



DXC REALTIME PROJECTS

AZ-900, DP - 203



JUNE 10, 2022
DXC TECHNOLOGY PVT. LTD

Name: Sai Kiran Anche

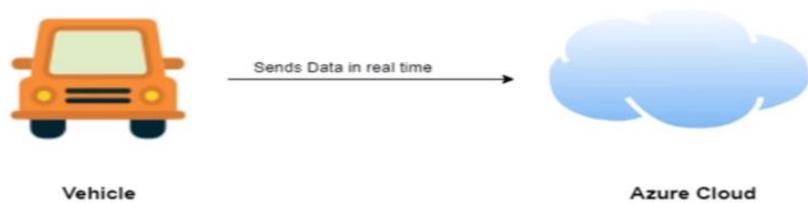
Reg No: DXC262AB12021

Project1 Name: Smart Vehicles

Date: June 10,2022

Project 1 : Connected Vehicles

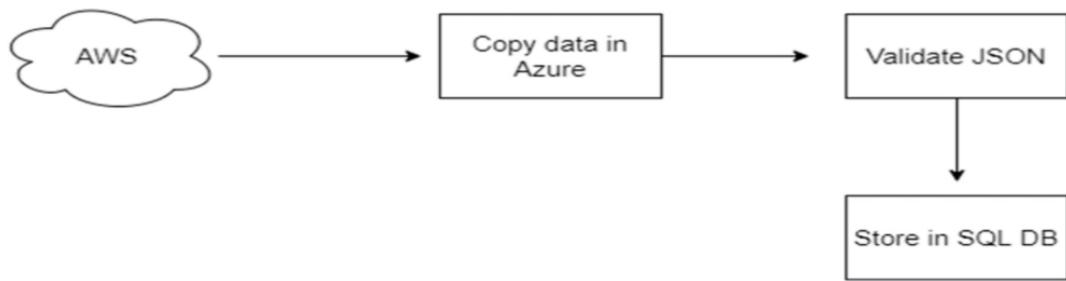
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



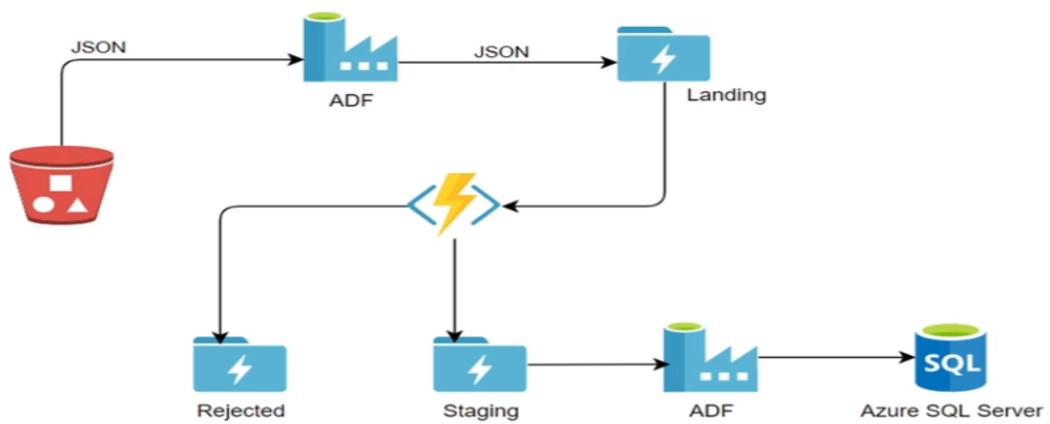
Project 1 : Connected Vehicles

- 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



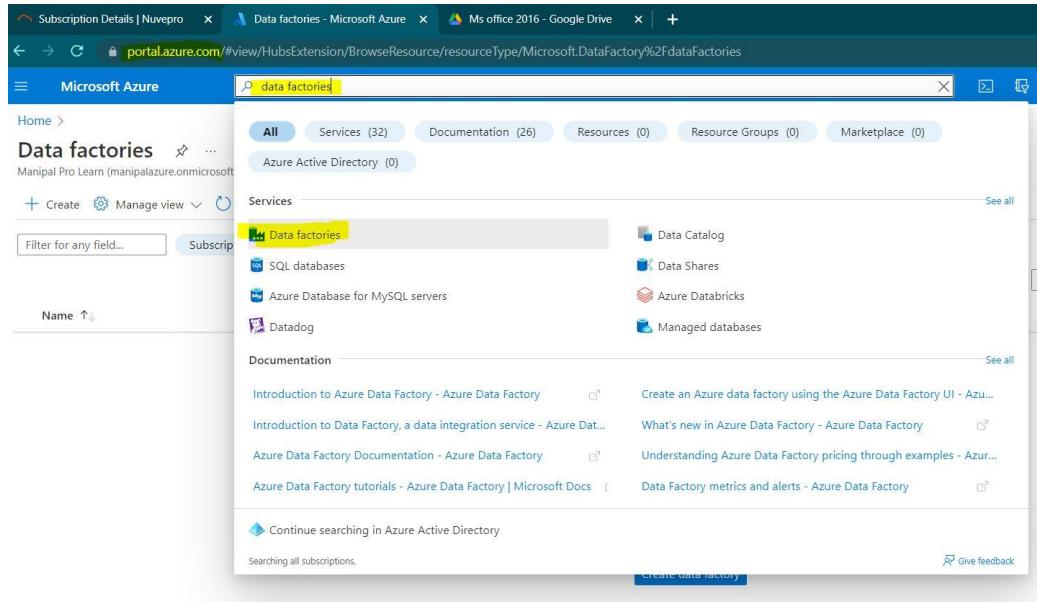
Project 1 : Connected Vehicles



Architecture Diagram for Connected Vehicle Project

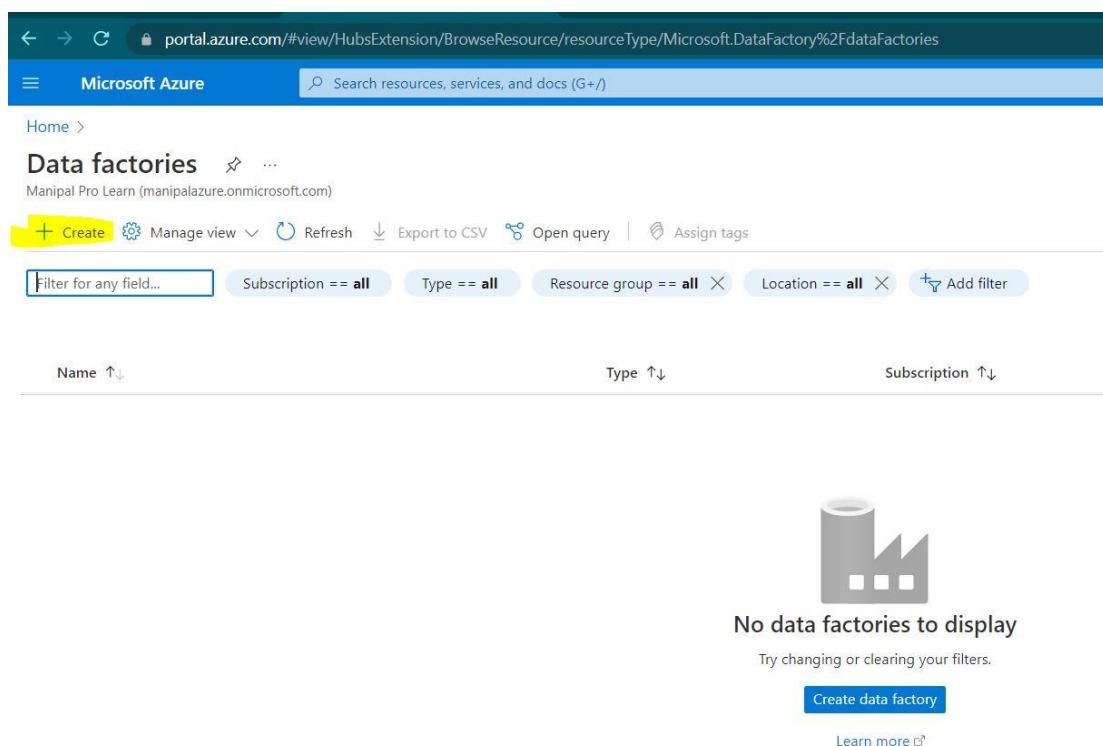
Practical Lab: Create Azure Data Factory Account For Data pipelines

Step1: Goto the URL: <https://portal.azure.com/#home> and search AzureDataFactory in the search bar.



The screenshot shows the Microsoft Azure portal interface. The address bar at the top displays the URL <https://portal.azure.com/#view/HubsExtension/BrowseResource/resourceType/Microsoft.DataFactory%2FdataFactories>. The search bar is filled with the text "data factories". The main content area is titled "Data factories" and shows a list of services. Under "Services", there are four items: "Data Catalog" (with a blue icon), "Data Shares" (with a blue icon), "Azure Databricks" (with a red icon), and "Managed databases" (with a blue icon). Below this, under "Documentation", there is a list of five links: "Introduction to Azure Data Factory - Azure Data Factory", "Introduction to Data Factory, a data integration service - Azure Data Factory", "Azure Data Factory Documentation - Azure Data Factory", "Azure Data Factory tutorials - Azure Data Factory | Microsoft Docs", and "Create an Azure data factory using the Azure Data Factory UI - Azu...". At the bottom right of the main content area, there is a blue button labeled "Create data factory".

Step2: click on the craete button.



The screenshot shows the Microsoft Azure portal interface, similar to the previous one but with a different view. The address bar at the top displays the URL <https://portal.azure.com/#view/HubsExtension/BrowseResource/resourceType/Microsoft.DataFactory%2FdataFactories>. The search bar is empty. The main content area is titled "Data factories" and shows a list of filters at the top: "Subscription == all", "Type == all", "Resource group == all", "Location == all", and "Add filter". Below these filters, there are three sorting options: "Name ↑↓", "Type ↑↓", and "Subscription ↑↓". In the center of the page, there is a large gray icon of a factory building. Below the icon, the text "No data factories to display" is displayed in bold. Underneath this text, there is a smaller message "Try changing or clearing your filters." and a blue "Create data factory" button. At the bottom right, there is a link "Learn more" with a small arrow icon.

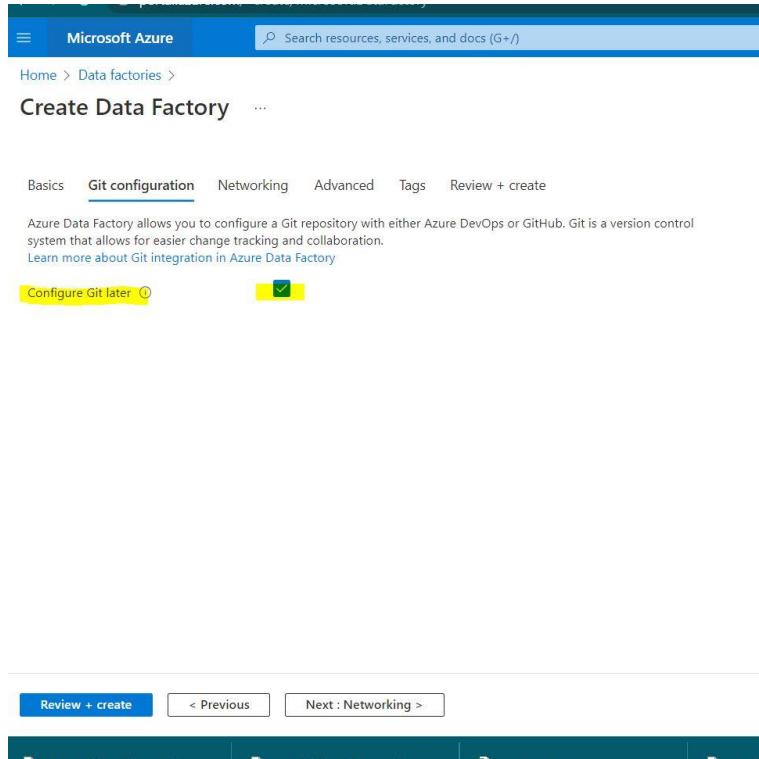
Step3: Create a new resource group for the GM motors DF.

The screenshot shows the 'Create Data Factory' wizard on the 'Basics' step. The 'Project details' section includes a 'Subscription' dropdown set to 'Azure-DXC262AB12Lab'. Below it is a 'Resource group' dropdown with '(New) GMmotors1' selected, which is highlighted with a yellow box. The 'Instance details' section shows 'Name' as 'Genericmotorsdata1', 'Region' as 'East US', and 'Version' as 'V2 (Recommended)'. At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next : Git configuration >'.

Step4: name the data factory and click on git configuration.

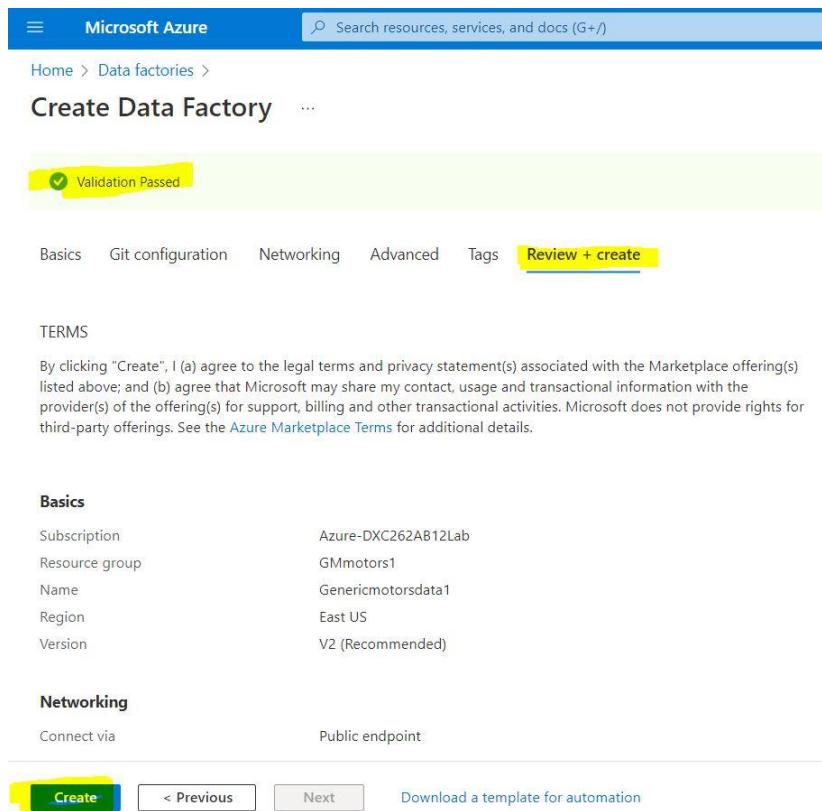
The screenshot shows the 'Create Data Factory' wizard on the 'Basics' step. A warning message at the top states: '⚠ Changes on this step may reset later selections you have made. Review all options prior to deployment.' The 'Project details' section includes a 'Subscription' dropdown set to 'Azure-DXC262AB12Lab' and a 'Resource group' dropdown with '(New) GMmotors1' selected, which is highlighted with a yellow box. The 'Instance details' section shows 'Name' as 'Genericmotorsdata1', 'Region' as 'East US', and 'Version' as 'V2 (Recommended)'. At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next : Git configuration >'.

Step5: click on configure git later and review +create.



The screenshot shows the Microsoft Azure 'Create Data Factory' interface. The top navigation bar includes 'Microsoft Azure' and a search bar. Below it, the breadcrumb navigation shows 'Home > Data factories > Create Data Factory'. The main content area has tabs for 'Basics', 'Git configuration' (which is selected and highlighted with a blue underline), 'Networking', 'Advanced', 'Tags', and 'Review + create'. A note below the tabs states: '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.' It also links to 'Learn more about Git integration in Azure Data Factory'. Two buttons are present: 'Configure Git later' (highlighted with a yellow box) and a green 'Next Step' button with a checkmark icon. At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next : Networking >'.

Step6: after successful validation click on create.



The screenshot shows the Microsoft Azure 'Create Data Factory' interface. The top navigation bar includes 'Microsoft Azure' and a search bar. Below it, the breadcrumb navigation shows 'Home > Data factories > Create Data Factory'. The main content area has tabs for 'Basics', 'Git configuration', 'Networking', 'Advanced', 'Tags', and 'Review + create' (which is selected and highlighted with a blue underline). A green banner at the top indicates 'Validation Passed' with a checkmark icon. Below the tabs, there is a section titled 'TERMS' with a note about agreeing to legal terms and privacy statements. Under the 'Basics' tab, the following details are listed: Subscription (Azure-DXC262AB12Lab), Resource group (GMMotors1), Name (Genericmotorsdata1), Region (East US), and Version (V2 (Recommended)). Under the 'Networking' tab, 'Connect via' is set to 'Public endpoint'. At the bottom, there are buttons for 'Create' (highlighted with a yellow box), '< Previous', 'Next', and 'Download a template for automation'.

Step7: The Data Factory is created.

Home >

 Microsoft.DataFactory-20220610153731 | Overview [↗](#) [...](#)

Deployment

[«](#) [Delete](#) [Cancel](#) [↑ Redeploy](#) [⟳ Refresh](#)

[Overview](#) [Inputs](#) [Outputs](#) [Template](#)

 We'd love your feedback! →

 Your deployment is complete

 Deployment name: Microsoft.DataFactory-20220610153731 Start time: 6/
Subscription: Azure-DXC262AB12Lab Correlation ID
Resource group: GMmotors1

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

[Next steps](#)

[Go to resource](#)

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

Step1: Open the Homepage pf the data Factory.

ie > Microsoft.DataFactory-20220610153731 >

Genericmotorsdata1 ⚡ ⭐ ...

Data factory (V2)

Search (Ctrl+ /) Delete

Overview

Resource group (move) : GMmotors1 Status : Succeeded Type : Data factory (V2)
Location : East US Getting started : Quick start
Subscription (move) : Azure-DXC262AB12Lab Subscription ID : 3a28cdce-3bd7-4219-858e-23ff20f8b998

Essentials

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Networking

Managed identities

Properties

Locks

Getting started

Open Azure Data Factory Studio Start authoring and monitoring your data pipelines and data flows. Open

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

Monitoring

PipelineRuns

ActivityRuns

Step2: click on home as we need to create blob storages.

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

Home > Microsoft.DataFactory-20220610153731 >

Genericmotorsdata1 ⚡ ⭐ ...

Data factory (V2)

Search (Ctrl+ /) Delete

Overview

Resource group (move) : GMmotors1 Status : Succeeded Type : Data factory (V2)
Location : East US Getting started : Quick start
Subscription (move) : Azure-DXC262AB12Lab Subscription ID : 3a28cdce-3bd7-4219-858e-23ff20f8b998

Essentials

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Networking

Managed identities

Properties

Locks

Getting started

Open Azure Data Factory Studio Start authoring and monitoring your data pipelines and data flows. Open

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

Monitoring

PipelineRuns

ActivityRuns

Step3: open storage accounts and click on create

The screenshot shows the Microsoft Azure Storage accounts page. At the top, there's a search bar and navigation links for Home, Storage accounts, and Manipal Pro Learn. Below the header are filter options for Subscription (all), Resource group (all), Location (all), and a 'Create' button. A table header row is visible with columns for Name, Type, Kind, and Resource. The main content area displays a message: 'No storage accounts to display'. It includes a small icon of a storage account, a brief description encouraging users to create one to store up to 500TB of data, and a blue 'Create storage account' button.

Step4: name the storage account and review it.

The screenshot shows the 'Create a storage account' Basics step. The page has a header with Microsoft Azure and a search bar. Below the header, there are navigation links for Home, Storage accounts, and Create a storage account. The main content area has tabs for Basics (selected), Advanced, Networking, Data protection, Encryption, Tags, and Review + create. Under the Basics tab, there's a 'Resource group' dropdown set to 'GMmotors1'. The 'Instance details' section contains fields for 'Storage account name' (set to 'connectedcars12'), 'Region' (set to '(US) East US'), 'Performance' (radio button selected for 'Standard'), and 'Redundancy' (dropdown set to 'Geo-redundant storage (GRS)'). A checkbox for 'Make read access to data available in the event of regional unavailability' is checked. At the bottom, there are buttons for 'Review + create' (highlighted in yellow) and 'Next : Advanced >'.

Step5: after successful validation click on create.

The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal. The 'Review + create' tab is selected. A green bar at the top indicates 'Validation passed'. The configuration table includes:

Blob soft delete	Enabled
Blob retention period in days	7
Container soft delete	Enabled
Container retention period in days	7
File share soft delete	Enabled
File share retention period in days	7
Versioning	Disabled
Blob change feed	Disabled
Version-level immutability support	Disabled

Encryption

Encryption type	Microsoft-managed keys (MMK)
Enable support for customer-managed keys	Blobs and files only
Enable infrastructure encryption	Disabled

At the bottom are buttons for 'Create' (highlighted in yellow), '< Previous', 'Next >', and 'Download a template for automation'.

Step6: our storage account creation is done.

The screenshot shows the 'connectedcars12_1654856639719 | Overview' page. The 'Overview' tab is selected. A message says 'Your deployment is complete' with a checkmark icon. Deployment details include:

- Deployment name: connectedcars12_1654856639719
- Subscription: Azure-DXC262AB12Lab
- Resource group: GMmotors1
- Start time: 6/10/2022, 3:54:07 PM
- Correlation ID: 768b654b-ea2b-4cdc-b127-d359e2587bae

Below are sections for 'Deployment details' (with a download link) and 'Next steps' (with a 'Go to resource' button highlighted in yellow).

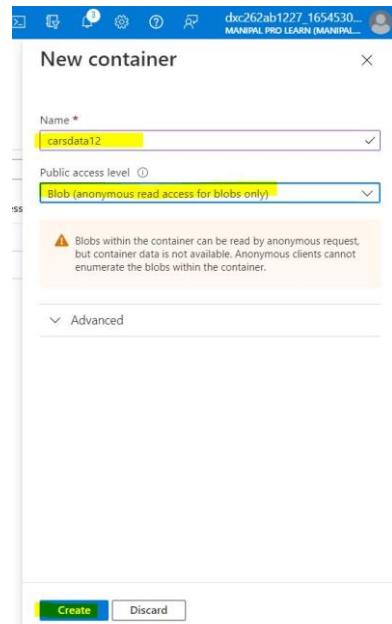
Step7: open storage accounts and click on containers.and create a container.

The screenshot shows the Azure Storage account 'connectedcars12' with the 'Containers' blade open. The left sidebar includes options like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, and Storage browser (preview). The 'Containers' option is selected and highlighted in yellow. At the top, there's a search bar and buttons for changing access level, restoring containers, refreshing, and deleting. A table lists existing containers, with '\$logs' being the only entry, showing its last modified date as 6/10/2022, 3:54:38.

Step8: name the container and give access level as blob.
and then create. It is the source folder.

The screenshot shows the 'New container' dialog. The 'Name' field is filled with 'car1iot'. The 'Public access level' dropdown is set to 'Blob (anonymous read access for blobs only)'. A warning message at the bottom states: '⚠️ Blobs within the container can be read by anonymous request, but container data is not available. Anonymous clients cannot enumerate the blobs within the container.' At the bottom, there are 'Create' and 'Discard' buttons, with 'Create' highlighted in yellow.

Step9:same way create destination blob too.

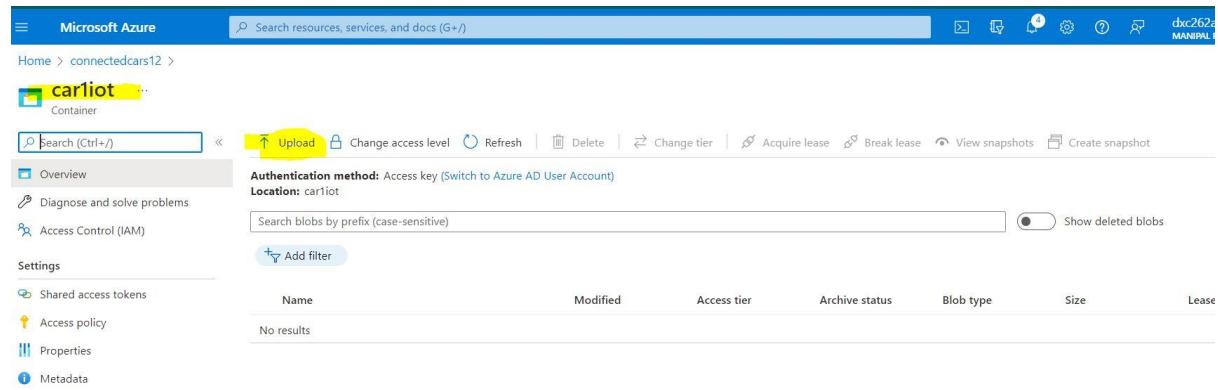


Step10: we can see the blob storages in the given screenshot.

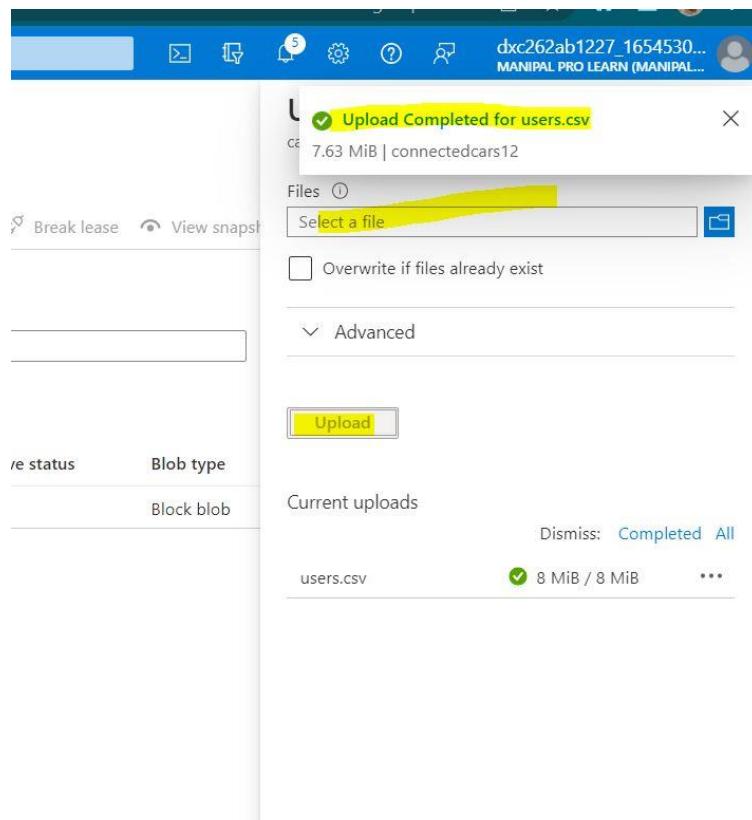
connectedcars12 | Containers

Name	Last modified	Public access level	Lease state
\$logs	6/10/2022, 3:54:38 PM	Private	Available
cariot	6/10/2022, 4:00:14 PM	Blob	Available
carsdata12	6/10/2022, 4:01:13 PM	Blob	Available

Step11: upload some data into the source blob storage.



The screenshot shows the Microsoft Azure Storage Explorer interface. The left sidebar shows the navigation path: Home > connectedcars12 > carliot. The main area is titled 'carliot' and shows it is a Container. Below this, there are tabs for Overview, Diagnose and solve problems, Access Control (IAM), Settings, Shared access tokens, Access policy, Properties, and Metadata. The 'Overview' tab is selected. At the top of the main content area, there are buttons for Upload, Change access level, Refresh, Delete, Change tier, Acquire lease, Break lease, View snapshots, and Create snapshot. Below these buttons, it says 'Authentication method: Access key (Switch to Azure AD User Account)' and 'Location: carliot'. There is a search bar for 'Search blobs by prefix (case-sensitive)' and a checkbox for 'Show deleted blobs'. A 'Add filter' button is also present. A table below lists blobs with columns for Name, Modified, Access tier, Archive status, Blob type, Size, and Lease. The table currently displays 'No results'.



The screenshot shows the Microsoft Azure Storage Explorer interface with a file upload dialog overlaid. The dialog title is 'Upload Completed for users.csv' and it shows the file size as 7.63 MiB and the container as connectedcars12. The dialog has a 'Files' section with a 'Select a file' input field, an unchecked 'Overwrite if files already exist' checkbox, and an 'Advanced' section with a 'Upload' button. In the background, the 'carliot' container overview is visible, showing a table with columns for 'Archive status' and 'Blob type', and a section for 'Current uploads' with a completed entry for 'users.csv' (8 MiB / 8 MiB).

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

Home > connectedcars12 > car1iot

Container

Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots Create snapshot

Authentication method: Access key (Switch to Azure AD User Account)
Location: car1iot

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

Add filter

Name	Modified	Access tier	Archive status	Blob type	Size
users.csv	6/10/2022, 4:02:45 PM	Hot (Inferred)		Block blob	7.63 MiB

Step12: goto data factory home page and click on ingest data.

Microsoft Azure | Genericmotorsdata1

adf.azure.com/en/home?factory=%2Fsubscriptions%2F3a28cdce-3bd7-4219-858e-23ff20f8b998%2FresourceGroups%2FGMmotors1%2Fproviders%2FMicrosoft.DataFactory%2F

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 Manage 8 packages

Discover more

Browse partners (preview) Pipeline templates SAP pipeline templates

Recent resources

No items to show

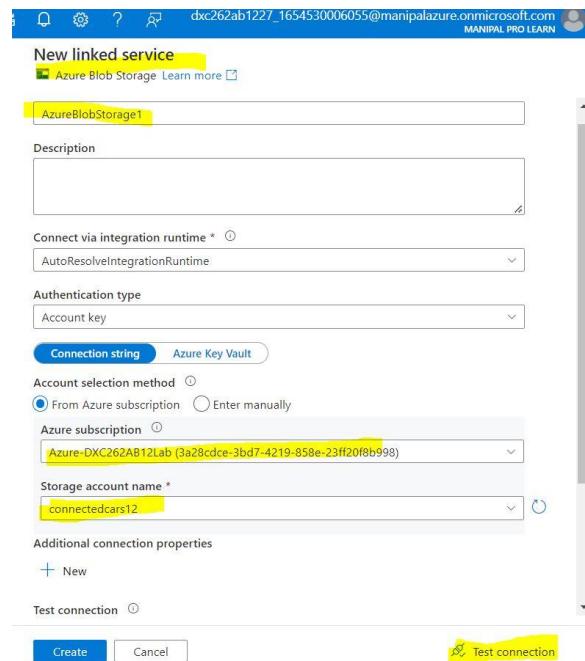
Step13: select Built-in-copy task and run once now. Then click next.

The screenshot shows the Microsoft Azure Copy Data tool interface. At the top, it says "Microsoft Azure | Genericmotorsdata1". Below that is a navigation bar with icons for notifications, settings, and help. The main title is "Copy Data tool". On the left, there's a vertical step-by-step navigation: 1 Properties (highlighted), 2 Source, 3 Target, 4 Settings, and 5 Review and finish. The main content area is titled "Properties" and describes using the tool to perform a one-time or scheduled data load from 90+ data sources. It includes a "Task type" section with two options: "Built-in copy task" (selected) and "Metadata-driven copy task". The "Built-in copy task" section states: "You will get single pipeline to copy data from 90+ data source easily." The "Metadata-driven copy task" section states: "You will get parameterized pipelines which can read metadata from an external store to load data at a large scale." Below this, there's a "Task cadence or task schedule *" section with three radio button options: "Run once now" (selected), "Schedule", and "Tumbling window". At the bottom, there are "Previous" and "Next >" buttons.

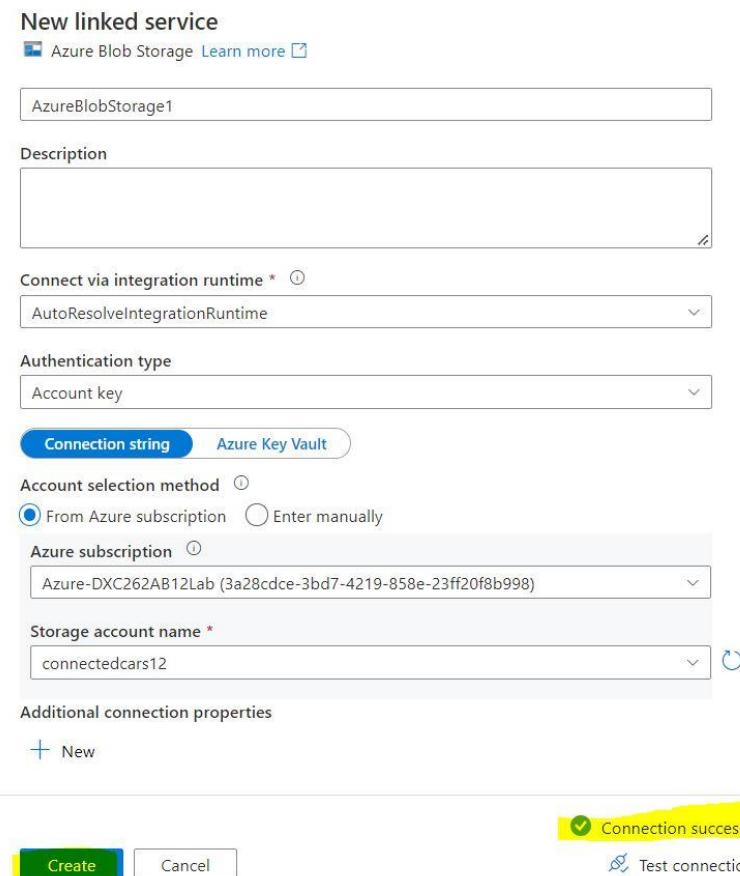
Step14: select source type and click on new connection.

The screenshot shows the Microsoft Azure Copy Data tool interface, specifically the "Source" step. The title is "ta tool". The main content area is titled "Source data store" and asks to specify the source data store for the copy task. It says: "Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store." There are two dropdown menus: "Source type" set to "Azure Blob Storage" and "Connection *" set to "Select...". A "New connection" button is also visible. The right side of the screen shows a preview of the target data store, "Genericmotorsdata1", with a table structure for "Cars" containing columns like "id", "name", "brand", "color", and "price".

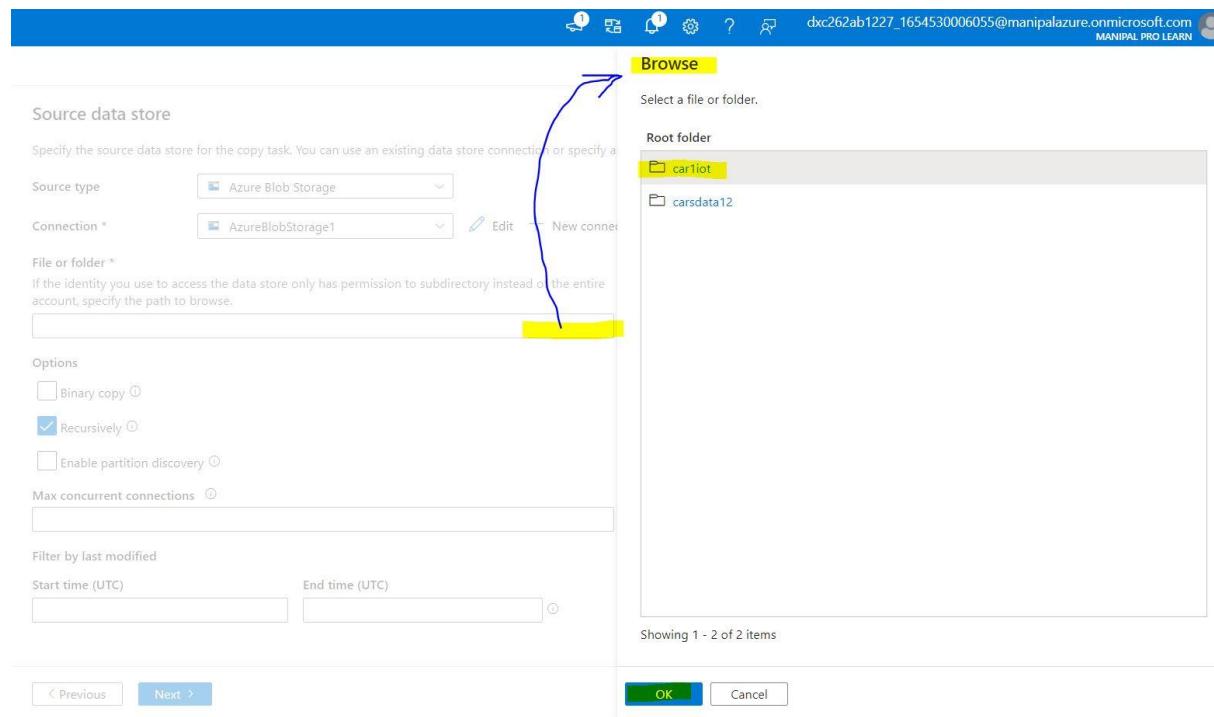
Step15: give the storage account details and test the connection.



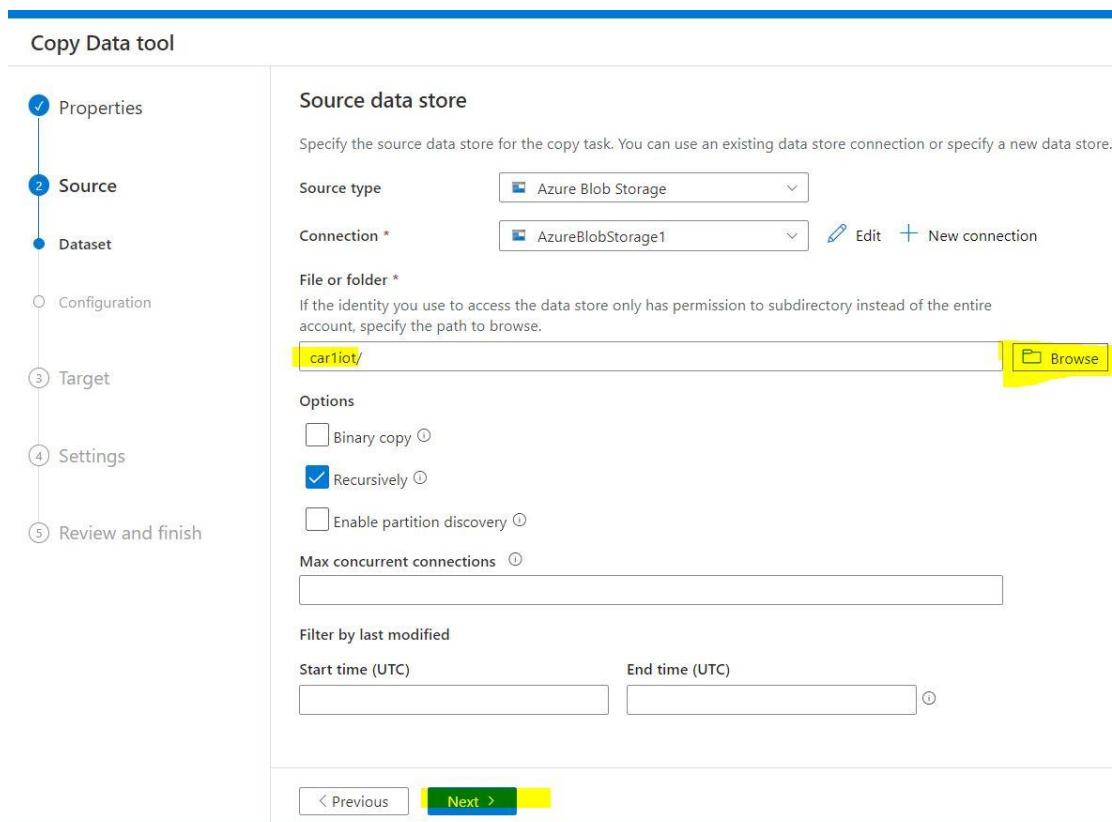
Step16: after the connection is successful click on create.



Step17: now select the source folder and click ok



Step18: Then click on Next.



Step19: select the file format and other settings.

Copy Data tool

Properties

Source

Dataset

Configuration

Target

Settings

Review and finish

File format settings

File format: DelimitedText

Column delimiter: Comma (,)

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

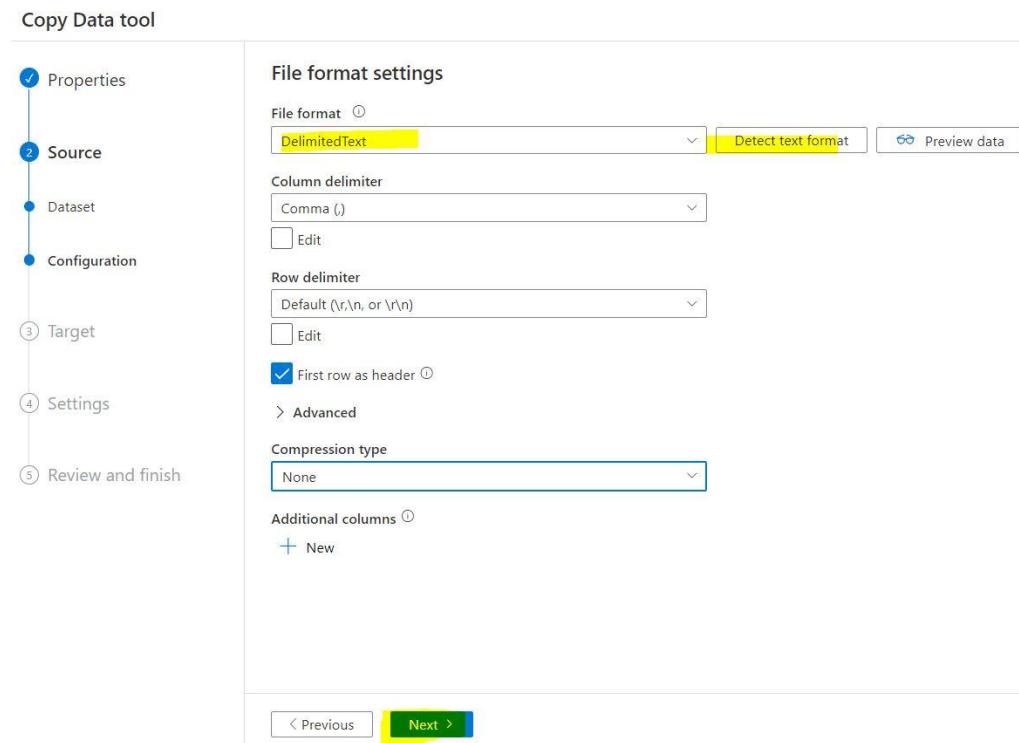
First row as header:

Advanced

Compression type: None

Additional columns: New

< Previous Next >



Step20: select the destination settings and update the information.

Microsoft Azure | Genericmotorsdata1

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 Blob Storage

Connection: AzureBlobStorage1

Folder path: carsdata12

File name: [empty]

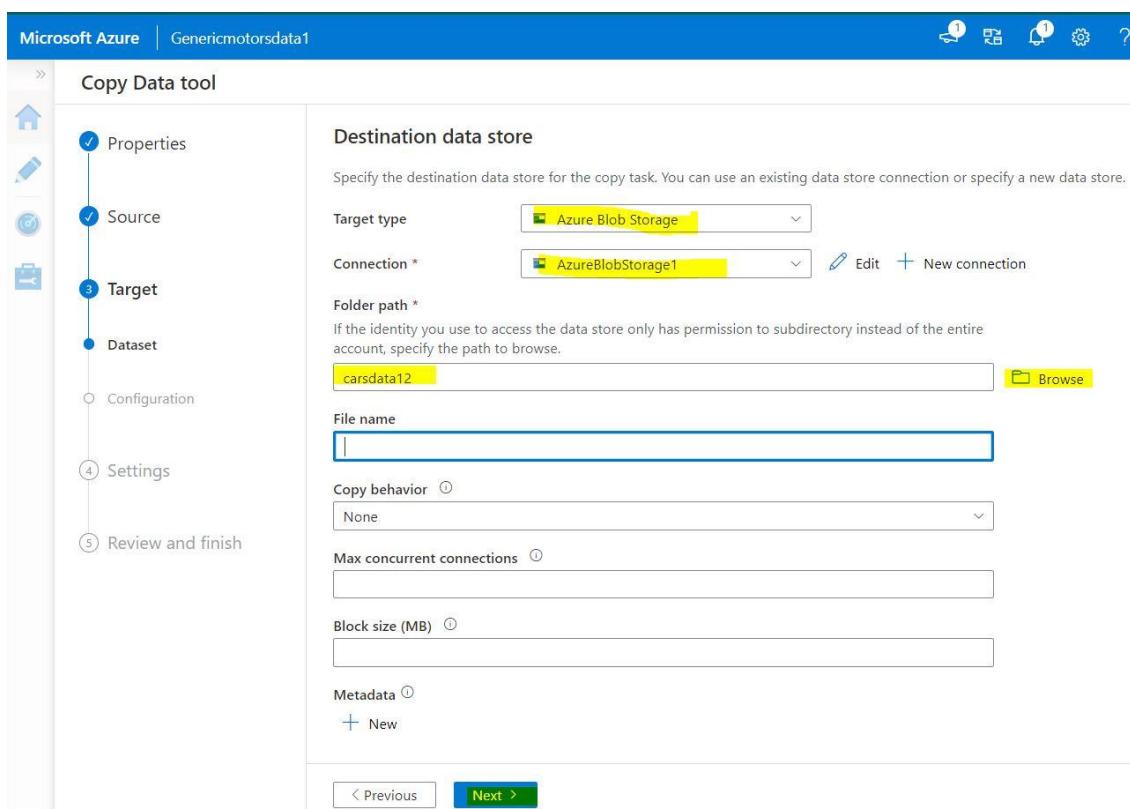
Copy behavior: None

Max concurrent connections: [empty]

Block size (MB): [empty]

Metadata: New

< Previous Next >



Step21: we can see the summary of the pipeline and click on next

Summary

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

A diagram illustrating the data flow of the pipeline. It shows two instances of 'Azure Blob Storage' with a central arrow pointing from the source to the target.

Properties

Task name: CopyPipeline_7de

Task description:

Source

Connection name: AzureBlobStorage1

Dataset name: SourceDataset_7de

Column delimiter: ,

Escape character: \

Quote char: "

First row as header: true

Container: car1iot

< Previous Next >

Step22: click on next.

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.

First row as header: true

Container: car1iot

Target

Connection name: AzureBlobStorage1

Dataset name: DestinationDataset_7de

Column delimiter: ,

Escape character: \

Quote char: "

Copy settings

Timeout: 7.00:00:00

Retry: 0

Retry interval (sec): 30

Secure output: false

Secure input: false

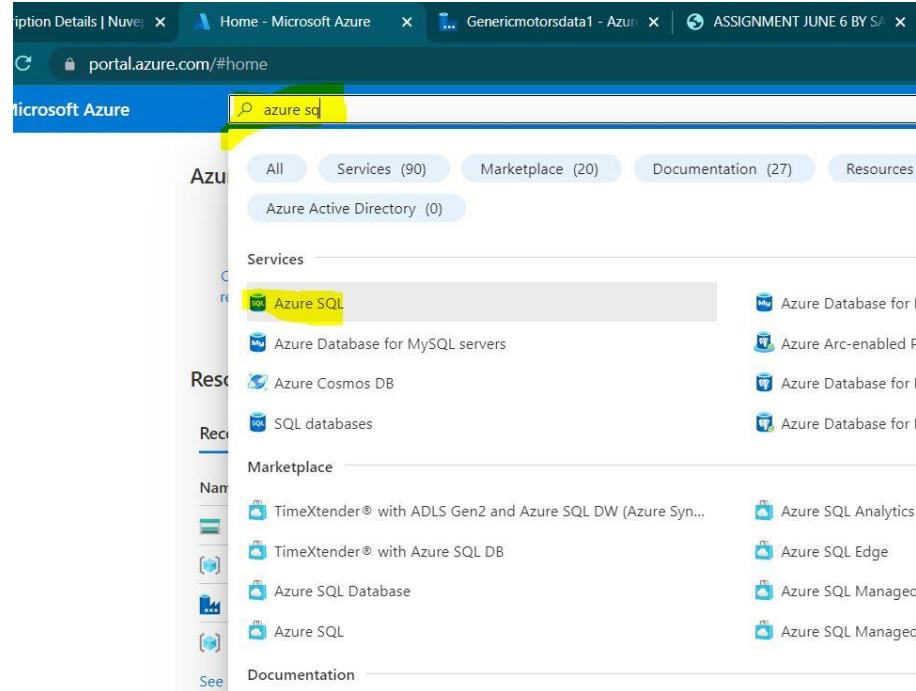
< Previous Next >

Step23: we can see that the deployment has been completed successfully

The screenshot shows the Microsoft Azure Copy Data tool interface. On the left, a sidebar lists steps: Properties, Source, Target, Settings, Review and finish, Review, and Deployment. The 'Review and finish' step is currently selected. The main area displays a flow diagram with two 'Azure Blob Storage' icons connected by a right-pointing arrow. Below the diagram, the text 'Deployment complete' is displayed. A table titled 'Deployment step' shows the status of three steps: 'Validating copy runtime environment' (Succeeded), 'Creating datasets' (Succeeded), 'Creating pipelines' (Succeeded), and 'Running pipelines' (Succeeded). A note at the bottom states: 'Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.' At the bottom of the screen, there are three buttons: 'Finish' (blue), 'Edit pipeline' (white), and 'Monitor' (yellow).

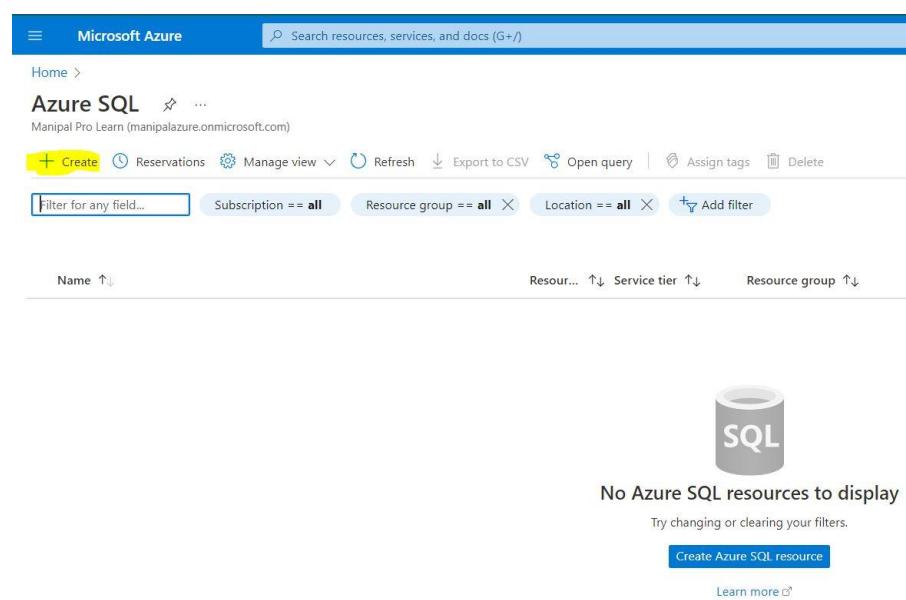
Practical Lab: Create Azure blob trigger logic Create Azure SQL Server and Database

Step1: got to portal.azure.com and search azure SQL.



The screenshot shows the Microsoft Azure portal homepage. The search bar at the top contains the text "azure sql". Below the search bar, there are several navigation tabs: All, Services (90), Marketplace (20), Documentation (27), and Resources. Under the Resources tab, the "Azure SQL" service is highlighted with a yellow box. Other services listed include Azure Database for MySQL servers, Azure Cosmos DB, SQL databases, Azure SQL Analytics, Azure SQL Edge, Azure SQL Managed, and Azure SQL Managed. The left sidebar shows categories like Compute, Storage, and Network.

Step2: open the home screen of azure SQL and click on create.



The screenshot shows the Azure SQL home screen. At the top, there is a search bar and a "Create" button, which is highlighted with a yellow box. Below the search bar, there are filters for Subscription, Resource group, and Location, along with an "Add filter" button. The main area displays a table with columns for Name, Resource..., Service tier, and Resource group. A large "SQL" icon is prominently displayed in the center. Below the icon, the text "No Azure SQL resources to display" is shown, followed by the instruction "Try changing or clearing your filters." and a "Create Azure SQL resource" button.

Step3: select SQL databases and click on Create.

Microsoft Azure

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

Home > Azure SQL >

Select SQL deployment option

Microsoft

Feedback

How do you plan to use the service?

SQL databases
Best for modern cloud applications. Hyperscale and serverless options are available.

Resource type
Single database

Create Show details

SQL managed instance
Best for most migrations ready.

Resource type
Single instance

Create Show d

Step4: select the resource group and give name to the database.

Microsoft Azure

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

Home >

Create SQL Database

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 * ⓘ
GMmotors1

Create new

Database details

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

Database name * ⓘ
gmcardata

Server * ⓘ
Select a server

Create new

The value must not be empty.

Want to use SQL elastic pool? ⓘ
Yes No

Compute + storage * ⓘ
Please select a server first.

Configure database

Backup storage redundancy

Review + create

Next : Networking >

Step5: Click on Networking

The screenshot shows the 'Create SQL Database' wizard on the 'Networking' step. At the top, it says 'Subscription *' with 'Azure-DXC262AB12Lab' selected. Below it, 'Resource group *' is set to 'GMmotors1'. Under 'Database details', 'Database name *' is 'gmcardata' and 'Server *' is '(new) gmcars12 (East US)'. A note asks if the user wants to use a SQL elastic pool, with 'No' selected. In the 'Compute + storage *' section, 'General Purpose' is chosen with 'Gen5, 2 vCores, 32 GB storage, zone redundant disabled'. The 'Configure database' link is visible. At the bottom, there are 'Review + create' and 'Next : Networking >' buttons.

Step6: click on review and create.

The screenshot shows the 'Create SQL Database' wizard on the 'Review + create' step. The 'Networking' tab is active. It displays 'Product details' for a SQL database by Microsoft, mentioning terms of use and privacy policy. The 'Estimated cost per month' is shown as '...'. Under 'Terms', it states that by clicking 'Create', the user agrees to legal terms and privacy statements, and that Microsoft may share contact, usage, and transactional information with third-party offerings. The 'Basics' summary includes: Subscription: Azure-DXC262AB12Lab; Resource group: GMmotors1; Region: East US; Database name: gmcardata; Server: (new) gmcars12; Authentication method: SQL authentication. At the bottom, there are 'Create' and '< Previous' buttons, along with a link to 'Download a template for automation'.

Step7: your SQL data base is created.

The screenshot shows the Microsoft Azure portal interface for managing a SQL database named 'gmcardsdata'. The URL in the address bar is <https://portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/3a28cdce-3bd7-4219-858e-23ff20f8b998/resourceGroups/GMmotors1/providers/Microsoft.Sql/servers/gmcars12/databases/gmcardsdata>.

Essentials

Resource group (move)	GMmotors1	Server name	gmcards12.database.windows.net
Status	Online	Elastic pool	No elastic pool
Location	East US	Connection strings	Show database connection strings
Subscription (move)	Azure-DXC262AB1Lab	Pricing tier	General Purpose: Gen5, 2 vCores
Subscription ID	3a28cdce-3bd7-4219-858e-23ff20f8b998	Earliest restore point	No restore point available
Tags (edit)	Click here to add tags		

Show data for last: 1 hour 24 hours 7 days Aggregation type: Max

Compute utilization

100%
90%
80%
70%
60%
50%
40%
30%
20%

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

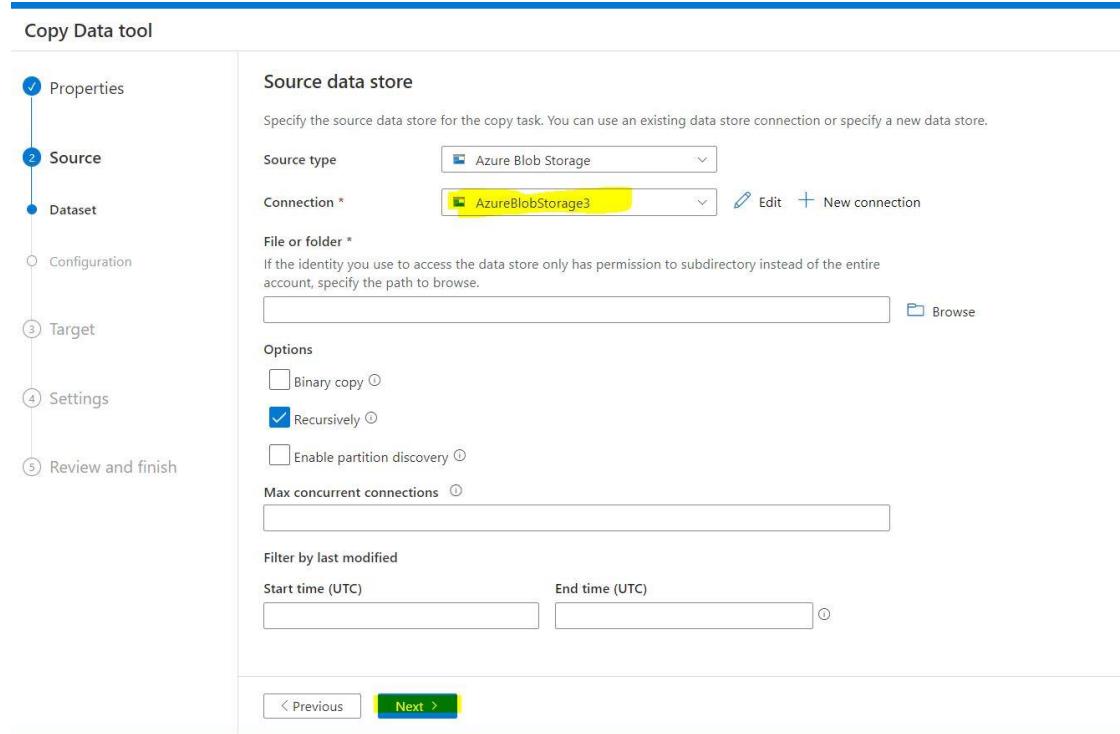
Step1: Go to the data factory studio and click on ingest.

The screenshot shows the Azure Data Factory studio interface. At the top, there is a message about setting up a Git repository. Below it, the data factory name 'Genericsmotorsdata1' is displayed. A 'New' button is visible. On the left, there are four main service icons: 'Ingest' (highlighted in yellow), 'Orchestrate', 'Transform data', and 'Configure SS'. Each icon has a brief description. Below these are links to 'Discover more' and other resources like 'Browse partners (preview)', 'Pipeline templates', and 'SAP pipeline templates'.

Step2: click on Built-in-copy task , run once and click next.

The screenshot shows the 'Copy Data tool' wizard. The left sidebar lists steps: 'Properties', 'Source', 'Target', 'Settings', and 'Review and finish'. The main area is titled 'Properties' and describes the Copy Data Tool. It shows two task types: 'Built-in copy task' (selected) and 'Metadata-driven copy task'. The 'Task type' section includes a note about getting a single pipeline for data from 90+ sources. Below this, 'Task cadence or task schedule' is set to 'Run once now'. Navigation buttons at the bottom include '< Previous' and 'Next >'.

Step3: select the source type and the connection and selet file folder.

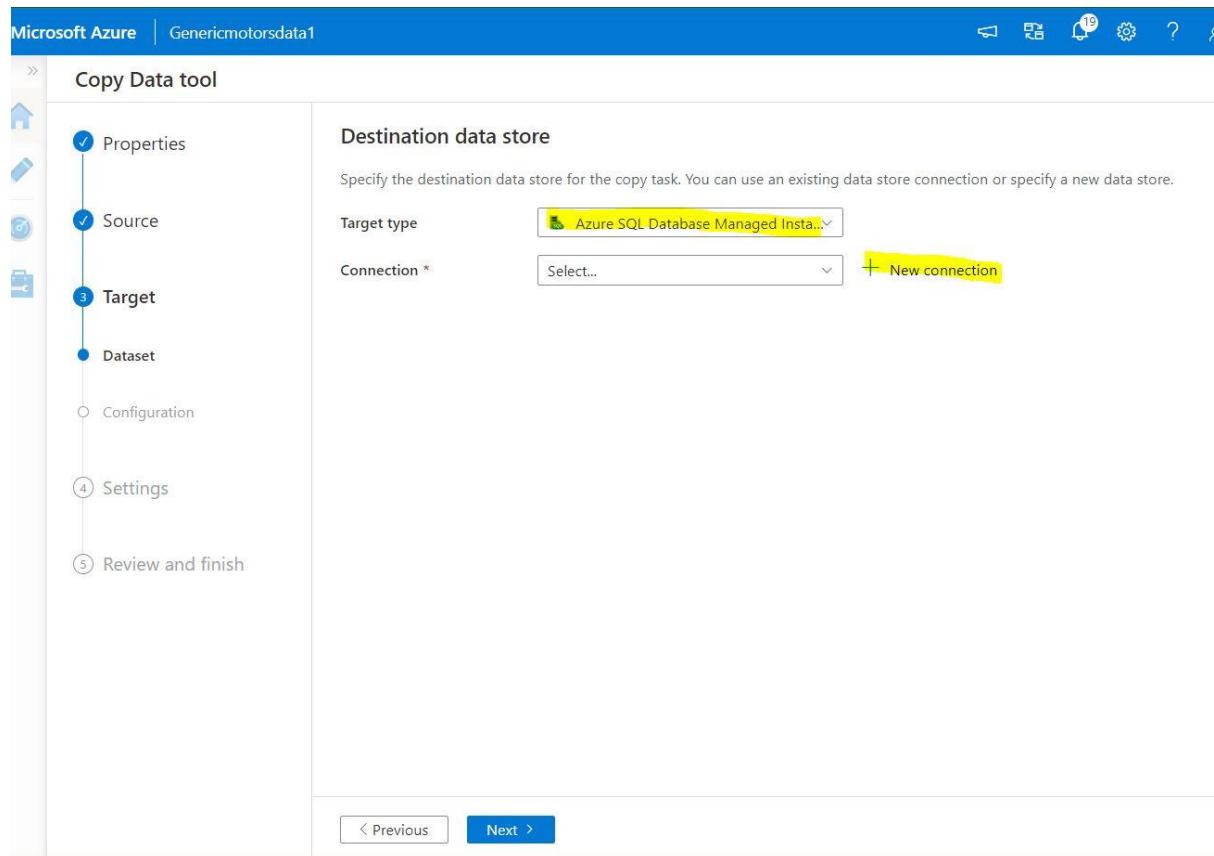


The screenshot shows the 'Copy Data tool' interface at the 'Source' step. On the left, a vertical navigation bar lists steps: Properties (selected), Source (highlighted in blue), Dataset, Configuration, Target, Settings, and Review and finish. The main panel is titled 'Source data store' and contains the following fields:

- Source type: Azure Blob Storage
- Connection *: AzureBlobStorage3 (highlighted)
- File or folder *: A text input field with a 'Browse' button.
- Options:
 - Binary copy
 - Recursively
 - Enable partition discovery
- Max concurrent connections: An empty text input field.
- Filter by last modified:
 - Start time (UTC): An empty text input field.
 - End time (UTC): An empty text input field.

At the bottom are 'Previous' and 'Next >' buttons, with 'Next >' being highlighted in blue.

Step4: Select the destination data store as Azure SQL

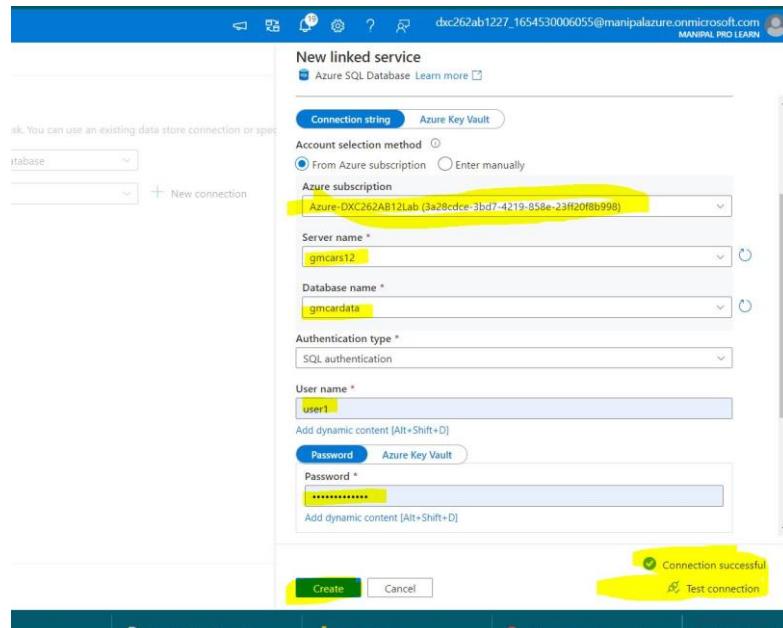


The screenshot shows the 'Copy Data tool' interface at the 'Target' step. The left navigation bar shows: Properties, Source (highlighted in blue), Target (selected), Dataset, Configuration, Settings, and Review and finish. The main panel is titled 'Destination data store' and contains the following fields:

- Target type: Azure SQL Database Managed Insta... (highlighted)
- Connection *: A dropdown menu showing 'Select...' and a '+ New connection' button (highlighted).

At the bottom are 'Previous' and 'Next >' buttons, with 'Next >' being highlighted in blue.

Step5: Click on new connection. Select the server name , database name , authentication type , give username and password. While checking the connection make sure that you allow the IP address of the pipeline will be allowed by the firewall of the database.



Step7: click on next.

The screenshot shows the 'Copy Data tool' interface. On the left, a vertical navigation bar lists steps: Properties, Source, Target, Dataset, Configuration, Settings, and Review and finish. The 'Target' step is currently selected and highlighted with a blue circle. The main workspace is titled 'Destination data store' and displays the configuration for a copy task. It specifies the target type as 'Azure SQL Database' and connects to 'AzureSqlDatabase1'. The source is set to 'Azure Blob Storage file' and the target is set to 'Azure Blob Storage file (auto-create)'. At the bottom, there is a checkbox for skipping column mapping and navigation buttons for previous, next, and cancel.

Step8: review and click on next.

The screenshot shows the 'Copy Data tool' interface in the Microsoft Azure portal. The left sidebar lists steps: Properties, Source, Target (selected), Dataset, Configuration, Settings, Review and finish, and Review. The main area is titled 'Column mapping' with the sub-section 'Table mappings (1)'. It shows a table with one row mapping from 'Source' (Azure Blob Storage file) to 'Target' (Azure Blob Storage file). The 'Column mappings' section shows three columns: Source, Type, Destination, and Type. The data is as follows:

Source	Type	Destination	Type
1004	String	1004	String
Akshay	String	Akshay	String
Software Engineer	String	Software Engineer	String

Below this is the 'Azure SQL Database sink properties' section, which includes a 'Pre-copy script' field and an 'Advanced' link. At the bottom are 'Previous' and 'Next >' buttons, and a 'Cancel' button on the right.

Step9: we can review the pipeline and then click on next.

The screenshot shows the 'Copy Data tool' interface in the Microsoft Azure portal, specifically the 'Review and finish' step. The left sidebar shows steps: Properties, Source, Target, Settings, Review and finish (highlighted in yellow), Review, and Deployment. The main area is titled 'Summary' and states: 'You are running pipeline to copy data from Azure Blob Storage to Azure SQL Database.' It contains two sections: 'Source' and 'Target'.

Source:

- Connection name: AzureBlobStorage3
- Dataset name: SourceDataset_ckd
- Column delimiter: ,
- Escape character: \
- Quote char: "
- First row as header: true
- File name: users.csv
- Container: source

Target:

- Connection name: AzureSqlDatabase1
- Dataset name: DestinationDataset_ckd
- Number of tables: 1

At the bottom are 'Previous' and 'Next >' buttons, and a 'Cancel' button on the right.

Step10: The pipeline deployment is completed.

Microsoft Azure | Genericmotorsdata1

Copy Data tool

Properties
Source
Target
Settings
Review and finish
Review
Deployment

Azure Blob Storage → Azure SQL Database

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 close Copy Data Tool.

Finish Edit pipeline Monitor

The screenshot shows the Microsoft Azure Copy Data tool interface. At the top, it displays 'Microsoft Azure | Genericmotorsdata1'. Below that is a navigation bar with icons for Home, New, Copy, Settings, Help, and a user icon. The main title 'Copy Data tool' is followed by a vertical list of steps: Properties (checkmark), Source (checkmark), Target (checkmark), Settings (checkmark), Review and finish (blue circle with a dollar sign), Review (blue circle), and Deployment (blue circle). To the right of the steps, there's a diagram showing 'Azure Blob Storage' with an arrow pointing to 'Azure SQL Database'. Below the diagram, the text 'Deployment complete' is displayed. A table titled 'Deployment step' lists four steps: 'Validating copy runtime environment' (Status: Succeeded), '> Creating datasets' (Status: Succeeded), '> Creating pipelines' (Status: Succeeded), and '> Running pipelines' (Status: Succeeded). At the bottom, a message says 'Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.' There are three buttons at the bottom: 'Finish' (blue), 'Edit pipeline' (white), and 'Monitor' (white).

Result:

Most of the project requirements have been fulfilled by the step by step procedures.

Conclusion:

All the cases have been solved successfully, it was so challenging when few of the pipelines got failed. After referring the error messages I'm able to run the pipelines successfully and had a great experience while learning the concepts in deep.

Apart of the demonstrations that were explained in the tutorials I'm able to perform a variety of new operations with a lot of enthusiasm and engagement in learning.

Name: Sai Kiran Anche

Reg No: DXC262AB12021

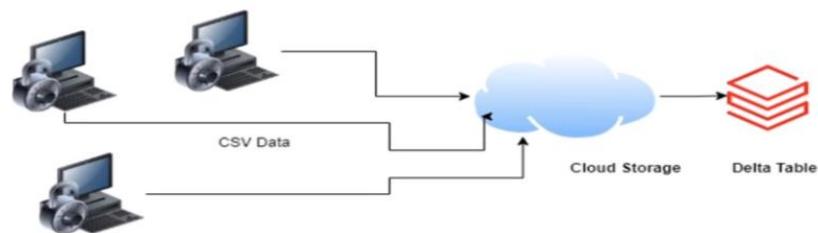
Project2 Name: AP Morgan Data Platform

Date: June 10,2022.

Project 2: AP Morgan Data Platform

Project 2 : AP Morgan

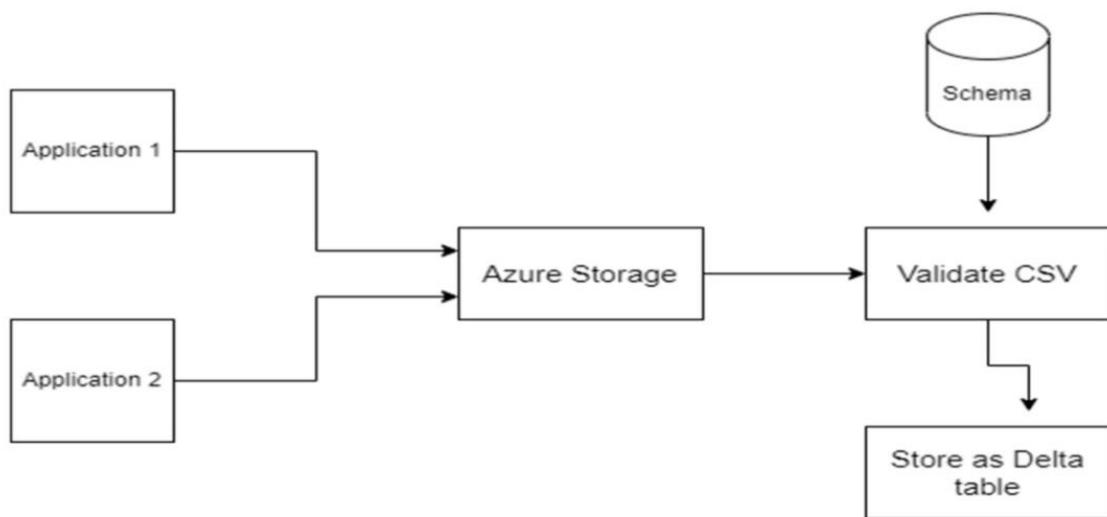
- Multiple Internal applications sends the data(huge size) in CSV format on daily basis in the cloud storage location. There are couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed data to be persisted as Delta table in Databricks for downstream system.



Project 2 : AP Morgan- High Level Detail

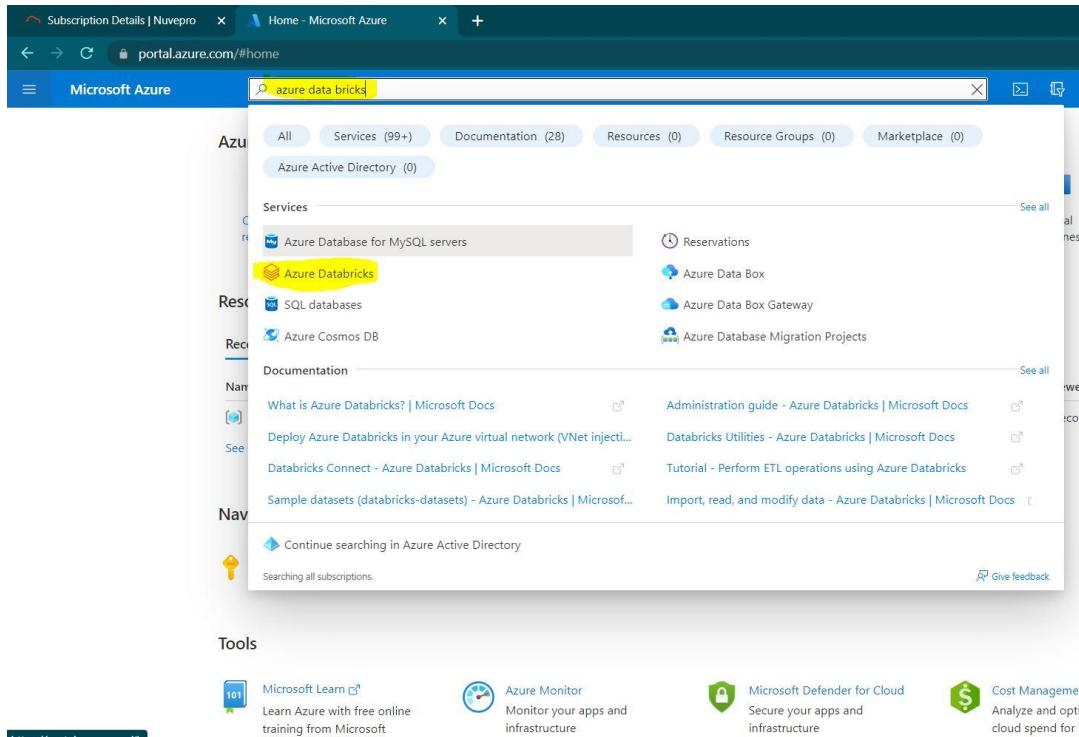
- Internal Application sends CSV file in Azure data lake storage.
- Validation needed to apply on this follows:
 - Check for duplicate rows. If it contains duplicate rows, file need to be rejected.
 - Need to validate the date format for all the date fields. Date column names and desired date format 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



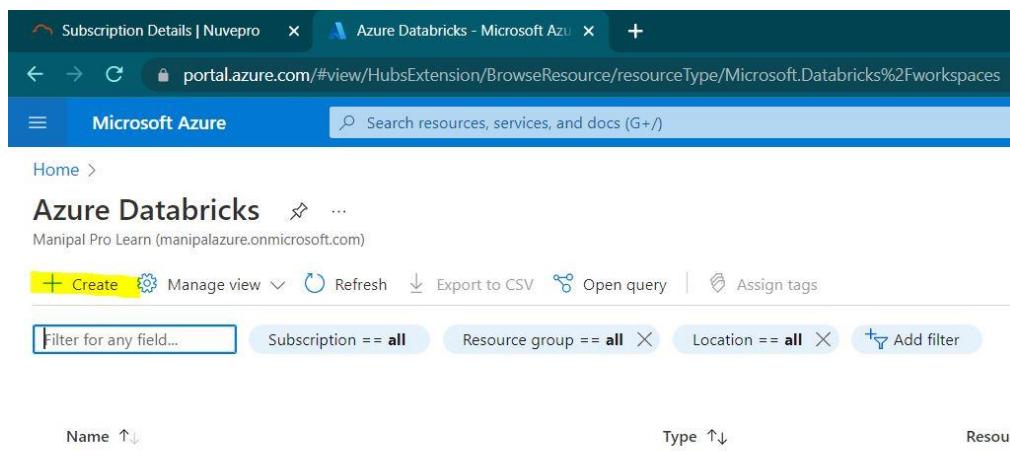
Practical Lab: Create a Databricks

Step1: Goto portal.azure.com and search for azure data bricks.



The screenshot shows the Microsoft Azure portal homepage. A search bar at the top contains the text "azure data bricks". Below the search bar, a navigation bar includes links for "All", "Services (99+)", "Documentation (28)", "Resources (0)", "Resource Groups (0)", and "Marketplace (0)". The main content area displays a list of services under "Services", with "Azure Database for MySQL servers" and "Azure Databricks" being the first two items. The "Azure Databricks" item is highlighted with a yellow box. Other services listed include "SQL databases", "Azure Cosmos DB", "Reservations", "Azure Data Box", "Azure Data Box Gateway", and "Azure Database Migration Projects". Below the services, there is a section for "Documentation" with several links related to Azure Databricks. At the bottom of the page, there are "Tools" sections for Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, and Cost Management.

Step2: click on create button.



The screenshot shows the Microsoft Azure portal with the URL "https://portal.azure.com/#view/HubsExtension/BrowseResource/resourceType/Microsoft.Databricks%2Fworkspaces" in the address bar. The page title is "Azure Databricks". The main content area shows a table with columns for "Name", "Type", and "Resource". At the top of the table, there are several filter and search options: "Create" (highlighted with a yellow box), "Manage view", "Refresh", "Export to CSV", "Open query", "Assign tags", "Filter for any field...", "Subscription == all", "Resource group == all", "Location == all", and "Add filter".



No azure databricks serv

Unlock insights from all your data and build artificial intelligence
Apache Spark environment in minutes, autoscale, and collaborate

Step3: Create a new resource group for the AP morgan group.

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 * ⓘ (New) AP_Morgan

Instance Details

Workspace name * Enter name for Databricks workspace

Region * West US

Pricing Tier * ⓘ Standard (Apache Spark, Secure with Azure AD)

Step4: give the name for the workspace and choose pricing tier as trial and later click 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 * ⓘ (New) AP_Morgan

Instance Details

Workspace name * apmorgan123

Region * West US

Pricing Tier * ⓘ Trial (Premium - 14 Days Free DBUs)

Step5: after successful validation click on create button.

The screenshot shows the 'Create an Azure Databricks workspace' wizard. At the top, there's a green banner with a checkmark icon and the text 'Validation Succeeded'. Below it, the 'Review + create' tab is selected. The 'Summary' section contains two tabs: 'Basics' and 'Networking'. Under 'Basics', the following details are listed:

Workspace name	apmorgan123
Subscription	Azure-DXC262AB12Lab
Resource group	AP_Morgan
Region	West US
Pricing Tier	trial

Under 'Networking', the following options are shown:

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)	No
Deploy Azure Databricks workspace in your own Virtual Network (VNet)	No

Under 'Advanced', the 'Enable Infrastructure Encryption' option is set to 'No'. At the bottom, there are three buttons: 'Create' (highlighted in yellow), '< Previous', and 'Download a template for automation'.

Step6: the deployment of the data bricks has been completed.

The screenshot shows the 'AP_Morgan_apmorgan123 | Overview' page. The left sidebar has tabs for 'Overview' (selected), 'Inputs', 'Outputs', and 'Template'. The main area displays deployment details. A prominent message says 'Your deployment is complete' with a checkmark icon. Below it, the deployment summary includes:

- Deployment name: AP_Morgan_apmorgan123
- Subscription: Azure-DXC262AB12Lab
- Resource group: AP_Morgan

At the bottom, there are links for 'Deployment details (Download)', 'Next steps', and a 'Go to resource' button.

Practical Lab: Create Cluster in Azure Databricks

Step1: goto overview of the data bricks deployment and click on launch workspace.

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and user account information. Below the navigation bar, the breadcrumb trail shows 'Home > AP_Morgan_apmorgan123 > apmorgan123'. The main content area displays the 'apmorgan123' Azure Databricks Service overview. The left sidebar contains sections for Overview, Activity log, Access control (IAM), Tags, Settings, Virtual Network Peerings, Encryption, Properties, Locks, Automation, Tasks (preview), Export template, Support + troubleshooting, and New Support Request. The right panel shows the 'Essentials' section with details such as Status: Active, Resource group: AP_Morgan, Location: West US, Subscription: Azure-DXC262AB12Lab, Managed Resource Group: databricks-rg-apmorgan123-lw23okv45yge, URL: https://adb-1849494649003159.19.azuredatabricks.net, and Pricing Tier: Trial (Premium - 14-Days Free DBUs). It also features a red 'Launch Workspace' button with a yellow arrow pointing to it. At the bottom, there are four buttons: Documentation, Getting Started, Import Data from File, and Import Data from Azure Storage.

Step2: choose an account to sign into the data bricks workspace

The screenshot shows the 'Sign In to Databricks' page. The header features the Azure Databricks logo and the text 'Azure Databricks'. The main content area has a light gray background with a central white box containing the 'Sign In to Databricks' heading. Below it, a message reads 'Sign in using Azure Active Directory Single Sign On.' A large button labeled 'Signing you in' with a circular progress icon is centered. At the bottom, a smaller message says 'Contact your site administrator to request access.'

Step3: Click on the create button in the menu and click on cluster

The screenshot shows the Microsoft Azure Databricks interface. On the left, there's a sidebar with icons for Notebook, Table, Cluster (which is highlighted), Job, Repo, and AutoML Experiment. The main content area has several cards: 'Data import' (with a red and blue icon), 'Partner Connect' (with a blue icon), and 'Guide: Quickstart tutorial' (with a yellow icon). Below these are sections for 'Last viewed', 'Documentation' (with links to 'Get started guide', 'Best practices', and 'Data guide'), 'Release notes' (with links to 'Runtime release notes', 'Azure Databricks preview releases', 'Platform release notes', and 'More release notes'), and 'Blog posts' (with links to 'Building ETL pipelines for the cybersecurity lakehouse with Delta Live Tables', 'Streaming Windows Event Logs into the Cybersecurity Lakehouse', and 'Speed Up Streaming Queries With Asynchronous State Checkpointing').

Step4: name the cluster, choose your required cluster mode and set termination time , then on create button.

The screenshot shows the 'Clusters / apmorgan123' configuration page. The 'Configuration' tab is active. It includes fields for 'Policy' (set to 'Unrestricted'), 'Cluster mode' (set to 'Single Node'), 'Databricks Runtime Version' (set to '10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)'), and 'Autopilot options' (checkbox for 'Terminate after [30] minutes of inactivity' is checked). Below these are sections for 'Node type' (set to 'Standard_DS3_v2') and 'DBU / hour' (set to '0.75'). At the bottom, there's a link to 'Advanced options'.

Step5: after successful creation of the cluster the green tick mark appears near the cluster name.

The screenshot shows the Microsoft Azure Databricks interface. At the top, it says "Microsoft Azure | Databricks". On the left is a sidebar with various icons: a cluster icon (highlighted with a red "NEW" badge), a plus sign, a document icon, a search icon, a cluster icon with a "NEW" badge, a gear icon, and a user icon. The main area shows a cluster named "apmorgan123" with a green checkmark icon next to it. Below the cluster name are tabs for Configuration, Notebooks (0), Libraries, Event log, Spark UI, Driver logs, Metrics, and Apps. Under Configuration, there are sections for Policy (Unrestricted), Cluster mode (Single Node), Databricks Runtime Version (10.4 LTS), and Autopilot options (checkbox for Photon Acceleration, checkbox for Terminate after 30 minutes). A Node type section shows "Standard_DS3_v2" with "14 GB Memory, 4 Cores". A DBU / hour value of "0.75" is also shown. At the bottom, there is a link to "Advanced options".

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

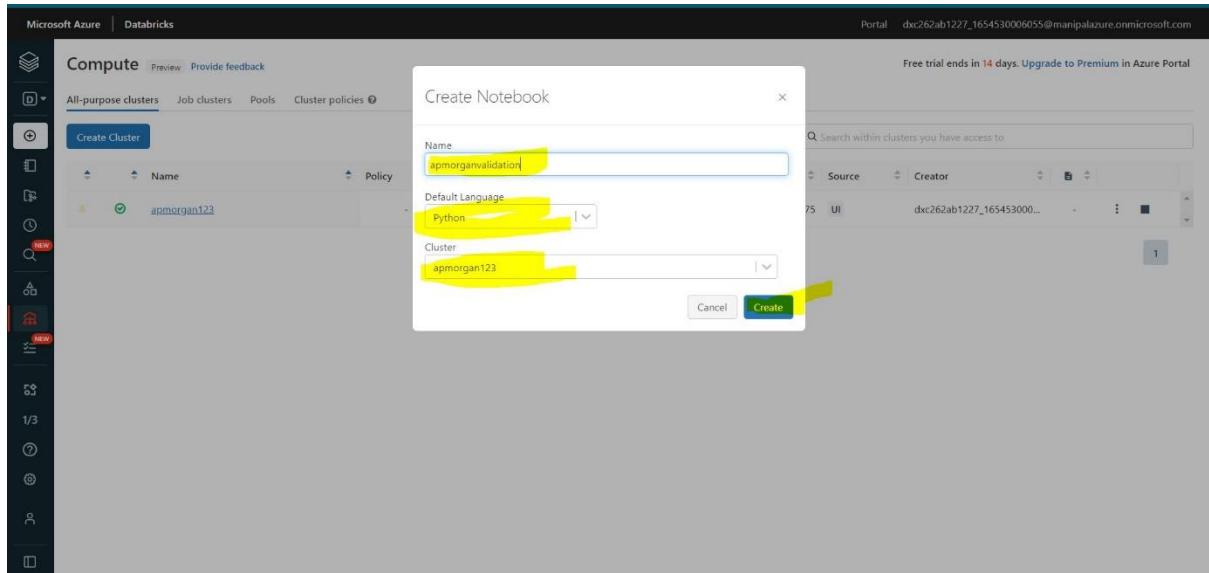
Step1: Go to create in the data bricks workspace.

The screenshot shows the Microsoft Azure Databricks Compute blade. At the top, there's a header with 'Microsoft Azure | Databricks' and a 'Portal' link. Below the header, it says 'Compute' with 'Preview' and 'Provide feedback' links. A navigation bar includes 'All-purpose clusters', 'Job clusters', 'Pools', and 'Cluster policies'. On the left, a sidebar has icons for 'Cluster', 'Table', 'Cluster', 'Job', 'Repo', and 'AutoML Experiment'. The main area displays a table with columns: Name, Policy, Runtime, Active memory, Active cores, Active DBU / h, Source, and Creator. One row is visible: 'apmorgan123' with values: 10.4, 14 GB, 4 cores, 0.75, UI, and 'dx262ab1227_1654530006055@manipalazure.onmicrosoft.com'. There are also 'Created by me' and 'Accessible by me' filters at the top of the table.

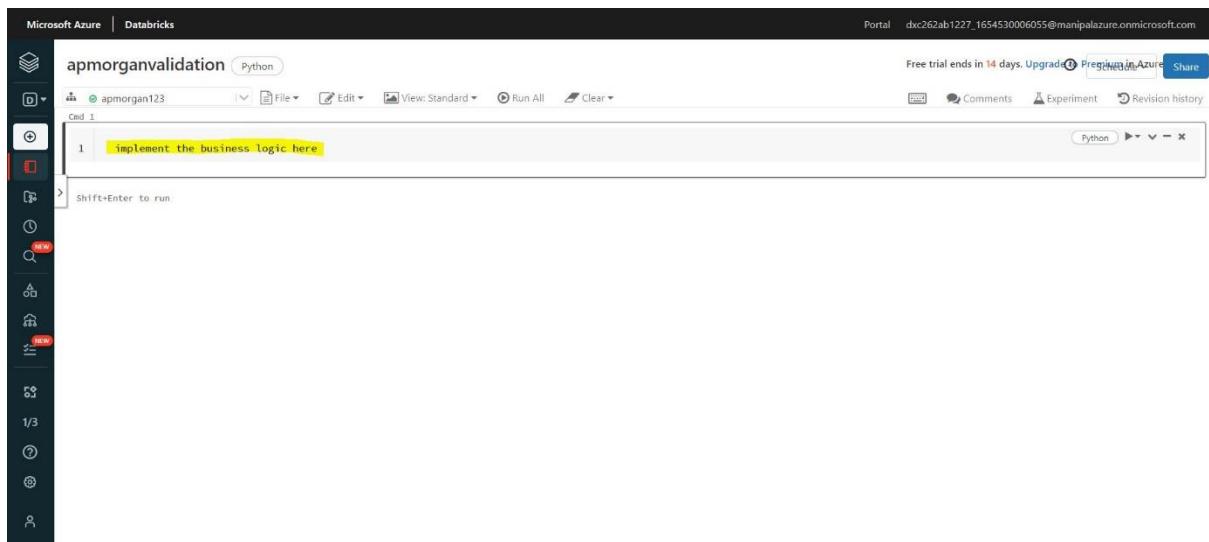
Step2: select notebook.

This screenshot is similar to the previous one, showing the Microsoft Azure Databricks Compute blade. The sidebar on the left now highlights the 'Notebook' icon. The main table data remains the same, showing the cluster 'apmorgan123' with its details. The navigation bar at the top still includes 'All-purpose clusters', 'Job clusters', 'Pools', and 'Cluster policies'.

Step3: name the notebook and assign the cluster to it.



Step4: You can now implement your business logic in the notebook provided.



Practical Lab: Azure Data Factory For AP Morgan

Step1: Goto portal.azure.com and search data factories.

The screenshot shows the Microsoft Azure portal interface. At the top, there are three tabs: 'Subscription Details | Nuvapro', 'Home - Microsoft Azure', and 'apmorganvalidation - Databricks'. Below the tabs, the address bar shows 'portal.azure.com/#home'. A search bar contains the text 'datafactory'. The main content area has a title 'Microsoft Azure' with a dropdown menu. Below it, a search bar also contains 'datafactory'. A navigation bar at the bottom includes links for 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboards'.

Step2: click on the create button.

The screenshot shows the 'Data factories' blade in the Microsoft Azure portal. The top navigation bar has 'Data factories' highlighted. Below the navigation bar, there are several buttons: '+ Create', 'Manage view', 'Refresh', 'Export to CSV', 'Open query', and 'Assign tags'. There is also a 'Filter for any field...' dropdown and filters for 'Subscription == all', 'Type == all', 'Resource group == all', and 'Location == all'. The main content area displays a message: 'No data factories to display. Try changing or clearing your filters.' It includes a 'Create data factory' button and a 'Learn more' link.

Step3: give the project details as well as the instance details and click on git configuration.

⚠ 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 *

Resource group *
Create new

Instance details

Name *

Region *

Version *

Review + create < Previous Next : Git configuration >

Step4: click on configure git later and later review and create.

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

Home > Data factories > Create Data Factory ...

Basics **Git configuration** Networking Advanced Tags Review + create

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 about Git integration in Azure Data Factory

Configure Git later

Review + create < Previous Next : Networking >

Step5: after successful validation click on create button.

The screenshot shows the 'Create Data Factory' wizard in the Microsoft Azure portal. At the top, there's a green banner with a checkmark and the text 'Validation Passed'. Below it, the 'Review + create' tab is selected. The 'TERMS' section contains a legal agreement. Under 'Basics', the configuration includes:

Subscription	Azure-DXC262AB12Lab
Resource group	AP_Morgan
Name	apmorgandata12
Region	East US
Version	V2 (Recommended)

Under 'Networking', 'Connect via' is set to 'Public endpoint'. At the bottom, there are 'Create', '< Previous', 'Next >', and 'Download a template for automation' buttons.

Step6: your data factory is created.

The screenshot shows the 'apmorgandata12' Data Factory overview page. The left sidebar has sections like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (Networking, Managed identities, Properties, Locks), Getting started (Quick start), Monitoring (Alerts, Metrics, Diagnostic settings, Logs), and Pipelines. The main area shows the 'Essentials' section with details:

Resource group (move)	: AP_Morgan	Type	: Data factory (V2)
Status	: Succeeded	Getting started	: Quick start
Location	: East US		
Subscription (move)	: Azure-DXC262AB12Lab		
Subscription ID	: 3a28cdce-3bd7-4219-858e-23ff20f8b998		

Below this are sections for 'Getting started' (with links to 'Open Azure Data Factory Studio' and 'Read documentation') and 'Monitoring' (with PipelineRuns and ActivityRuns charts).

Practical Lab: Create Azure Databricks Linked Service in ADF

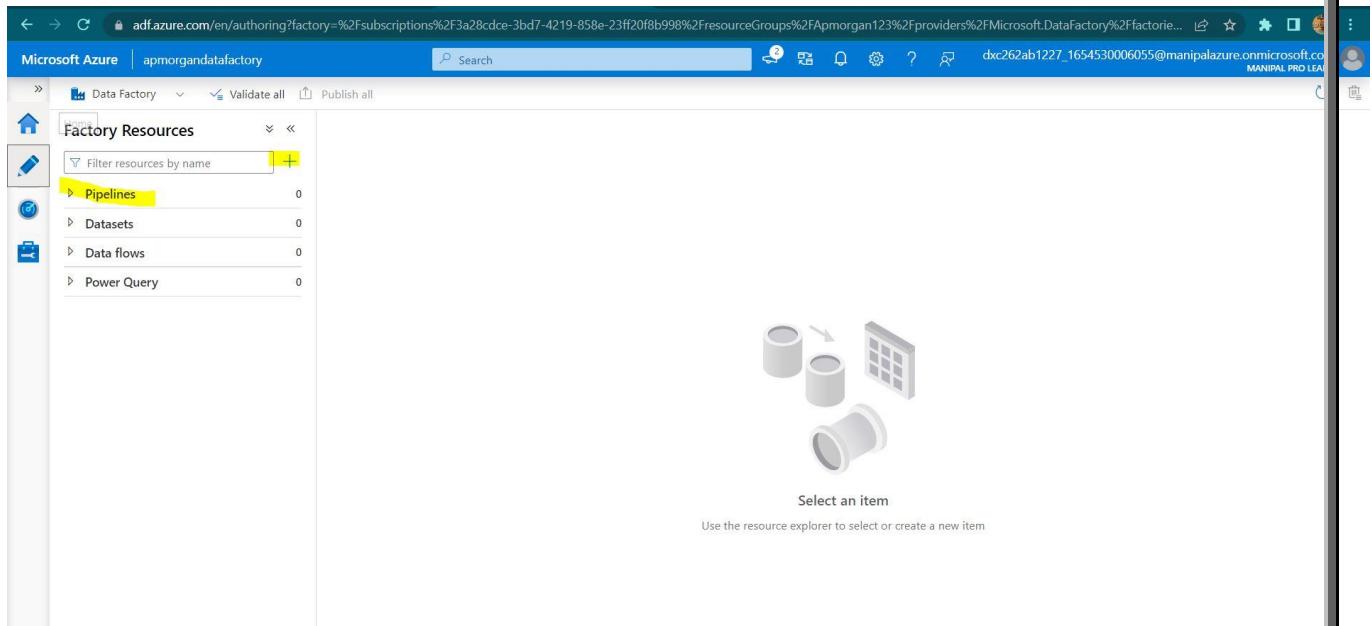
Step1: open the overview page of the data factory and click on the azure data factory studio.

The screenshot shows the Microsoft Azure portal interface for a Data Factory named 'apmorgandata12'. The left sidebar contains navigation links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Networking, Managed identities, Properties, Locks, Getting started (Quick start, Diagnostic settings, Logs), Monitoring (Alerts, Metrics), and PipelineRuns. The main content area displays 'Getting started' with a 'Open Azure Data Factory Studio' button and a 'Read documentation' section. Below this is a 'Monitoring' section with two charts: 'PipelineRuns' and 'ActivityRuns', both showing zero activity over the specified time period.

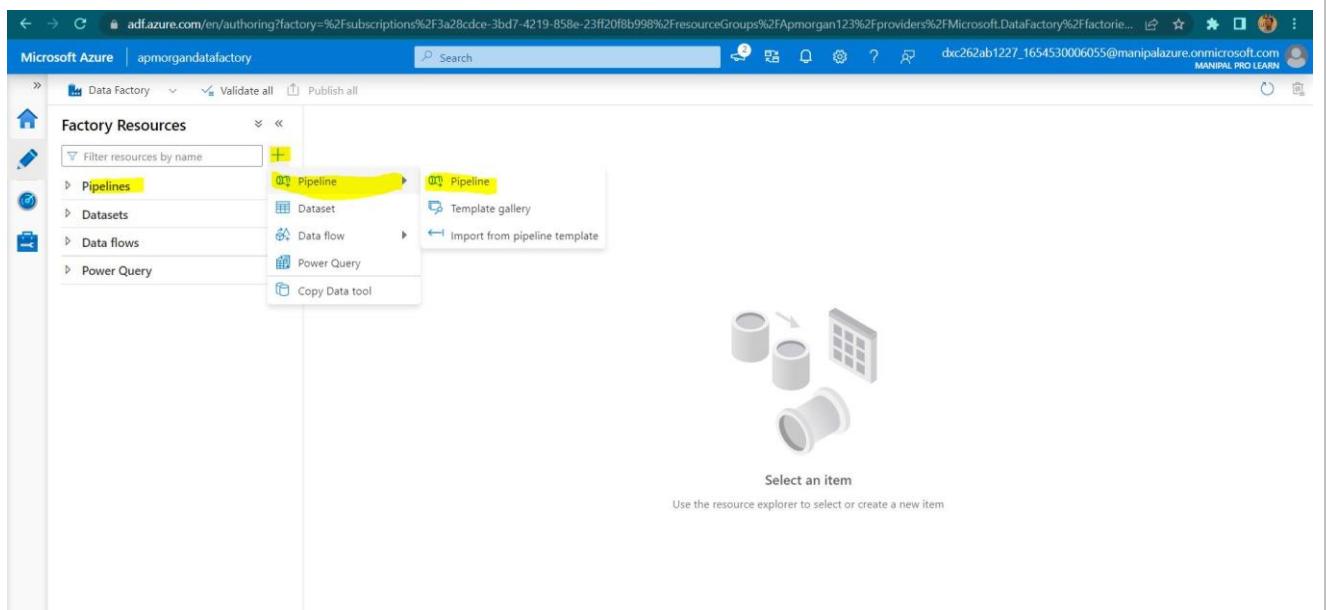
Step2: Click on compose button.

The screenshot shows the Azure Data Factory studio interface for the 'apmorgandatafactory' data factory. The top navigation bar includes 'Subscription Details | Navopro', 'apmorgandatafactory - Microsoft', 'apmorgandatafactory - Azure DevOps', 'apmorgandatafactory - Databricks', and 'Inbox (10,000) - jdkjankarsh11'. The main content area features a large central diagram illustrating data flow between various components like databases, tables, and cloud storage. Below the diagram are four primary buttons: 'Ingest' (Copy data at scale once or on a schedule), 'Orchestrate' (Code-free data pipelines), 'Transform data' (Transform your data using data flows), and 'Configure SSIS' (Manage & run your SSIS packages in the cloud). At the bottom, there are sections for 'Discover more' (Browse partners (preview), Pipeline templates, SAP pipeline templates) and 'Recent resources'.

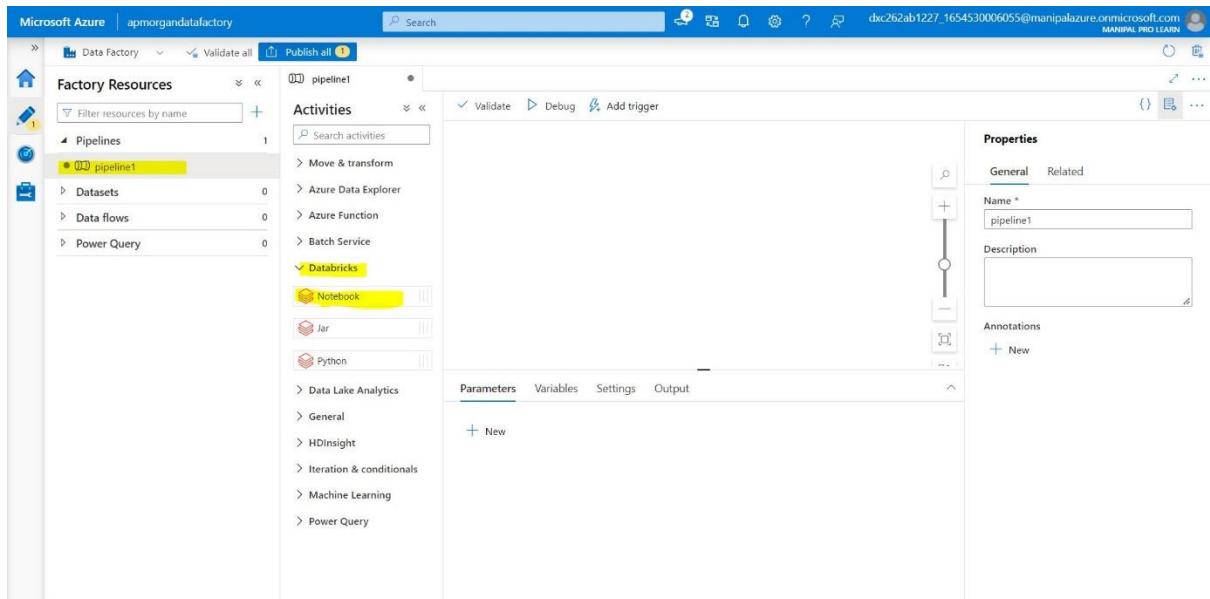
Step3: click on the + symbol



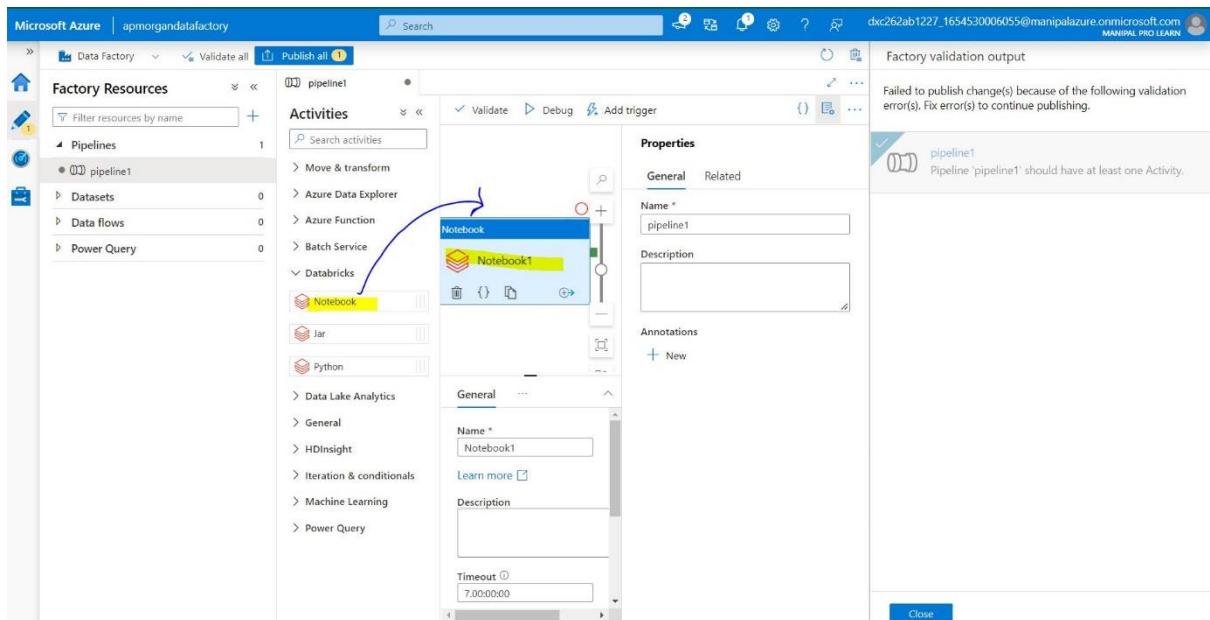
Step4: click on pipeline> choose pipeline again



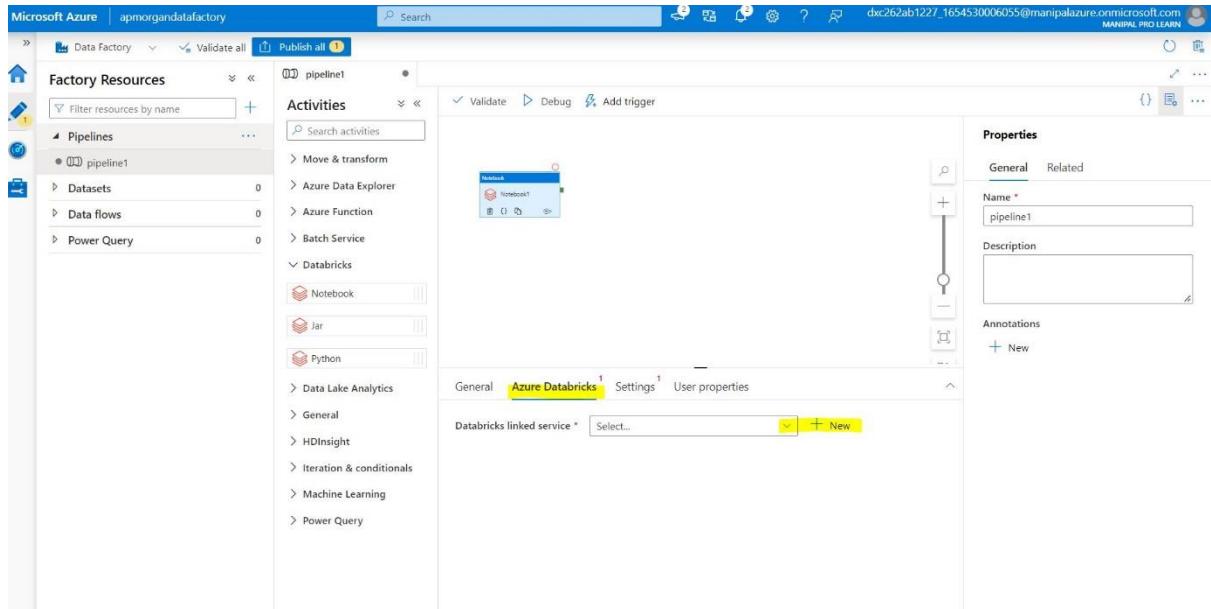
Step5: Choose databricks , Notebook under the activities.



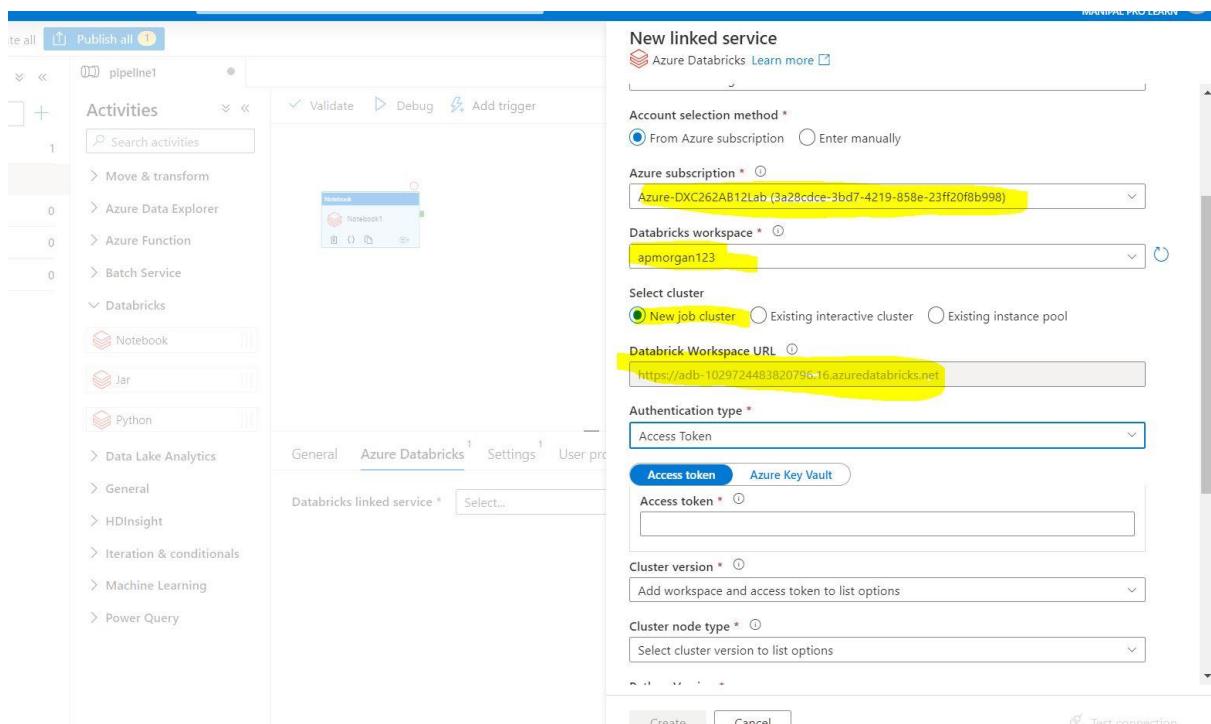
Step6: drag and drop the notebook in the space provided.



Step7: click on azure data bricks and select the new option near data bricks linked service.



Step8: select the workspace , azure subscription, databricks workspace URL. Select authentication type as Access token.



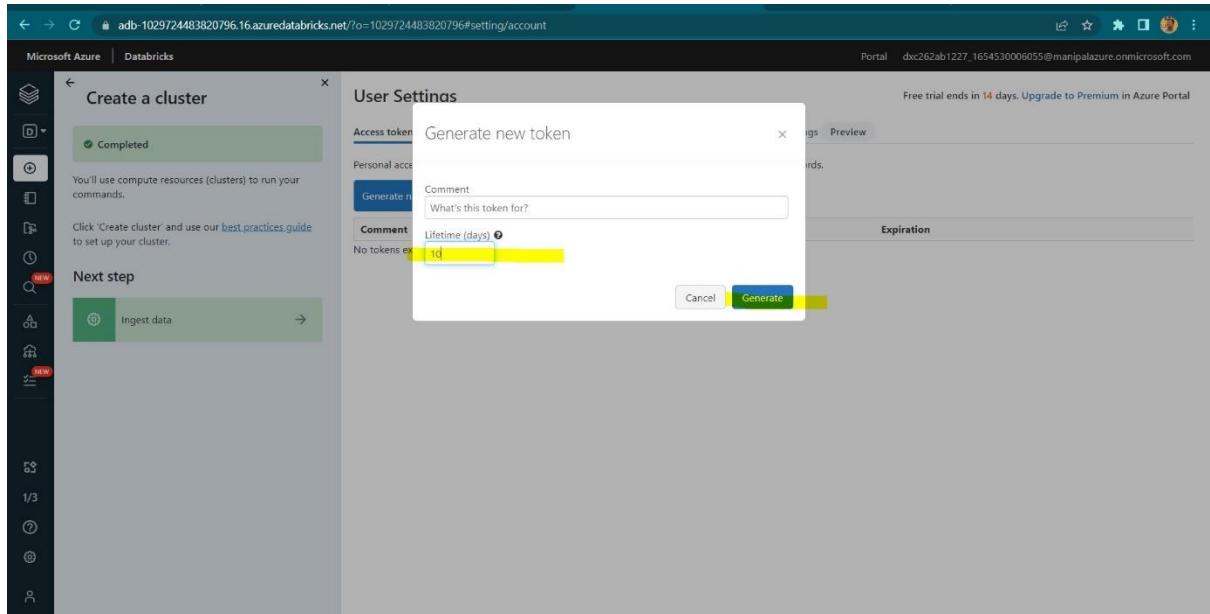
Step9: open the data bricks workspace , click on settings , user settings.

The screenshot shows the Microsoft Azure Databricks workspace. On the left, there's a sidebar with various icons and a progress bar indicating '1/3' completed. The main area has a title 'apmorgan234' and a tab for 'Python'. Below the title, it says 'Completed' and provides instructions to use compute resources (clusters) to run commands. A 'Next step' section shows a green button labeled 'Ingest data'. At the bottom of the main area, there's a 'User Settings' button highlighted with a yellow box. The URL in the browser is adb-1029724483820796.16.azuredatabricks.net/?o=1029724483820796#notebook/629522128091127/command/629522128091128.

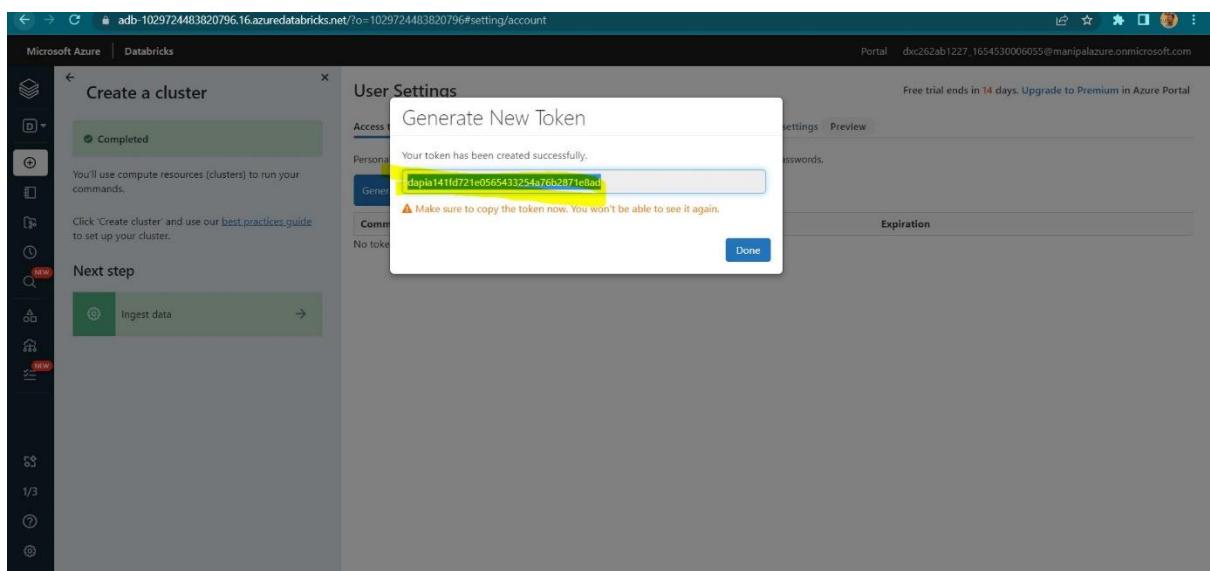
Step10: click on generate new token.

This screenshot is similar to the previous one, showing the 'User Settings' page in the Databricks workspace. The 'Generate new token' button is highlighted with a yellow box. The URL in the browser is adb-1029724483820796.16.azuredatabricks.net/?o=1029724483820796#setting/account. The rest of the interface is identical to the previous screenshot.

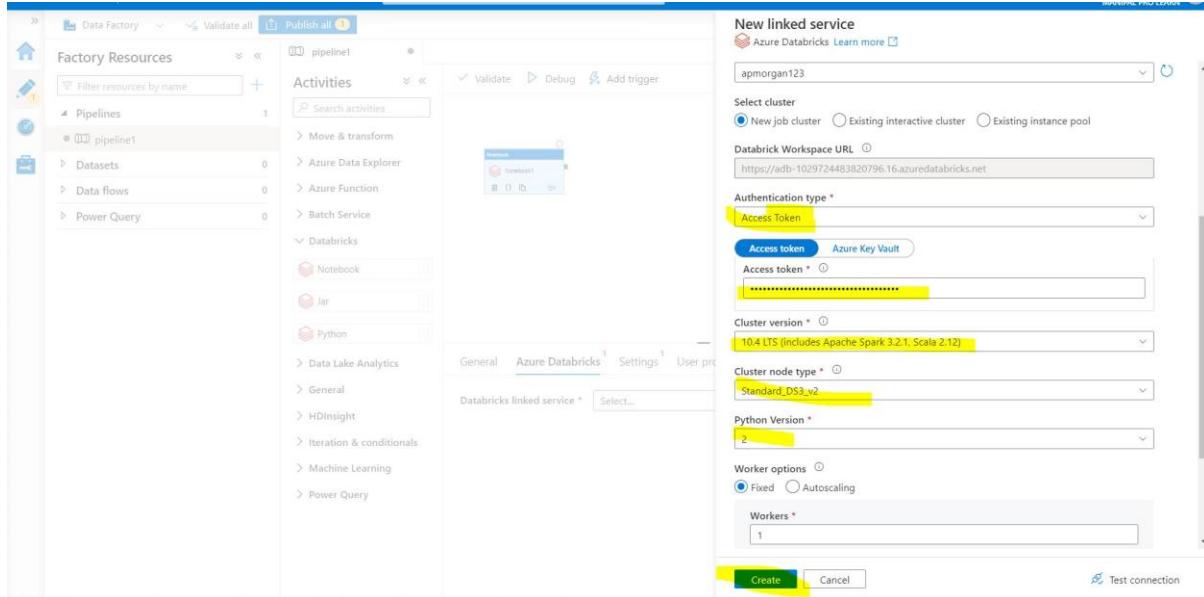
Step11: give the life time of the token and click on generate.



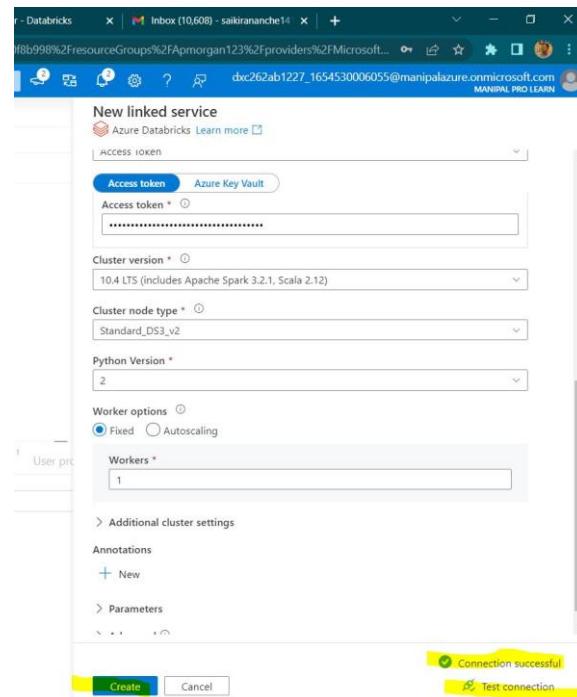
Step12: copy the generated access token.



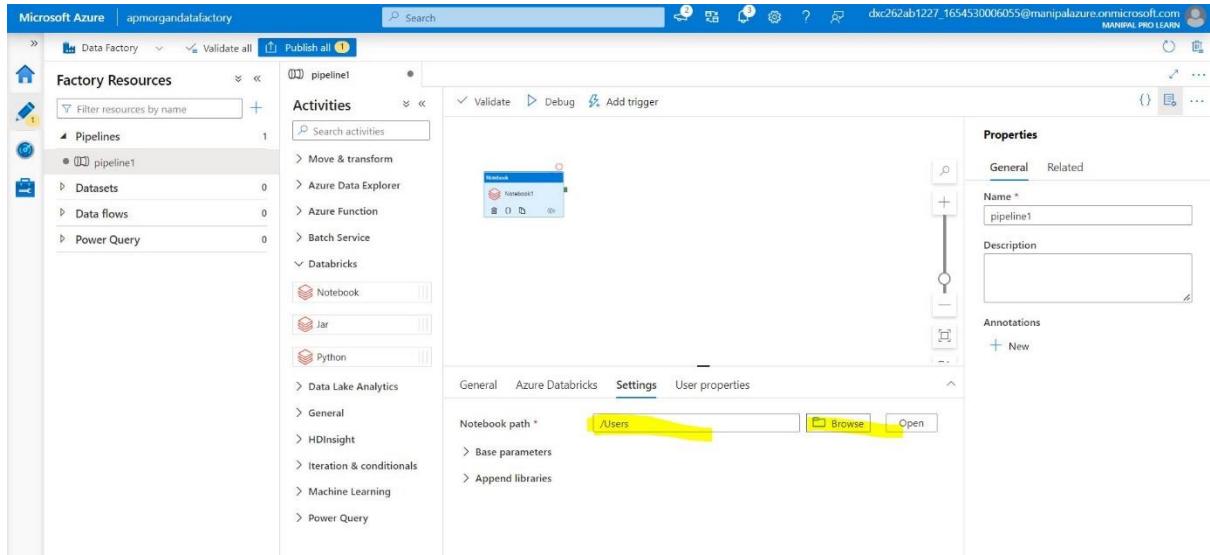
Step13: put the access token and select cluster version,cluster node type, version etc and test the connection.



Step14: after the connection gets success click on create.

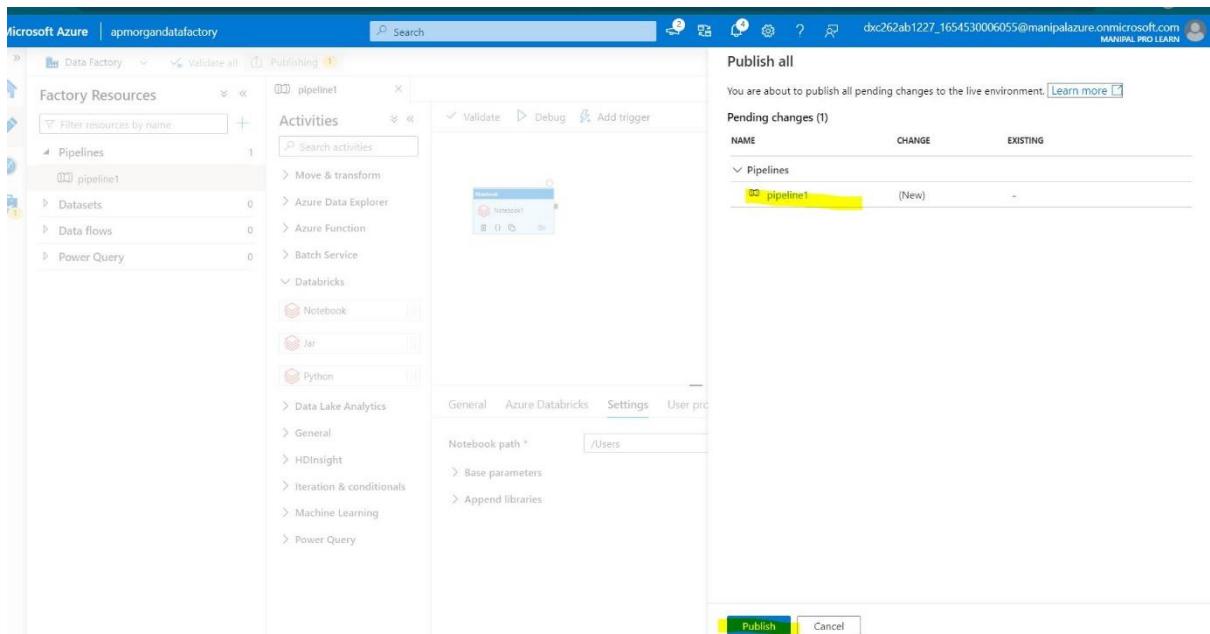


Step15: select the proper path of the notebook.



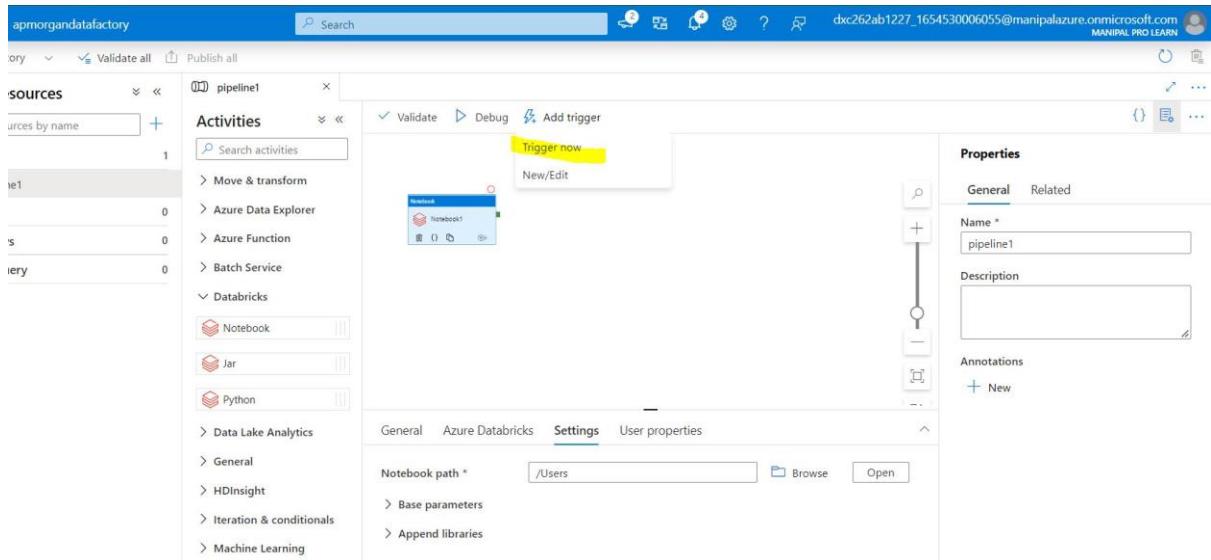
The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar is open, showing 'Pipelines' with one item named 'pipeline1'. In the main workspace, there is a single activity node labeled 'Notebook'. To the right, the 'Properties' panel is open for 'pipeline1', with the 'General' tab selected. The 'Name' field is set to 'pipeline1'. The 'Settings' tab is active, showing the 'Notebook path' field with the value '/Users' highlighted by a yellow box. Below it are 'Base parameters' and 'Append libraries' sections. At the bottom of the Properties panel, there are 'Annotations' and a '+ New' button.

Step16: publish the pipeline in-order to save.

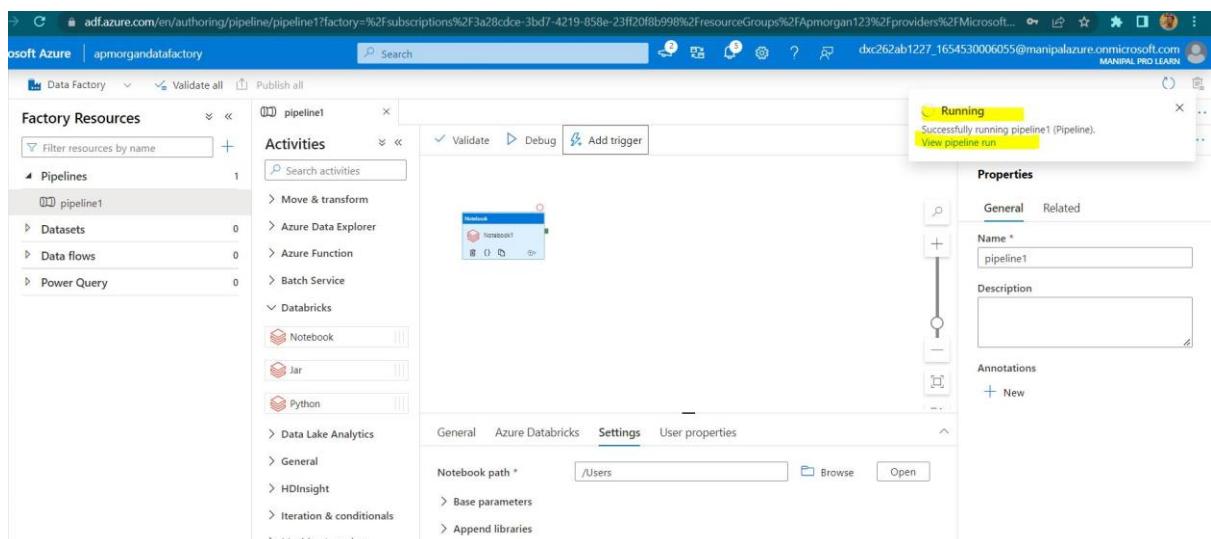


The screenshot shows the Microsoft Azure Data Factory pipeline editor with the 'Publishing' tab selected. The top bar indicates 'Publishing 1'. The 'Pending changes (1)' section shows a table with one row: NAME: pipeline1, CHANGE: (New), EXISTING: -. The 'Activities' pane shows a single 'Notebook' activity. The 'Properties' panel is partially visible on the right. At the bottom, there are 'Publish' and 'Cancel' buttons, with 'Publish' highlighted by a yellow box.

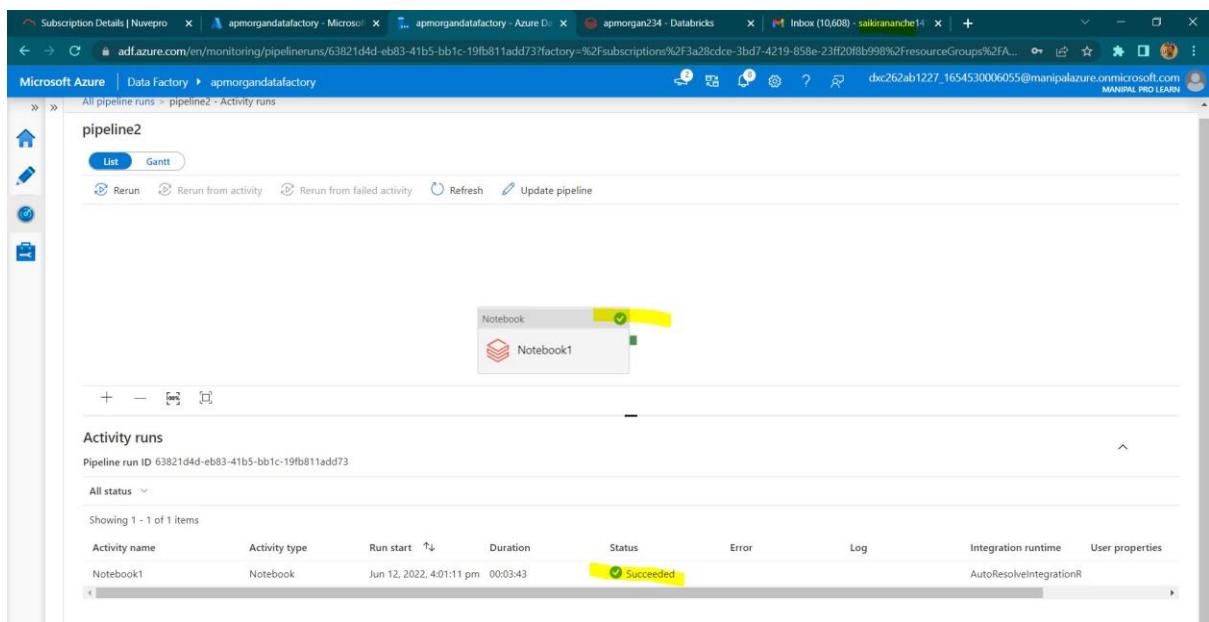
Step17: trigger the pipeline by clicking on add trigger >trigger now.



Step18: You can see the notification running.



Step19: you can see your pipeline run by opting the monitor Option in the DF and hence the azure data bricks based ADF service is created.



The screenshot shows the Microsoft Azure Data Factory interface for monitoring pipeline runs. The top navigation bar includes links for Subscription Details, Nuvepro, apmorgandatafactory - Microsoft Edge, apmorgandatafactory - Azure Databricks, apmorgan234 - Databricks, and Inbox (10,608) - sakuranche14. The main title is "Microsoft Azure | Data Factory > apmorgandatafactory". Below the title, the path "All pipeline runs > pipeline2 - Activity runs" is shown. On the left, there's a sidebar with icons for Home, Create, Monitor, and Help. The main content area is titled "pipeline2" and shows a "List" view. At the top of the list, there are buttons for "Rerun", "Rerun from activity", "Rerun from failed activity", "Refresh", and "Update pipeline". The first item in the list is a "Notebook" activity named "Notebook1", which has a green checkmark icon indicating it succeeded. Below this, the "Activity runs" section displays a table with one row:

Activity name	Activity type	Run start	Duration	Status	Error	Log	Integration runtime	User properties
Notebook1	Notebook	Jun 12, 2022, 4:01:11 pm	00:03:43	Succeeded			AutoResolveIntegrationR	

Result:

Most of the project requirements have been fulfilled by the step by step procedures.

Conclusion:

All the cases have been solved successfully, it was so challenging when few of the pipelines got failed. After referring the error messages I'm able to run the pipelines successfully and had a great experience while learning the concepts in deep.

Apart of the demonstrations that were explained in the tutorials I'm able to perform a variety of new operations with a lot of enthusiasm and engagement in learning.