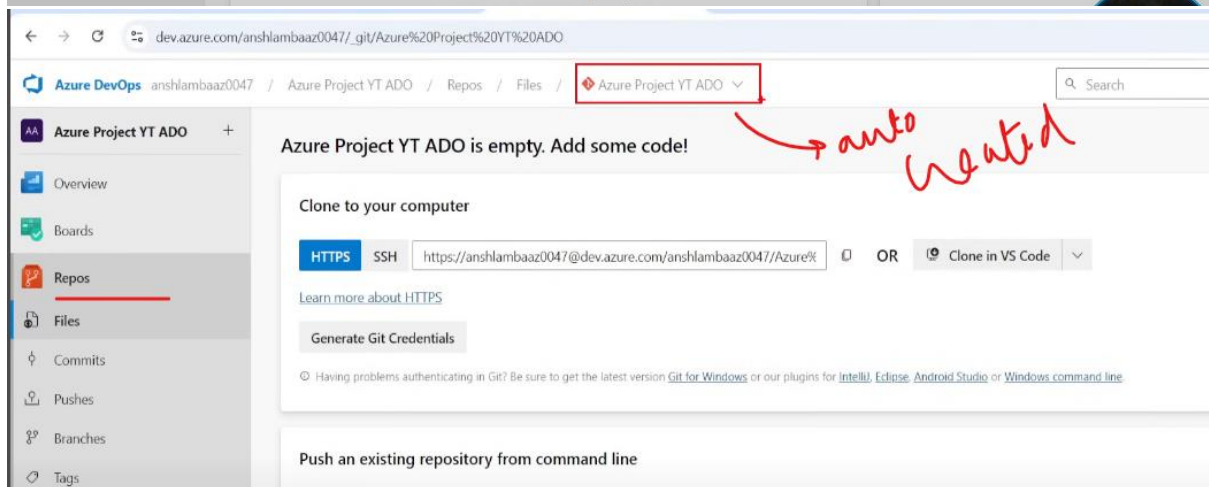
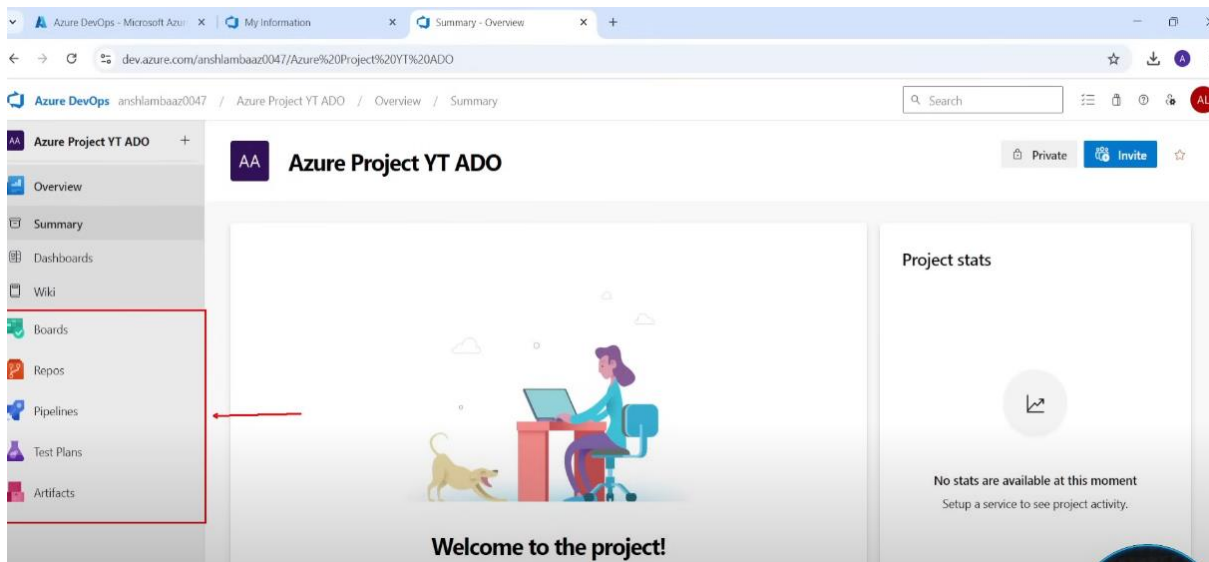
**Overview**  
Name  
anshlambaed  
Use the new URL: <https://dev.azure.com/anshlambaed/>  
[Learn more about URLs](#)  
Privacy URL  
[Learn more about the Privacy URL](#)  
Description  
Add organization description  
Time zone  
UTC  
Geography  
Canada

**Organization Settings**  
anshlambaed  
Search Settings

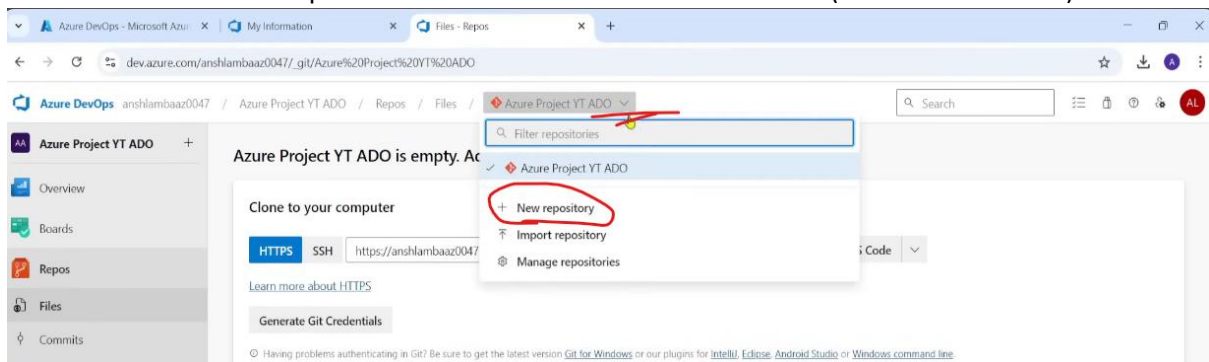
**Projects**  
Filter projects  
+ New project  
Total 0  
Name ↑ Description Last Updated Process Visibility

**Azure DevOps**  
anshlambaaz0047  
anshlambaaz  
View organization

**Create a project to get started**  
Project name \*  
Azure Project YT ADO  
Description  
Visibility  
Public  
Private  
Public projects are disabled for your organization. You can turn on public visibility with [organization policies](#).  
Advanced  
Version control Git  
Work item process Basic  
**Create project**



A Repo will be created automatically when we create a project with the same name.  
Better to create a new Repo with same name as that of ADF instance (real world scenario)

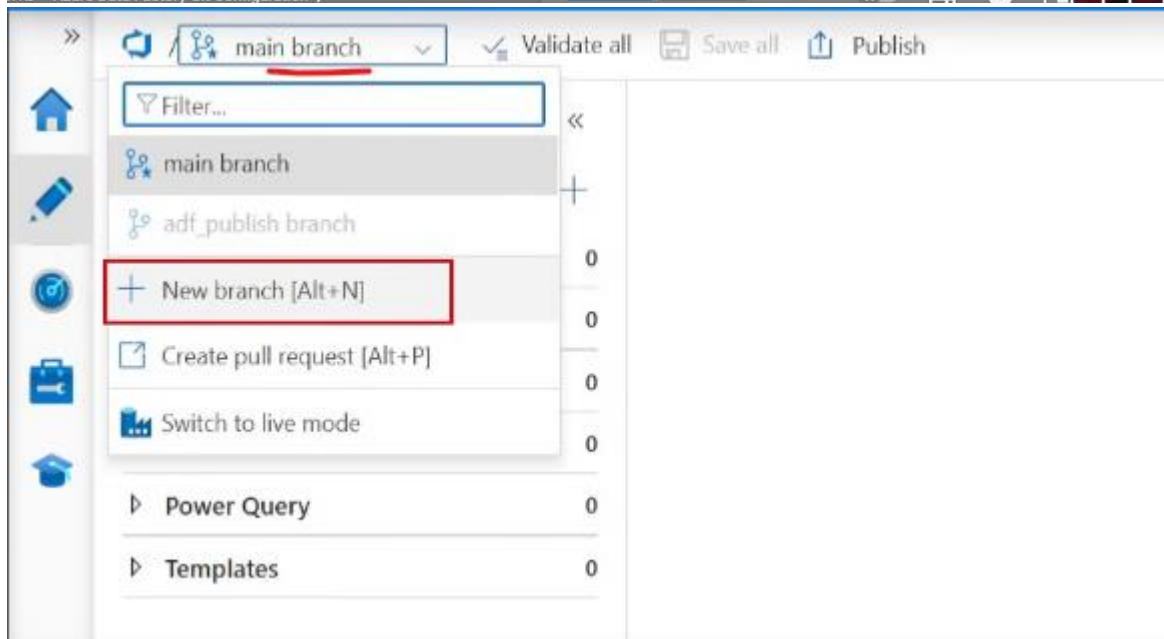
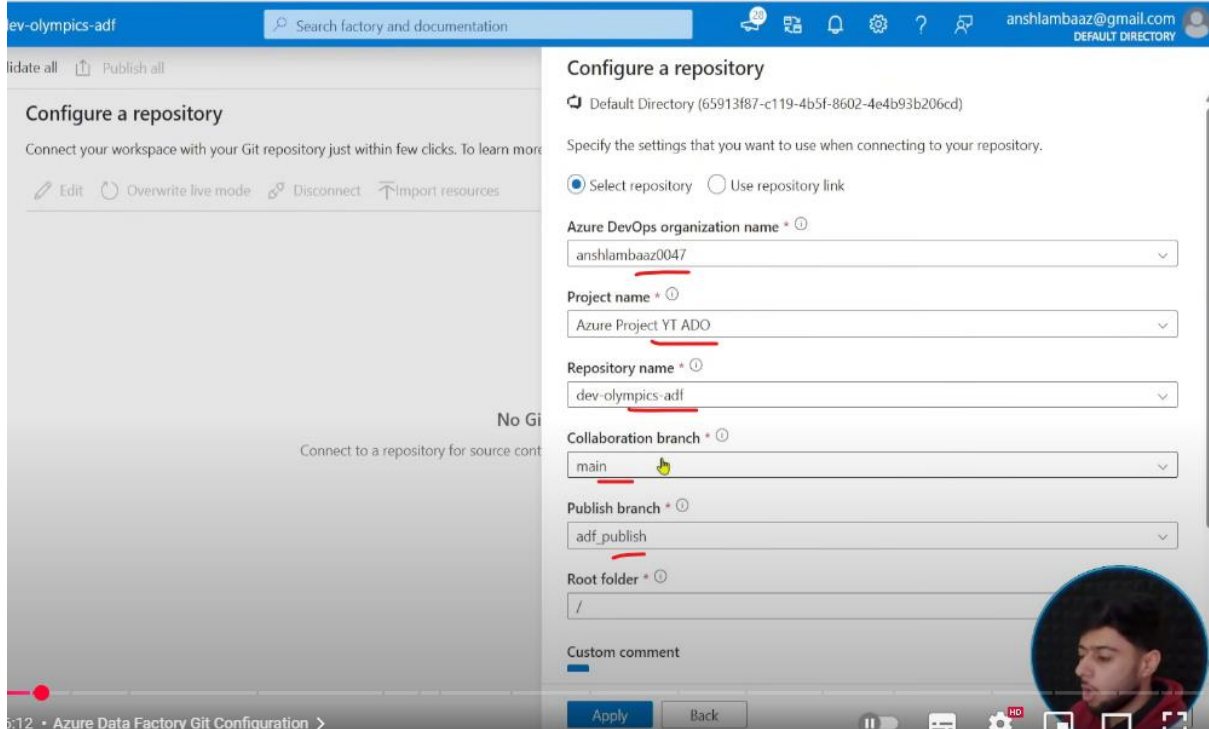
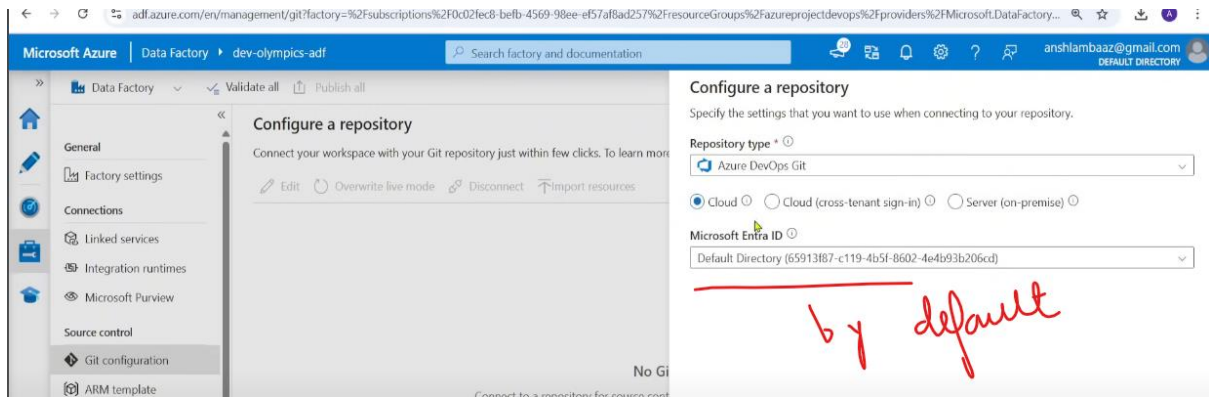




The first screenshot shows the 'Create a repository' dialog box in Azure DevOps. The repository type is 'Git', the name is 'dev-olympics-adf', and 'Add a README' is checked. The second screenshot shows the 'Files' view of the 'dev-olympics-adf' repository. A red box highlights the 'README.md' file, and the handwritten text 'nothing is here' is written over it. The third screenshot shows the 'Branches' view, with a red box highlighting the 'main' branch.

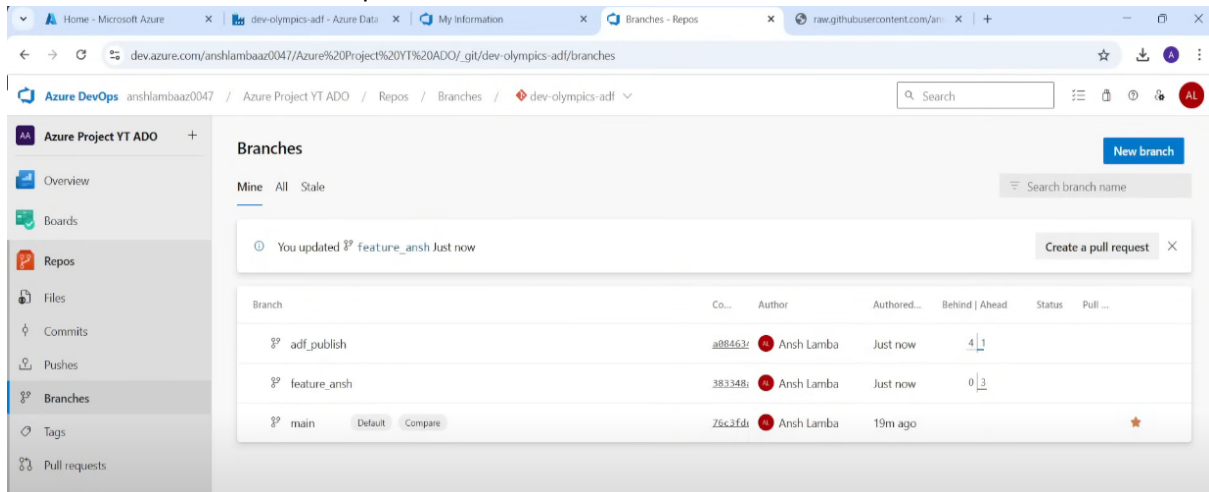
Now go to ADF to integrate version control

The screenshot shows the Microsoft Azure Data Factory 'dev-olympics-adf' workspace. The 'Connections' tab is selected in the left sidebar. The 'Configure a repository' dialog box is open, showing 'Azure DevOps Git' as the selected repository type. A red box highlights the 'Connections' tab in the sidebar.





## Create a Feature branch on top of main branch

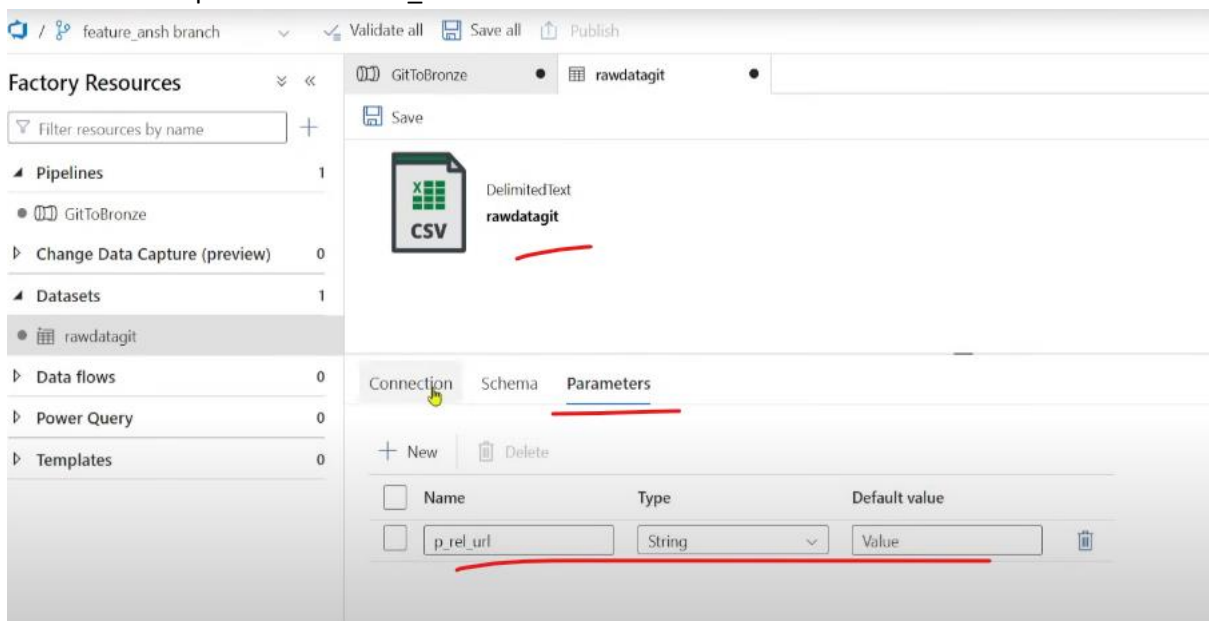


The screenshot shows the Azure DevOps web interface for the 'dev-olympics-adf' repository. The 'Branches' tab is selected, displaying a list of branches. A notification at the top indicates 'You updated feature\_anish Just now'. The table below shows the following branches:

Branch	Co...	Author	Authored...	Behind   Ahead	Status	Pull ...
adf_publish	a08463r	Ansh Lamba	Just now	4   1		
feature_anish	383348r	Ansh Lamba	Just now	0   3		
main	76c2fdu	Ansh Lamba	19m ago			

Buttons for 'New branch', 'Search branch name', and 'Create a pull request' are visible.

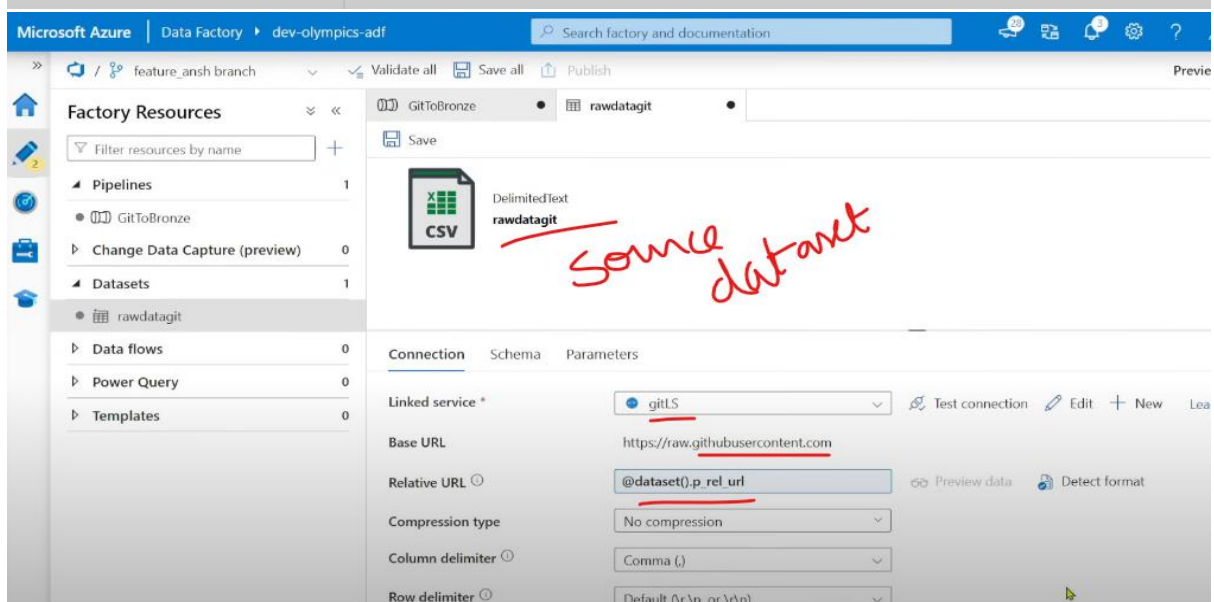
Now create a ADF pipeline to pull the 3 files from GitHub, when we save our changes in FR all the artifacts will be present in feature\_anish branch



The screenshot shows the configuration for a pipeline named 'rawdatagit' in the 'feature\_anish' branch. The 'Parameters' tab is selected, showing a table with one parameter:

Name	Type	Default value
p_rel_url	String	Value

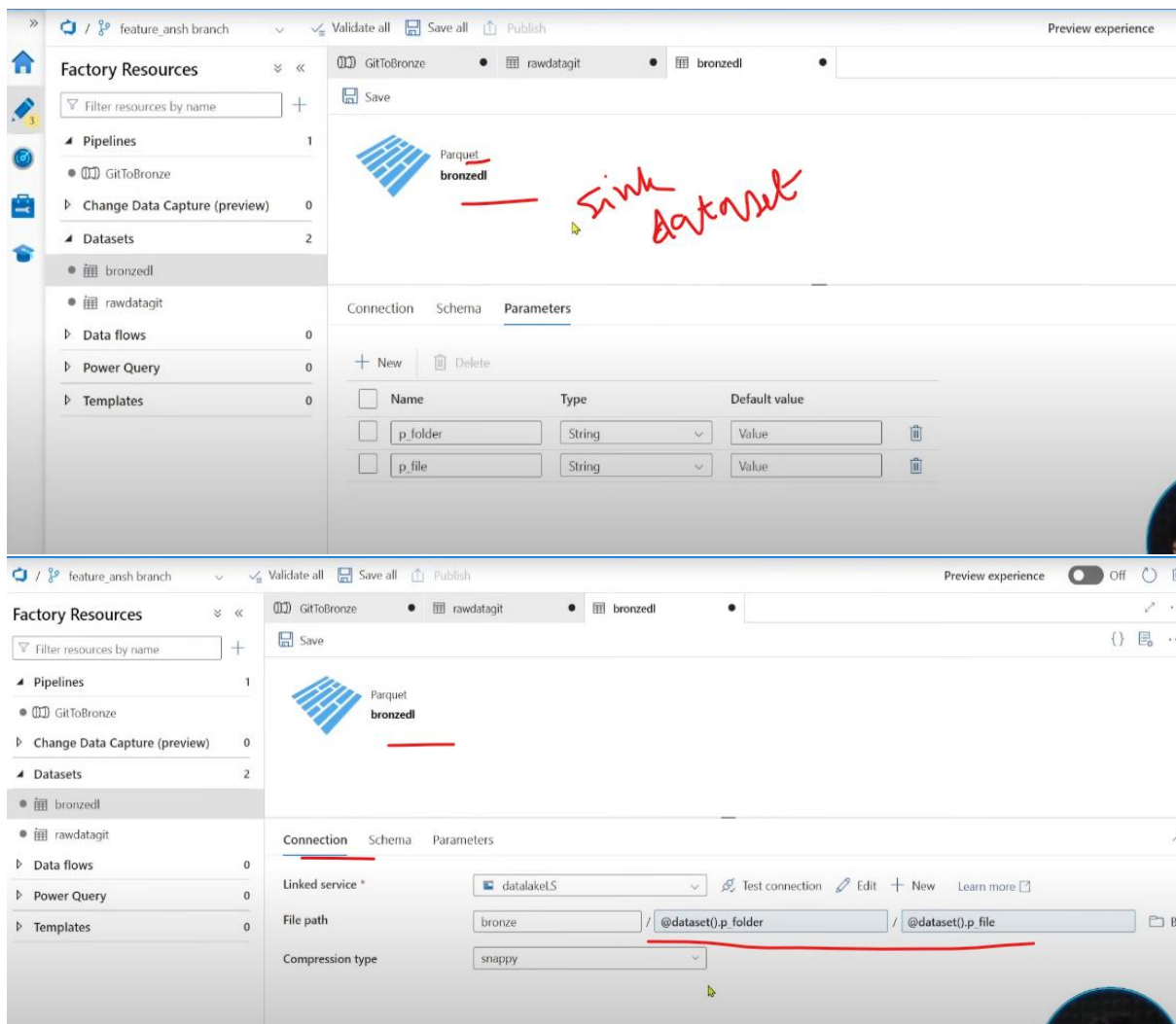
Handwritten red annotations include 'rawdatagit' and 'Source dataset'.



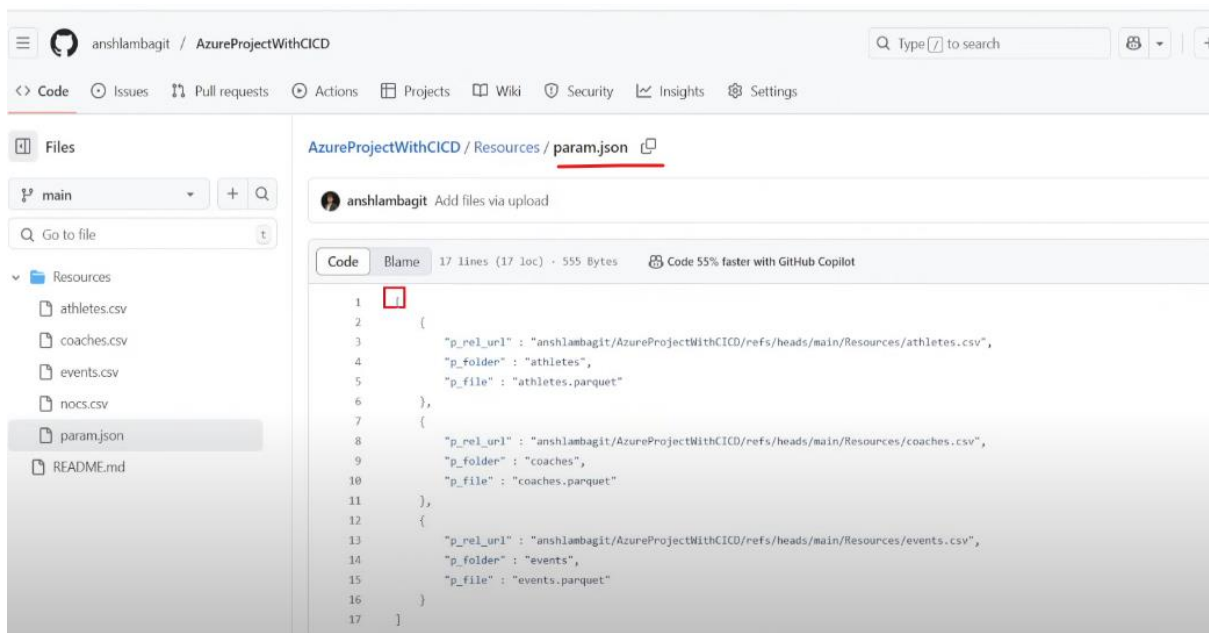
The screenshot shows the configuration for a pipeline named 'rawdatagit' in the 'feature\_anish' branch. The 'Connection' tab is selected, showing the following settings:

- Linked service: gitLS
- Base URL: https://raw.githubusercontent.com
- Relative URL: @dataset().p\_rel\_url
- Compression type: No compression
- Column delimiter: Comma (,)
- Row delimiter: Default (\r\n, or \r\n)

Handwritten red annotations include 'rawdatagit' and 'Source dataset'.



Best way to pass the arrays to the pipeline is via a JSON file holding details of all parameters rather than passing from pipeline parameters.



Keep the file in bronze folder

Validate all Save all Publish Preview experience

GitToBronze rawdatagit bronzedl

Activities

look

General

Lookup

LookupJSON

Copy data

GitToBronzeCopy

Trigger test runs of the current pipeline without publishing your changes to the service

General Settings User properties

Source dataset \* jsonDS

Open + New Preview data Learn more

First row only

File path type

File path in dataset Wildcard file path List of files

Filter by last modified

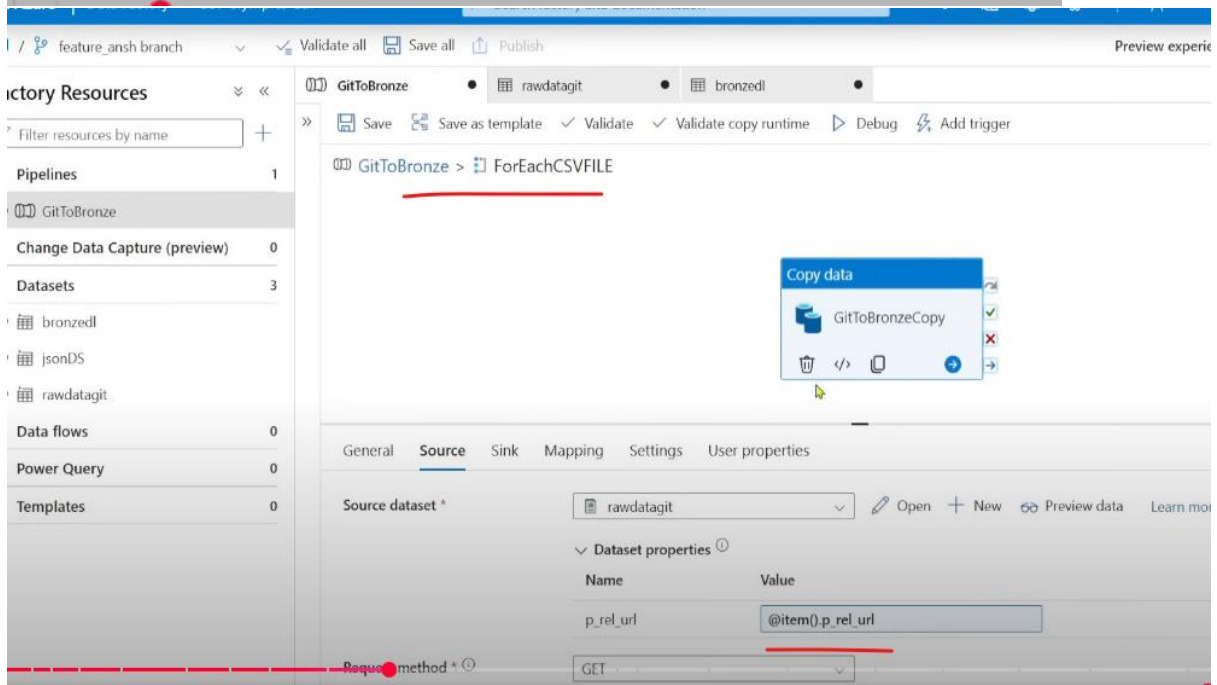
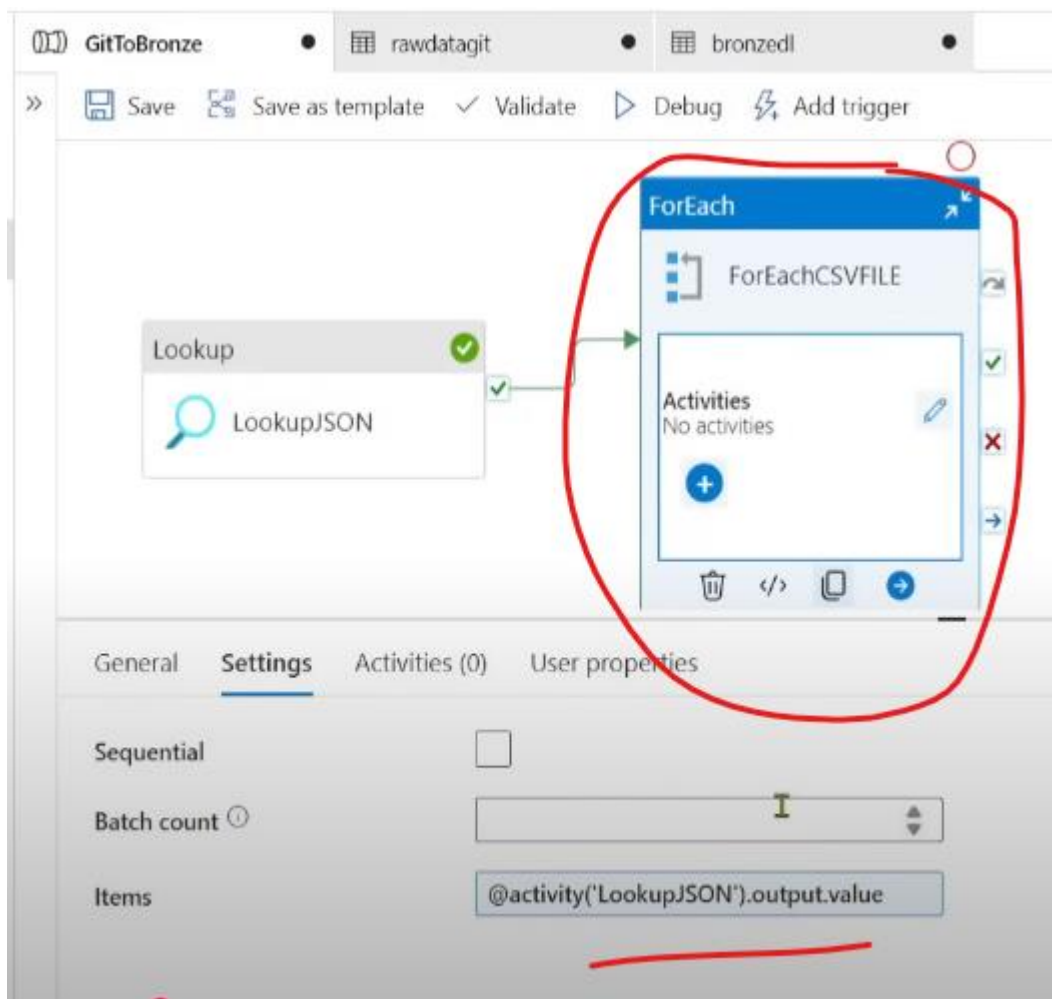
Start time (UTC) End time (UTC)

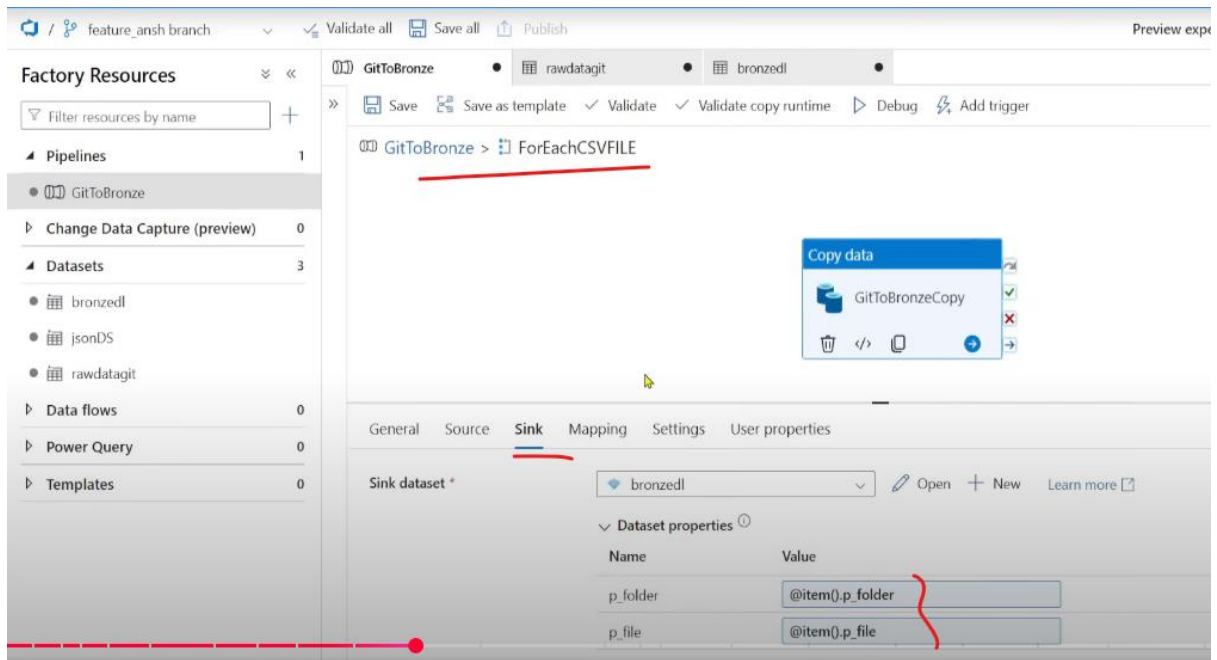
Recursively

## Output

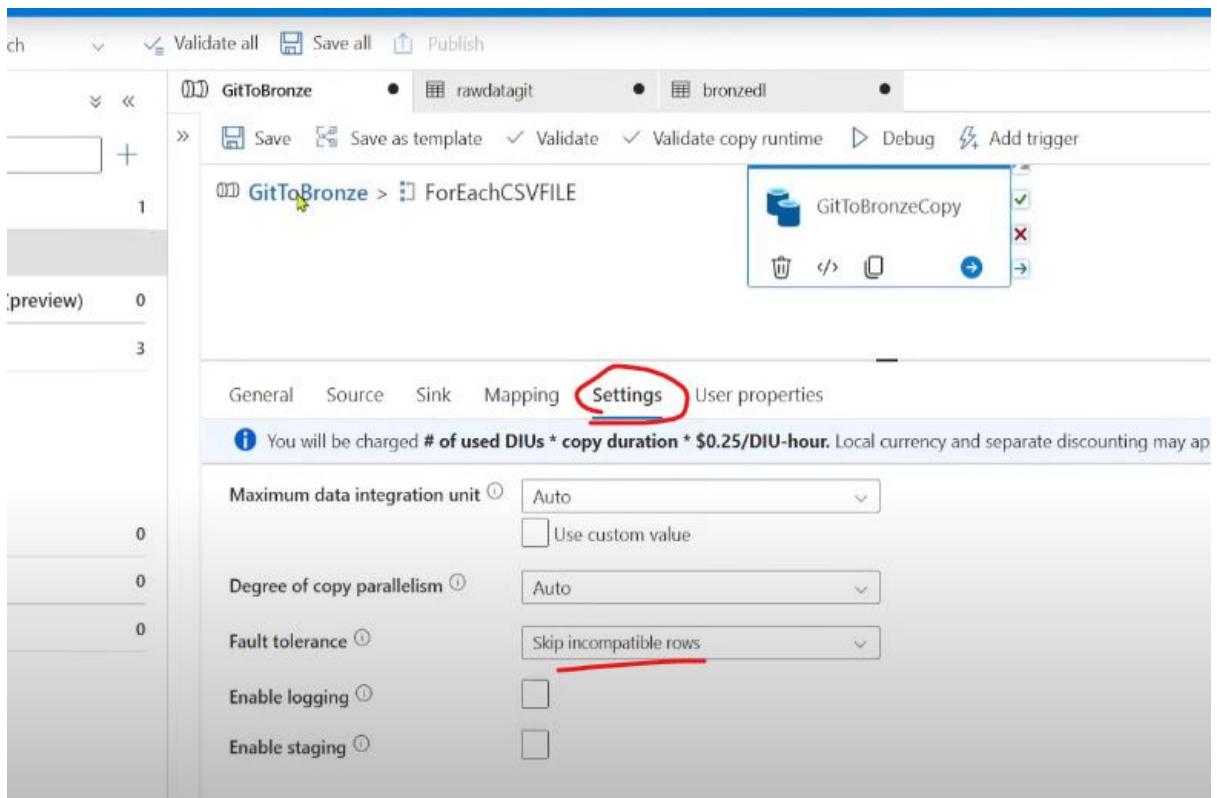
Copy to clipboard

```
{
  "count": 3,
  "value": [
    {
      "p_rel_url": "anshlambagit/AzureProjectWithCICD/refs/heads/main/Resources/athletes.csv",
      "p_folder": "athletes",
      "p_file": "athletes.parquet"
    },
    {
      "p_rel_url": "anshlambagit/AzureProjectWithCICD/refs/heads/main/Resources/coaches.csv",
      "p_folder": "coaches",
      "p_file": "coaches.parquet"
    },
    {
      "p_rel_url": "anshlambagit/AzureProjectWithCICD/refs/heads/main/Resources/events.csv",
      "p_folder": "events",
      "p_file": "events.parquet"
    }
  ],
  "effectiveIntegrationRuntime": "AutoResolveIntegrationRuntime (Canada Central)",
  "billingReference": {
    "activityType": "PipelineActivity",
    "billingReference": {}
  }
}
```





Run the pipeline → failure is happening for one of the file since its containing more than 36 columns  
Enable Fault Tolerance



Create one more pipeline for the remaining one file (nocs data) ingestion from Data Lake  
Scenarios(adding complexity):

rojdl | Containers > source >

raw\_data/nocs.csv ...

Authentication method: Access key (Switch to Microsoft Entra user account)  
Location: source / raw\_data

Search blobs by prefix (case-sensitive)

Show deleted objects

Name

[-]

nocs.csv

Overview Versions Edit Generate SAS

```

1 code,country,country_long,tag,note
2 AFG,Afghanistan,Afghanistan,afghanistan,P
3 AHO,Netherlands Antilles,Netherlands Antilles,netherlands-antilles,H
4 AIN,AIN,AIN,ain,P
5 ALB,Albania,Albania,albania,P
6 ALG,Algeria,Algeria,algeria,P
7 AND,Andorra,Andorra,andorra,P
8 ANG,Angola,Angola,angola,P
9 ANT,Antigua and Barbuda,Antigua and Barbuda,antigua-and-barbuda,P
10 ARG,Argentina,Argentina,argentina,P
11 ARM,Armenia,Armenia,armenia,P
12 ARU,Aruba,Aruba,aruba,P
13 ASA,American Samoa,American Samoa,american-samoa,P

```

1) Make sure that the file has 5 columns

Home > olympicsprojdl | Containers >

source ...

Authentication method: Access key (Switch to Microsoft Entra user account)  
Location: source / raw\_data

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[-]						
athletes.csv	22/3/2025, 12:38:39 ...	Hot (Inferred)		Block blob	6.89 MiB	Available
coaches.csv	22/3/2025, 12:38:39 ...	Hot (Inferred)		Block blob	91.28 KiB	Available
events.csv	22/3/2025, 12:38:39 ...	Hot (Inferred)		Block blob	30.88 KiB	Available
nocs.csv	22/3/2025, 9:52:11 am	Hot (Inferred)		Block blob	8 KiB	Available

Successfully uploaded blob(s)  
Successfully uploaded 3 blob(s).

2) We have to process only the nocs file but I want the count of other files also

feature\_ansh branch

Validate all Save all Publish

Factory Resources

Filter resources by name

Pipelines 2

GitToBronze

DataLakeIngestion

Change Data Capture (preview) 0

Datasets 4

bronzedl

datalakeDS

jsonDS

rawdatagit

Data flows 0

Power Query 0

Templates 0

GitToBronze rawdatagit bronzedl DataLakeIngestion

Save Save as template Validate Debug Add trigger

Get Metadata

Get Metadata Folder

General Settings User properties

Dataset \* datalakeDS

Field list \*

Argument

Child items

Start time (UTC)

End time (UTC)

Filter by last modified



### Output

Copy to clipboard

```
{
  "childItems": [
    {
      "name": "athletes.csv",
      "type": "File"
    },
    {
      "name": "coaches.csv",
      "type": "File"
    },
    {
      "name": "events.csv",
      "type": "File"
    },
    {
      "name": "nocs.csv",
      "type": "File"
    }
  ]
}
```

GitToBronze rawdatagit bronzedl DataLakeIngestion para\_DS

Save Save as template Validate Debug Add trigger

Get Metadata

Get Metadata Folder

ForEach

ForEachFile

General

Settings

Activities (1)

User properties

Sequential

Batch count

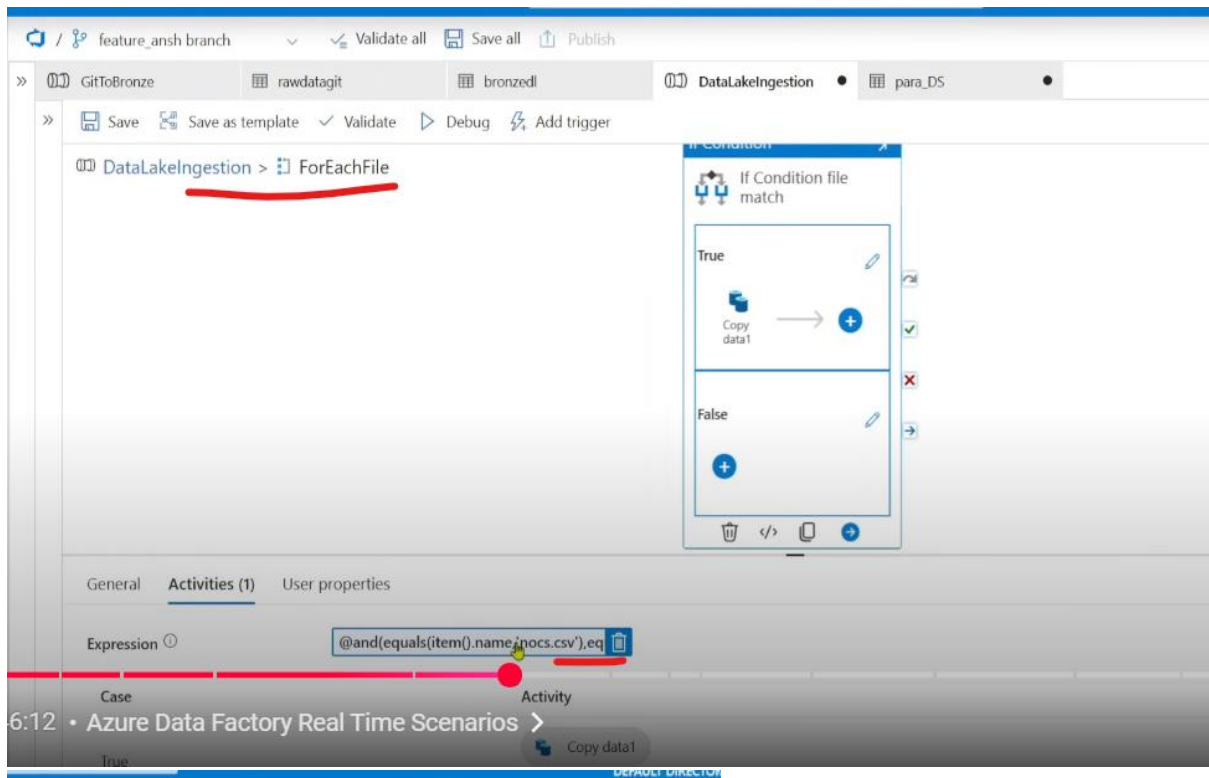
Items

@activity('Get Metadata Folder').out...

## Pipeline expression builder

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

```
@activity('Get Metadata Folder').output.childItems
```

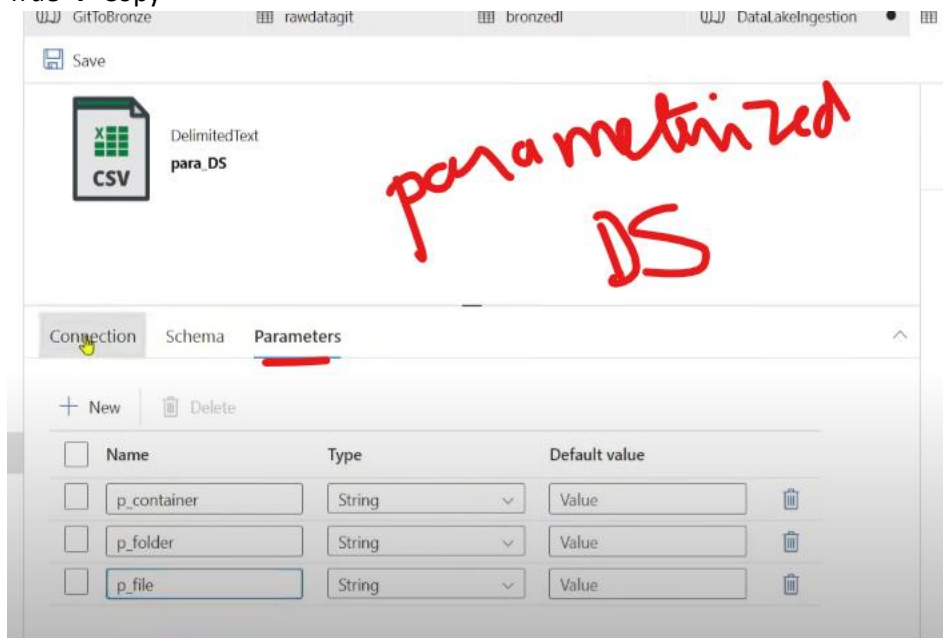


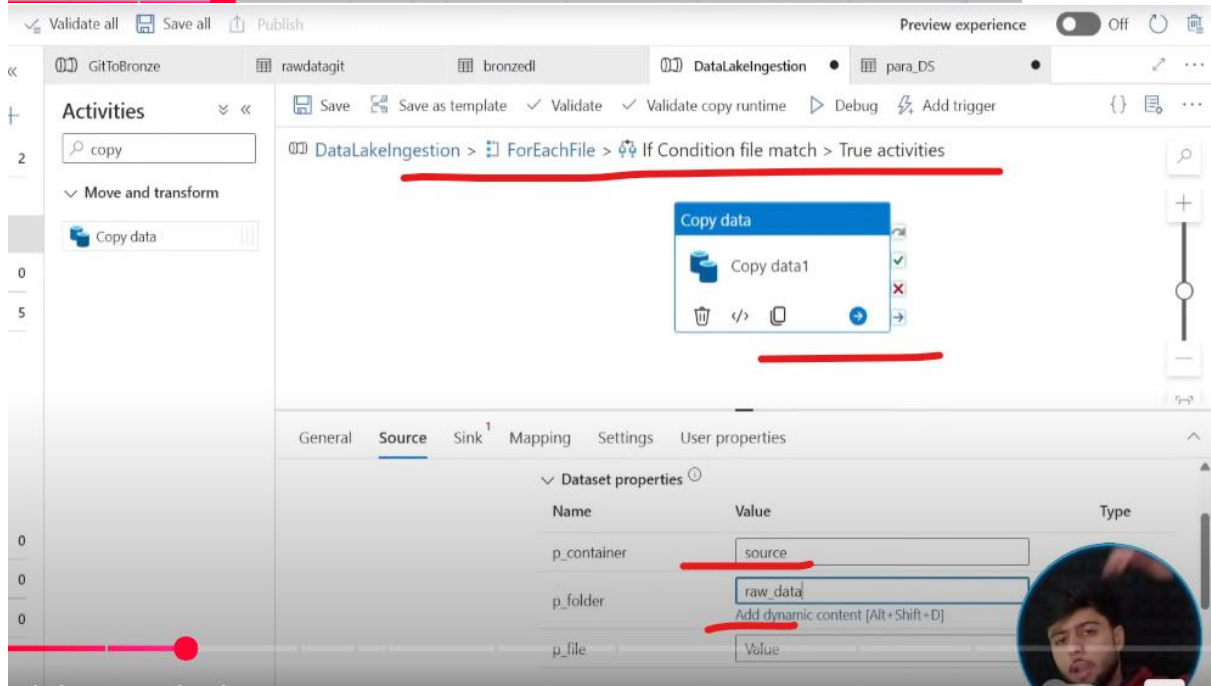
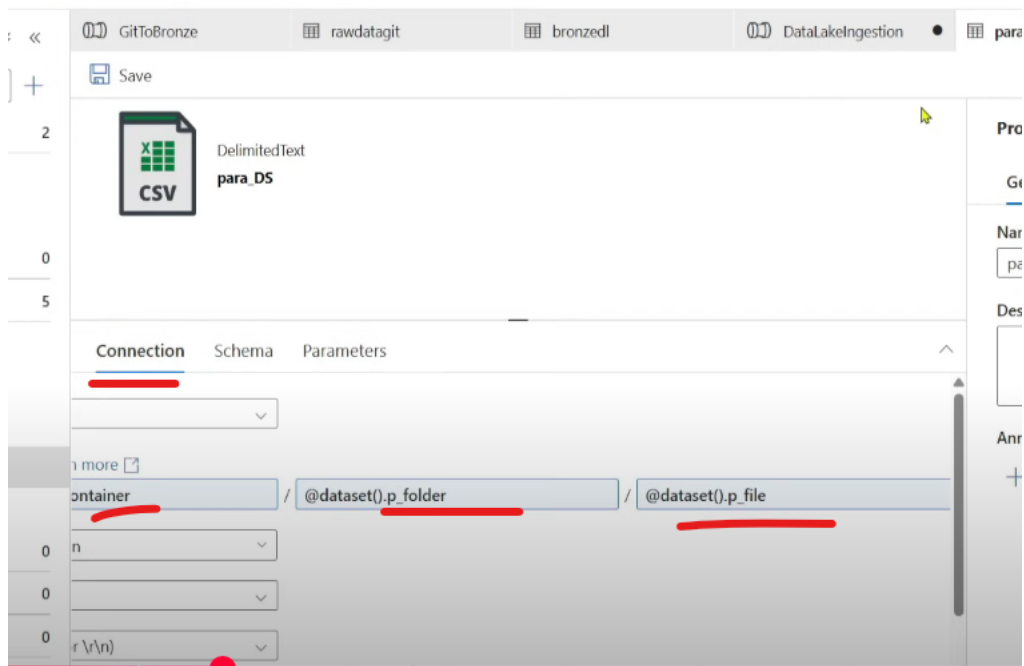
### Pipeline expression builder

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

```
@and(equals(item().name,'nocs.csv'),equals(item().type,'File'))
```

True → Copy





ite all Save all Publish Preview experience Off

GitToBronze rawdatagit bronzedl DataLakeIngestion para\_DS

activities copy

Move and transform Copy data

DataLakeIngestion > ForEachFile > If Condition file match > True activities

Copy data

Copy data1

General Source Sink<sup>1</sup> Mapping Settings User properties

p\_folder raw\_data string

p\_file @item().name string

File path type ☒ File path in dataset ☐ Wildcard file path ☐ List of files

Start time (UTC) End time (UTC)

Filter by last modified

Save the current resource without validation to your git repository

Save Save as template Validate Validate copy runtime Debug Add trigger

ForEachFile > If Condition file match > True activities

Copy data

Copy data1

General Source Sink Mapping Settings User properties

Sink dataset \* para\_DS Open + New Learn more

Dataset properties

Name	Value
p_container	bronze
p_folder	nocs
p_file	@item().name

Lets create 3 variables → one to hold the names of the files apart from nocs, one to hold the count of items in the previous variable/array, one to convert the list of names in the array to string

The image shows two screenshots from the Azure Data Factory portal. The top screenshot displays a workflow with a 'Get Metadata' activity (Get Metadata Folder) connected to a 'ForEach' activity (ForEachFile). The bottom screenshot shows the 'Variables' tab of the workflow editor, listing three variables: 'v\_file\_array' (Array), 'v\_files\_no' (Integer), and 'v\_file\_names' (String). A red bracket highlights these variables. Below the list, the 'Append variable' activity is shown, with the variable 'v\_file\_array' selected and the value '@item().name' entered. A red line underlines the path 'DataLakeIngestion > ForEachFile > If Condition file match - False activities' in the workflow tree. The 'Settings' tab for the 'Append variable' activity is also shown, with 'v\_file\_array' in the 'Name' field and '@item().name' in the 'Value' field. A disclaimer at the bottom states: 'Disclaimer: Append Variables Only Supports Adding To 'Array' Type Variables'.

Get Metadata  
Get Metadata Folder

ForEach  
ForEachFile

Parameters Variables Settings Output

+ New Delete

Name	Type	Default value
v_file_array	Array	Value
v_files_no	Integer	Value
v_file_names	String	Value

Save Save as template Validate Debug Add trigger

DataLakeIngestion > ForEachFile > If Condition file match - False activities

Append variable

Append variable files

v\_file\_array

@item().name

Disclaimer: Append Variables Only Supports Adding To 'Array' Type Variables

The top screenshot shows a pipeline with a 'Get Metadata' activity followed by a 'ForEach' loop containing 'ForEachFile'. Two 'Set variable' activities are connected to the 'ForEach' loop: 'Set File Names' and 'Set file numbers'. The bottom screenshot shows the 'Settings' tab for the 'Set file numbers' activity, where the variable type is 'Pipeline variable', the name is 'v\_file\_names', and the value is '@string(variables('v\_file\_array'))'. The 'Activities' pane on the left shows the 'Set variable' activity selected.

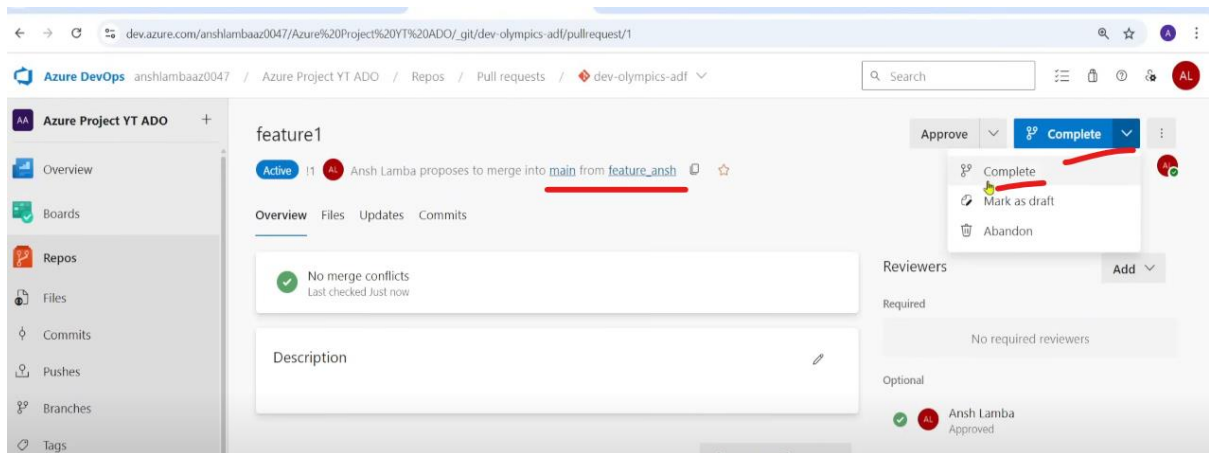
Note both the set variable activities are connected with on completion of ForEach activity  
Run the pipeline

The left screenshot shows the output for the 'v\_file\_names' variable, which is an array of file names: ['athletes.csv', 'events.csv', 'coaches.csv']. The right screenshot shows the output for the 'v\_files\_no' variable, which is the count of files: 3.

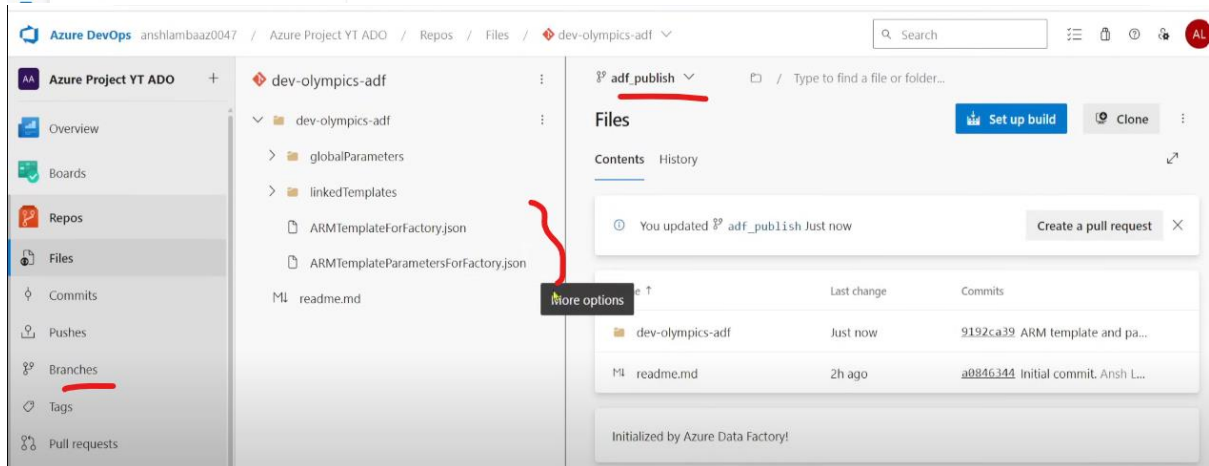
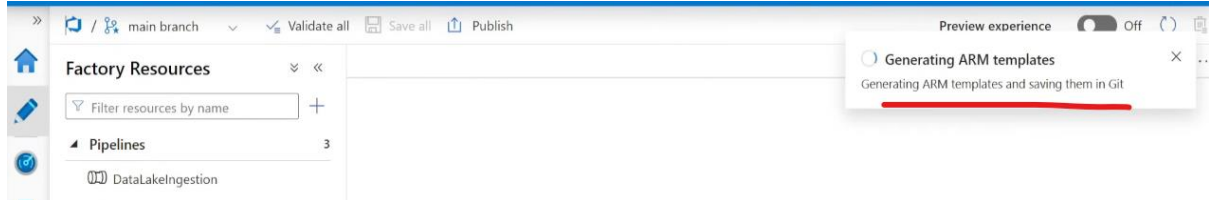
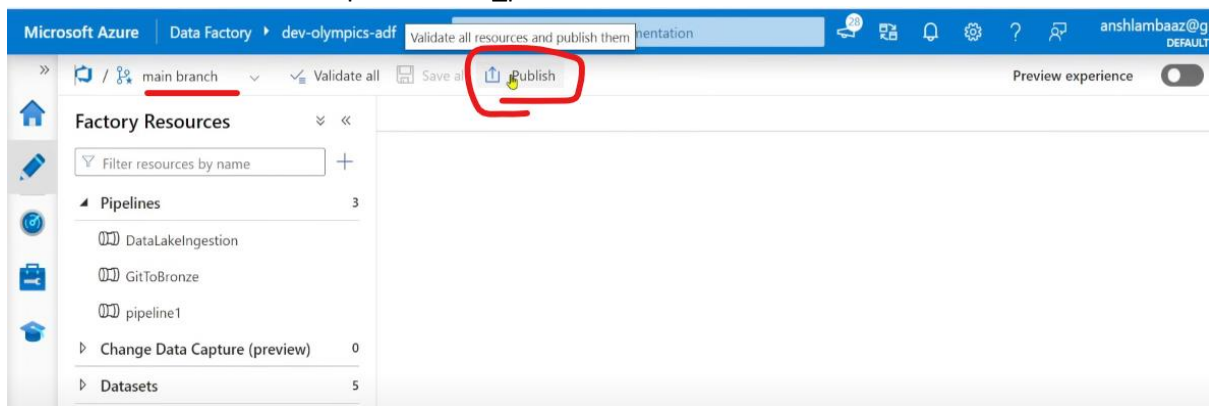
Save all the changes in the FR

Create a PR and complete and merge the FR with main branch



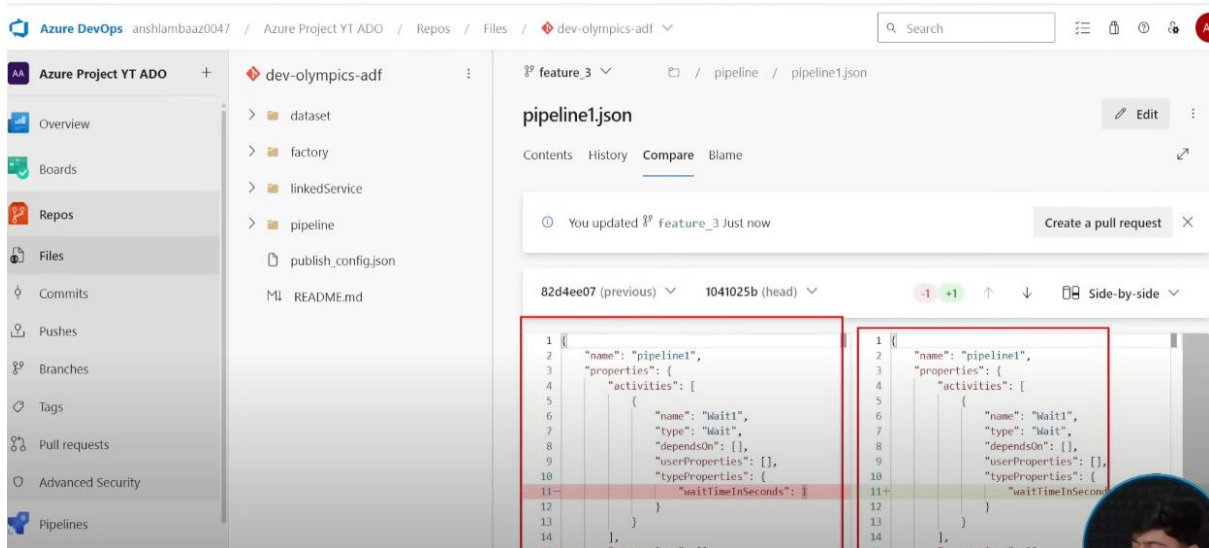


post merge the main branch will be updated with the new changes, then publish to ADF Live mode, ADF bts will create ARM templates in `adf_publish` branch

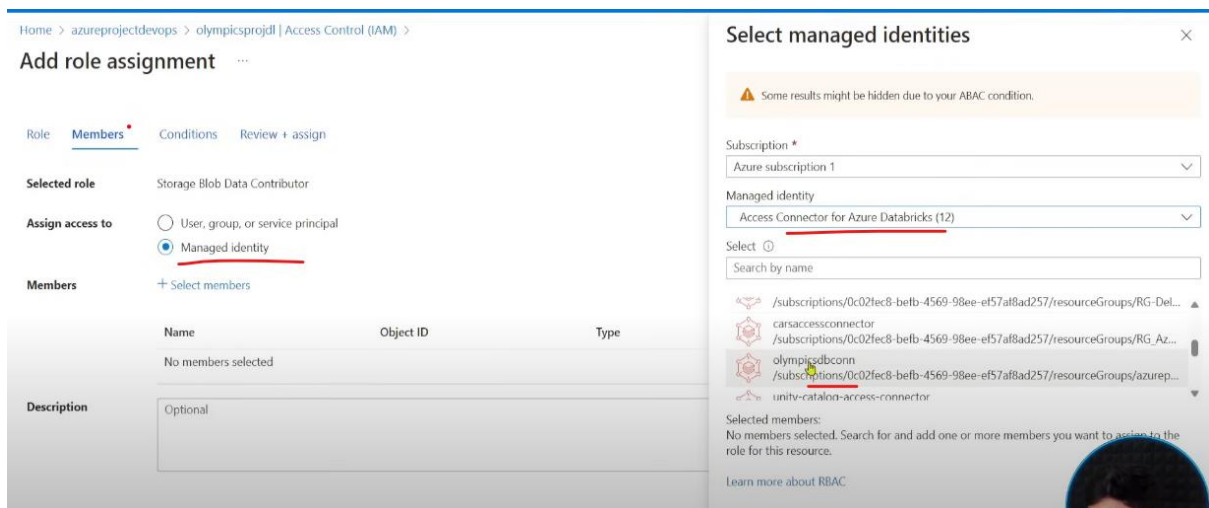


This is the crux of ADF artifacts & also help us in the deployment

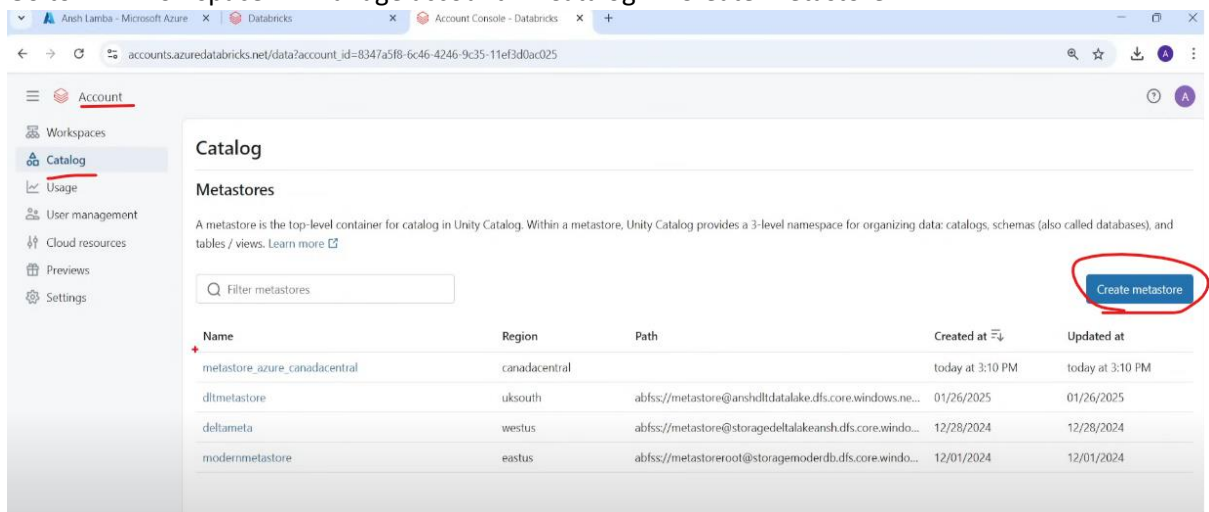
Additional : You can see the changes done on FR with the main branch using Compare option



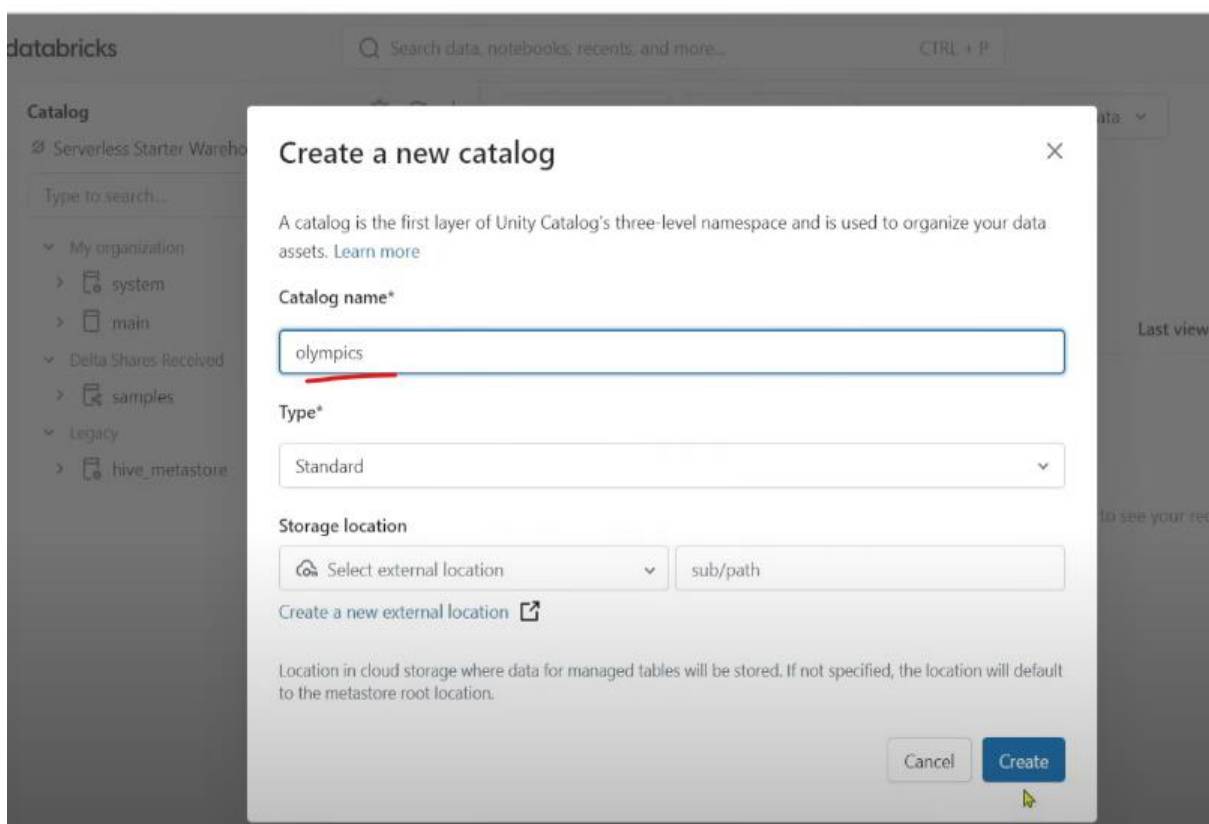
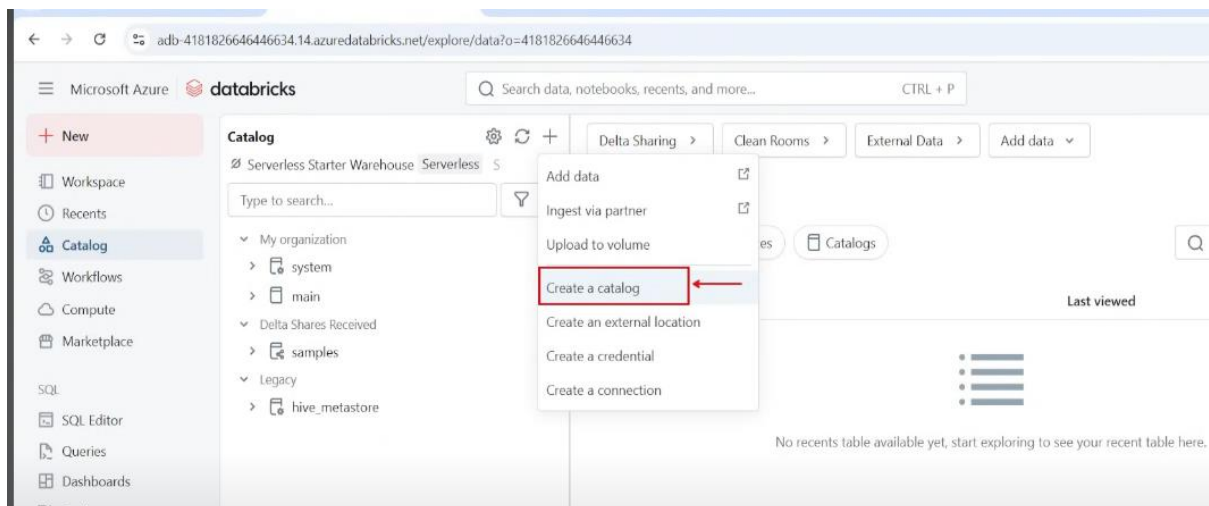
Create ADB workspace → Create a container in ADLS for Unity Metastore (we can store metadata & managed table data for multiple ADB workspaces) → Create Azure Databricks Connector and give SBC access to it



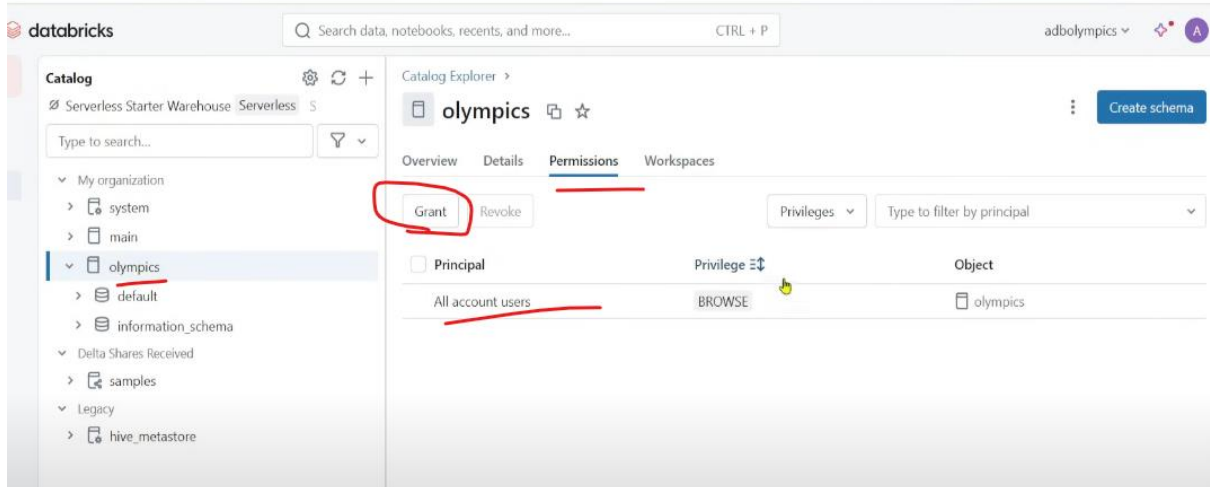
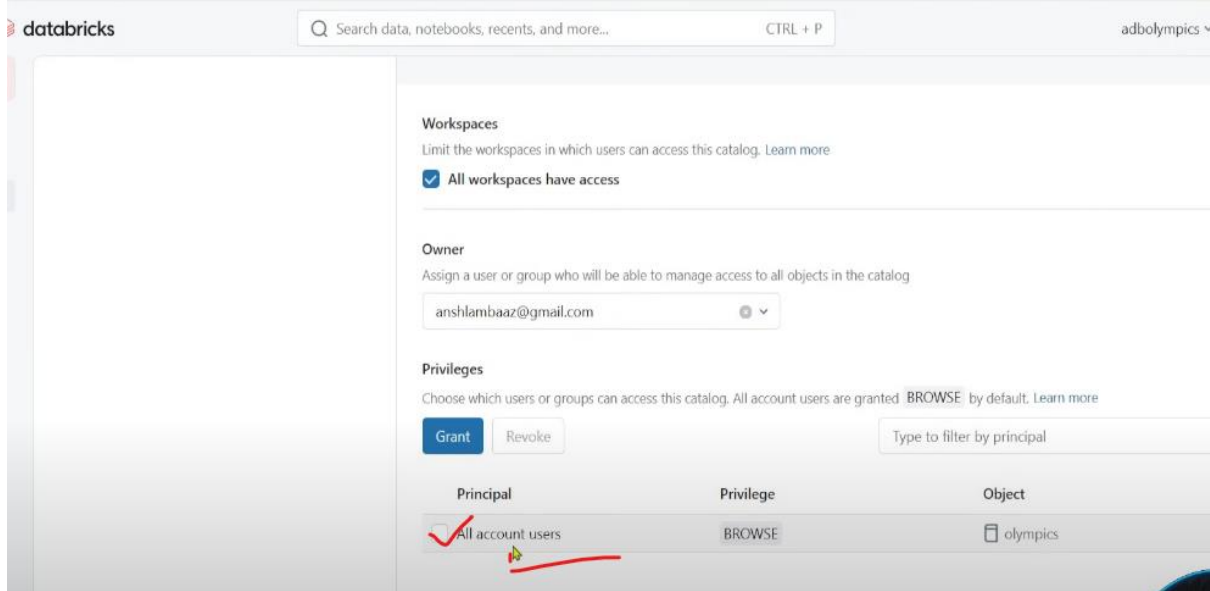
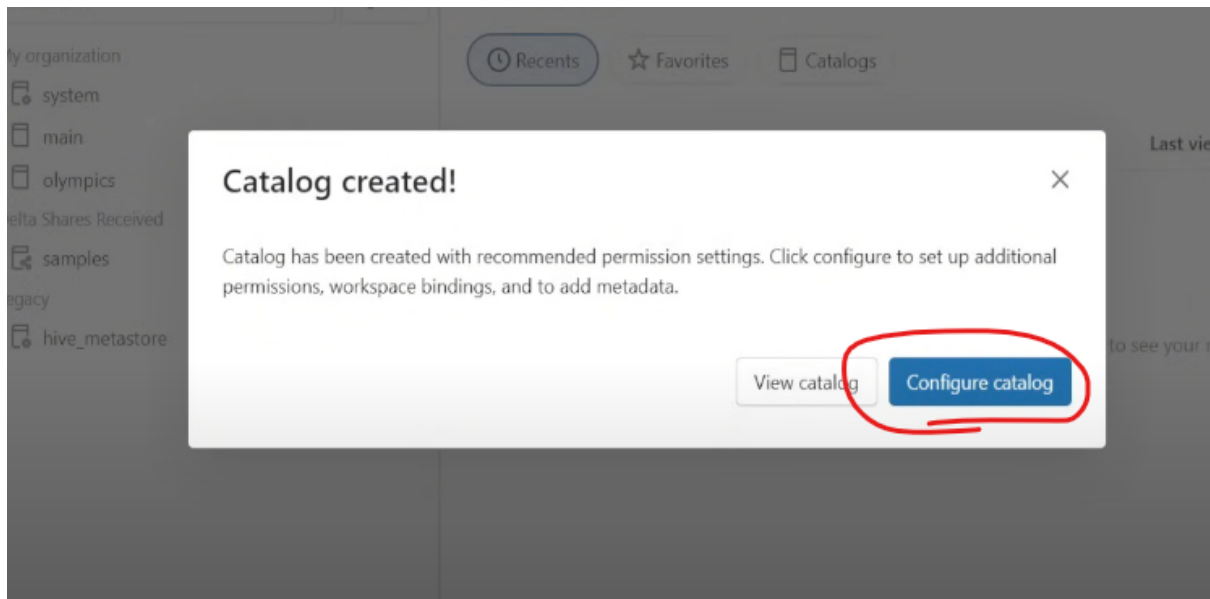
Go to ADB workspace → Manage account → Catalog → Create Metastore



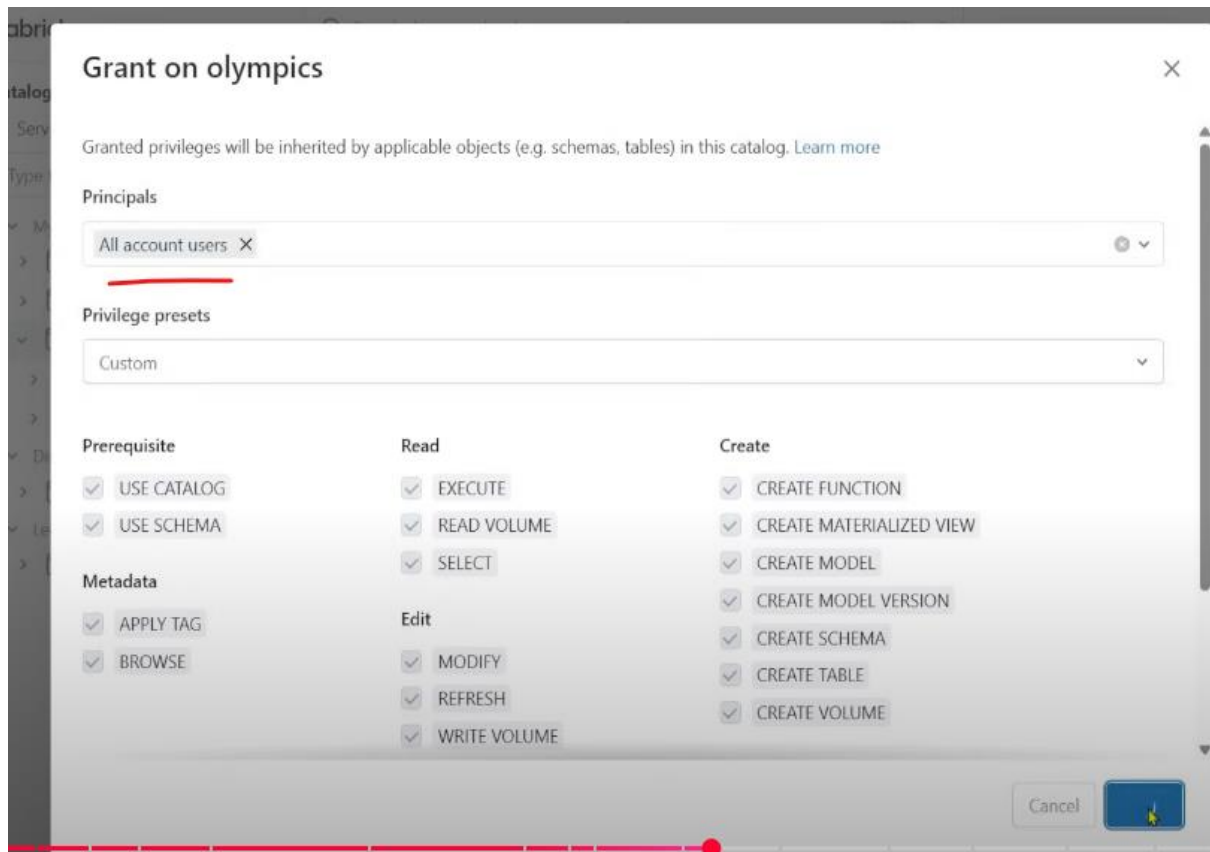
Create a Catalog



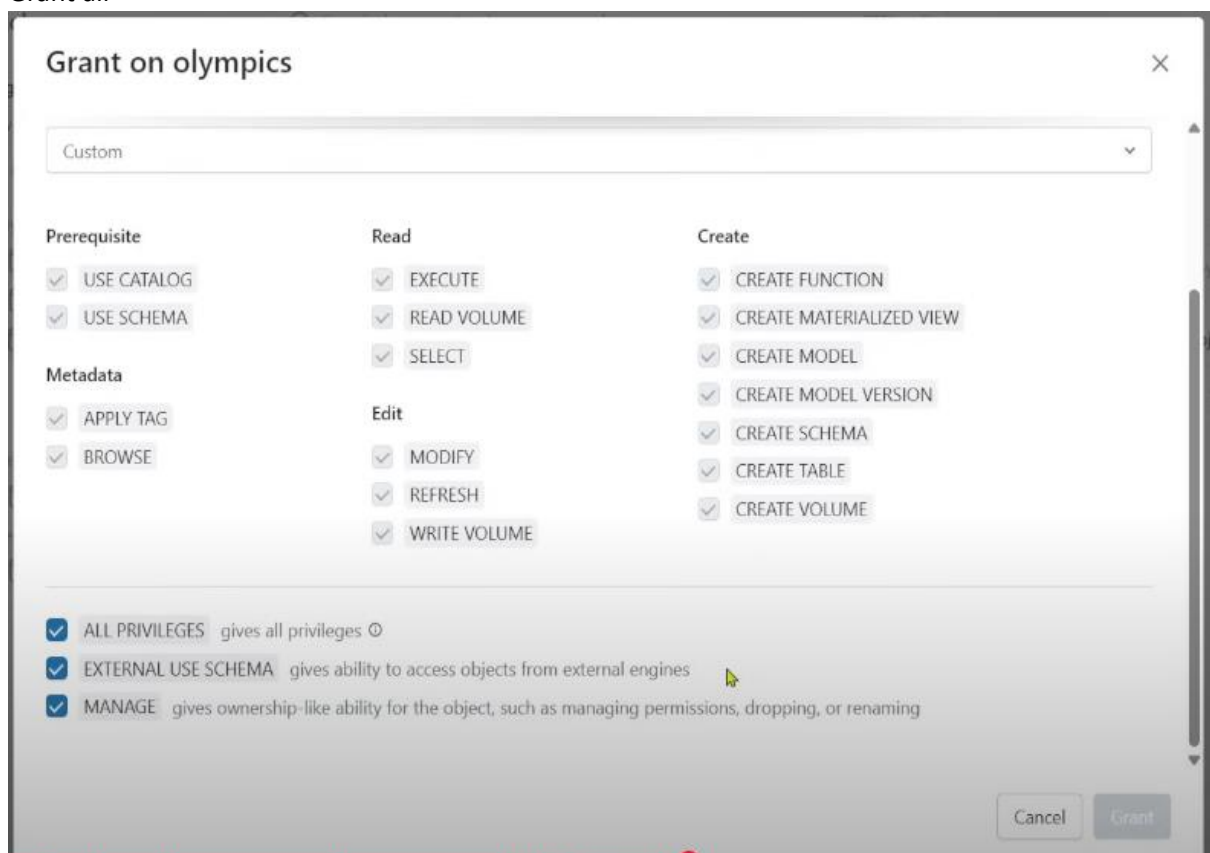
Location is not mandatory as we have given location while creating the Metastore



Click on Grant



Grant all



Search data, notebooks, recents, and more... CTRL + P | adbolympics

Catalog Explorer > olympics

Overview Details Permissions Workspaces

Grant Revoke Privileges Type to filter by principal

Principal	Privilege	Object
All account users	ALL PRIVILEGES	olympics
All account users	BROWSE	olympics
<input type="checkbox"/> All account users	EXTERNAL USE SCHEMA	olympics
All account users	MANAGE	olympics

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P

+ New Workspace Recents Catalog Workflows Compute Marketplace SQL SQL Editor Queries Dashboards Genie

Catalog Serverless Starter Warehouse Serverless S

Type to search... olympics

My organization system main olympics default information\_schema Delta Shares Received samples Legacy hive\_metastore

Catalog Explorer > olympics

Overview Details Permissions Workspaces

Specify which workspaces can have access to this catalog.

☒ All workspaces have access

By default all the workspaces have access to the catalog

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P | adbolympics

+ New Workspace Recents Catalog Workflows Compute Marketplace SQL SQL Editor Queries Dashboards Alerts

Catalog Serverless Starter Warehouse Serverless S

Type to search... olympics

My organization system main olympics default information\_schema Delta Shares Received samples Legacy hive\_metastore

Catalog Explorer > olympics

Overview Details Permissions Workspaces

Specify which workspaces can have access to this catalog.

☐ All workspaces have access

Assign to workspaces Manage Access Level Revoke Filter workspaces...

Workspace name	Workspace id	Access Level	Resource group	Subscription id
adbolympics	4181826646446634	Read & Write	azureprojectdevops	0c02fec8-befb-4569-9...

choose in workspace

Now in order to access bronze, silver and gold containers create External Location  
Go to External Data → Create 3 external locations for bronze, silver and gold containers



**External location name\***

**URL\*** ⓘ

Enter the bucket path that you want to use as the external location. Note: This must be an ADLS Gen2 storage account with a hierarchical namespace

 Copy from DBFS ▾

**Storage credential\*** [Learn more](#) ⓘ

Provide a storage credential capable of accessing the URL

**Credential Type\***

**Access connector ID** [Learn more](#) ⓘ

**User assigned managed identity ID (optional)**

Storage credential is the fancy name for Databricks Connector

Now lets create a schema

Microsoft Azure **databricks**  CTRL + P

**Silver\_Entities** Python ▾ ☆

File Edit View Run Help Last edit was 4 minutes ago

**Catalog**

Type to search...

For you All

- My organization
  - system
  - main
  - olympics**
  - default
  - information\_schema
- Delta Shares Received
  - samples
- Legacy
  - hive\_metastore

1

start typing or generate with AI (Ctrl + I)...

[Shift+Enter] to run and move to next cell  
[Ctrl+Shift+P] to open the command palette  
[Esc H] to see all keyboard shortcuts

Open in Catalog Explorer

Copy catalog path

Add to favorites

Microsoft Azure databricks

Search data, notebooks, recents, and more... CTRL + P

adbolympics

**Catalog**

Serverless Starter Warehouse Serverless

Type to search...

- My organization
  - system
  - main
  - olympics
    - default
    - information\_schema
- Delta Shares Received
  - samples
- Legacy
  - hive\_metastore

**Catalog Explorer**

olympics

Overview Details Permissions Workspaces

Description

AI generate Add

Filter schemas 2 schemas

Name	Owner	Created at
default	anshlambaaz@gmail.com	Mar 22, 2025, 03:29 PM
information_schema	System user	Mar 22, 2025, 03:29 PM

About this catalog

Owner Ansh Lamba

Tags

Add tags

Create schema

databricks

Search data, notebooks, recents, and more... CTRL + P

## Create a new schema

A schema is the second layer of Unity Catalog's three-level namespace and organizes tables and views. [Learn more](#)

Schema name\*

silver

Storage location

Select external location sub/path

Create a new external location

Location in cloud storage where data for managed tables will be stored. If the storage location is not specified, it will default to catalog location first and then metastore root location if catalog location is not specified as well.

Comment

Cancel

Microsoft Azure databricks

Search data, notebooks, recents, and more... CTRL + P

Silver\_Entities Python

File Edit View Run Help Last edit was 5 minutes ago

Run all

**Catalog**

Type to search...

For you All

- My organization
  - system
  - main
  - olympics
    - default
    - information\_schema
    - silver
- Delta Shares Received
  - samples
- Legacy
  - hive\_metastore

silver olympics

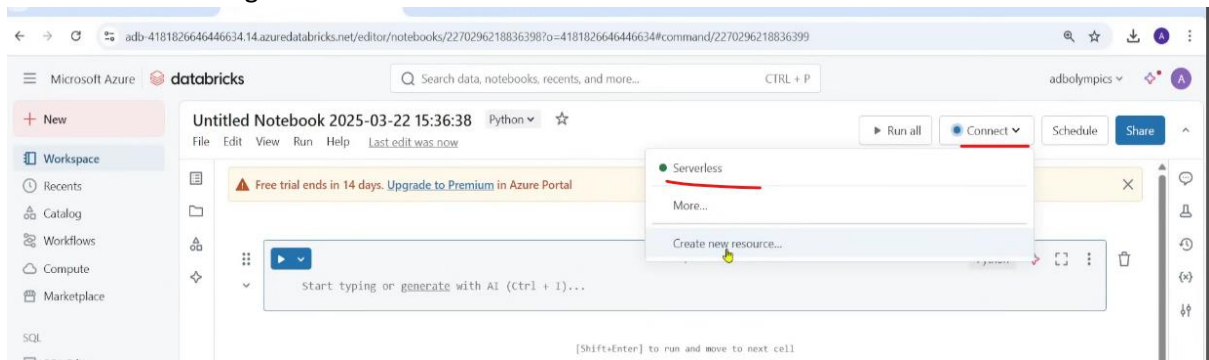
Owner Ansh Lamba

Updated 28 seconds ago

Start typing or generate with AI (Ctrl + I)...

[Shift+Enter] to run and move to next cell  
[Ctrl+Shift+P] to open the command palette  
[Esc H] to see all keyboard shortcuts

Now lets start coding



Databricks Serverless is always on, its like a Starter Pool where one machine is always on