

DXC - RealTime Projects

Name: Lakshmi Naidu Thatikonda

Reg No: dxcab1226

Project1 Name: Smart Vehicles

Date: 10/06/2022

Project 1 : Connected Vehicles

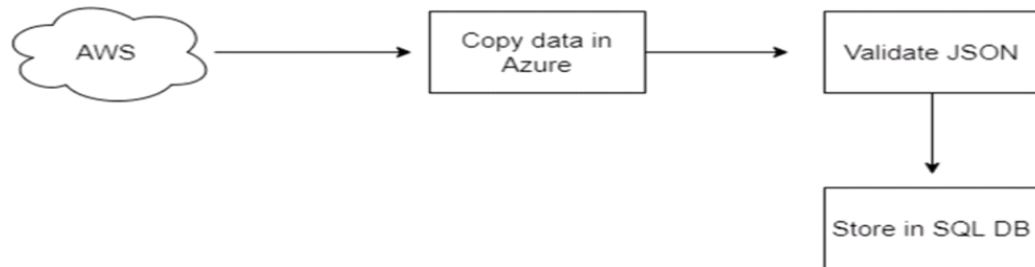
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



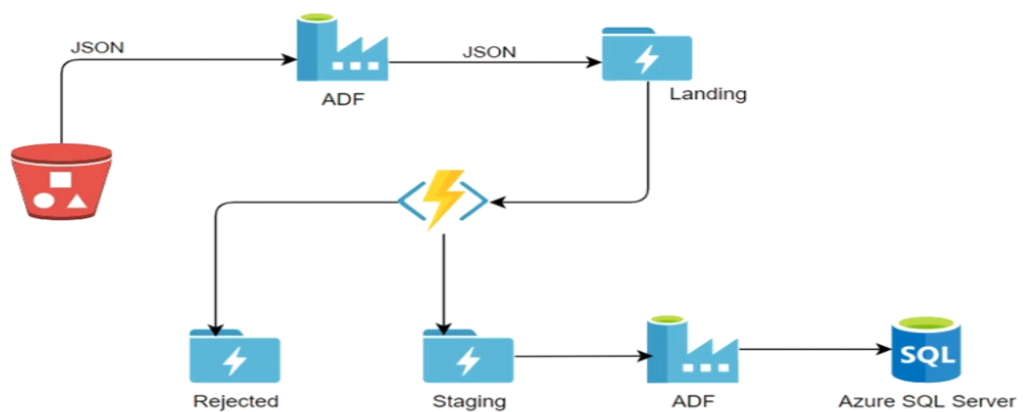
Project 1 : Connected Vehicles

- Vehicle has third party IoT device which will send the telemetry data (in JSON format) over the AWS cloud.
- You need to move data from third party AWS to General Motors Azure cloud.
- You need to validate the JSON sometime it could be incomplete or wrong JSON which need to be rejected.
- Once JSON got validated this data would be stored in the SQL database which will be further utilized by data science team.

Project 1 : Connected Vehicles



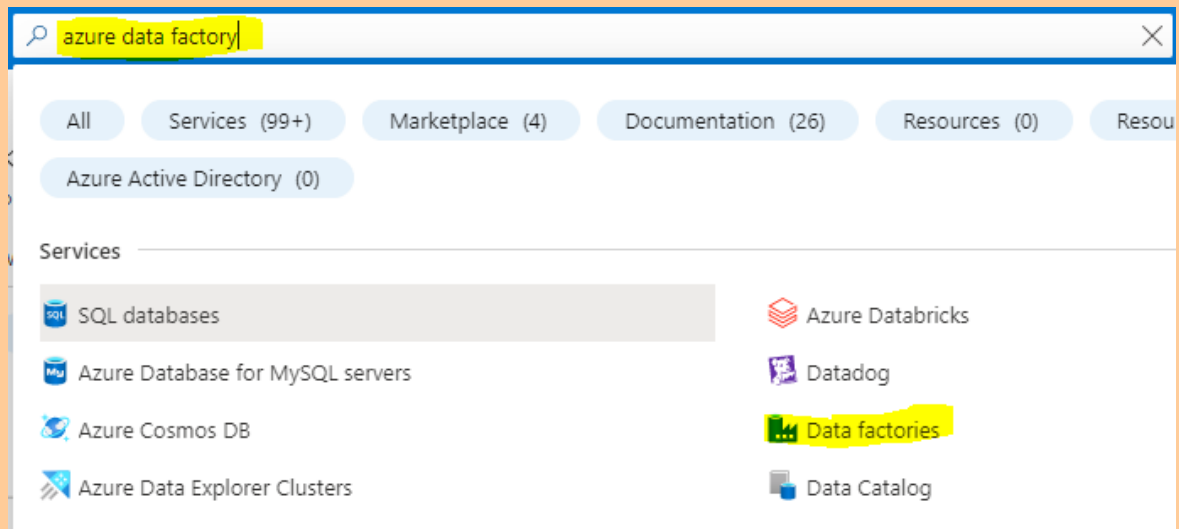
Project 1 : Connected Vehicles



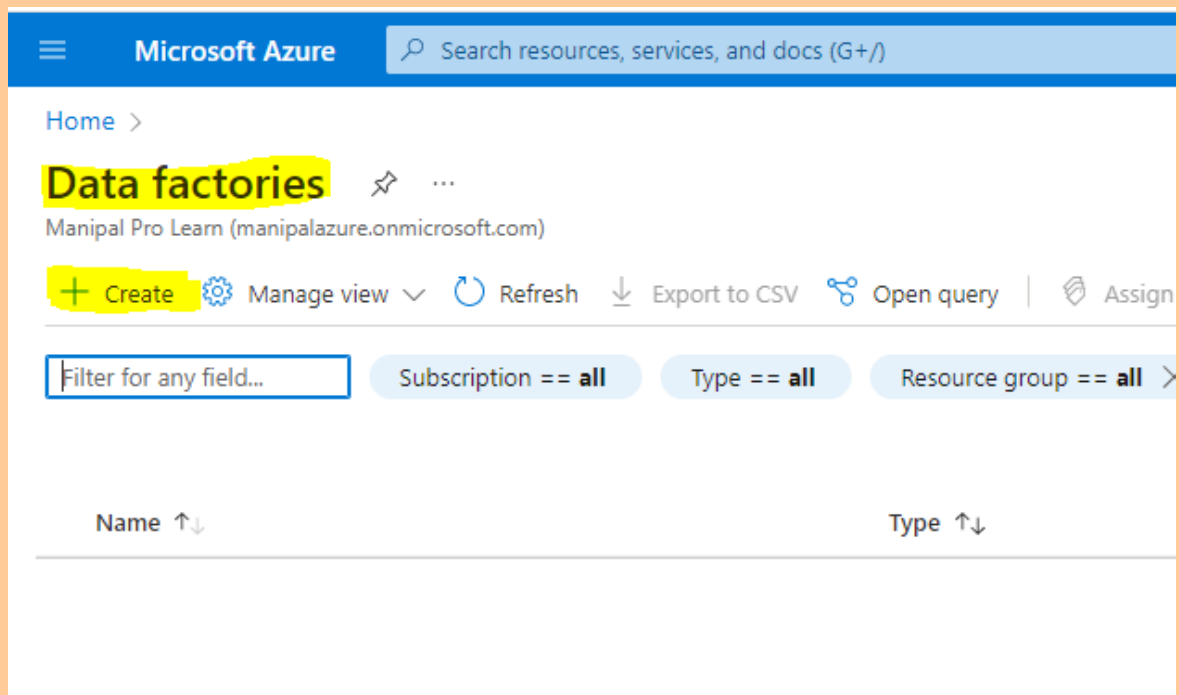
Architecture Diagram for Connected Vehicle Project

Practical Lab: Create **Azure Data Factory** Account For Data pipelines

Step 1: Go to search bar click "Data Factory" and click on it.



Step 2: Click on "Create".



Step 3: Fill the details like "Resource Group" and "Data Factory Name"

Microsoft Azure

Search resources, services, and docs (G+/I)

[Home](#) > [Data factories](#) >

Create Data Factory

Basics

Git configuration

Networking

Advanced

Tags

Review + create

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-DXC262AB12Lab

Resource group * ⓘ

dxccorg235

Create new

Instance details

Name * ⓘ

dxccdemodatafact235

Region * ⓘ

East US

Version * ⓘ

V2 (Recommended)

Review + create

< Previous

Next : Git configuration >

Step 4: No need to set all the highlighted portions. Click on "review and Create".

Microsoft Azure

Search resources, services, and docs (G+)

[Home](#) > [Data factories](#) >

Create Data Factory

✓ Validation Passed

Basics

Git configuration

Networking

Advanced

Tags

Review + create

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; and (b) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

Basics

Subscription	Azure-DXC262AB12Lab
Resource group	dxccorg235
Name	dxccdemodatafact235
Region	East US

Create

< Previous

Next >

[Download a template for automation](#)

Step 5: And the Data Factory is being Deployed.

Microsoft.DataFactory-20220610160619 | Overview

Deployment

Search (Ctrl+J)

Delete Cancel Redeploy Refresh

Overview

Inputs

Outputs

Template

✓ We'd love your feedback! →

Deployment is in progress

Deployment name: Microsoft.DataFactory-20220610160619

Subscription: Azure-DXC262AB12Lab

Resource group: dxccorg235

Start time: 6/10/2022, 4:16:35 PM

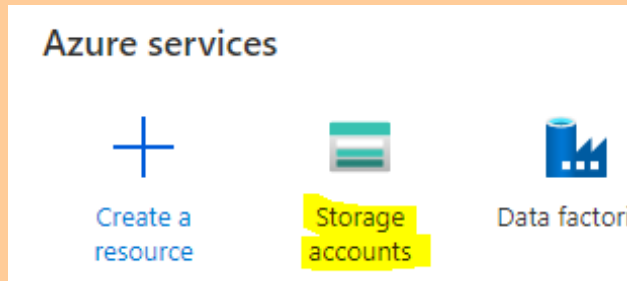
Correlation ID: d06f697a-c7cb-4852-9ad7-8ed1dfdfef3da

Deployment details (Download)

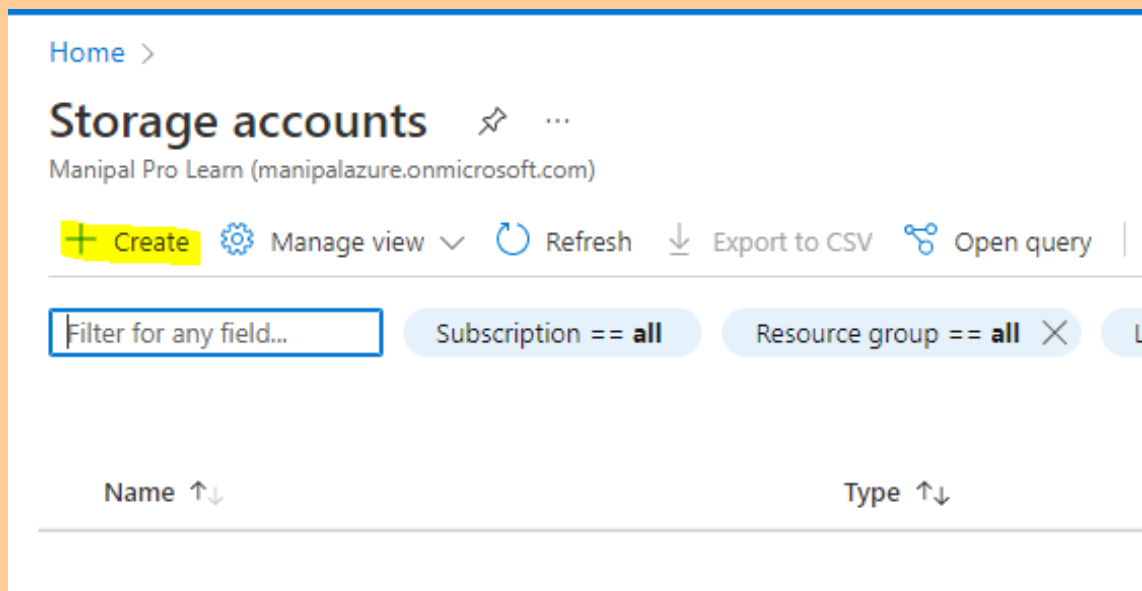
Resource	Type	Status	Operation details
No results.			

Practical Lab: Create **ADF Pipeline** End to end pipeline with triggers enabled

Step 1: Click on "Azure Data Storage" on home screen of Azure.



Step 2: Click on "Create".



Step 3: On the next screen Give all the details and "Click: Adavanced".

Microsoft Azure

Search resources, services, and docs (G+)

[Home](#) > [Storage accounts](#) >

Create a storage account

Basics

Advanced

Networking

Data protection

Encryption

Tags

Review + create

Subscription *

Azure-DXC262AB12Lab

Resource group *

dxccorg235

[Create new](#)

Instance details

If you need to create a legacy storage account type, please click [here](#).

Storage account name ⓘ *

dxcdemdstorage235

Region ⓘ *


(US) East US

Review + create

< Previous

Next : Advanced >


Step 4: Go on till the "Review and create" tab by filling the details required and click on "Create".

 Microsoft Azure

Search resources, services, and docs (G+/)

Home > Storage accounts >

Create a storage account ...

 Validation passed

Basics

Advanced

Networking

Data protection

Encryption

Tags

Review + create

Deployment model

Resource manager

Performance

Standard

Replication

Read-access geo-redundant storage (RA-GRS)

Advanced

Secure transfer

Enabled

Allow storage account key access

Enabled

Allow cross-tenant replication

Enabled

Default to Azure Active Directory authorization in the Azure portal

Disabled

Blob public access

Enabled

Minimum TLS version

Version 1.2

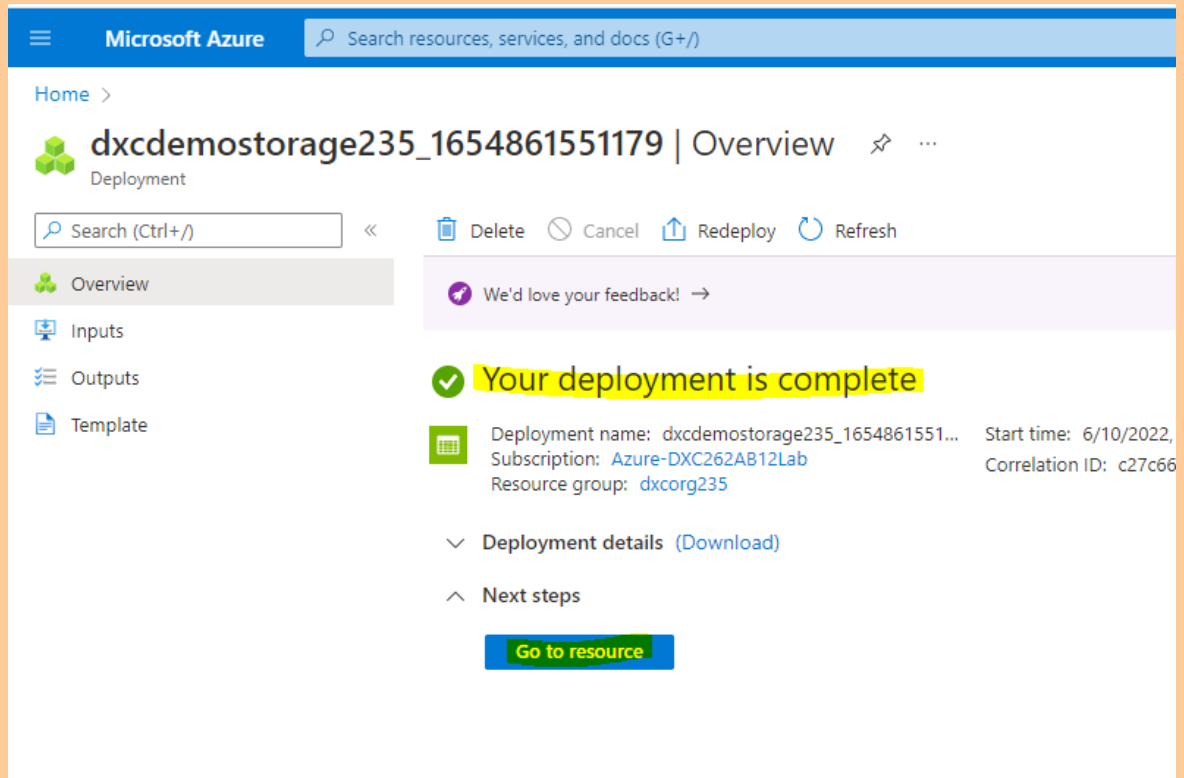
Create

< Previous

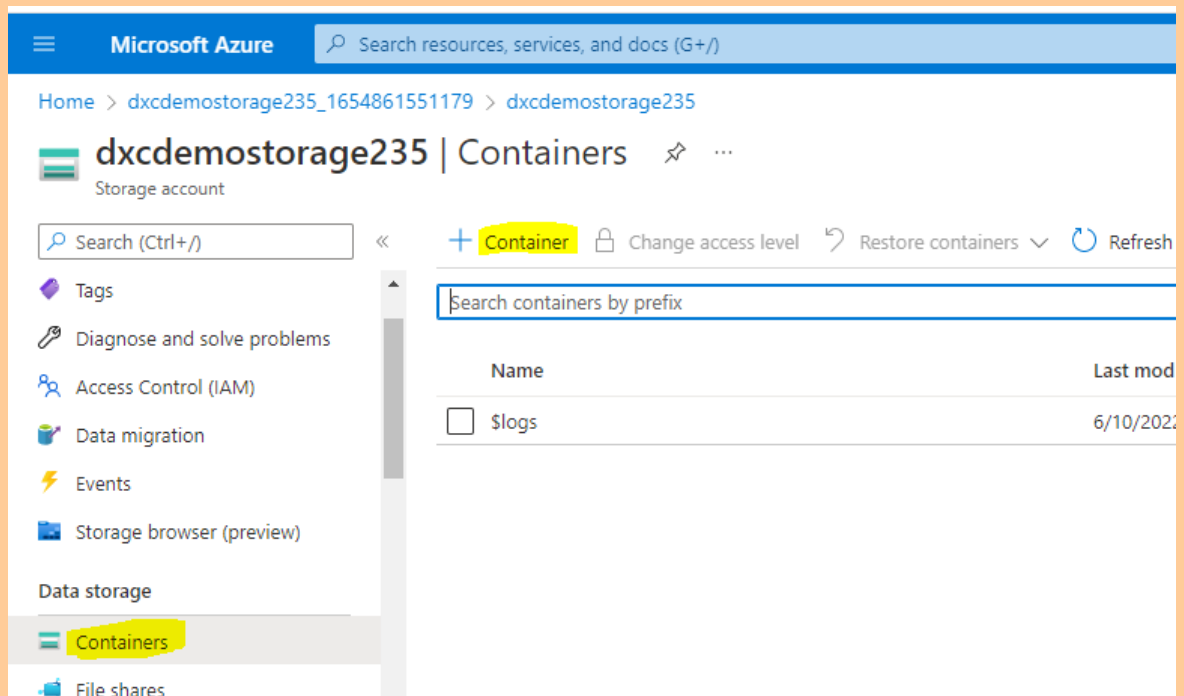
Next >

Download a template for automation

Step 5: Once Deployment is done, Click on "Go to resource".



Step 6: After goint to resource click on "Container->+Container".



Step 7: Then give the name for blob as "sourceblob" and click on "Create".

Refresh | Delete

Last modified	
6/10/2022, 5:16:33 PM	

New container

Name *
sourceblob

Public access level ⓘ
Blob (anonymous read access for blobs only)

⚠ Blobs within the container can be read by anonymous users but container data is not available. Anonymous users cannot enumerate the blobs within the container.

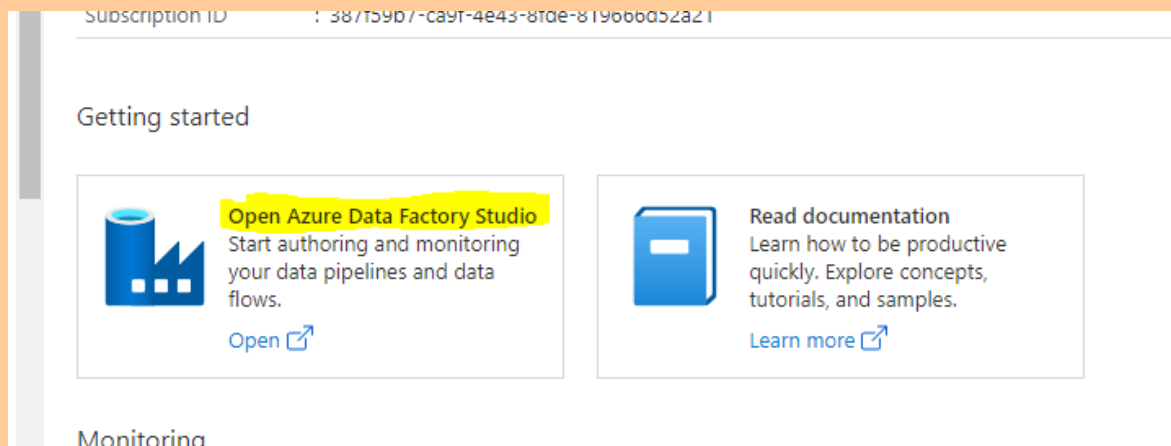
Advanced

Create Discard

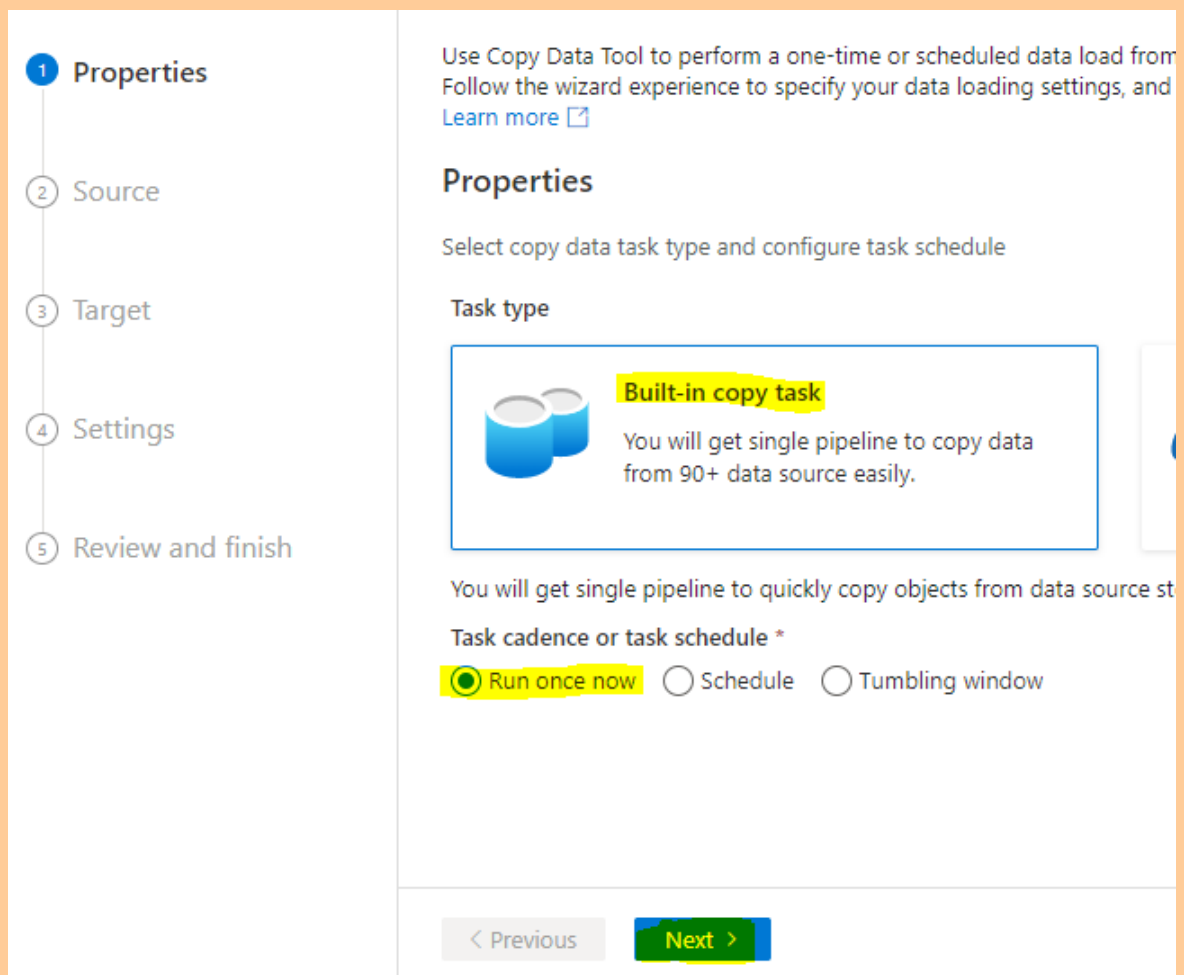
Step 8: Similarly create the destination blob as "destinationblob". Both blobs are created.

+ Container	Change access level	Restore containers	Refresh
Search containers by prefix			
Name		Last modified	
<input type="checkbox"/> \$logs		6/10/2022, 5:16:33 PM	
<input type="checkbox"/> destinationblob		6/10/2022, 5:16:33 PM	
<input type="checkbox"/> sourceblob		6/10/2022, 5:16:33 PM	

Step 9: Go to Data factory and click on "Azure Data Factory Studio" and then click on "Ingest".



Step 10: Click on "Built-in copy task". And click on "Next".



Step 11: Click on the source as "Azure Blob Storage".

✓ Properties

2 Source

• Dataset

○ Configuration

3 Target


4 Settings

5 Review and finish

Source data store

Specify the source data store for the copy task. You can use an existing data

Source type

 Azure Blob Storage

Connection *



Select...

< Previous

Next >


Step 12: Click on "New connection" and create "Connection".

New linked service

 Azure Blob Storage [Learn more](#) 

AzureBlobStorage1

Description


Connect via integration runtime * 

AutoResolveIntegrationRuntime


Authentication type

Account key

Connection string Azure Key Vault

Account selection method 

☒ From Azure subscription ☐ Enter manually

Azure subscription 

Azure-DXC262AB12Lab (387f59b7-ca9f-4e43-8fde-

Storage account name *

dxcdemostorage235

Create Cancel

Step 13: Click on "Browse" and "Next".

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source type

Azure Blob Storage

Connection *

AzureBlobStorage1

Edit

+ New connection

File or folder *

If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.

Browse

Options

☐ Binary copy ⓘ

☒ Recursively ⓘ

☐ Enable partition discovery ⓘ

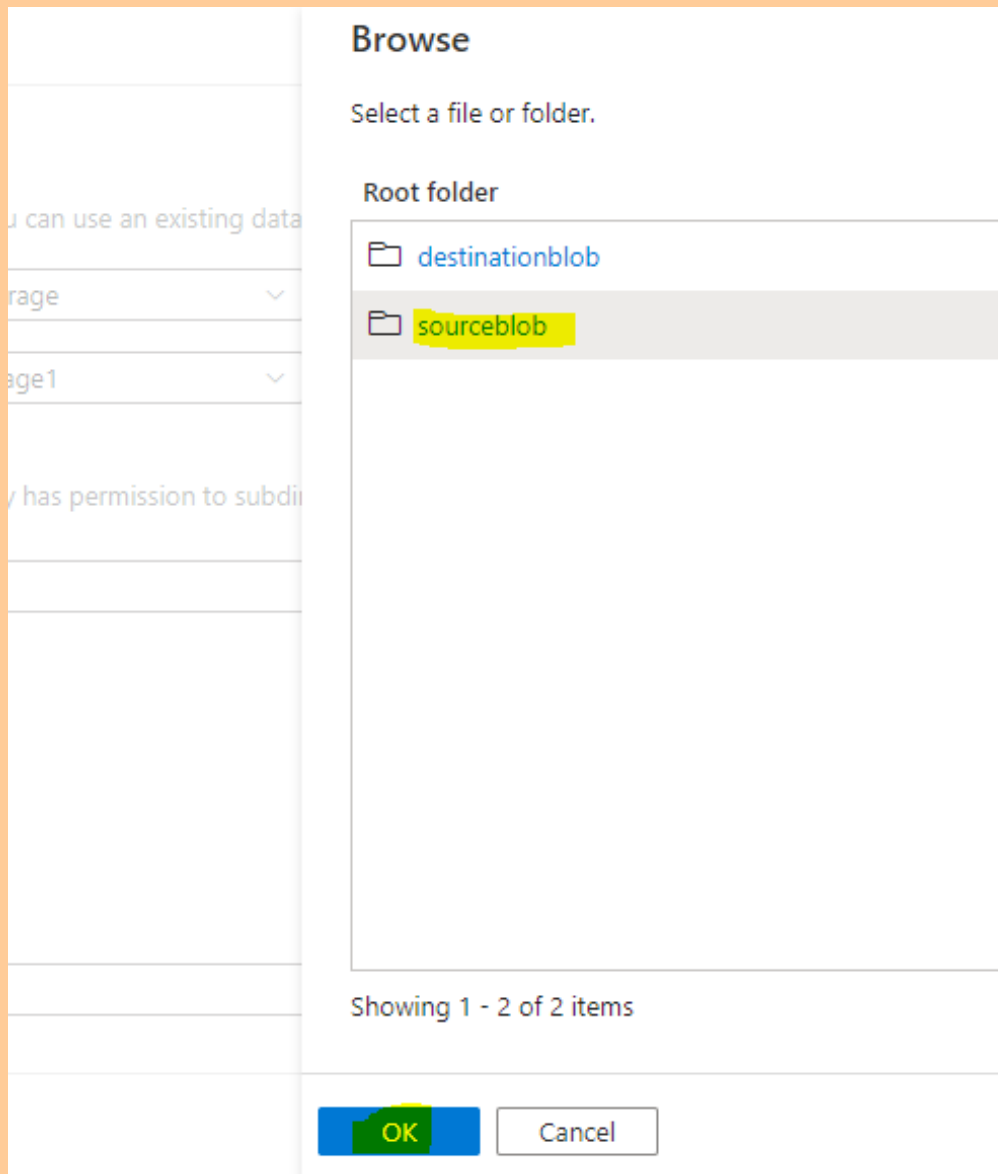
Max concurrent connections ⓘ

Please see task model for details

< Previous

Next >

Step 14: Click on sourceblob and "ok".



Step 14: In Configuration follow the steps shown below.

Microsoft Azure | dxcdemodatafact235

Copy Data tool

Home

Pencil

Source

Dataset

Configuration

3 Target

4 Settings

5 Review and finish

File format settings

File format ⓘ
DelimitedText

Column delimiter
Comma (,)
☐ Edit

Row delimiter
Default (\r\n, or \r\n)
☐ Edit

☒ First row as header ⓘ

> Advanced

Compression type
None

Additional columns ⓘ
[+ New](#)

< Previous Next >

Step 15: For target data set follow the steps shown.

Microsoft Azure | dxcdemodatafact235

Copy Data tool

Home

Pencil

Source

3 Target

Dataset

Configuration

4 Settings

5 Review and finish

Destination data store

Specify the destination data store for the copy task. You can use an existing

Target type
Azure Blob Storage

Connection *
Select...

< Previous Next >

New linked service

Azure Blob Storage [Learn more](#)

AzureBlobStorage2

Description

Connect via integration runtime * ⓘ
AutoResolveIntegrationRuntime

Authentication type
Account key

Connection string Azure Key Vault

Account selection method ⓘ
☒ From Azure subscription ☐ Enter manually

Azure subscription ⓘ
Azure-DXC262AB12Lab (987f59b7-ca9f-4e43-8fde-819660d52a21)

Storage account name *
dxcdemostorage235

Create Cancel

Step 16: Select destinationblob for the target.

Copy Data tool

Properties ✓
Source ✓
Target 3
Dataset
Configuration
Settings 4
Review and finish 5

Destination data store
Specify the destination data store for the copy task. You can use an existing connection.

Target type: Azure Blob Storage
Connection *: AzureBlobStorage2

Folder path *
If the identity you use to access the data store only has permission to subdirectory, specify the path to browse.

File name

Copy behavior ⓘ
None

Max concurrent connections ⓘ

Block size (MB) ⓘ

< Previous Next >

Browse
Select a file or folder.

Root folder

- destinationblob
- sourceblob

Showing 1 - 2 of 2 items

OK Cancel

Step 17: Follow this for Return Configuration.

Copy Data tool

Properties ✓
Source ✓
Target 3
Dataset
Configuration
Settings 4
Review and finish 5

File format settings

File format ⓘ
DelimitedText

Column delimiter
Comma (,)

☐ Edit

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

☐ Edit

☐ Add header to file ⓘ

> Advanced

Compression type
None

Max rows per file

< Previous Next >

Step 18: In the settings click on "next".

Settings

Enter name and description for the copy data task, more options for data movement

Task name *

CopyPipeline_yhq

Task description

Data consistency verification

☐

Fault tolerance ⓘ

Enable logging ⓘ

☐

Enable staging ⓘ

☐

> Advanced

< Previous

Next >

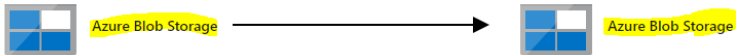
Step 19: Review and click on "Ok".

Copy Data tool

- Properties
- Source
- Target
- Settings
- Review and finish**
- Review
- Deployment

Summary

You are running pipeline to copy data from Azure Blob Storage to Azure Blob Storage.



Source: Azure Blob Storage → Target: Azure Blob Storage

Properties

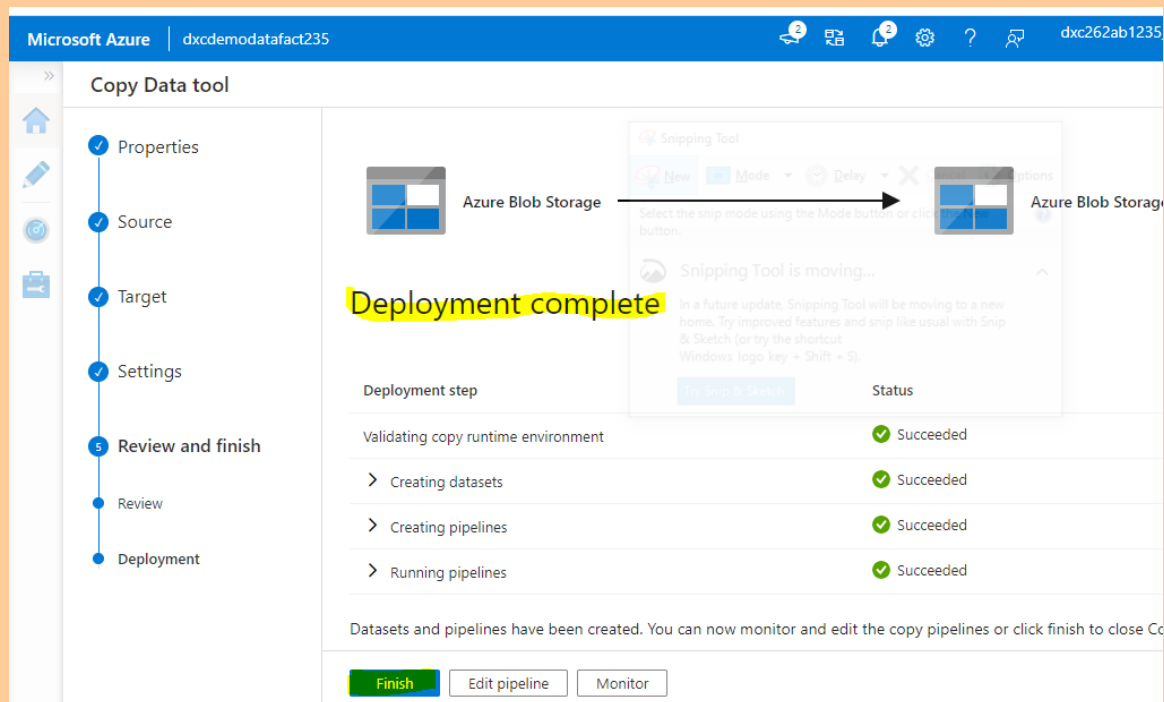
Task name	CopyPipeline_yhq	Edit
Task description		
Source		
Connection name	AzureBlobStorage1	
Dataset name	SourceDataset_yhq	
Column delimiter	,	
Escape character	\	
Quote char	"	

[Edit](#)

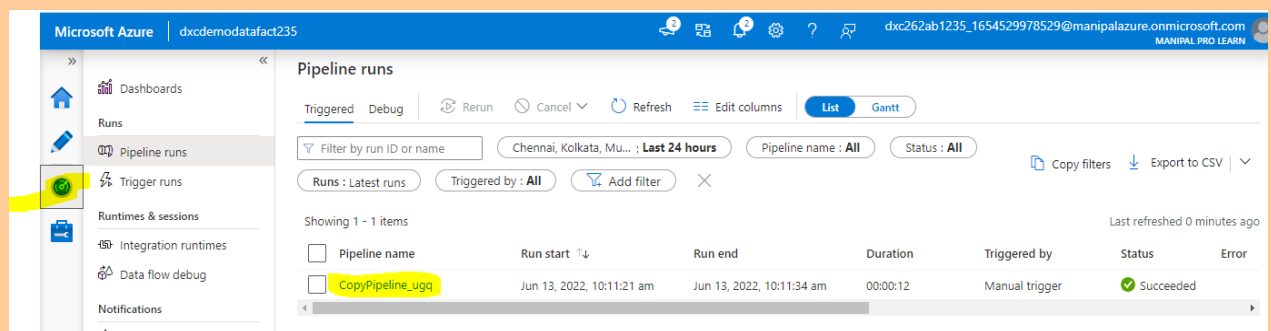
< Previous

Next >

Step 20: Pipeline is created click on "Finish".

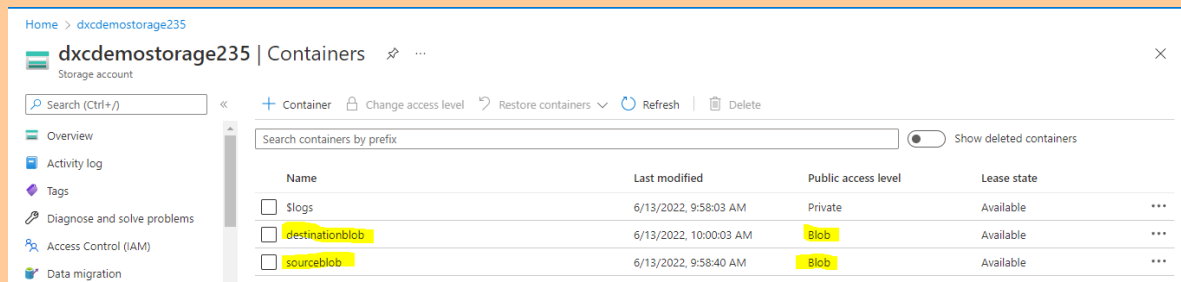


Step 21: Now, enable the pipeline.

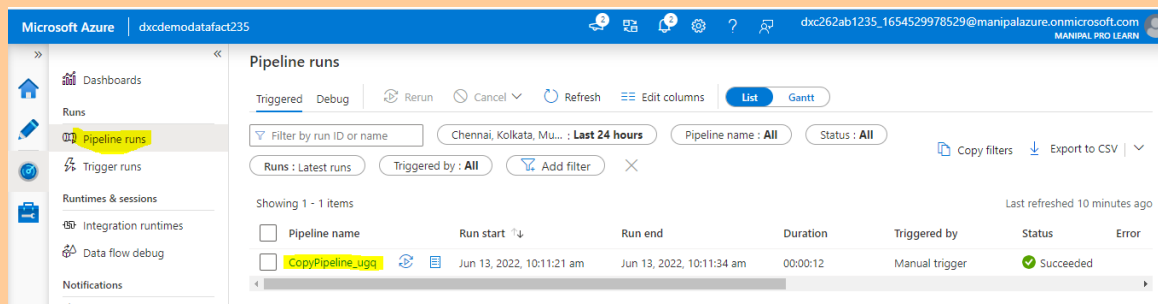


Practical Lab: Create Azure blob trigger logic

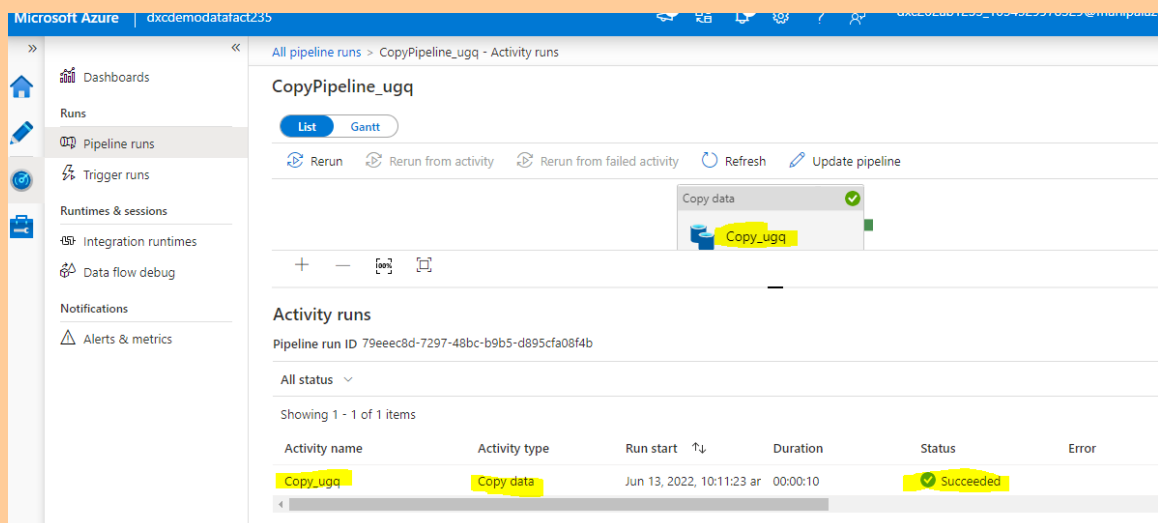
Step 1: Create a Blob Storage Container as shown.



Step 2: Click on "Pipeline Runs" and then click on the the pipeline with "Blob trigger".

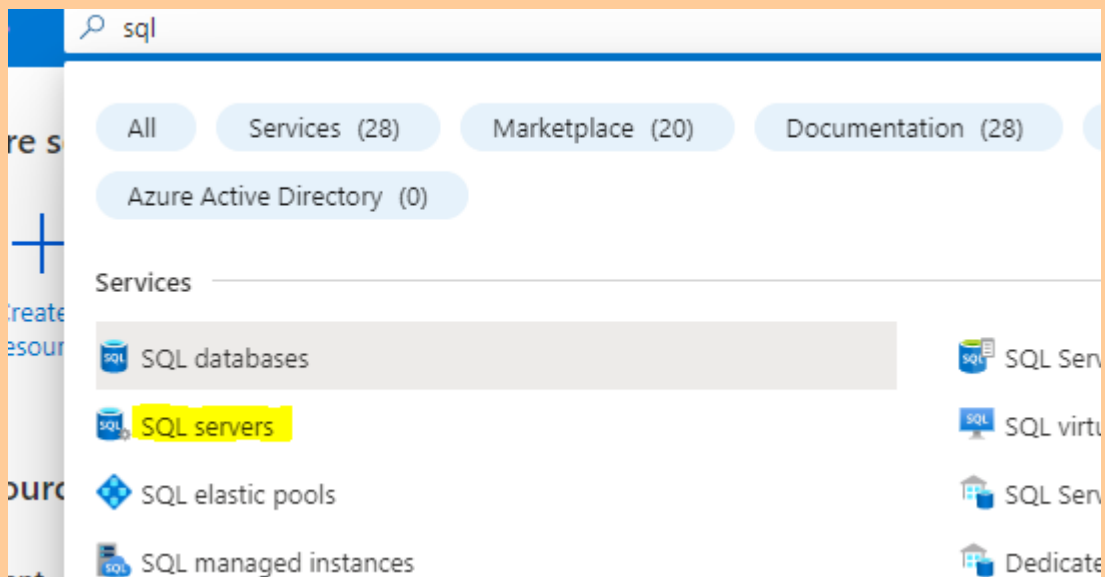


Step 3: And the Blob trigger is ran.

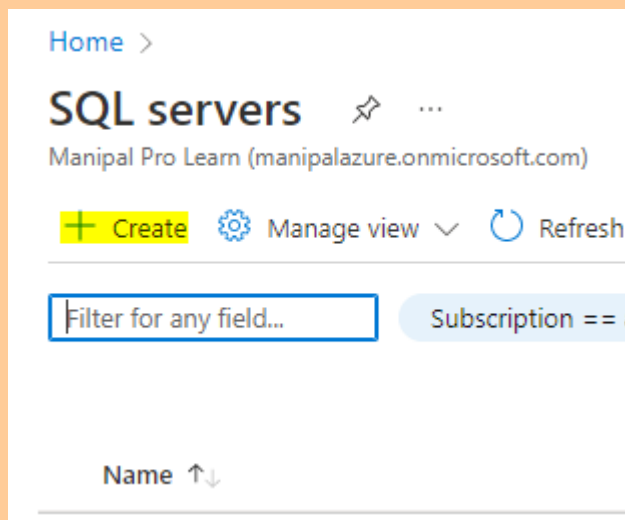


Practical Lab: Create Azure SQL Server and Database

Step 1: Click on "Sql Server" in Azure open page.



Step 2: Click on "Create".



Step 3: Fill all the details and click on "Next".

Home > SQL servers >

Create SQL Database Server

Microsoft

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

Subscription * ⓘ Azure-DXC262AB12Lab

Resource group * ⓘ dxcorg235

[Create new](#)

Server details

Enter required settings for this server, including providing a name and location.

Server name * dxcdemoqlserver

Location * (US) East US

Authentication

Select your preferred authentication methods for accessing this server. Create a server admin login and password to access your server with SQL authentication; select only Azure AD authentication [Learn more](#) or using an existing Azure

- ✓ Server name should not contain reserved words.
- ✓ The specified server name is available.
- ✓ Your server name can't start or end with hyphens '-', nor '-' in third and fourth places of the name.

[Review + create](#) [Next: Networking >](#)

Step 4: Now click on "Create".

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

Home > SQL servers >

Create SQL Database Server

Microsoft

Basics Networking Additional settings Tags **Review + create**

Product details

SQL Database Server
by Microsoft
[Terms of use](#) | [Privacy policy](#)

Estimated cost per month
No additional charges

Terms

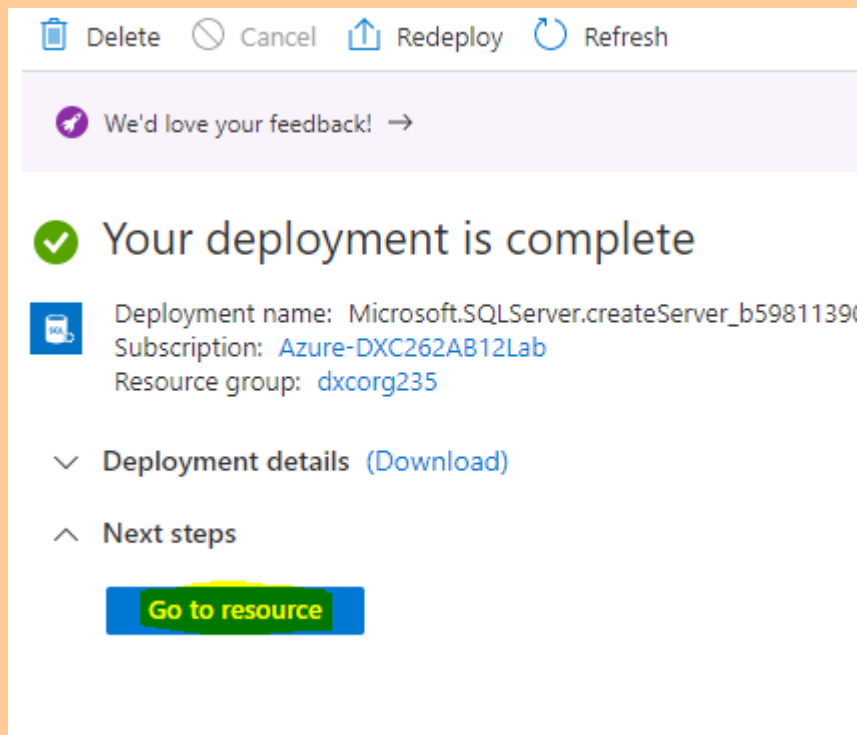
By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed below; (b) agree to the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see [Azure Marketplace Terms of Use](#).

Basics

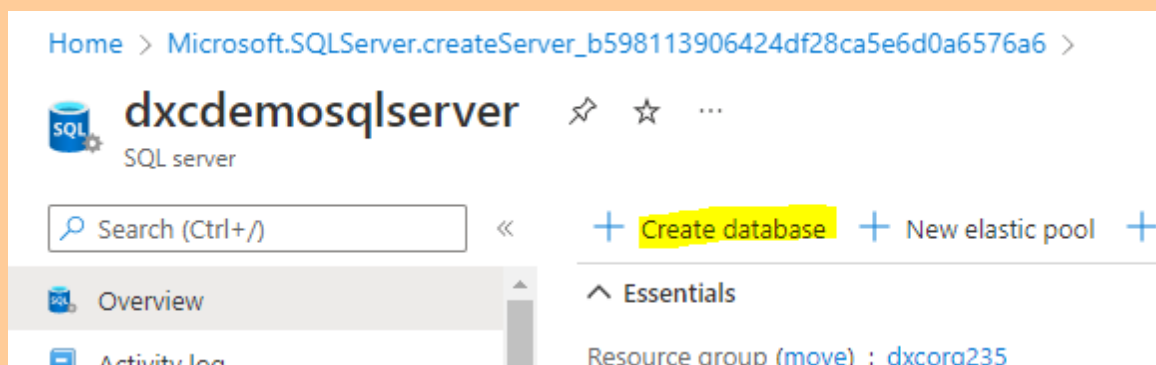
Subscription	Azure-DXC262AB12Lab
Resource group	dxcorg235
Server name	dxcdemoqlserver

[Create](#) [< Previous](#) [Download a template for automation](#)

Step 5: After clicking on "Deployment" will be done. Now click on "Go to Resource".



Step 6: Now after going to the resource, Now click on "Create Database".



Step 7: Fill all the data and click on "Next".

Home > Microsoft.SqlServer.createServer_b598113906424df28ca5e6d0a6576a6 > dxcdemosqlserver >

Create SQL Database ...

Microsoft

Subscription ⓘ Azure-DXC262AB12Lab

Resource group ⓘ dxccorg235

Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name * dxcdemodb ✓

Server ⓘ Select a server

Want to use SQL elastic pool? ⓘ ☐ Yes ☒ No

Compute + storage * ⓘ Please select a server first.
[Configure database](#)

[Review + create](#) [Next : Networking >](#)

Step 8: And now click on "Create".

Microsoft Azure

Search resources, services, and docs (G+)

[Home](#) >

Create SQL Database

Microsoft

[Terms](#)

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my confidential transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see [Azure Marketplace](#).

[Basics](#)

Subscription	Azure-DXC262AB12Lab
Resource group	dxccorg235
Region	eastus
Database name	dxcdemodb
Server	dxcdemosqlserver
Compute + storage	General Purpose: Gen5, 2 vCores, 32 GB storage, zone redundant disabled
Backup storage redundancy	Geo-redundant backup storage

[Networking](#)


Allow Azure services and resources to	No
---------------------------------------	----

Create

< Previous

[Download a template for automation](#)

Step 9: Now the deployment is done.

 **Microsoft.SQLDatabase.newDatabaseExistingServer_4e40ee9fac93**
Deployment


<< Delete Cancel Redeploy Refresh


Overview


Inputs


Outputs


Template

 We'd love your feedback! →

 **Your deployment is complete**

 Deployment name: Microsoft.SQLDatabase.newDatabaseExistingSe... Start
Subscription: [Azure-DXC262AB12Lab](#) Corre
Resource group: [dxccorg235](#)

 **Deployment details** [\(Download\)](#)

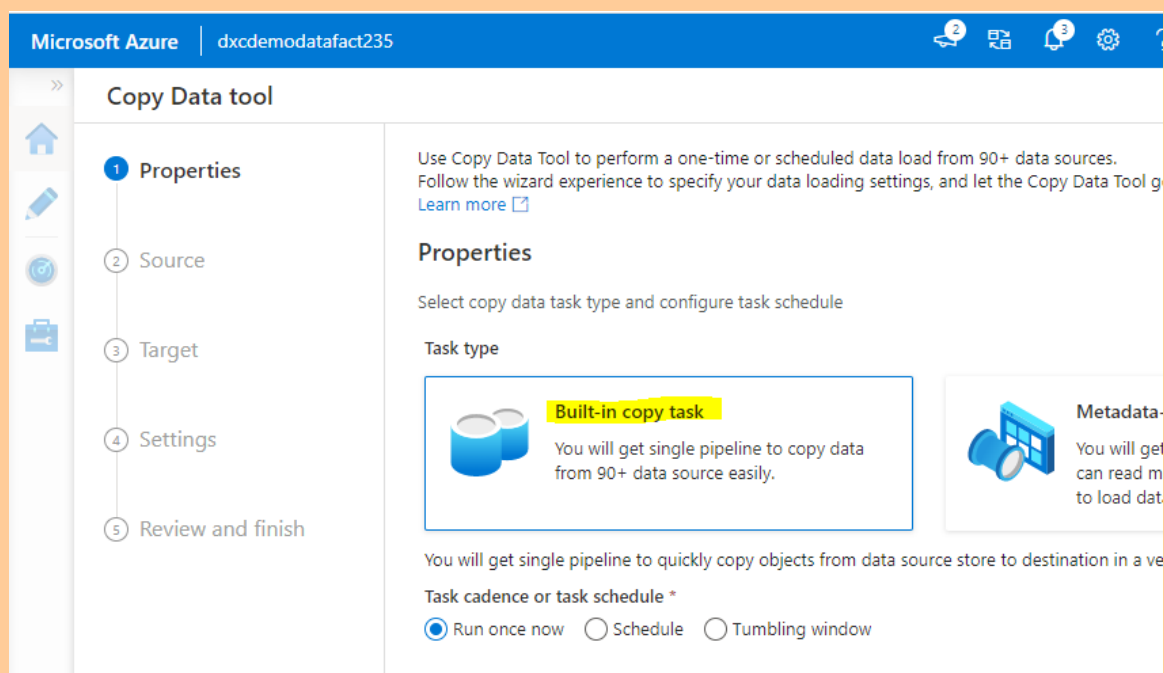
 **Next steps**

Go to resource

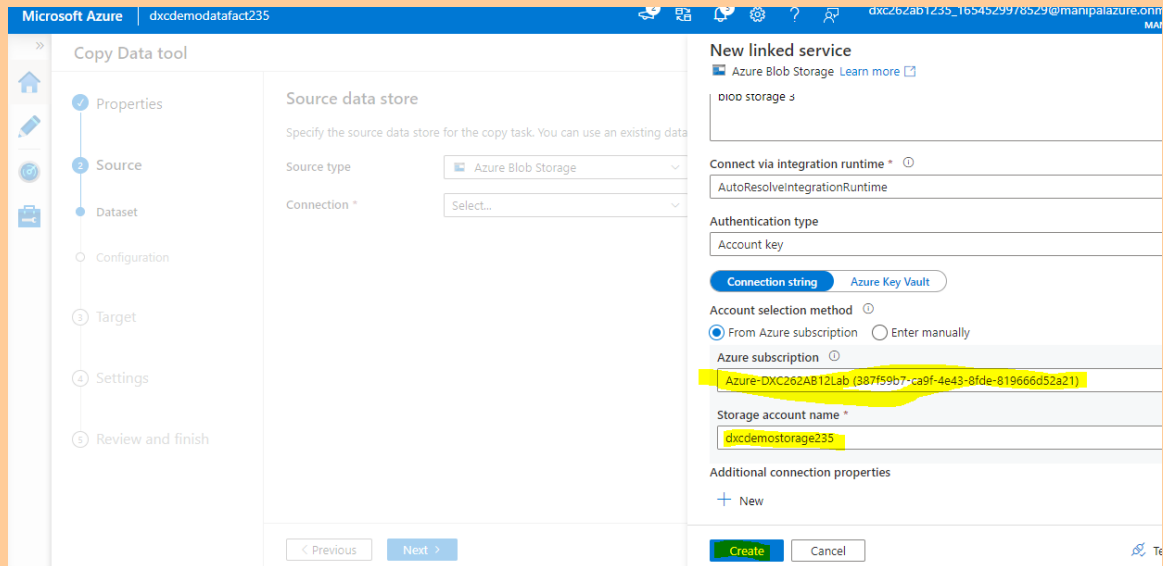
Practical Lab: Add another pipelines for moving data from Staging to **SQL DB**

1. To solve this firstly we need a data storage account, data factory and a sql database. we can create them as shown in above steps.
2. After that we need a source storage container to hold blob data. And to move this data to SQL db we need to use pipeline. So we can use Data factory to do this

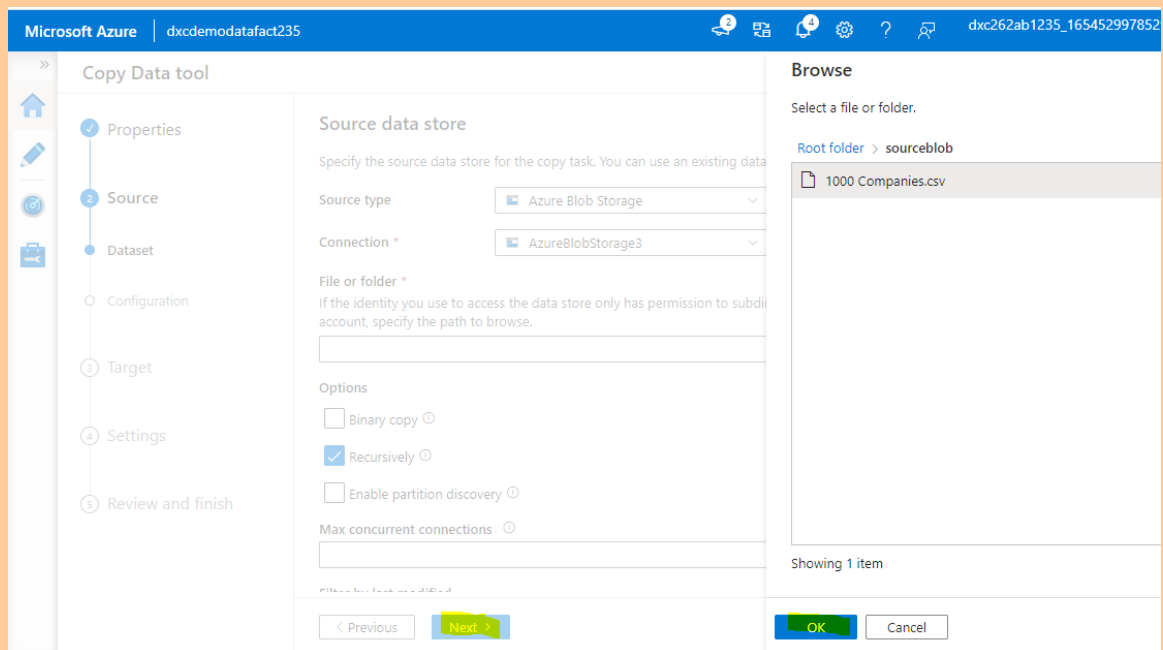
Step 1: Go to Data Factory->Ingest->Built-in Copy.



Step 2: Fill the source details of the source container in the storage account. And Click on "Create".



Step 3: Browse the source file and then click on "Ok" and click on "Next".



Step 4: Now select target as Sql Db and fill details and click on "Ok" and click on "Next".

Microsoft Azure | dxcdemodatafact235

Copy Data tool

Properties

Source

Target

Dataset

Configuration

Settings

Review and finish

Destination data store

Specify the destination data store for the copy task. You can use an existing

Target type: Azure SQL Database

Connection *: Select...

New linked service

Azure SQL Database [Learn more](#)

☒ From Azure subscription ☐ Enter manually

Azure subscription: Azure-DXC262AB12Lab (387f59b7-ca9f-4e43-8fde-819666)

Server name *: dxcdemoqlserver

Database name *: dxcdemodb

Authentication type *: SQL authentication

User name *: Lucky

Password:

Add dynamic content [Alt+Shift+D]

Create Cancel

Microsoft Azure | dxcdemodatafact235

Copy Data tool

Properties

Source

Target

Dataset

Configuration

Settings

Review and finish

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Target type: Azure SQL Database

Connection *: AzureSqlDatabase3 [Edit](#) [+ New connection](#)

Source: Azure Blob Storage file

Target: [Use existing table](#)

Azure Blob Storage file (auto-create)

☐ Skip column mapping for all tables

Previous Next

Step 5: Check Summary and click on "Next".

Microsoft Azure | dxcdemodatafact235

Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Target
- 4 Settings**
- 5 Review and finish

Settings

Enter name and description for the copy data task, more options for data movement

Task name *

Task description

Data consistency verification ☐

Fault tolerance

Enable logging ☐

Enable staging ☐

> Advanced



< Previous **Next >**

Step 6: Now the pipeline is created and now review it and click on "Finish" it. And the it is Deployed.

Microsoft Azure | dxcdemodatafact235

Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Target
- ✓ Settings
- 5 Review and finish**
- Review
- Deployment

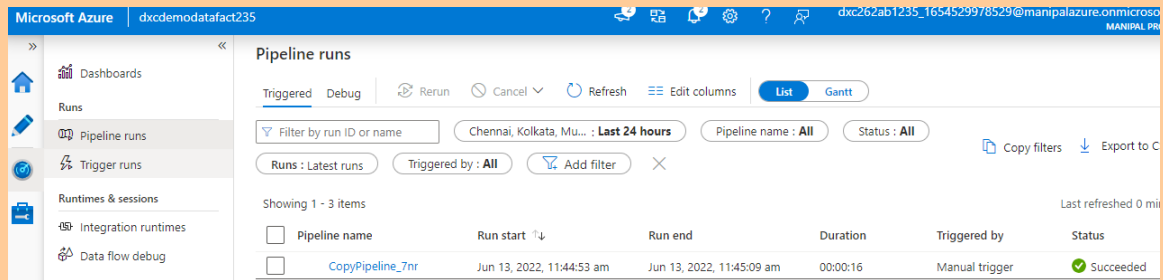
 Azure Blob Storage →  Azure SQL

Deployment complete

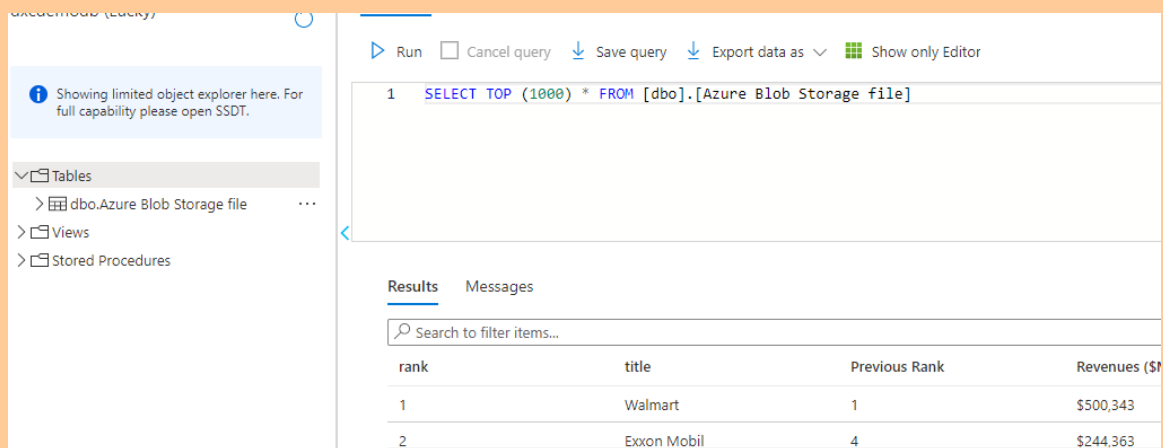
Deployment step	Status
Validating copy runtime environment	✓ Succeeded
> Creating datasets	✓ Succeeded
> Creating pipelines	✓ Succeeded
> Running pipelines	✓ Succeeded

Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to

Finish Edit pipeline Monitor



Step 7: Now navigate to sql database Query page to check the data.



Result: By using all the above shown steps and tools we can create pipeline with the help of a Data factory and can use them to validate and copy data into a Sql Database.

Conclusion: Here, we have stored, validated and copied Blob data to the SQL Database.

Name: Lakshmi Naidu Thatikonda

Reg No: dxcab1226

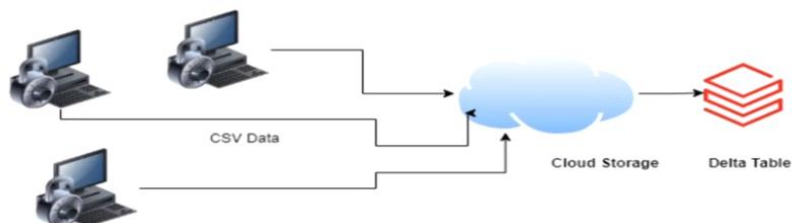
Project2 Name: AP Morgan Data Platform

Date: 10/06/2022

Project 2: AP Morgan Data Platform

Project 2 : AP Morgan

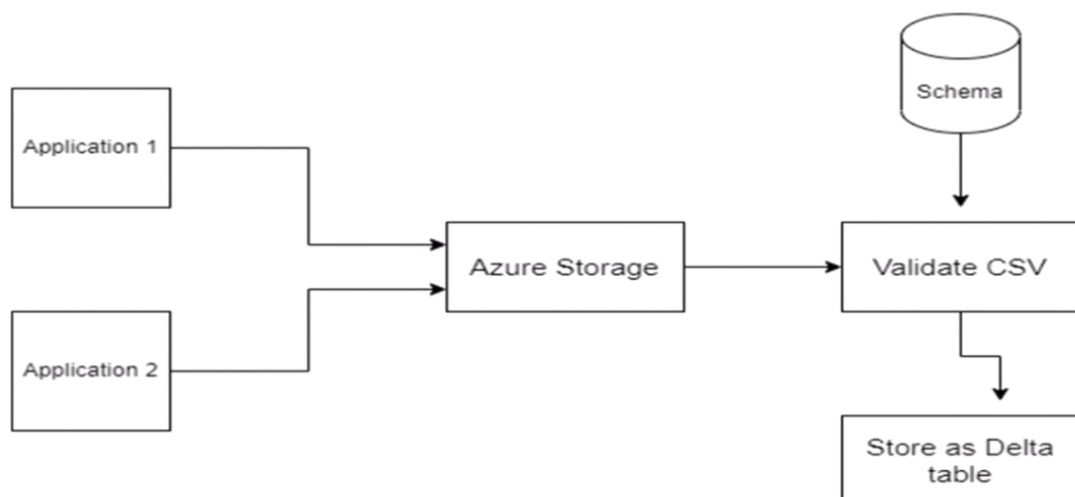
- Multiple Internal applications sends the data(huge size) in CSV format on daily basis in the cloud storage location. There are couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed data to be persisted as Delta table in Databricks for downstream system.



Project 2 : AP Morgan- High Level Detail

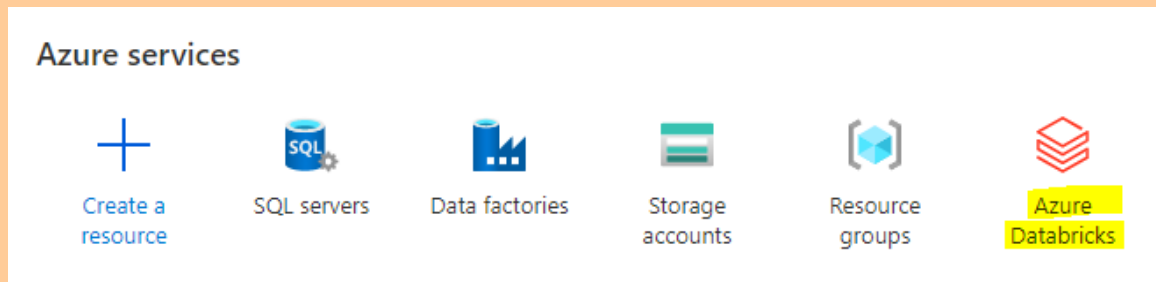
- Internal Application sends CSV file in Azure data lake storage.
- Validation needed to apply on this follows:
 - Check for duplicate rows. If it contains duplicate rows, file need to be rejected.
 - Need to validate the date format for all the date fields. Date column names and desired date format is stored in a Azure SQL server. If validation fails file will be rejected.
- Move all the rejected files to Reject folder.
- Move all the passed files to Staging folder.
- Write the passed files as the Delta table in the Azure Databricks

Project 2 : AP Morgan

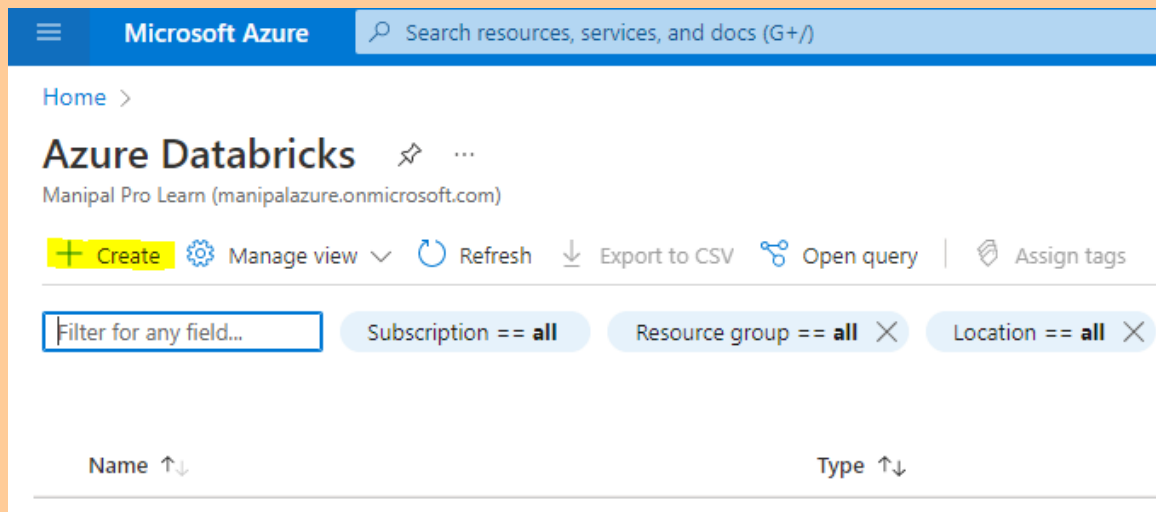


Practical Lab: Create a **Databricks**

Step 1: Login to Azure Portal and select "Data Bricks".



Step 2: Click on "+create".



Step 3: Fill all the required details and click on "Next" and finally on Review and Create.

[Home](#) > [Azure Databricks](#) >

Create an Azure Databricks workspace ...

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

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-DXC262AB12Lab

Resource group * ⓘ

dxccorg235

[Create new](#)

Instance Details

Workspace name *

dxccdemobricks

Region *

East US

Pricing Tier * ⓘ


Standard (Apache Spark, Secure with Azure AD)

Review + create

< Previous

Next : Networking >


Step 4: Click on "Create".

 Microsoft Azure

Search resources, services, and docs (G+)

Home > Azure Databricks >

Create an Azure Databricks workspace ...

 Validation Succeeded

Basics

Networking

Advanced

Tags

Review + create

Summary

Basics

Workspace name	dxcdemobricks
Subscription	Azure-DXC262AB12Lab
Resource group	dxccorg235
Region	East US
Pricing Tier	standard

Networking





Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)	No
Deploy Azure Databricks workspace in	No


Create


< Previous

[Download a template for automation](#)


Step 5: Deployment done. Now click on "Go to Resource".

 Delete  Cancel  Redeploy  Refresh

 We'd love your feedback! →



Your deployment is complete




Deployment name: dxcorg235_dxcdemobricks


Subscription: [Azure-DXC262AB12Lab](#)

Resource group: dxcorg235

Start time: 6/13/20

Correlation ID: e6b

 Deployment details [\(Download\)](#)


 Next steps

[Go to resource](#)

Step 6: Click on "Launch Workspace".

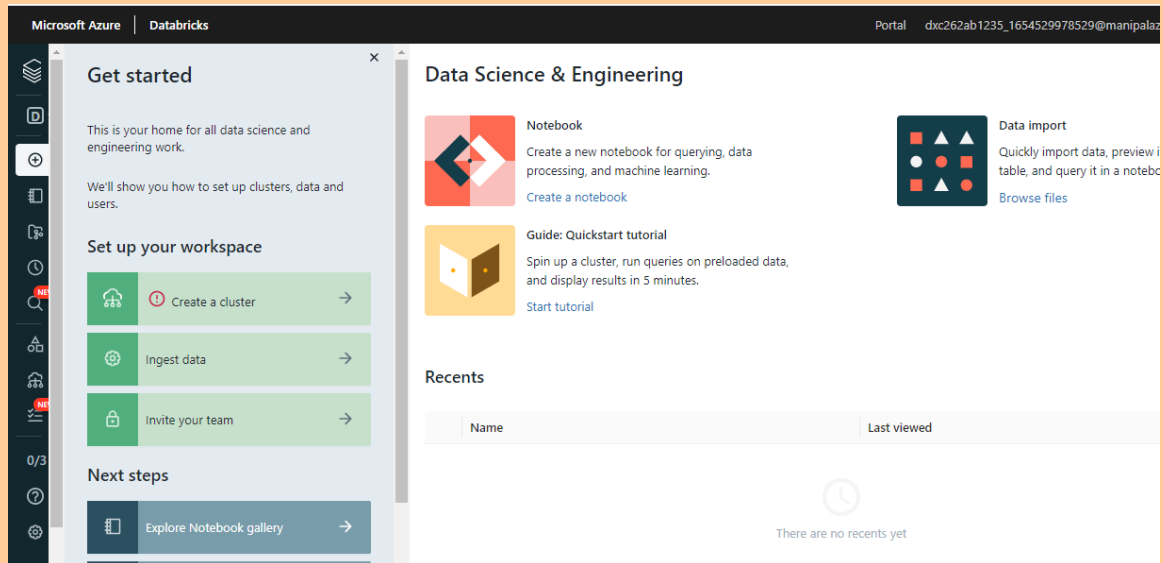
: 387f59b7-ca9f-4e43-8fde-819666d52a21

: [Click here to add tags](#)



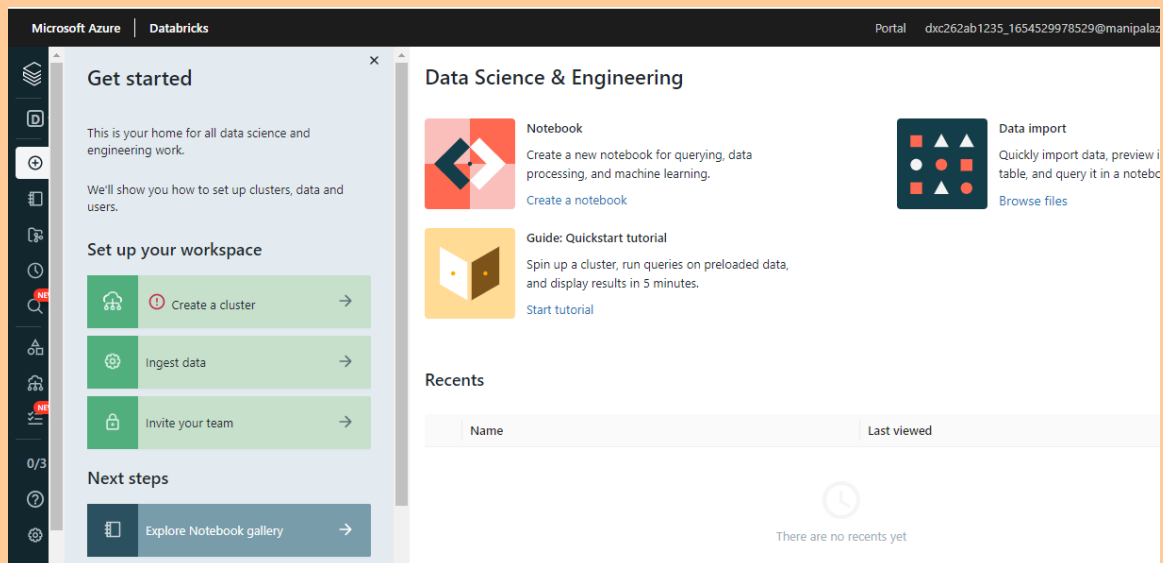
[Launch Workspace](#)

Step 7: Sign into Databricks and Azure Data Bricks is created.

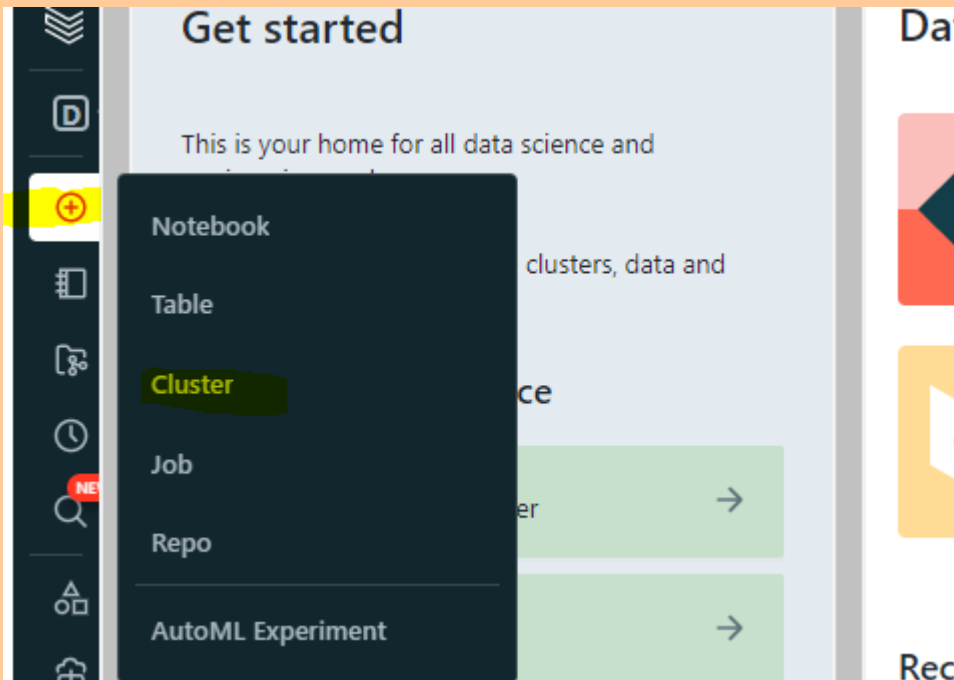


Practical Lab: Create **Cluster** in Azure Databricks

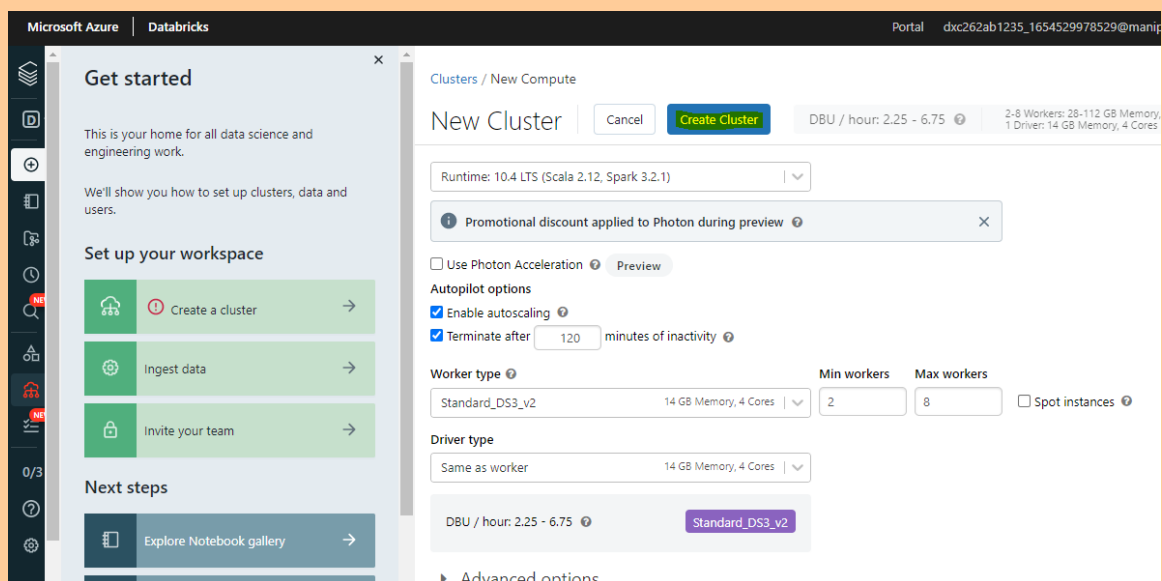
Step 1: Open Data Bricks Workspace.



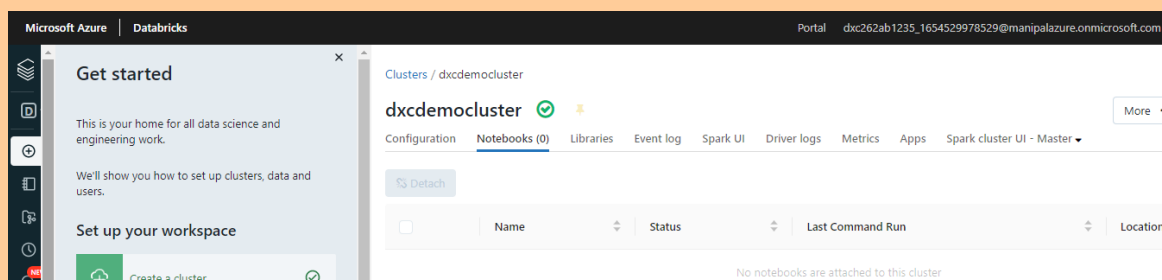
Step 2: Click on "+Cluster".



Step 3: Fill all the necessary information for the cluster and then Click on "Create Cluster".

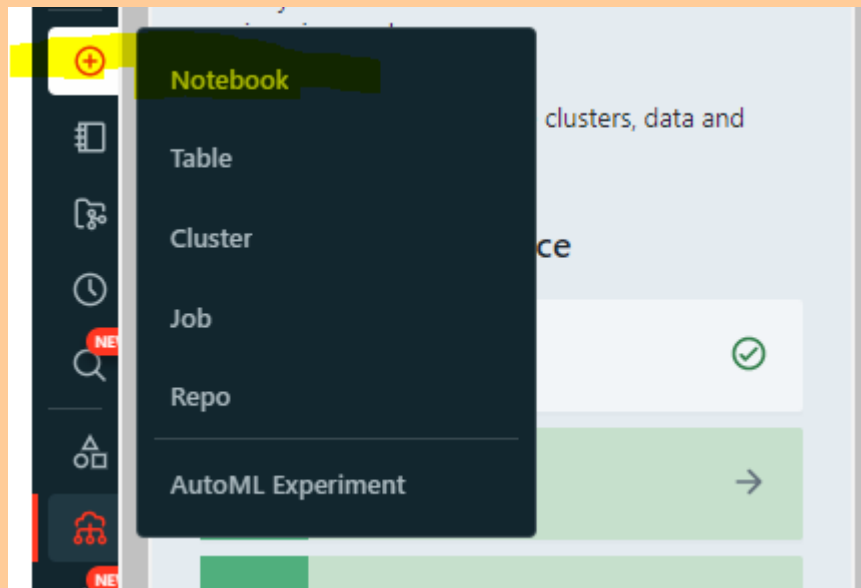


Step 4: Cluster is created.

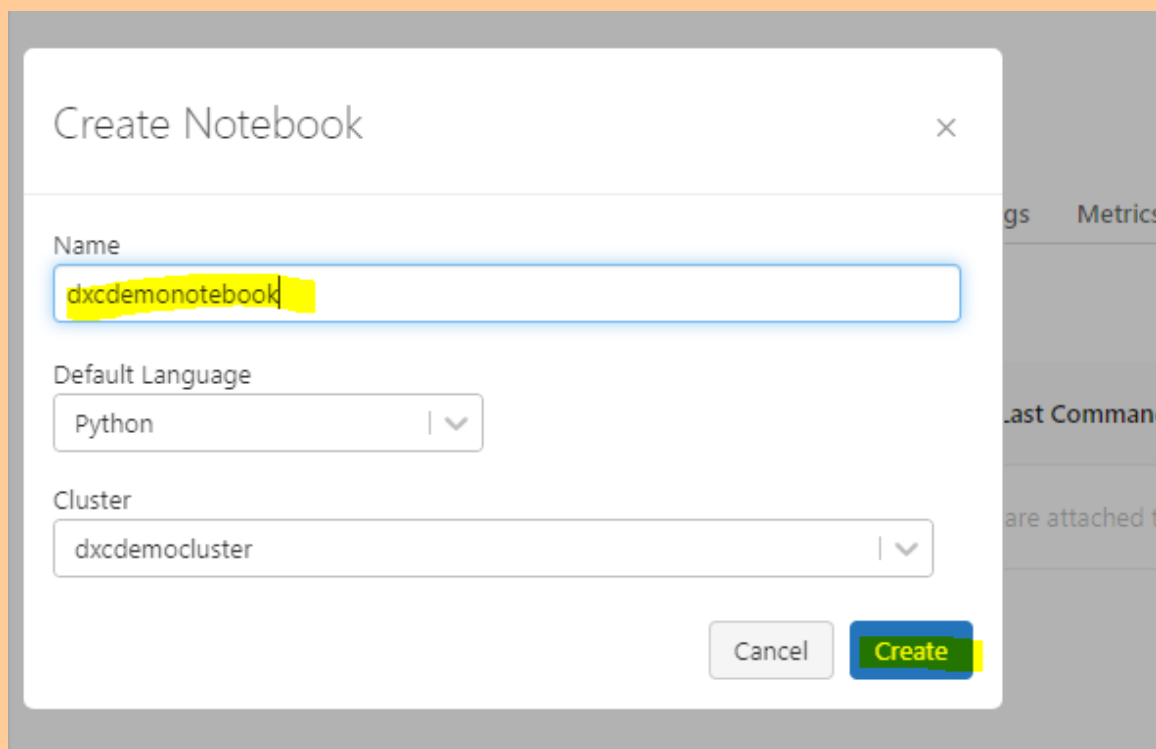


Practical Lab: Add notebook in Databricks and Implement the Business Logic

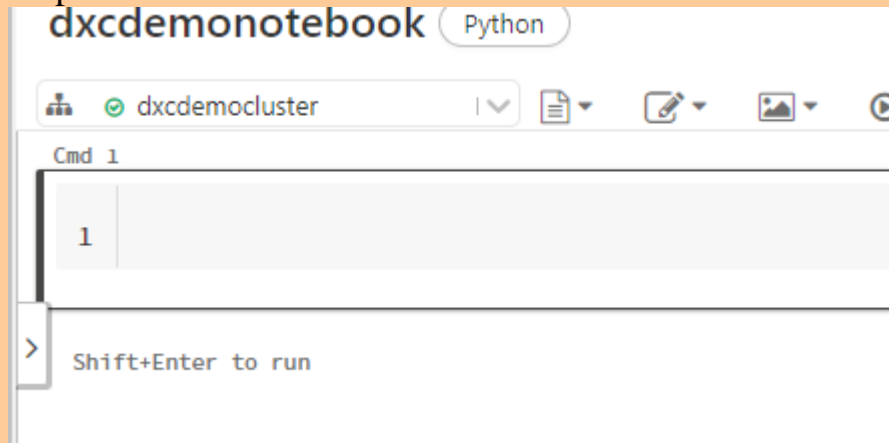
Step1: We need a notebook to perform operations with data available in DataBricks. So Click on "+notebook" in Data Bricks Workspace.



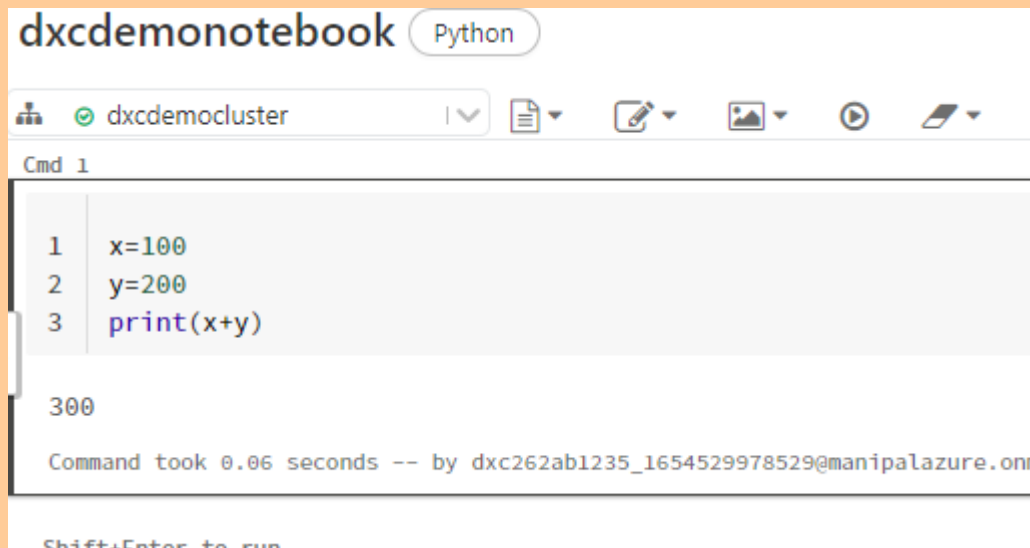
Step 2: Fill the name and Click on "Create".



Step 3: Now the notebook is created.



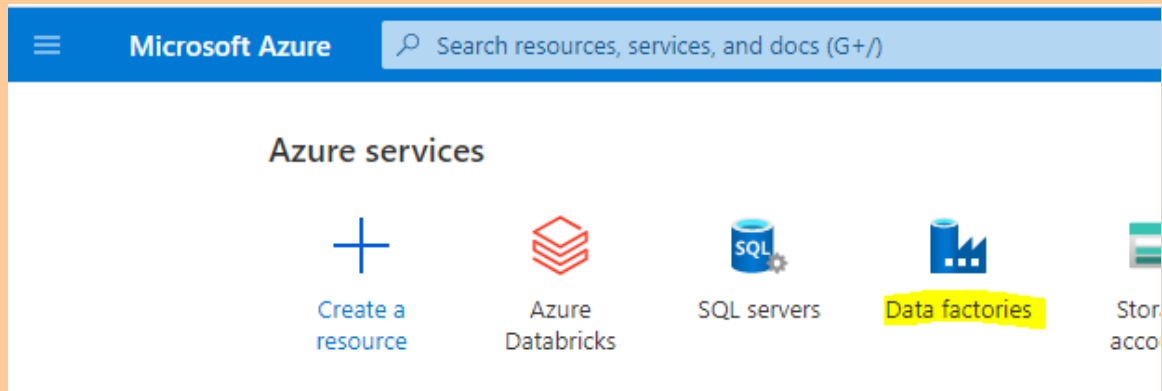
Step 4: And then Now perform some operations in the notebook.



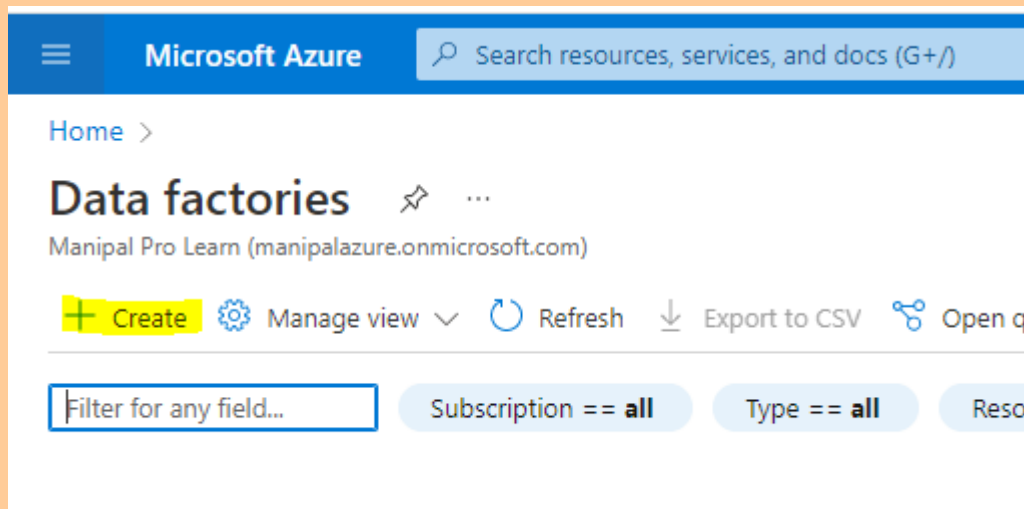
And now we have note book which can be attached to Data Factory and execute.

Practical Lab: **Azure Data Factory** For AP Morgan

Step 1: Go to Azure Portal and click on "Data Factories".



Step 2: Click on "+Create".



Step 3: Fill all required fields and then "Review+Create".

Microsoft Azure

Search resources, services, and docs (G+)

[Home](#) > [Data factories](#) >

Create Data Factory ...

⚠

Changes on this step may reset later selections you have made. Review all options prior to deployment.

Basics

Git configuration

Networking

Advanced

Tags

Review + create

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-DXC262AB12Lab

Resource group * ⓘ

dxccorg235

[Create new](#)

Instance details

Name * ⓘ

dxcdemodatafactory

Region * ⓘ

East US

Version * ⓘ

V2 (Recommended)

Review + create

< Previous

Next : Git configuration >

Step 4: Click on Create to start Deployment.

Create Data Factory ...

✓ Validation Passed

Basics Git configuration Networking Advanced Tags Review + create

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; and (b) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

Basics

Subscription	Azure-DXC262AB12Lab
Resource group	dxccorg235
Name	dxcdemodatafactory
Region	East US

Create

< Previous


Next

[Download a template for automation](#)





Step 5: Deployment is done. And click on "Go to resource".


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


Home >


 **Microsoft.DataFactory-20220613123152 | Overview** ⚙ ...


Deployment


<<  Delete  Cancel  Redeploy  Refresh

 Overview


 Inputs

 Outputs

 Template

 We'd love your feedback! →

✓ Your deployment is complete

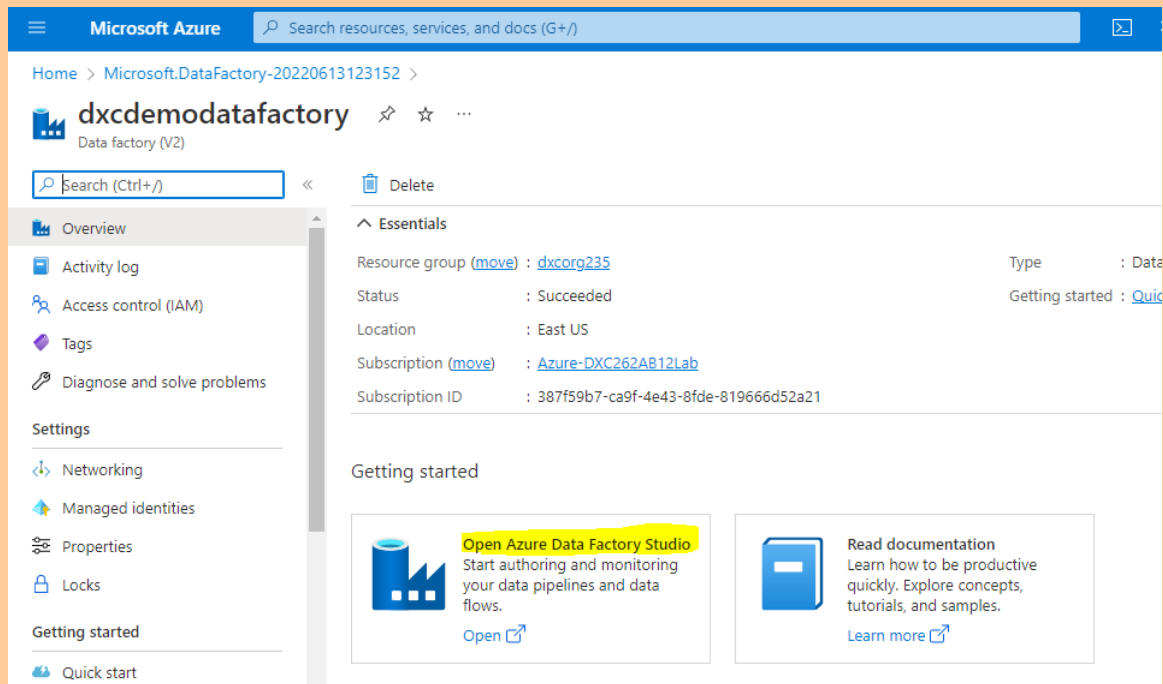
 Deployment name: Microsoft.DataFactory-2022061312... Start time:
Subscription: [Azure-DXC262AB12Lab](#) Correlation
Resource group: [dxccorg235](#)

∨ Deployment details (Download)

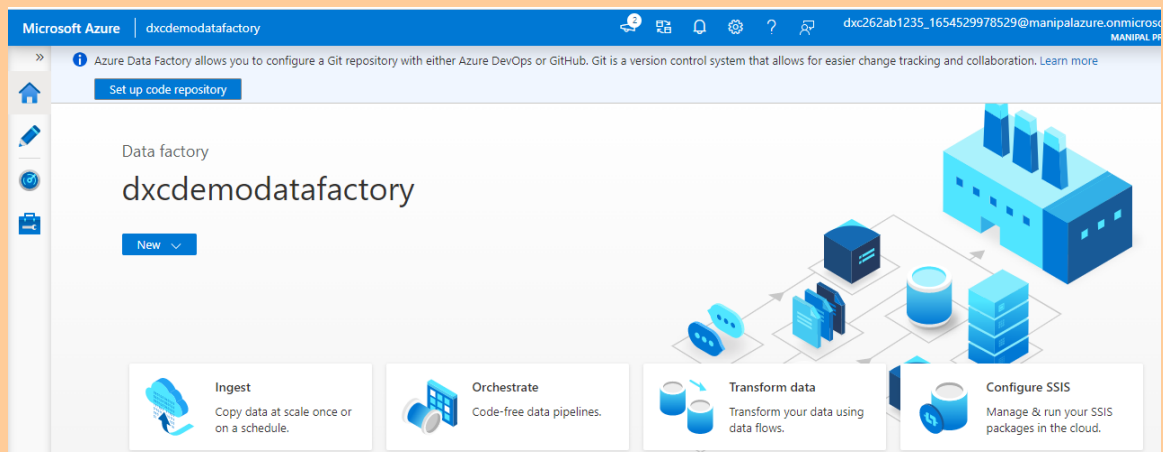
∧ Next steps

Go to resource

Step 6: Now click on "Open Azure Data Factory".



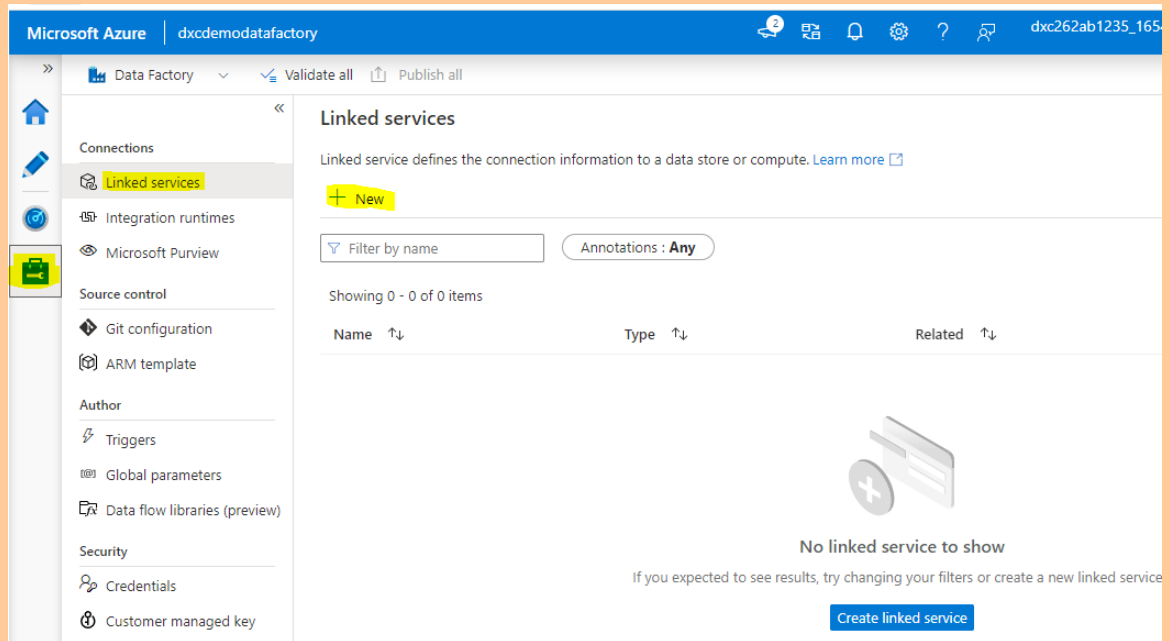
Step 7: And the Azure data Factory is created.



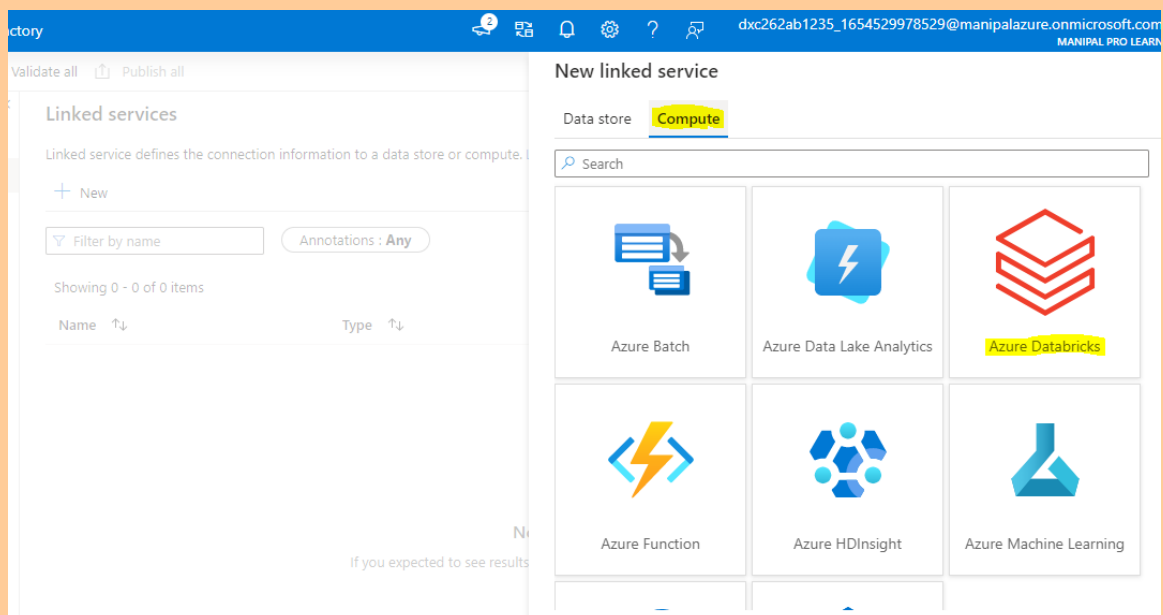
Practical Lab: Create Azure Databricks Linked Service in ADF

1. Here, we should create a pipeline to establish a connection between Data Bricks notebook and Data Factory.

Step 1: Click on "Manage" and select "Linked services" under "Connections" and click on "+add".






Step 2: Click on Compute and then select "Azure Databricks".



Step 3: Fill all the required fields to link bata bricks service.

New linked service


 Azure Databricks [Learn more](#) 

Workspace namespace 

dxcdemobricks

Select cluster

☒ New job cluster ☐ Existing interactive cluster ☐ Existing instance pool


Databrick Workspace URL 


https://adb-4027006216227962.2.azure.databricks.net


Authentication type *

Access Token

Access token Azure Key Vault

Access token * 

Cluster version * 

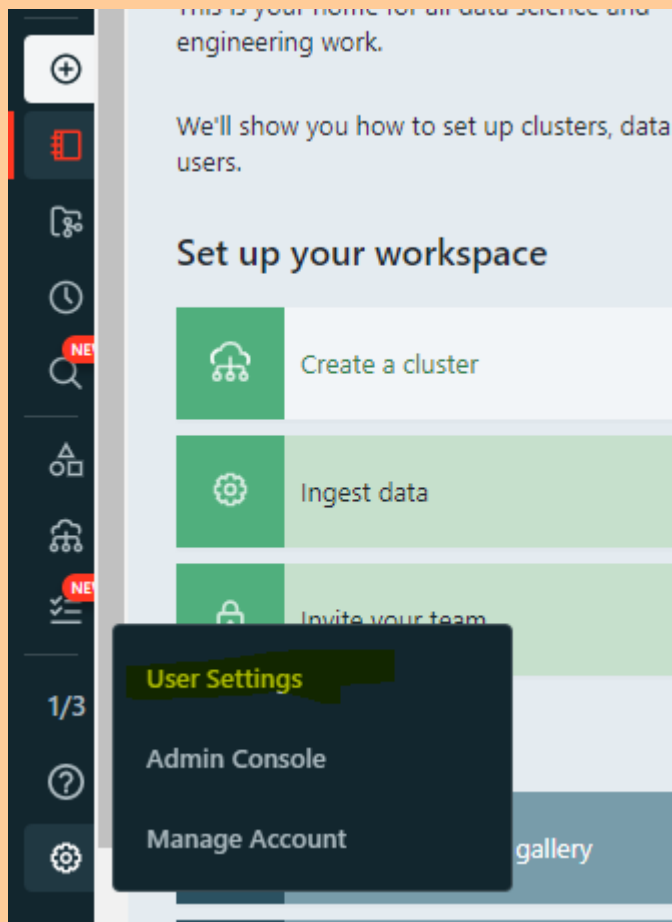
Cluster node type * 

Select cluster version to list options

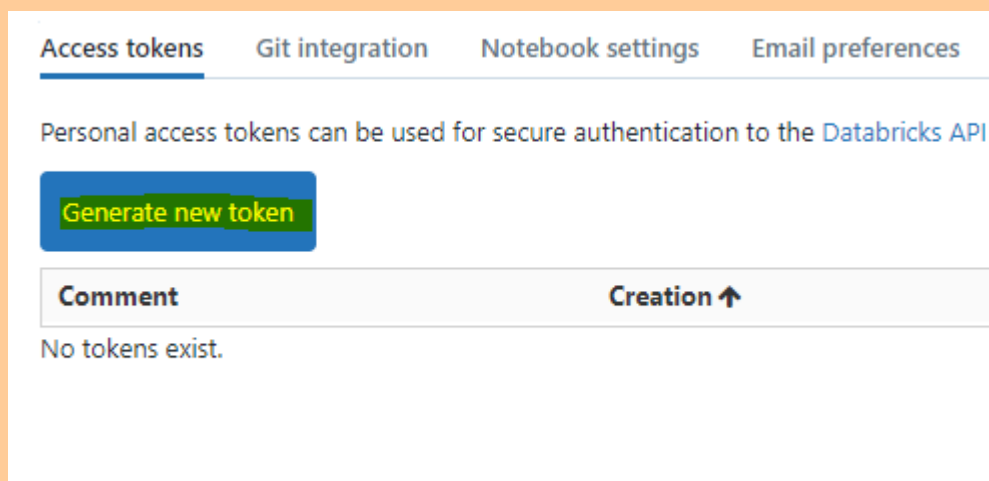
Create Back Test connection Cancel

Here to get the access token let's take help from Databricks.

Move to "User Settings" in "Settings" of "DataBricks".



Click on "Generate new Token".



Click on "Generate".

Generate new token

Comment

pro

Lifetime (days) ⓘ

90

CancelGenerate

Token is generated.

Personal access tokens can be used for secure authentication to the [Databricks API](#) instead of passwords.

Generate new token

Comment	Creation ↑	Expiration
	2022-06-13 12:56:53 IST	2022-09-11 12:56:53 IST

Step 4: Click on "Create" after filling the required details.

Linked services

Linked service defines the connection information to a data store or compute.

+ New

Filter by name Annotations : Any

Showing 0 - 0 of 0 items

Name ↑↓	Type ↑↓
---------	---------

No results found. If you expected to see results, check the filters.

Azure Databricks Learn more

Databricks Workspace URL ⓘ

https://adb-4027006216227962.2.azuredatabricks.net

Authentication type *

Access Token

Access token Azure Key Vault

Access token * ⓘ

.....

Cluster version * ⓘ

10.1 (includes Apache Spark 3.2.0, Scala 2.12)

Cluster node type * ⓘ

Standard_DS3_v2

Python Version *

2

Worker options ⓘ

☒ Fixed ☐ Autoscaling

CreateBack

Test connect


Step 5: Azure data bricks is created.

Linked service defines the connection information to a data store or compute. [Learn more](#)

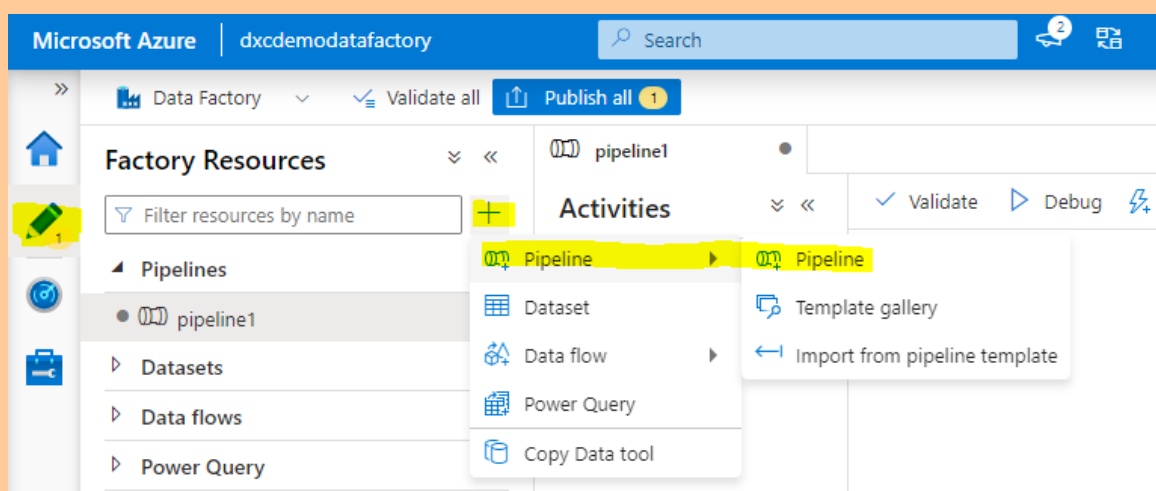
+ New

Filter by name Annotations : Any

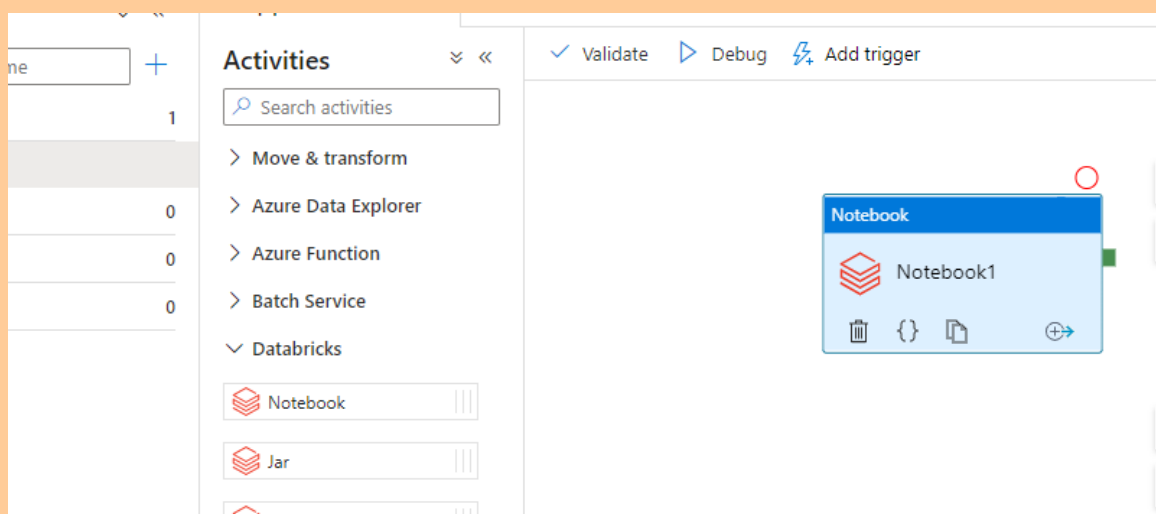
Showing 1 - 1 of 1 items

Name ↑↓	Type ↑↓	Related ↑↓
 AzureDatabricks1	Azure Databricks	0

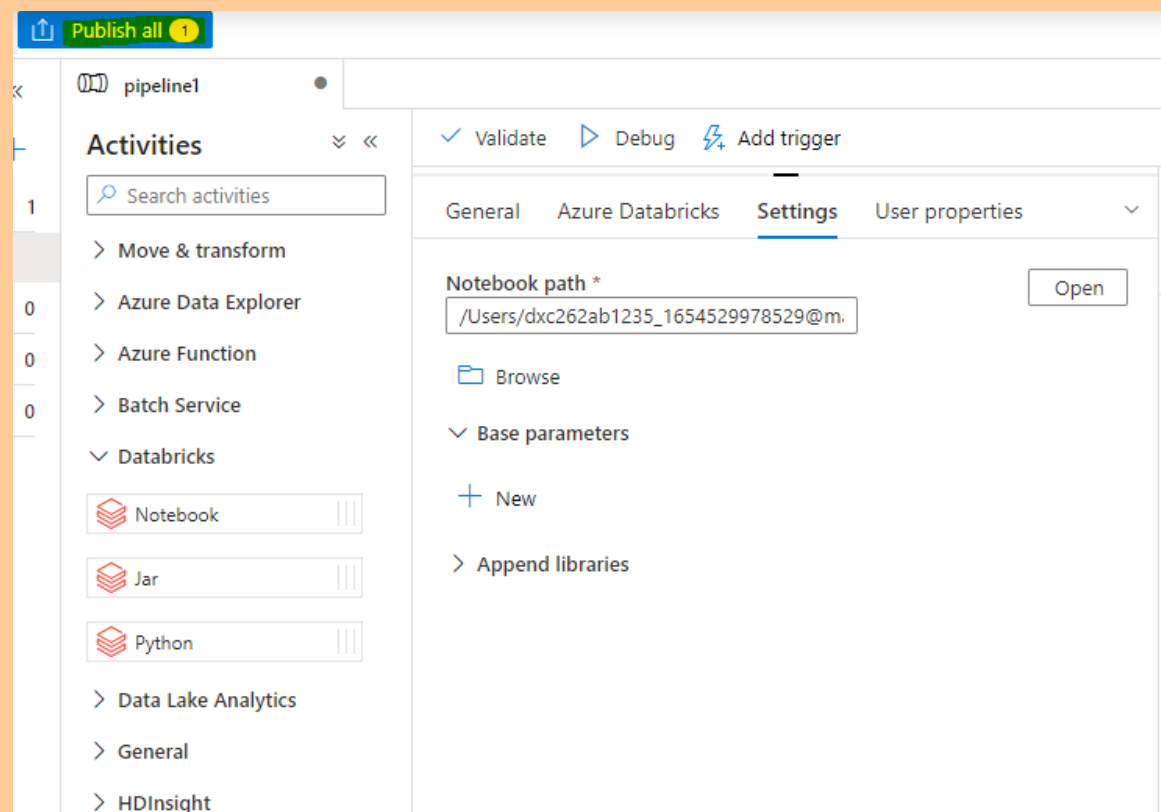
Step 6: And here we should create a pipeline. To do that, click on "Author" -> "Add Resource" to create a pipeline in Data Factory.



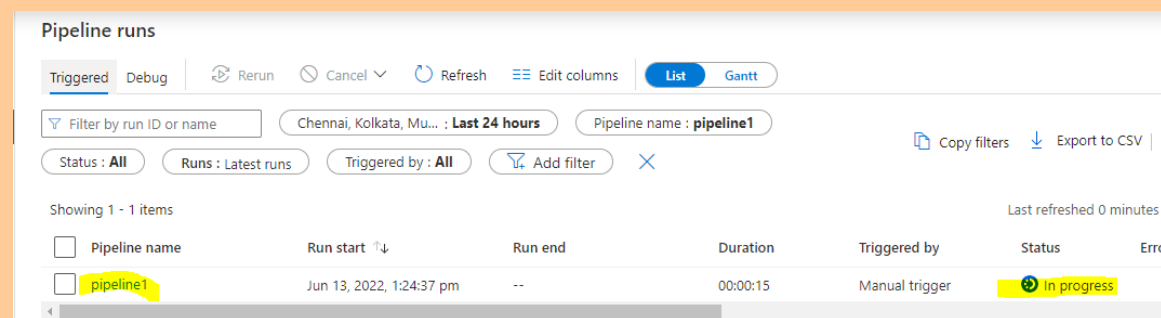
Step 7: After that pipeline is created and now click on "Pipelines" and select the pipeline and move to the "Activity" block and then select the "Databricks" drop down and then drag "Notebook" on to the pipeline workspace.



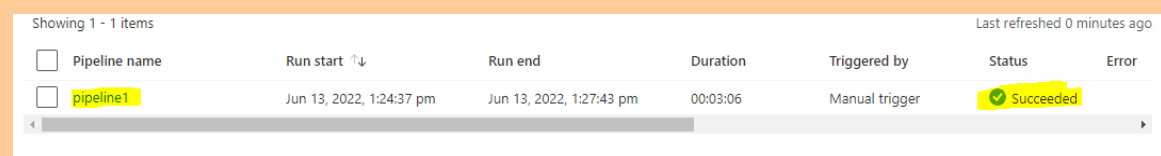
Step 8: Fill all required details and Click on "Publish all" to publish the pipeline and we should run the pipeline by using trigger.



Step 9: Now trigger and monitor the Pipeline created.



Step 10: Pipelining is successfully executed.



Result: We can observe that, we have successfully link and trigger an Azure DataBricks notebook using Data Factory.

Conclusion: We can see and establish the relation between Azure Data Factory and Azure Data Bricks.

References: <https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers>