

---

# DXC REALTIME PROJECTS

---

AZ-900, DP - 203



JUNE 10, 2022  
[DXC TECHNOLOGY]

**Name:** Bhogadi Naga Iswarya Lakshmi

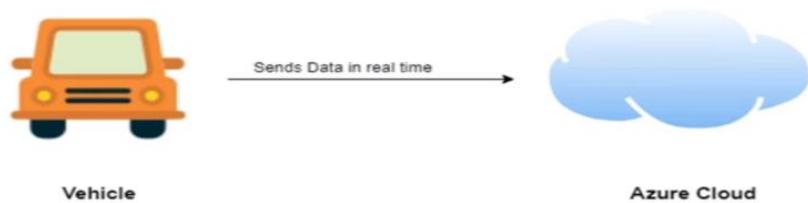
**Reg No:** DXC262AB12005

**Project1 Name:** Smart Vehicles

**Date:**10-06-2022

## Project 1 : Connected Vehicles

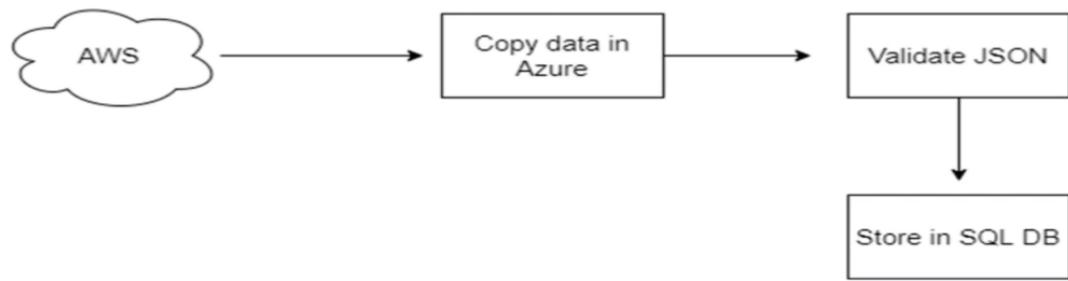
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



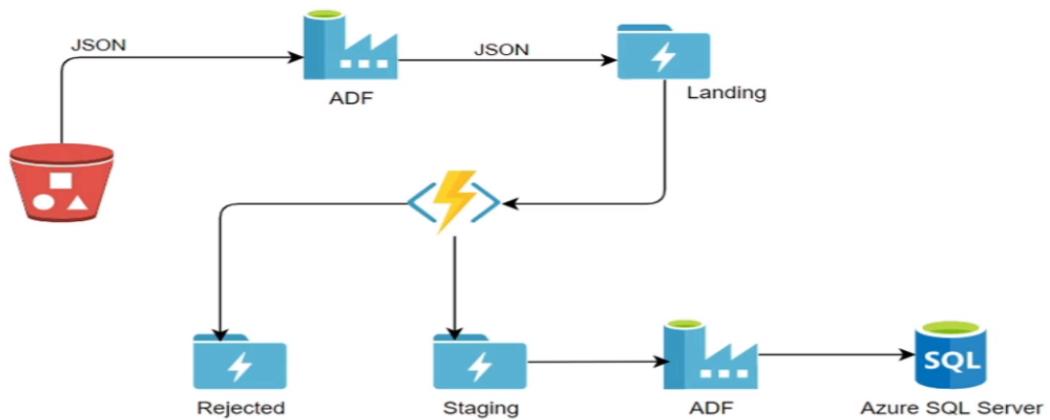
## Project 1 : Connected Vehicles

- 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

Azure Data Factory is a cloud-based data integration service that orchestrates and automates the movement and transformation of data.

**Step-1:** First open your Microsoft Azure account and search for Data Factories.

The screenshot shows the Microsoft Azure portal interface. The search bar at the top contains the query "data factories". Below the search bar, there are several tabs: All, Services (32), Documentation (27), Resources (0), Resource Groups (0), and Marketplace (0). The "Services" tab is selected. Under the "Services" heading, "Data factories" is highlighted with a blue selection bar. Other service options listed include Data Catalog, Data Shares, Azure Databricks, and Managed databases. To the right of the main search area, there is a sidebar titled "See all services" with a large arrow icon pointing right.

## Step 2: Click on create to create a new data factory account.

The screenshot shows the "Data factories" blade in the Microsoft Azure portal. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information (dx262ab1205\_1654530... MANIPAL PRO LEARN (MANIPAL...)). Below the navigation, the breadcrumb path is "Home > Data factories". The main content area displays a table header with columns: Name, Type, Subscription, Resource group, and Location. A message "No data factories to display" is centered, along with a note "Try changing or clearing your filters." and a "Create data factory" button.

## Step 3: Complete the Project and Instance details.

The screenshot shows the "Create Data Factory" blade in the Microsoft Azure portal. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information (dx262ab1205\_1654530... MANIPAL PRO LEARN (MANIPAL...)). The breadcrumb path is "Home > Data factories > Create Data Factory". The main form is divided into two sections: "Project details" and "Instance details".

**Project details:**

- Subscription: Azure-DXC262AB12Lab
- Resource group: dxcrg2317 (with a "Create new" option)

**Instance details:**

- Name: vehiclesdf123
- Region: East US
- Version: V2 (Recommended)

At the bottom of the form are three buttons: "Review + create", "< Previous", and "Next : Git configuration >".

## Step 4: After completing those details continue to Git configuration.

The screenshot shows the 'Create Data Factory' wizard on the 'Git configuration' tab. At the top, there's a navigation bar with 'Microsoft Azure' and a search bar. Below it, the breadcrumb path is 'Home > Data factories > Create Data Factory'. The main content area has tabs for 'Basics', 'Git configuration' (which is selected), 'Networking', 'Advanced', 'Tags', and 'Review + create'. A note says '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.' Below this, there's a checkbox labeled 'Configure Git later' with a checked status. At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next : Networking >'.

Step 5: After step-4, continue the completion of details of Networking ,Advanced ,Tags and finally goes to Review and Create and also completes validation process.

The screenshot shows the 'Create Data Factory' wizard on the 'Review + create' tab. At the top, there's a navigation bar with 'Microsoft Azure' and a search bar. Below it, the breadcrumb path is 'Home > Data factories > Create Data Factory'. The main content area shows a green banner with a checkmark icon and the text 'Validation Passed'. Below this, there's a 'TERMS' section with a note about agreeing to legal terms and privacy statements. The 'Basics' section shows subscription 'Azure-DXC262AB12Lab' and resource group 'dxcrg2317'. At the bottom, there are buttons for 'Create' (which is highlighted in blue), '< Previous', 'Next', and 'Download a template for automation'.

Step 6: After completing validation process, Click on create and Deployment process starts.

The screenshot shows the Microsoft Azure Data Factory Deployment Overview page. At the top, there's a search bar and a navigation bar with icons for Home, Search, Delete, Cancel, Redeploy, Refresh, and Help. The main title is "Microsoft.DataFactory-20220610162942 | Overview". On the left, a sidebar has tabs for Overview, Inputs, Outputs, and Template, with "Overview" selected. A feedback link "We'd love your feedback!" is present. The main content area displays deployment details: "Deployment is in progress", "Deployment name: Microsoft.DataFactory-20220610162942", "Subscription: Azure-DXC262AB12Lab", "Resource group: dxrcg2317", "Start time: 6/10/2022, 4:35:37 PM", and "Correlation ID: d9b03e83-0cbd-4fb8-85c8-35fb43ad8852". Below this, a table titled "Deployment details (Download)" shows "No results." under columns for Resource, Type, Status, and Operation details.

Step 7: After completing deployment process, click on goto resource.

The screenshot shows the Microsoft Azure Data Factory Deployment Overview page after deployment completion. The main title is "Microsoft.DataFactory-20220610162942 | Overview". The deployment status is now "Your deployment is complete". The deployment details show the same information as before: "Deployment name: Microsoft.DataFactory-20220610162942", "Subscription: Azure-DXC262AB12Lab", "Resource group: dxrcg2317", "Start time: 6/10/2022, 4:35:37 PM", and "Correlation ID: d9b03e83-0cbd-4fb8-85c8-35fb43ad8852". Below the deployment details, there are sections for "Deployment details (Download)", "Next steps", and a prominent blue "Go to resource" button. To the right, there are two side-panel cards: "Cost Management" (with a green dollar sign icon) and "Microsoft Defender for Cloud" (with a shield icon).

Step 8: Then ,click on Open Azure Data Factory Studio and start authoring & monitoring of your data pipelines.

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

### Practical Lab: Create Azure blob trigger logic

Blob source(Input)->ADF->Blob Destination(Output).

**Step-1:** First open your Microsoft Azure account and search for Storage Accounts.

The screenshot shows the Microsoft Azure search interface. The search bar at the top contains the query "storage accounts". Below the search bar, there are several tabs: All, Services (19), Marketplace (1), Documentation (28), Resources (0), and Resource Groups (0). The "All" tab is selected. Under the "Services" section, "Storage accounts" is highlighted. Other services listed include Azure Active Directory (0), Genomics accounts, Storage accounts (classic), Integration accounts, Automation Accounts, Lab accounts, Batch accounts, Storage Explorer, and Marketplace. There are also sections for Recent, Favorties, and Marketplace. At the bottom, there are links for "Storage account overview - Azure Storage | Microsoft Docs" and "Introduction to Azure Storage - Cloud storage on Azure | Microsoft...". A "See all" link is visible on the right side.

## Step 2: Click on create to create a new storage account.

The screenshot shows the Microsoft Azure Storage accounts list page. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information (dx262ab1205\_1654530... MANIPAL PRO LEARN (MANIPAL...)). Below the navigation is a breadcrumb trail: Home > Storage accounts. The main area displays a table with columns: Name, Type, Kind, Resource group, Location, and Subscription. A "Create" button is located at the top left of the table. Filter options include "Subscription == all", "Resource group == all", "Location == all", and "Add filter". Below the table, a message states "No storage accounts to display". A descriptive text block explains how to create a storage account, mentioning Blob storage account and access tiers. A "Give feedback" link is at the bottom right.

## Step 3: Complete the Project and Instance details.

The screenshot shows the Microsoft Azure "Create a storage account" wizard, specifically the "Basics" step. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information (dx262ab1205\_1654530... MANIPAL PRO LEARN (MANIPAL...)). Below the navigation is a breadcrumb trail: Home > Storage accounts > Create a storage account. The main area has tabs for Basics, Advanced, Networking, Data protection, Encryption, Tags, and Review + create. The Basics tab is selected. A descriptive text block explains that Azure Storage is a Microsoft-managed service providing cloud storage. It mentions that the cost of your storage account depends on usage and options chosen. A "Project details" section asks for the subscription to create the new storage account. A dropdown menu shows "Azure-DXC262AB12Lab". Navigation buttons at the bottom include "Review + create" (highlighted in blue), "< Previous", and "Next : Advanced >".

## Step 4: After completing those details continue to Advanced.

The screenshot shows the 'Create a storage account' wizard on the 'Advanced' tab. At the top, there's a note: 'Certain options have been disabled by default due to the combination of storage account performance, redundancy, and region.' Below this, the 'Security' section is visible, containing three checked checkboxes: 'Require secure transfer for REST API operations', 'Enable blob public access', and 'Enable storage account key access'. At the bottom of the screen are navigation buttons: 'Review + create' (highlighted in blue), '< Previous', and 'Next : Networking >'.

Step 5: After step-4, continue the completion of details of Networking, Data protection, Encryption, Tags and finally goes to Review and Create and also completes validation process.

The screenshot shows the 'Create a storage account' wizard on the 'Review + create' tab. A green banner at the top indicates 'Validation passed'. Below it, the 'Basics' section displays the following configuration details:

Subscription	Azure-DXC262AB12Lab
Resource Group	dxcrg2317
Location	eastus
Storage account name	vehicles231
Deployment model	Resource manager
Performance	Standard
Replication	Read-access geo-redundant storage (RA-GRS)

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

Step 6: After completing validation process, Click on create and Deployment process starts.

The screenshot shows the Microsoft Azure Deployment Overview page for a resource named 'vehicles231\_1654860223171'. The status bar indicates 'Deployment is in progress'. Deployment details include a name, start time (6/10/2022, 4:53:53 PM), subscription, correlation ID, and resource group. A table shows the resource 'vehicles231' with type 'Microsoft.Storage/storageAccounts' and status 'Accepted'. A sidebar on the right provides links to Microsoft Defender for Cloud, free tutorials, and work with experts.

Step 7: After completing deployment process, click on goto resource.

The screenshot shows the Microsoft Azure Deployment Overview page for the same resource. The status bar now indicates 'Your deployment is complete'. Deployment details remain the same. A 'Next steps' section contains a 'Go to resource' button. A sidebar on the right provides links to Cost Management and Microsoft Defender for Cloud.

Step 8: Click on Goto resource and click containers

The screenshot shows the Microsoft Azure Storage account 'vehicles231' Containers page. It lists a single container named '\$logs' with details: Last modified (6/12/2022, 9:42:35 AM), Public access level (Private), and Lease state (Available). The sidebar on the left shows other storage options like Activity log, Tags, and Data migration.

Click on +container to create a new container to store the data.

The screenshot shows the Microsoft Azure Storage account interface for the 'vehicles231' account. The left sidebar has 'Containers' selected under 'Data storage'. The main area displays a list of existing containers: '\$logs' (Last modified: 6/12/2022, 9:42:35 AM, Public access level: Private). A 'New container' dialog is open on the right, prompting for a name ('\$newcontainer') and setting the public access level to 'Private (no anonymous access)'. At the bottom right of the dialog are 'Create' and 'Discard' buttons.

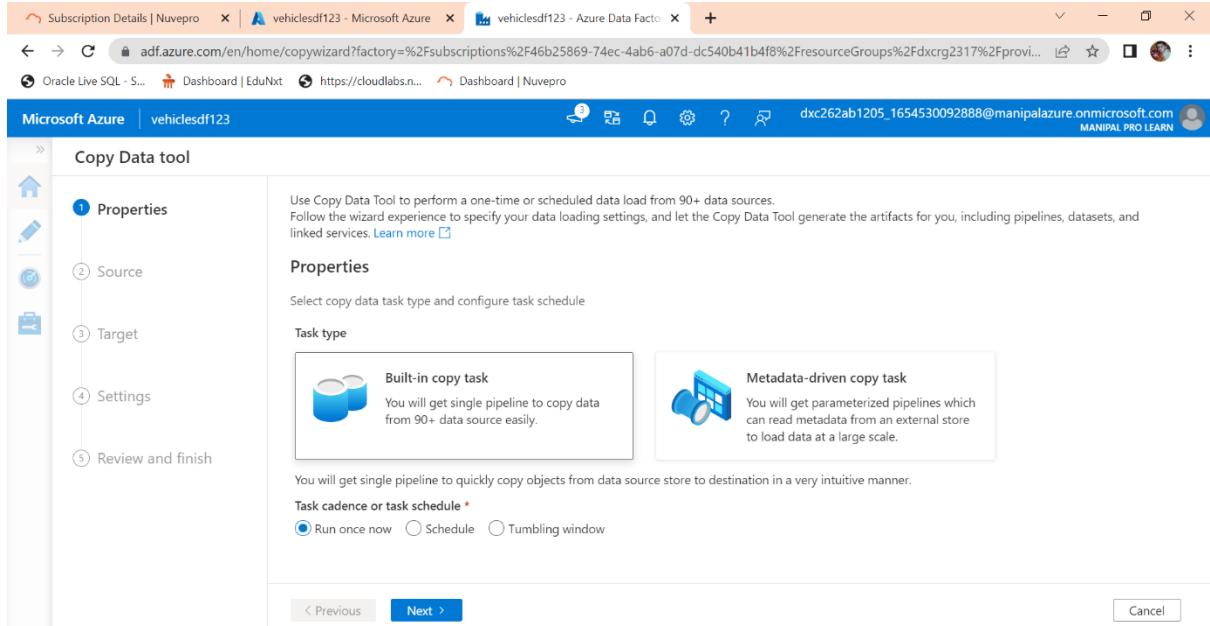
Create a new source container.

The screenshot shows the Microsoft Azure Storage account interface for the 'vehicles231' account. The left sidebar has 'Containers' selected under 'Data storage'. The main area displays a list of existing containers: '\$logs' and 'source' (Last modified: 6/12/2022, 9:55:18 AM, both Private and Available). A 'Show deleted containers' toggle switch is visible. At the bottom right of the list area are 'Create' and 'Discard' buttons.

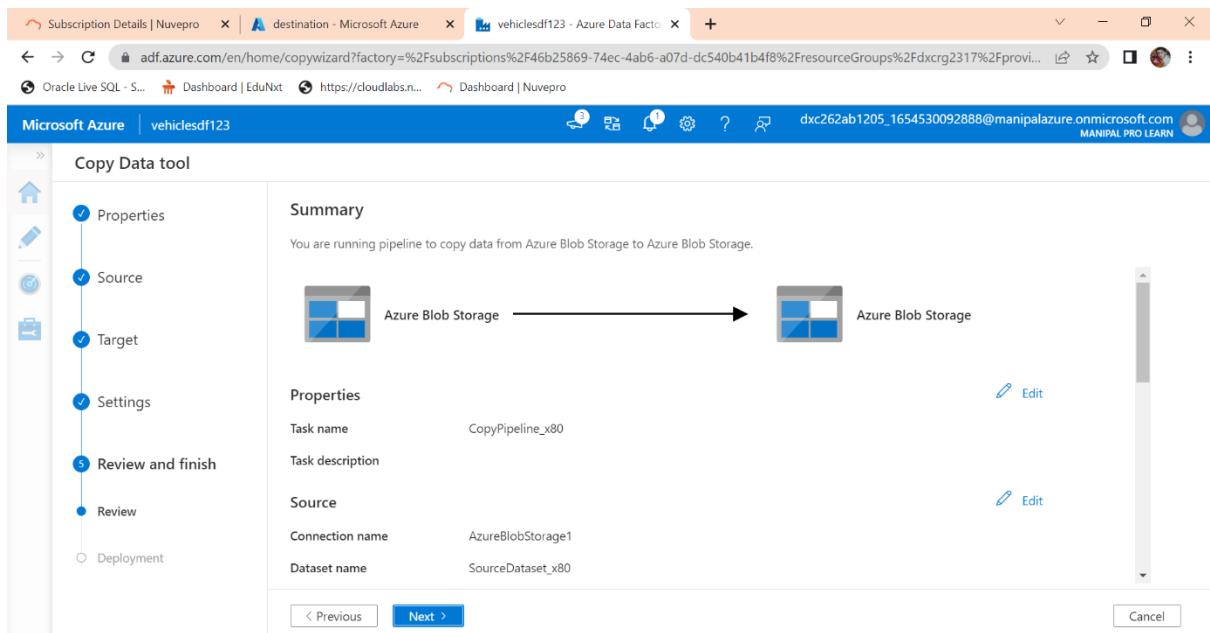
Upload data to source container

The screenshot shows the Microsoft Azure Storage account interface for the 'source' container. The left sidebar has 'source' selected under 'Containers'. The main area shows a list of blobs: 'guests.csv' (Modified: 6/12/2022, 10:12:48 AM, Access tier: Hot (Inferred), Archive status: Not yet archived, Blob type: Block blob, Size: 83 B). At the top, there are buttons for 'Upload', 'Change access level', 'Refresh', 'Delete', 'Change tier', 'Acquire lease', 'Break lease', 'View snapshots', and 'Add filter'.

Similarly create a destination container ,Open datafactory account and start using copy tool.



Complete the details of properties ,source, target ,settings and come to review+finish.



After completing the copy process, guests.csv file present in destination from source container.

Subscription Details | Nuvepro

destination - Microsoft Azure

vehiclesdf123 - Azure Data Facto...

portal.azure.com/#view/Microsoft\_Azure\_Storage/ContainerMenuBlade/~/overview/storageAccountName%2Fsubscriptions%2F46b25869-74ec-4ab6-a07d-d...

Oracle Live SQL - ... Dashboard | EduNxt https://cloudlabs.n... Dashboard | Nuvepro

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

Home > vehicles231 >

**destination** Container

Search (Ctrl+/) Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots ...

Overview Diagnose and solve problems Access Control (IAM) Settings Shared access tokens Access policy Properties Metadata

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

Location: destination

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

Add filter

Name	Modified	Access tier	Archive status	Blob type	Size
guests.txt	6/12/2022, 10:47:11 ...	Hot (Inferred)		Block blob	71 B

At final stage the pipelines are trigger enabled.

All pipeline runs > CopyPipeline\_x80

Pipeline runs

Run group ID: be84fbef-626d-4519-b343-5769a9d62427

Refresh Edit columns

Showing 1 - 2 of 2 items

In start	Run end	Duration	Triggered by	Triggered by type	Status	Error
Jun 12, 2022, 10:48:38 am	Jun 12, 2022, 10:48:47 am	00:00:09	Manual trigger	Manual	Succeeded	
Jun 12, 2022, 10:47:01 am	Jun 12, 2022, 10:47:13 am	00:00:12	Manual trigger	Manual	Succeeded	

## Practical Lab: Create Azure SQL Server and Database

Microsoft Azure

Search: sql

All Services (28) Marketplace (20) Documentation (28) Resources (0) Resource Groups (0)

Azure Active Directory (0)

Services

- SQL databases
- SQL servers
- SQL elastic pools
- SQL managed instances
- Marketplace
- Zumero - Sync SQL Server data with offline SQLite
- SQLStream 6.0.0.1
- Azure Synapse Analytics
- Continue searching in Azure Active Directory

See all → services

See all →

Subscriptions Resource groups All resources Dashboard

Searching all subscriptions.

https://portal.azure.com/#blade/HubsExtension/BrowseResourceBlade/resourceType/Microsoft.Sql%2fservers%2fdatabases

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

Home > SQL databases > Create SQL Database ... Microsoft

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

Create a SQL database with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more](#)

**Did you know?** new users in Azure can create a free Azure SQL Database and use it for 12 months using Azure free account? [Learn more](#)

#### 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 \*

[Review + create](#)

[Next : Networking >](#)

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

Home > SQL databases > Create SQL Database > Create SQL Database Server ... Microsoft

**Server details**

Enter required settings for this server, including providing a name and location. This server will be created in the same subscription and resource group as your database.

Server name \*  .database.windows.net

Location \*

**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](#) using an existing Azure AD user, group, or application as Azure AD admin [Learn more](#), or select both SQL and Azure AD authentication.

Authentication method  Use SQL authentication

[OK](#)

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

Home > SQL databases > Create SQL Database ... Microsoft

**Networking**

Allow Azure services and resources to access this server	No
Private endpoint	None
Minimum TLS version	1.2
Connection Policy	Default

**Security**

Identity	Not enabled
Service principal (preview)	Off
Transparent data encryption	Service-managed key selected
Advanced data security	Not now
Sql Ledger(Database)	Disabled
Digest Storage	Disabled

[Create](#)

[< Previous](#)

[Download a template for automation](#)

The screenshot shows two main sections of the Microsoft Azure portal.

**Top Section (Deployment Overview):**

- Header: Microsoft Azure, Search resources, services, and docs (G+), various icons.
- Resource Name: Microsoft.SQLDatabase.newDatabaseNewServer\_6f8aa8d03b7049039adf1 | Overview
- Actions: Delete, Cancel, Redeploy, Refresh.
- Status: Your deployment is complete.
- Deployment details: Deployment name: Microsoft.SQLDatabase.newDatabaseNewServer\_6f8aa8d03b7049039adf1, Subscription: Azure-DXC262AB12Lab, Resource group: dxcrg2317. Start time: 6/12/2022, 11:22:54 AM. Correlation ID: a0c8e7cb-3ed4-40b5-9225-415e67c7d36d.
- Next steps: Go to resource.

**Bottom Section (Database Overview):**

- Header: Microsoft Azure, Search resources, services, and docs (G+), various icons.
- Resource Name: Home > Microsoft.SQLDatabase.newDatabaseNewServer\_6f8aa8d03b7049039adf1 > sql231 (server232/sql231) SQL database
- Actions: Copy, Restore, Export, Set server firewall, Delete, Connect with..., Feedback.
- Status: This database was just created. Do you need any help [getting started?](#)
- Essentials:**
  - Resource group: [move](#) dxcrg2317
  - Status: Online
  - Location: East US
  - Subscription: [move](#) Azure-DXC262AB12Lab
  - Subscription ID: 46b25869-74ec-4ab6-a07d-dc540b41b4f8
  - Tags: [edit](#) Click here to add tags
- Server name: server232.database.windows.net
- Connection strings: [Show database connection strings](#)
- Pricing tier: General Purpose - Serverless, Gen5, 1 vCore
- Auto-pause delay: 1 hour
- Earliest restore point: No restore point available
- Show data for last: 1 hour, 24 hours, 7 days
- Aggregation type: Max

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

Open datafactory account and start using copy tool.

The screenshot shows the Microsoft Azure Copy Data tool wizard. The title bar indicates the current step is 'Properties'. The left sidebar lists steps 1 through 5: Properties, Source, Target, Settings, Review and finish. Step 1 is highlighted with a blue circle. The main area contains a brief description of the Copy Data Tool and its purpose. It then asks to select a copy data task type and configure a task schedule. Two options are shown: 'Built-in copy task' (selected) and 'Metadata-driven copy task'. The 'Built-in copy task' option is described as providing a single pipeline to copy data from 90+ data sources. The 'Metadata-driven copy task' option is described as providing parameterized pipelines that can read metadata from an external store to load data at a large scale. Below these options, there is a section for 'Task cadence or task schedule \*' with three radio button options: 'Run once now' (selected), 'Schedule', and 'Tumbling window'. At the bottom of the page are 'Previous' and 'Next >' buttons, and a 'Cancel' button.

Complete the details of properties ,source, target ,settings and come to review+finish.

The screenshot shows the Microsoft Azure Copy Data tool wizard. The title bar indicates the current step is 'Summary'. The left sidebar lists steps 1 through 5: Properties, Source, Target, Settings, Review and finish. Steps 1, 2, and 3 are highlighted with blue circles. The main area displays a summary of the pipeline configuration. It shows a flow from 'Azure Blob Storage' to 'Azure Blob Storage'. The 'Properties' section includes the task name 'CopyPipeline\_x80' and task description. The 'Source' section includes the connection name 'AzureBlobStorage1' and dataset name 'SourceDataset\_x80'. There are 'Edit' buttons next to each of these items. At the bottom of the page are 'Previous' and 'Next >' buttons, and a 'Cancel' button.

After completing the copy process, guests.csv file present in destination from source container.

The screenshot shows the Microsoft Azure portal interface with two tabs open:

- destination - Microsoft Azure**: This tab displays the storage container named "destination". It includes a search bar, navigation buttons, and a toolbar with options like Upload, Change access level, Refresh, Delete, Change tier, Acquire lease, Break lease, View snapshots, and more. A sidebar on the left provides links for Overview, Diagnose and solve problems, Access Control (IAM), Settings (Shared access tokens, Access policy, Properties, Metadata), and a blob list table.
- vehiclesdf123 - Azure Data Fact...**: This tab shows the overview of a storage account. It includes a search bar, navigation buttons, and a toolbar with icons for Home, Storage accounts, Containers, and Blobs. The main content area displays the storage account's properties and metrics.

**destination Container**

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

**Location:** destination

**Search blobs by prefix (case-sensitive):** guests.txt

**Show deleted blobs:**

Name	Modified	Access tier	Archive status	Blob type	Size
guests.txt	6/12/2022, 10:47:11 ...	Hot (Inferred)		Block blob	71 B

**Pipeline runs**

Run group ID: be84fbef-626d-4519-b343-5769a9d62427

Showing 1 - 2 of 2 items

Run start	Run end	Duration	Triggered by	Triggered by type	Status	Error
Jun 12, 2022, 10:48:38 am	Jun 12, 2022, 10:48:47 am	00:00:09	Manual trigger	Manual	<span style="color: green;">Succeeded</span>	
Jun 12, 2022, 10:47:01 am	Jun 12, 2022, 10:47:13 am	00:00:12	Manual trigger	Manual	<span style="color: green;">Succeeded</span>	

**Result:** In this project a new pipeline is created and transferred data from source to destination by using data tool by using data Azure Data Factory.

**Conclusion:** Data is successfully validated and stored using Microsoft Azure.

Name:Bhogadi Naga Iswarya Lakshmi

Reg No: DXC262AB12005

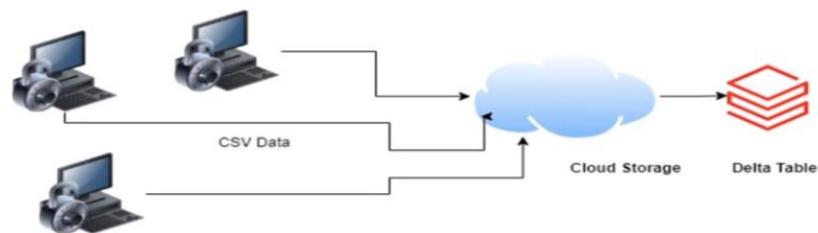
**Project2 Name: AP Morgan Data Platform**

Date: 10-06-2022

# **Project 2: AP Morgan Data Platform**

## Project 2 : AP Morgan

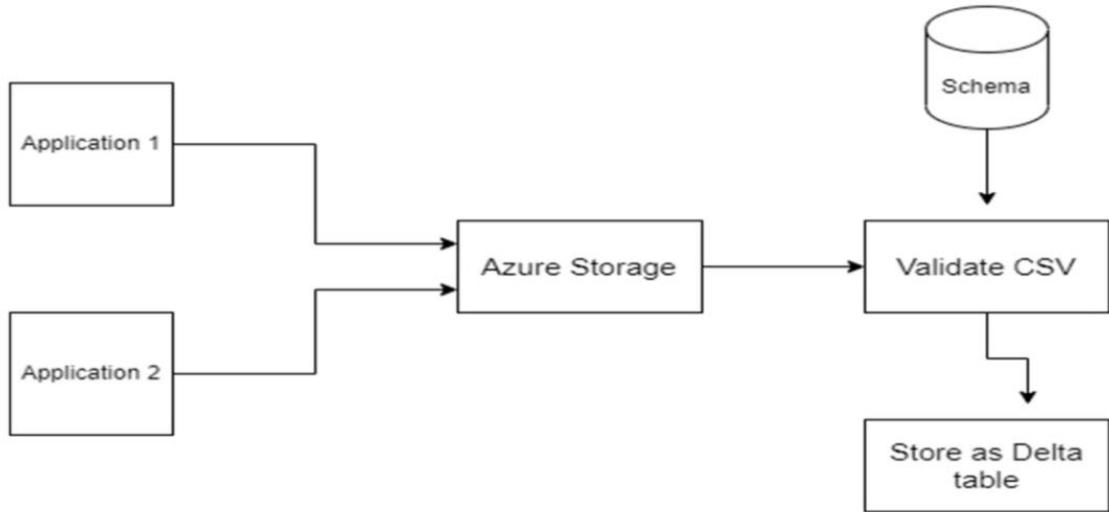
- Multiple Internal applications sends the data(huge size) in CSV format on daily basis in the cloud storage location. There are couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed data to be persisted as Delta table in Databricks for downstream system.



## 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

Azure Databricks is a Scalable analytics in Azure based on Apache Spark Workflows and workspace for data users Native integration with other Azure services.

Step 1: First open your Microsoft Azure account and search for Data Bricks.

The screenshot shows the Microsoft Azure portal's search interface. The search bar at the top contains the text "data bricks". Below the search bar, a navigation bar includes links for "All", "Services (33)", "Marketplace (9)", "Documentation (28)", "Resources (0)", and "Resource Groups (0)". The main content area displays search results under the "Services" category. The first result, "Azure Databricks", is highlighted with a blue selection bar. Other visible service icons include Datadog, Data Catalog, Data factories, Data Shares, Unravel for Azure Databricks, Trifactor for Azure, and Lakehouse Monitor. On the left side, there is a sidebar with sections for "Create a resource", "Recent", "Resources", and "Marketplace". The "Recent" section lists items like "vehiclesdf12", "vehicles231", and "dxcrg2317". The "Resources" section has tabs for "Recent" and "Favorited". The "Marketplace" section lists various Azure services. At the bottom of the page, there is a footer with links for "Give feedback" and "Search all subscriptions".

Step 2: Click on create to create a new Databricks account.

Microsoft Azure Search resources, services, and docs (G+/-) dx262ab1205\_1654530... MANIPAL PRO LEARN (MANIPAL)

Home > Azure Databricks ...

Azure Databricks (manipalazure.onmicrosoft.com)

+ Create Manage view Refresh Export to CSV Open query Assign tags

Filter for any field... Subscription == all Resource group == all Location == all Add filter

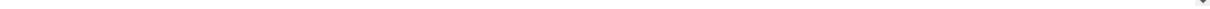
No grouping List view

Name ↑ Type ↑ Resource group ↑ Location ↑ Subscription ↑

No azure databricks services to display

Unlock insights from all your data and build artificial intelligence (AI) solutions with Azure Databricks, set up your Apache Spark environment in minutes, autoscale, and collaborate on shared projects in an interactive workspace.

Give feedback



## Step 3: Complete the Project and Instance details.

Microsoft Azure Search resources, services, and docs (G+/-) dx262ab1205\_1654530... MANIPAL PRO LEARN (MANIPAL)

Home > Azure Databricks > Create an Azure Databricks workspace ...

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 \* dxcrg2317 Create new

Instance Details

Workspace name \* APMorgan23

Region \* East US

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

Review + create < Previous Next : Networking >



Step 4: After step-3, continue the completion of details of Networking, Advanced ,Tags and finally goes to Review and Create and also completes validation process.

Microsoft Azure Search resources, services, and docs (G+/-) dx262ab1205\_1654530... MANIPAL PRO LEARN (MANIPAL)

Home > Azure Databricks > Create an Azure Databricks workspace ...

Validation Succeeded

Basics Networking Advanced Tags Review + create

Summary

Basics

Workspace name	APMorgan23
Subscription	Azure-DXC262AB12Lab
Resource group	dxcrg2317
Region	East US
Pricing Tier	standard

Networking

Create < Previous Download a template for automation



**Step 5: After completing validation process, Click on create and Deployment process starts.**

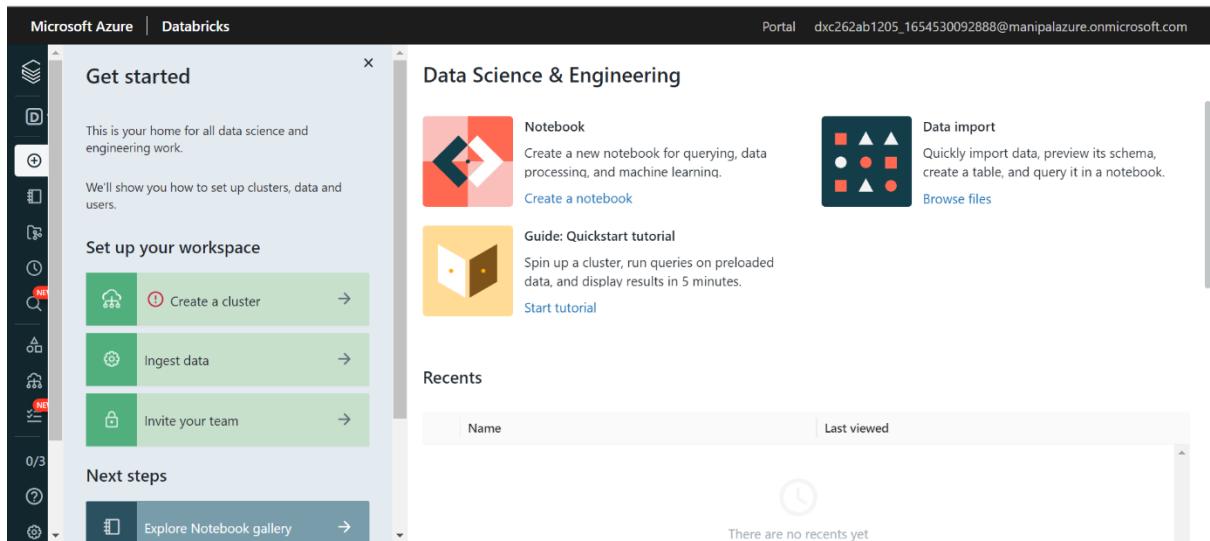
The screenshot shows the Microsoft Azure Overview page for a resource group named 'dxcrg2317\_APMargin23'. The main header bar includes the Microsoft Azure logo, a search bar, and various navigation icons. Below the header, the title 'dxcrg2317\_APMargin23 | Overview' is displayed, along with a 'Deployment' icon. A prominent message box in the top right corner states 'Deployment is in progress...' with the subtext 'Deployment to resource group 'dxcrg2317' is in progress.' Below this, there are tabs for 'Overview', 'Inputs', 'Outputs', and 'Template'. The 'Overview' tab is selected. A large central area displays deployment details, including the deployment name 'dxcrg2317\_APMargin23', subscription information 'Azure-DXC262AB12Lab', and resource group 'dxcrg2317'. It also shows the start time '6/10/2022, 5:46:54 PM' and correlation ID 'e8f2e6e8-37e5-4d3c-b065-c5c520b45188'. A section titled 'Deployment details' with a download link is present, followed by a table with columns 'Resource', 'Type', 'Status', and 'Operation details', which is currently empty ('No results').

**Step 6: After completing deployment process, click on goto resource.**

The screenshot shows the Microsoft Azure Overview page for the same resource group 'dxcrg2317\_APMargin23'. The main header bar and title are identical to the previous screenshot. The message box now indicates 'Your deployment is complete'. The deployment details remain the same. A new section titled 'Next steps' is visible, containing a blue 'Go to resource' button. To the right of the main content area, there are two promotional cards: 'Cost Management' (with a green dollar sign icon) and 'Microsoft Defender for Cloud' (with a shield icon). Both cards include descriptive text and a 'Go to [service]' link.

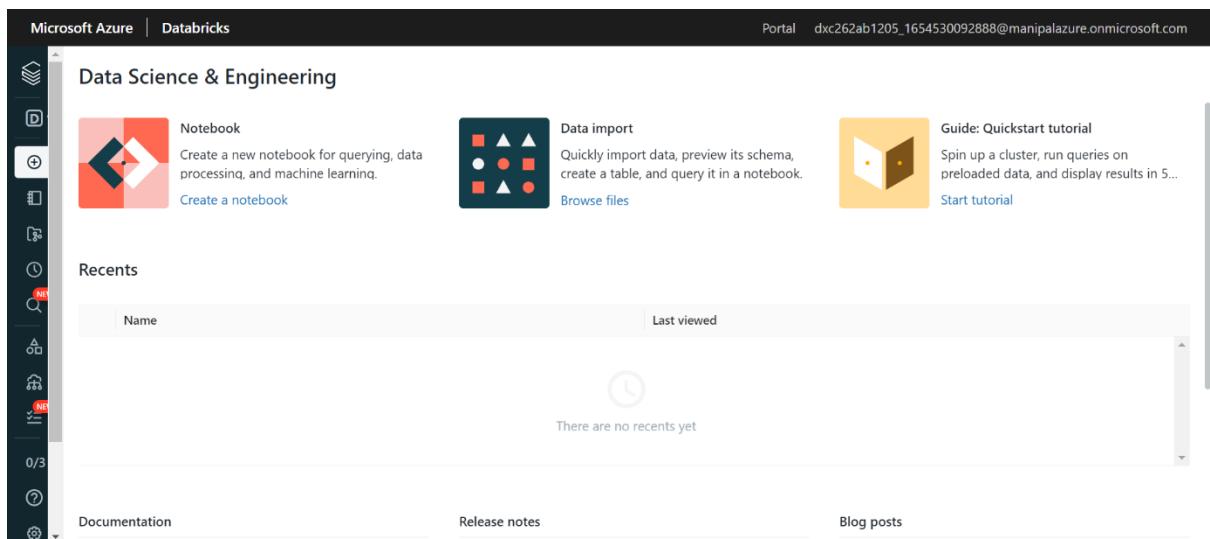
The screenshot shows the Microsoft Azure Overview page for an Azure Databricks service named 'APMargin23'. The header bar includes the Microsoft Azure logo, a search bar, and various navigation icons. The title 'APMargin23 | Overview' is displayed, along with an 'Azure Databricks Service' icon. A 'Delete' button is visible. The left sidebar contains a navigation menu with sections like 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Settings', 'Virtual Network Peerings', 'Encryption', 'Properties', 'Locks', 'Automation', 'Tasks (preview)', and 'Export template'. The 'Overview' section is selected. The main content area displays 'Essentials' information, including status 'Active', resource group 'dxcrg2317', location 'East US', subscription 'Azure-DXC262AB12Lab', and a URL 'https://adb-7716123982395834.14.azuredatabricks.net'. There is also a 'Managed Resource Group' entry for 'databricks-rg\_APMargin23-3j6jagen7lq4e'. A 'JSON View' button is located at the top right of the essentials table. At the bottom center of the page is the red Databricks logo.

## Step 7: Click on overview and launch workspace.



## Practical Lab: Create Cluster in Azure Databricks

### Step 1: Open Azure Databricks



### Step 2: Click on create, and create a cluster.

The screenshot shows the Microsoft Azure Databricks home page. On the left, there's a sidebar with various navigation options like Data Science & Machine Learning, Workspace, Repos, Recents, Search, Data, Compute, Workflows, and Help. A red box highlights the 'Cluster' option under the Compute section. The main content area has sections for Notebook, Table, Cluster, Job, Repo, and AutoML Experiment. There are also links for Data import, Guide: Quickstart tutorial, and Start tutorial. At the bottom, there are Release notes and Blog posts.

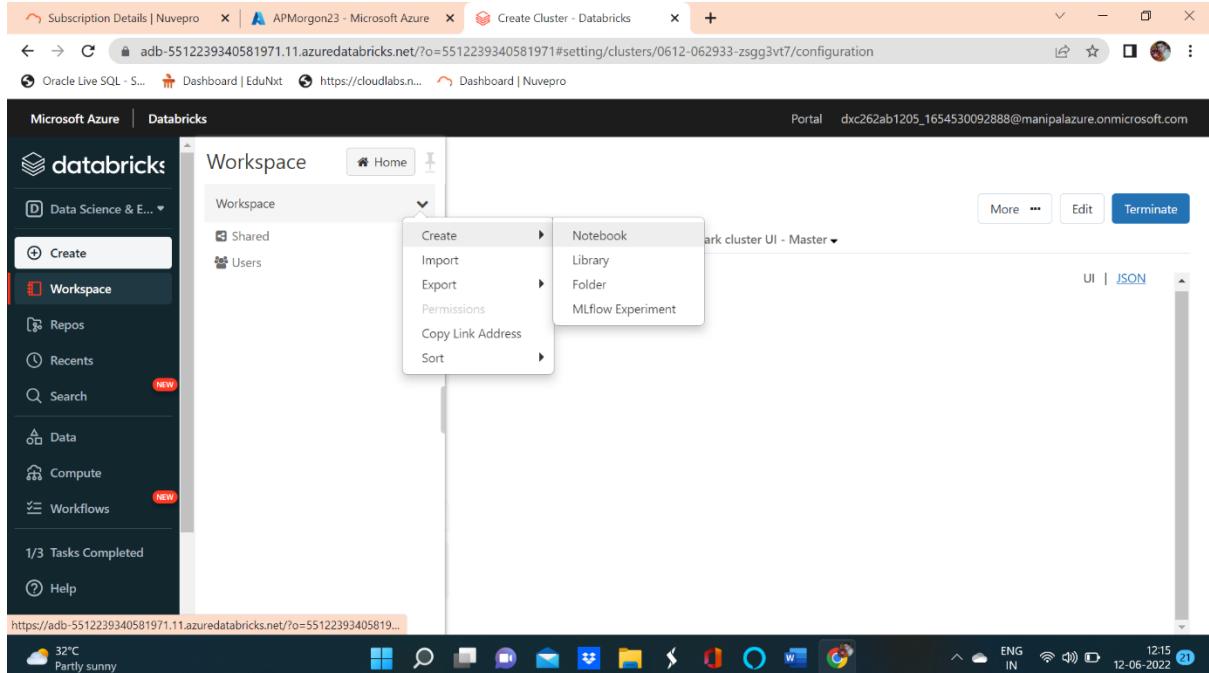
Step 3: Complete all the data & click on create cluster.

This screenshot shows the 'New Cluster' configuration page. It includes fields for Databricks runtime version (set to 10.4 LTS), Autopilot options (terminating after 30 minutes of inactivity), Node type (Standard\_DS3\_v2), and DBU / hour (0.75). A promotional discount message is displayed: '50% promotional discount applied to Photon during preview'. The 'Create Cluster' button is at the top right.

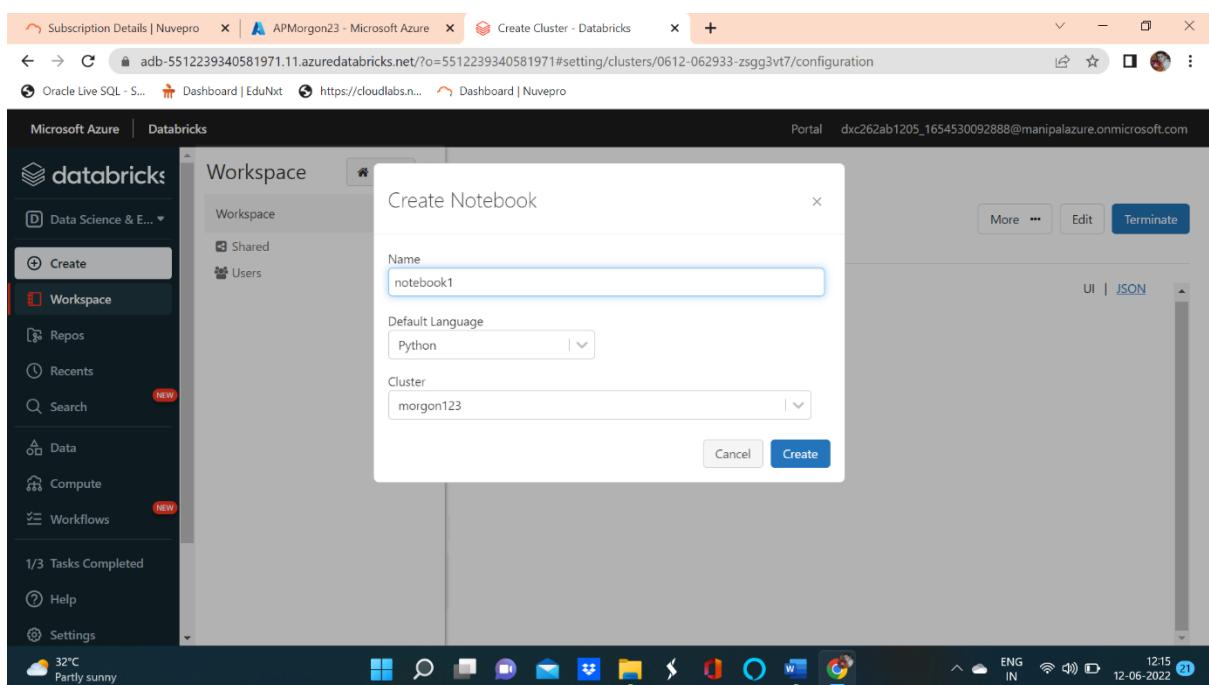
Step 4: it configures all the data and creates cluster.

This screenshot shows the configuration page for the 'morgon123' cluster. It displays the same settings as the previous screen: Cluster mode (Single Node), Databricks Runtime Version (10.4 LTS), Autopilot options (terminating after 30 minutes of inactivity), Node type (Standard\_DS3\_v2), and DBU / hour (0.75). The 'Edit' and 'Terminate' buttons are visible at the top right.

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



The screenshot shows the Microsoft Azure Databricks workspace interface. On the left, there's a sidebar with various options like 'Data Science & E...', 'Create', 'Workspace', 'Repos', 'Recents', 'Search', 'Data', 'Compute', 'Workflows', and 'Help'. The 'Create' button is highlighted. In the main area, there's a 'Workspace' section with 'Shared' and 'Users' options. A context menu is open over the 'Create' button, with 'Notebook' being the selected option. The URL in the browser bar is <https://adb-5512239340581971.11.azuredatabricks.net/?o=5512239340581971#setting/clusters/0612-062933-zsgg3vt7/configuration>. The status bar at the bottom shows the date as 12-06-2022 and the time as 12:15.



This screenshot shows the 'Create Notebook' dialog box in the Microsoft Azure Databricks workspace. The dialog has fields for 'Name' (set to 'notebook1'), 'Default Language' (set to 'Python'), and 'Cluster' (set to 'morgen123'). There are 'Cancel' and 'Create' buttons at the bottom. The background shows the same workspace interface as the previous screenshot, with the 'Create' menu still open. The URL in the browser bar is the same as the previous screenshot. The status bar at the bottom shows the date as 12-06-2022 and the time as 12:15.

The screenshot shows a Microsoft Azure Databricks notebook titled "notebook1" running Python code. The code in Cmd 1 is:

```
x=100  
y=200  
print(x+y)
```

The output is 300. A command took 0.04 seconds. Cmd 2 has a single digit '1' entered. The status bar at the bottom shows the date and time as 12-06-2022 12:16:42 PM.

## Practical Lab: Azure Data Factory For AP Morgan

### Practical Lab: Create Azure Databricks Linked Service in ADF

Go to azure home page and search for Data Factory

The screenshot shows the Microsoft Azure portal homepage. Under "Azure services", the "Data factories" service is highlighted. Other services shown include Azure Synapse Analytics, Azure Active Directory, Virtual networks, and Virtual machines. Below this, there are sections for "Resources" (Recent and Favorite), "Useful links" (Overview, Get started, Documentation), and "Last Viewed" (Azure Databricks Service, Resource group). At the bottom, there are links for "Navigate" (Subscriptions, Resource groups, All resources, Dashboard) and "Tools" (Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, Cost Management).

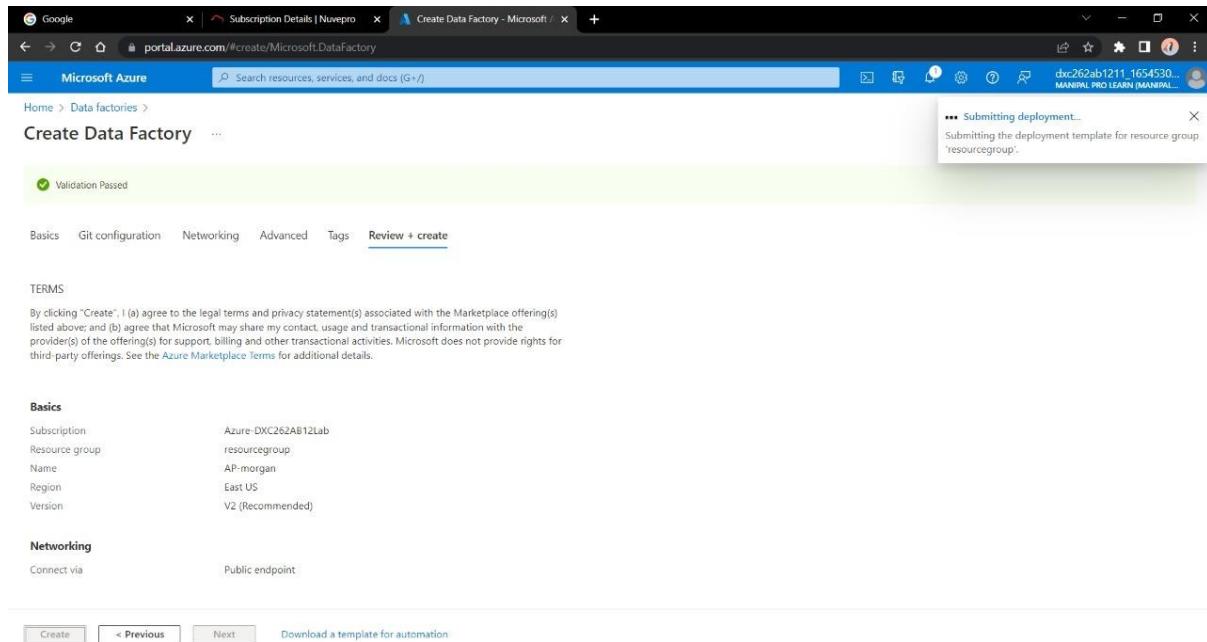
Create a new account on Azure Data Factory

The screenshot shows the Microsoft Azure portal interface. The title bar includes tabs for 'Google', 'Subscription Details | Nuvepro', and 'Data factories - Microsoft Azure'. The main content area is titled 'Data factories' and shows a message: 'No data factories to display. Try changing or clearing your filters.' Below this, there is a 'Create data factory' button and a 'Learn more' link. The page includes standard Azure navigation elements like a search bar, filter buttons for 'Subscription', 'Type', 'Resource group', and 'Location', and sorting options for 'Name', 'Type', 'Subscription', 'Resource group', and 'Location'.

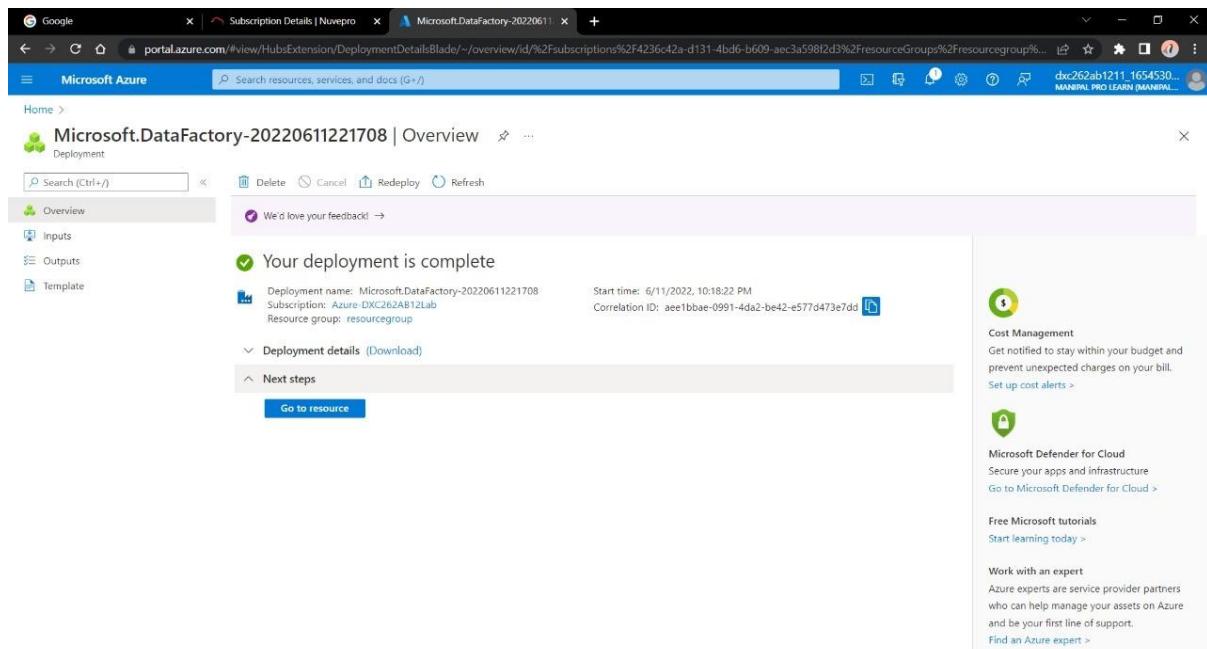
Enter the details for the factory account

The screenshot shows the 'Create Data Factory' wizard in the 'Basics' step. The title bar includes tabs for 'Google', 'Subscription Details | Nuvepro', and 'Create Data Factory - Microsoft Azure'. The main content area is titled 'Create Data Factory' and shows the 'Basics' tab selected. It includes sections for 'Project details' (Subscription: Azure-DXC262AB12Lab, Resource group: resourcegroup), 'Instance details' (Name: AP-morgan, Region: East US, Version: V2 (Recommended)), and navigation buttons for 'Review + create', '< Previous', and 'Next : Git configuration >'.

Validate and deploy the Data Factory



After the deployment is done , Navigate to “Go to resource”



Now open the Data Factory studio

The screenshot shows the Azure Data Factory Overview page for the 'AP-morgan' resource. The left sidebar contains navigation links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (Networking, Managed identities, Properties, Locks), and Getting started (Quick start, Monitoring, Alerts, Metrics, Diagnostic settings, Logs). The main content area displays the 'Essentials' section with details: Resource group (move) : resourcegroup, Status : Succeeded, Location : East US, Subscription (move) : Azure-DXC262AB12Lab, and Subscription ID : 4236c42a-d131-4bd6-b609-aec3a598f2d3. It also includes 'Getting started' options to Open Azure Data Factory Studio or Read documentation. The 'Monitoring' section shows PipelineRuns and ActivityRuns with one run each.

We need to create a pipeline that will connect Data Bricks notebook with Data Factory.

The screenshot shows the Azure Data Factory Factory Resources page. The left sidebar lists Factory Resources: Pipelines, Datasets, Data flows, and Power Query. A dropdown menu is open over the 'Pipelines' item, showing options: Pipeline (selected), Pipeline gallery, Import from pipeline template, and Copy Data tool. The main workspace is titled 'Select an item' with the sub-instruction 'Use the resource explorer to select or create a new item'. There is also a small icon of three cylinders connected by arrows.

Create a new Pipeline and drag&drop the notebook tab from Azure Databricks dropdown into the pipeline workspace.

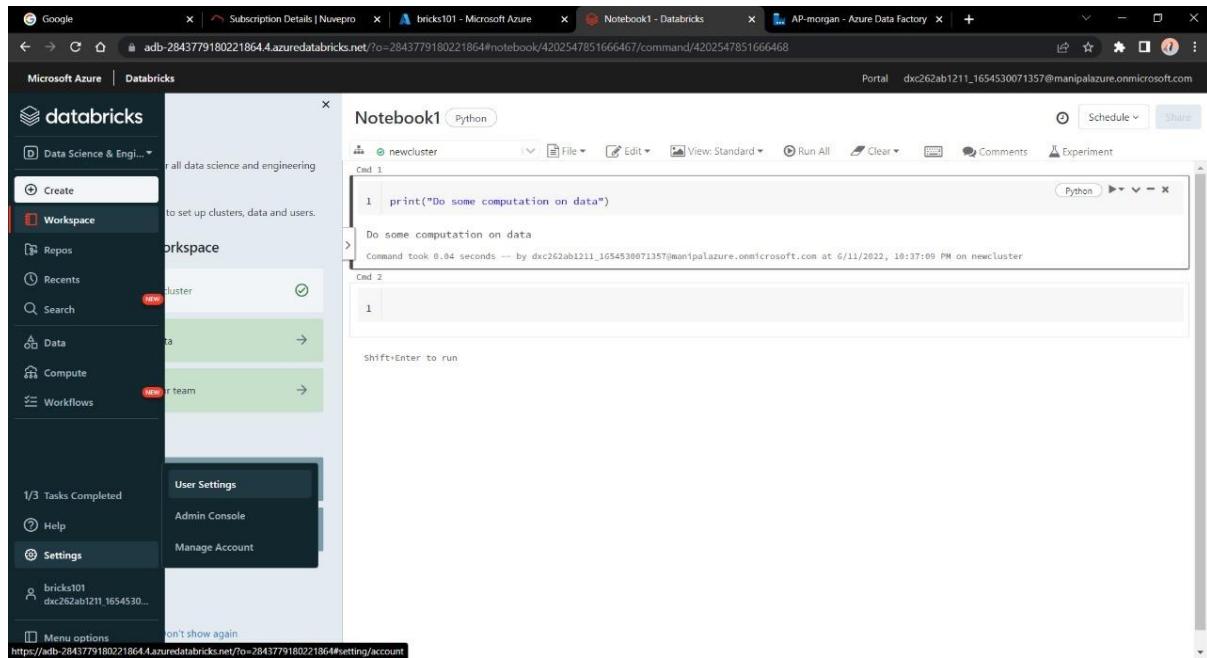
The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Datasets' (0), 'Data flows' (0), and 'Power Query' (0). The main area displays a 'pipeline1' pipeline with a single 'Notebook' activity. The 'Activities' tab is selected, showing options like 'Move & transform', 'Azure Data Explorer', 'Azure Function', 'Batch Service', and 'Databricks'. The 'Databricks' section is expanded, showing 'Notebook', 'Jar', and 'Python' options. The 'Notebook' activity is selected, and its properties are shown in the 'Properties' panel on the right. The 'General' tab is selected, showing the name 'pipeline1' and a description field.

## Create a new Linked service for databricks

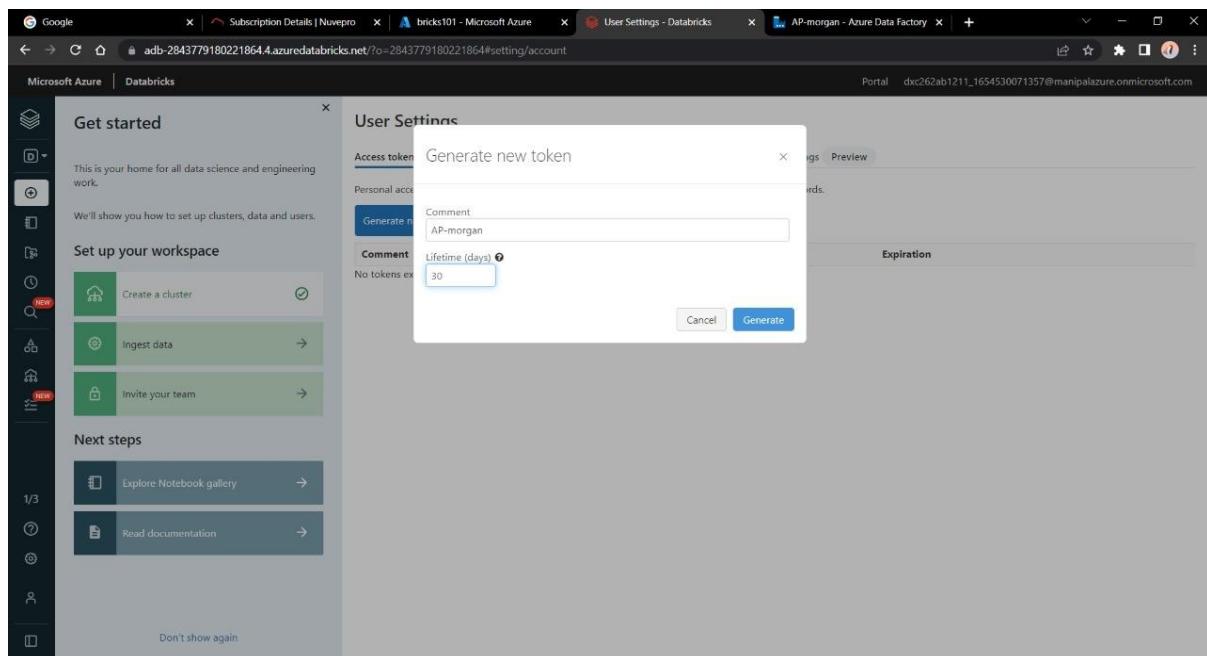
The screenshot shows the Microsoft Azure Data Factory pipeline editor with the 'Databricks' linked service configuration dialog open. The 'General' tab is selected. Under 'Databricks linked service', there is a dropdown menu labeled 'Select...'. The 'Azure Databricks' section contains fields for 'Account selection method' (set to 'From Azure subscription'), 'Azure subscription' (set to 'Azure-DXC262A012Lab (4236c42a-d131-4bd6-b609-aec3a598f2d3)'), 'Databricks workspace' (set to 'bricks101'), and 'Select cluster' (radio button for 'Existing interactive cluster' selected). The 'Databrick Workspace URL' field contains the URL 'https://adb-2843779180221864.4.azuredatabricks.net'. The 'Authentication type' section shows 'Access Token' selected, with an 'Access token' input field containing a placeholder 'I'. The 'Annotations' section has a '+ New' button. At the bottom are 'Create' and 'Cancel' buttons, and a 'Test connection' link.

We need the access token of Data bricks account in order to access it.

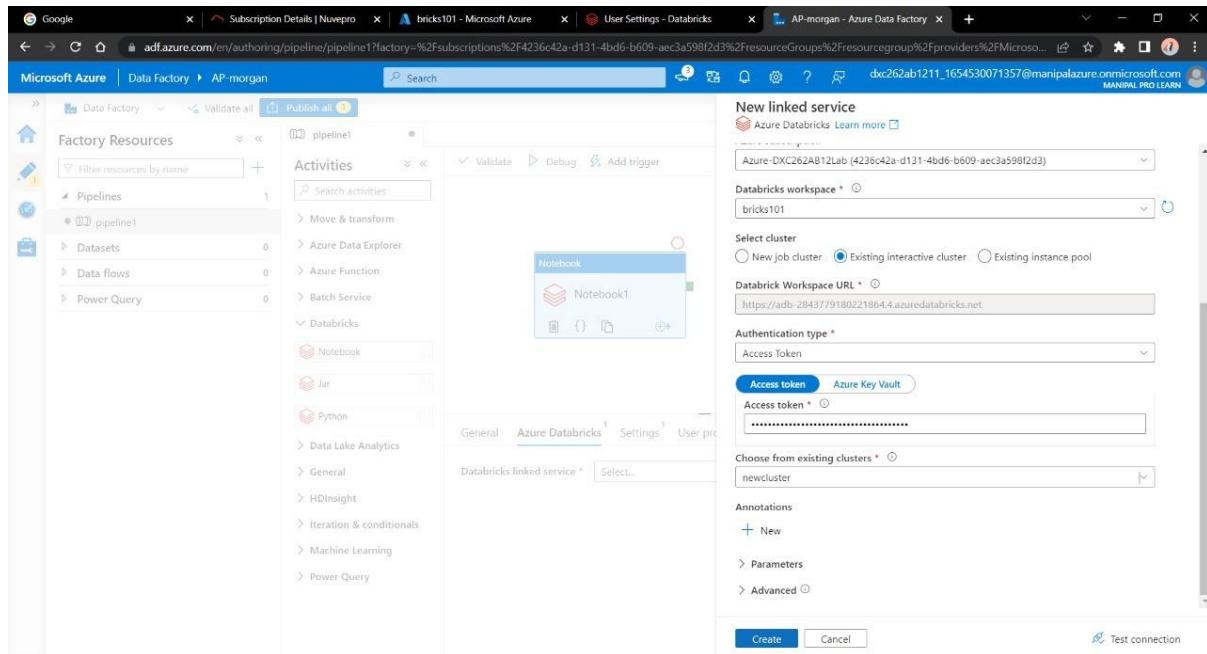
Navigate to Data bricks workspace and go to Users in settings



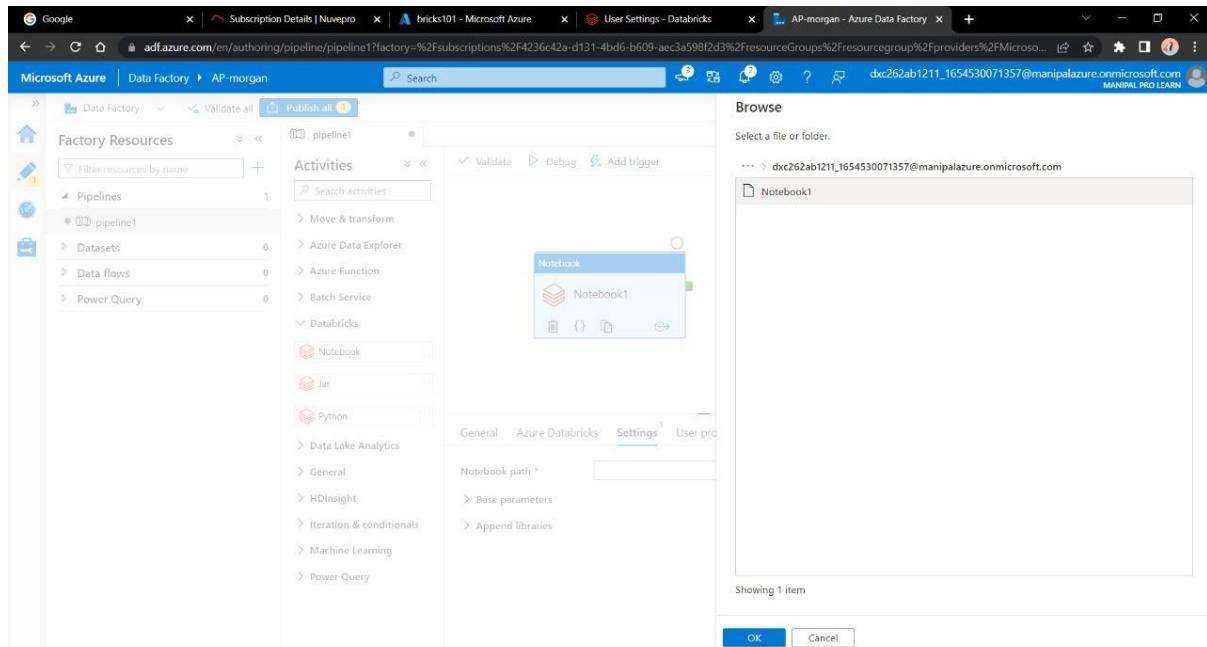
Click on “Generate token” and mention a small quick description name for the token



After that a tab will open containing your access token , copy it and paste it in the azure factory data connection form

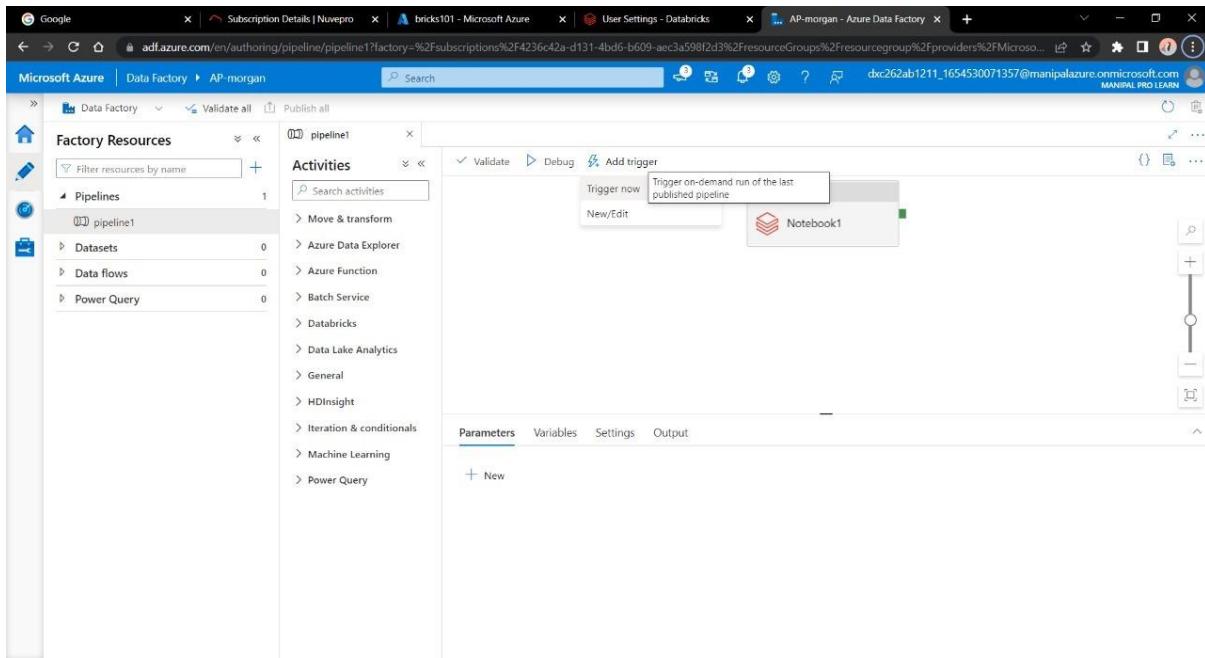


Choose the right cluster and click on create

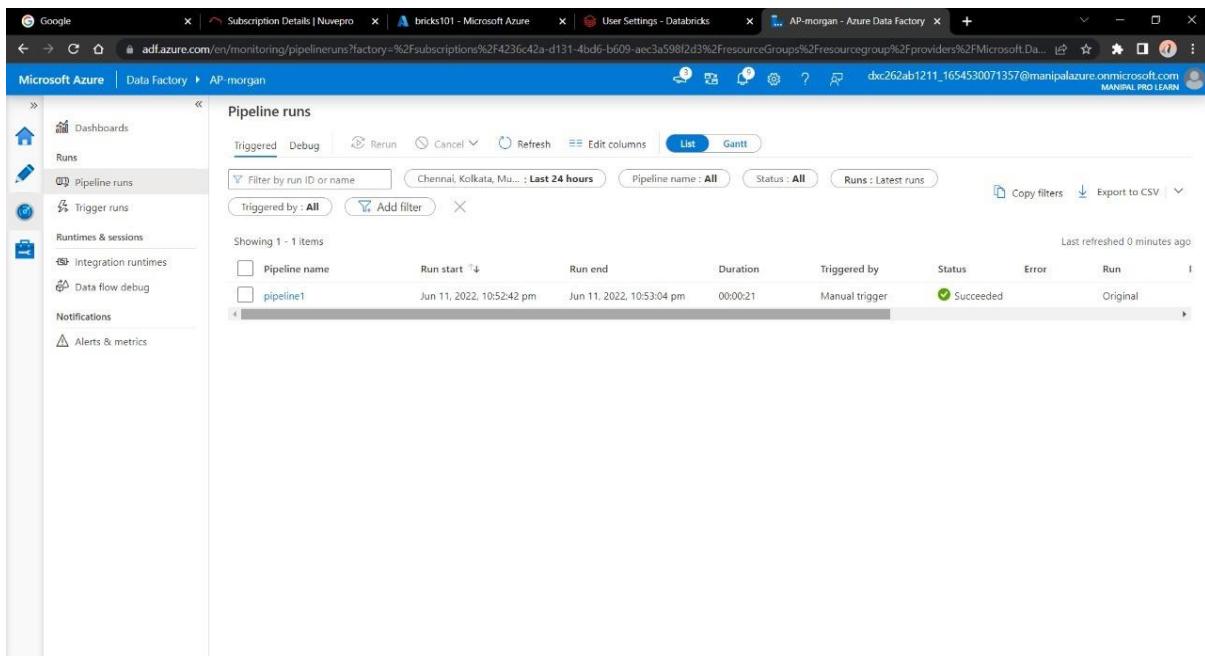


Now choose the notebook in the pipeline notebook tab setting in order to trigger it.

Publish the pipeline in order to trigger it , After publication is done click on “Trigger now”



Go to Monitor window in order to check it's execution



Here we can see we have successfully triggered a linked notebook of Data bricks from Data Factory.

**Result:** : In this project a data is able to link and trigger azure DataBricks notebook.

**Conclusion:** Azure Data Factory linked with Azure Data bricks.