

---

# DXC REALTIME PROJECTS

---

AZ-900, DP - 203



**JUNE 10, 2022**  
DXC TECHNOLOGY PVT.LTD.

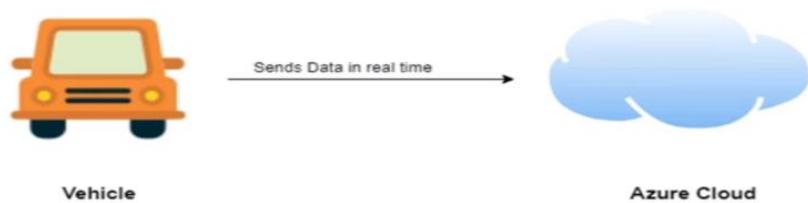
Name: Aishee Bhattacharya

Reg No: DXC262AB1204

Date: 12th June, 2022

## Project 1 : Connected Vehicles

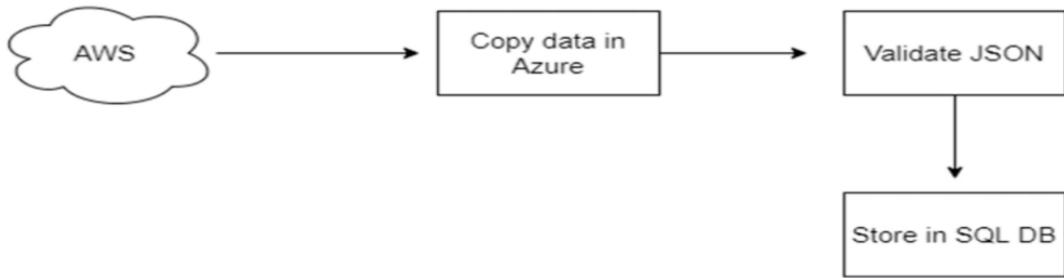
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



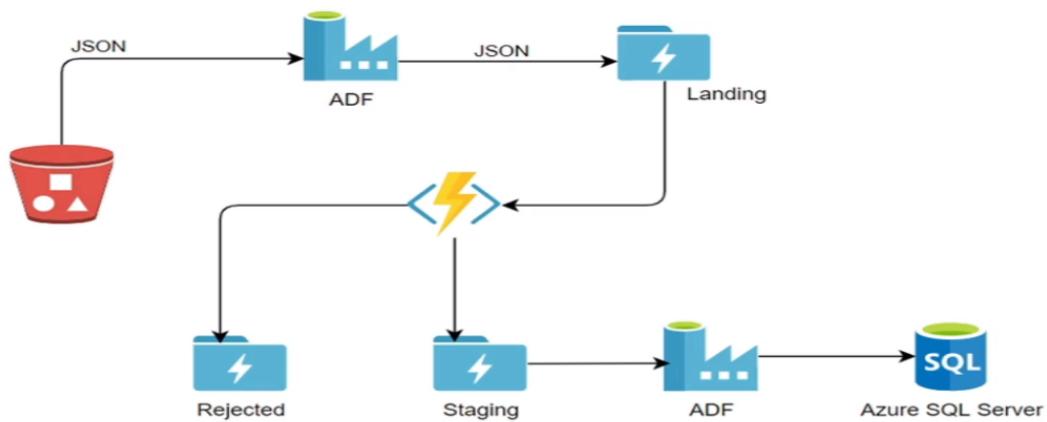
## Project 1 : Connected Vehicles

- 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

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

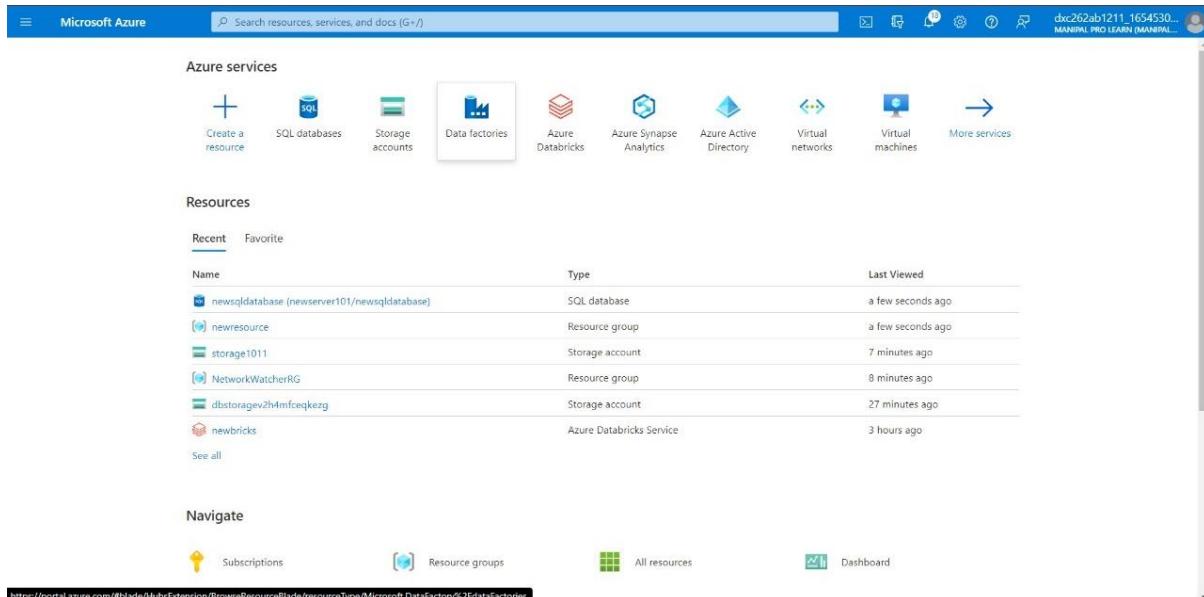
Practical Lab: Create **Azure blob trigger logic**

Practical Lab: Create **Azure SQL Server and Database**

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

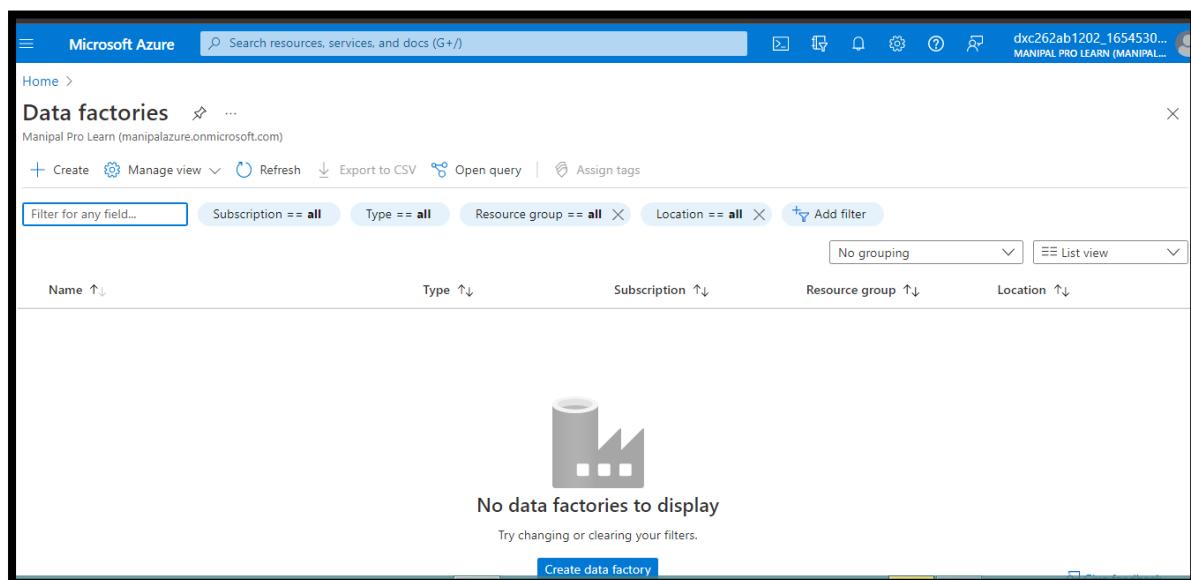
In order fulfill the above mentioned cases we need to follow the steps below:

## Step 1:



The screenshot shows the Microsoft Azure portal homepage. At the top, there's a search bar and a navigation bar with icons for account management and help. Below the search bar, the 'Azure services' section is visible, featuring icons for Create a resource, SQL databases, Storage accounts, Data factories (which is highlighted with a white box), Azure Databricks, Azure Synapse Analytics, Azure Active Directory, Virtual networks, Virtual machines, and More services. Under the 'Data factories' icon, there's a sub-section titled 'Resources' with tabs for 'Recent' and 'Favorite'. It lists several resources: newsqldatabase (SQL database), newresource (Resource group), storage1011 (Storage account), NetworkWatcherRG (Resource group), dbstoragev2h4mfcceqkezg (Storage account), and newbricks (Azure Databricks Service). Each item has a small icon, its name, type, and last viewed time. Below this is a 'Navigate' section with links for Subscriptions, Resource groups, All resources, and Dashboard.

## Step 2:



The screenshot shows the 'Data factories' blade in the Microsoft Azure portal. At the top, there's a back arrow labeled 'Home', the title 'Data factories', and three dots for more options. Below the title, it says 'Manipal Pro Learn (manipalazure.onmicrosoft.com)'. There are buttons for 'Create', 'Manage view', 'Refresh', 'Export to CSV', 'Open query', and 'Assign tags'. A filter bar at the top allows filtering by 'Subscription', 'Type', 'Resource group', and 'Location', with each set to 'all'. Below the filters, there are sorting options for 'Name', 'Type', 'Subscription', 'Resource group', and 'Location'. The main area displays a message: 'No data factories to display' with a factory icon, followed by the text 'Try changing or clearing your filters.' and a blue 'Create data factory' button.

## Step 3:

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

Home > Data factories > Create Data Factory ...

Basics Git configuration Networking Advanced Tags Review + create

**Project details**

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

Subscription \* Azure-DXC262AB12Lab

Resource group \* newresource Create new

**Instance details**

Name \* newdata-factory

Region \* East US

Version \* V2 (Recommended)

Review + create < Previous Next : Git configuration >

## Step 4:

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

Home > Data factories > Create Data Factory ...

Validation Passed

I listed above, and (b) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

**Basics**

Subscription	Azure-DXC262AB12Lab
Resource group	DXCProj
Name	automobile
Region	East US
Version	V2 (Recommended)

**Networking**

Connect via	Public endpoint
-------------	-----------------

Submitting deployment...  
Submitting the deployment template for resource group 'DXCProj'.

## Step 5:

The screenshot shows the Microsoft Azure portal interface for a Data Factory named 'automobile'. The left sidebar contains navigation links such as Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Networking, Managed identities, Properties, Locks, Getting started, and Quick start. The main content area is titled 'Getting started' and includes two cards: 'Open Azure Data Factory Studio' (with a 'Read documentation' link) and 'Read documentation' (with a 'Learn more' link). Below this is a 'Monitoring' section with two cards: 'PipelineRuns' (showing 100 runs) and 'ActivityRuns' (showing 100 runs). The top right corner shows the user's email (dx262ab1202\_1654530...) and name (MANIPAL PRO LEARN).

## Step 6:

The screenshot shows the Azure Data Factory studio interface for the 'automobile' data factory. The top navigation bar includes 'Set up code repository'. The main area features a large 3D isometric diagram of a factory building with various data storage and processing components. Below the diagram are four cards: '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.). A 'Discover more' button is located at the bottom left.

## Step 7:

The screenshot shows the Microsoft Azure Storage accounts list page. At the top, there is a search bar and several navigation icons. Below the header, there are filter options: 'Subscription == all', 'Resource group == all', and 'Location == all'. The main table lists one storage account:

Name	Type	Kind	Resource group	Location	Subscription
dbstorageytqqj4llwxr4	Storage account	BlobStorage	databricks-rg-dxcdemodbs2...	East US	Azure-DXC262AB12Lab

At the bottom of the page, there are navigation links for 'Page' and '1 of 1', and a note 'Showing 1 to 1 of 1 records.'

## Step 8:

The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal. The validation status is 'Validation passed'. The 'Review + create' tab is selected. The 'Basics' section contains the following configuration:

Subscription	Azure-DXC262AB12Lab
Resource Group	NetworkWatcherRG
Location	eastus
Storage account name	storage1011
Deployment model	Resource manager
Performance	Standard
Replication	Read-access geo-redundant storage (RA-GRS)

The 'Advanced' section contains the following settings:

Secure transfer	Enabled
Allow storage account key access	Enabled
Allow cross-tenant replication	Enabled
Default to Azure Active Directory authorization in the Azure portal	Disabled
Block public access	Enabled

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

## Step 9:

The screenshot shows the Microsoft Azure Storage Overview page for a resource named "storage1011\_1654861967486". The main message is "Your deployment is complete". Deployment details include a name ("storage1011\_1654861967486"), subscription ("Azure DVC262AB1Lab"), and resource group ("NetworkWatcherRG"). The deployment started at 6/10/2022, 5:23:21 PM. A "Go to resource" button is present. On the right, there are promotional cards for Cost Management, Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

## Step 10:

The screenshot shows the Microsoft Azure Storage Containers page for the "storage1011" account. The left sidebar lists various storage management options like Overview, Activity log, Tags, and Container creation (+ Container). The main area displays a table of existing containers, with one row for "Logs" showing the last modified date as 6/10/2022, 5:23:53 PM, Public access level as Private, and Lease state as Available. A search bar at the top allows filtering by prefix.

## Step 12:

The screenshot shows the Microsoft Azure Storage account interface for 'storage1011'. On the left, the navigation menu includes 'Overview', 'Activity log', 'Tags', 'Diagnose and solve problems', 'Access Control (IAM)', 'Data migration', 'Events', and 'Storage browser (preview)'. Under 'Data storage', 'Containers' is selected. In the center, the 'Containers' blade displays a table with columns: Name, Last modified, and Public access level. It shows two existing containers: '\$logs' (Last modified 6/10/2022, 5:23:53 PM, Private) and 'source' (Last modified 6/10/2022, 5:26:01 PM, Private). A 'New container' dialog is open on the right, prompting for a name ('source') and public access level ('Private (no anonymous access)').

## Step 13:

The screenshot shows the same Microsoft Azure Storage account interface for 'storage1011'. The 'Containers' blade now lists three containers: '\$logs', 'source', and another unnamed container (Last modified 6/10/2022, 5:23:53 PM, Private). The 'source' container has a 'Lease state' of 'Available'.

## Step 14:

The screenshot shows the Microsoft Azure Storage Explorer interface. On the left, a sidebar menu includes 'Overview', 'Diagnose and solve problems', 'Access Control (IAM)', 'Shared access tokens', 'Access policy', 'Properties', and 'Metadata'. The main area displays a table of blobs in the 'source' container. One blob, 'transactions.csv', is highlighted. The table columns are 'Name', 'Modified', 'Access tier', 'Archive status', and 'Blob type'. The blob details show it was modified on 6/10/2022 at 5:27:16 PM, is in the 'Hot (Inferred)' tier, and is a 'Block blob'. On the right, a file upload dialog is open, showing a successful upload of 'transactions.csv' from 'MANUAL PRO LEARN (MANUAL...)'. The file size is 880.14 KB.

## Step 15:

The screenshot shows the Microsoft Azure portal homepage. The 'Azure services' section is visible, with 'SQL databases' selected. Other options include 'Create a resource', 'Storage accounts', 'Azure Active Directory', 'Virtual networks', 'Virtual machines', and 'More services'. Below this, the 'Resources' section shows recent resources like 'storage1011', 'NetworkWatcherRG', and 'newresource'. The 'SQL databases' blade contains sections for 'Free training from Microsoft', 'Last Viewed' (with items last viewed 3 minutes ago, 4 minutes ago, 23 minutes ago, 32 minutes ago, and 3 hours ago), and 'Useful links' (including 'Overview', 'Get started', 'Documentation', and 'Pricing'). At the bottom, there's a 'Navigate' bar with links to 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'.

## Step 16:

The screenshot shows the 'Create SQL Database' wizard in the Microsoft Azure portal. The 'Basics' step is selected. Key configuration details include:

- Subscription:** Azure-DXC262AB12Lab
- Resource group:** newresource
- Database name:** newsqldatabase
- Server:** newserver101 (East US)
- Compute + storage:** General Purpose (Gen5, 2 vCores, 32 GB storage, zone redundant disabled)
- Backup storage redundancy:** (not explicitly shown in the screenshot)

At the bottom, there are 'Review + create' and 'Next : Networking >' buttons.

## Step 17:

The screenshot shows the 'Networking' step of the 'Create SQL Database' wizard. The configuration includes:

- Firewall rules:** Allow Azure services and resources to access this server (Yes)
- Add current client IP address:** (Yes)
- Private endpoints:** A table showing private endpoint connections for the server.

At the bottom, there are 'Review + create', '< Previous', and 'Next : Security >' buttons.

## Step 18:

The screenshot shows the Microsoft Azure Deployment Overview page for a deployment named "Microsoft.SQLDatabase.newDatabaseExistingServer\_b925ac6280474c34". The status is "Your deployment is complete". Deployment details include a start time of 6/10/2022, 5:30:01 PM, and a correlation ID of ddb35c5f-d325-444d-ab1f-ae1a58039696. A "Go to resource" button is visible. On the right, there are promotional cards for Cost Management, Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

## Step 19:

The screenshot shows the Microsoft Azure Copy Data tool wizard on the "Properties" step. It explains how to use the tool to perform a one-time or scheduled data load from 90+ data sources. It offers two task types: "Built-in copy task" (single pipeline) and "Metadata-driven copy task" (parameterized pipelines). Task cadence options include "Run once now", "Schedule", and "Tumbling window". Navigation buttons at the bottom are "Previous", "Next", and "Cancel".

## Step 20:

The screenshot shows the Microsoft Azure Copy Data tool interface. On the left, a navigation pane lists steps: Properties, Source (selected), Dataset, Configuration, Target, Settings, Review and finish. The main area is titled 'Source data store' and shows the configuration for a copy task. Under 'Source type', 'Azure Blob Storage' is selected. The 'Connection' dropdown is set to 'Select...' and has a '+ New connection' button. To the right, a 'New linked service' panel is open, titled 'Azure Blob Storage'. It includes fields for 'Name' (set to 'AzureBlobStorage1'), 'Description', 'Connect via integration runtime' (set to 'AutoResolveIntegrationRuntime'), 'Authentication type' (set to 'Account key'), and 'Azure subscription' (set to 'Azure-DX262A812Lab (4236c42a-d131-4bd6-b609-aec3a598f2d3)'). Below these, 'Storage account name' is set to 'storage1011'. A 'Test connection' button is visible at the bottom right.

## Step 21:

This screenshot continues from Step 20, showing the 'Source' configuration step. The 'File or folder' field contains 'source/transactions.csv'. Under 'Options', 'Recursively' is checked. Other options like 'Binary copy' and 'Enable partition discovery' are available but unchecked. The 'Max concurrent connections' field is empty. At the bottom, there are 'Start time (UTC)' and 'End time (UTC)' fields with dropdown menus. Navigation buttons 'Previous' and 'Next' are at the bottom, along with a 'Cancel' button.

## Step 22:

The screenshot shows the Microsoft Azure Copy Data tool interface. The left sidebar lists steps: Properties, Source, Dataset, Configuration, Target, Settings, and Review and finish. The 'Source' step is selected. The main panel displays 'File format settings' for a 'DelimitedText' file format. It includes fields for 'Column delimiter' (Comma (,), checked), 'Row delimiter' (Default (\r\n, or \n\r)), 'First row as header' (checked), and 'Compression type' (None). Below these are sections for 'Advanced' and 'Additional columns'. At the bottom are 'Previous' and 'Next >' buttons, and a 'Cancel' button.

## Step 23:

The screenshot shows the Microsoft Azure Copy Data tool interface. The left sidebar lists steps: Properties, Source, Target, Dataset, Configuration, Settings, and Review and finish. The 'Target' step is selected. The main panel displays 'Destination data store' settings. Under 'Target type', 'Azure SQL Database' is selected. Under 'Connection', 'Select...' is chosen. To the right, the 'New linked service' section is open, showing 'Azure SQL Database' as the connection type. 'From Azure subscription' is selected. The 'Azure subscription' dropdown shows 'Azure-DXC262AB12Lab (4236c42a-d131-4bd6-b609-aec3a598f2d3)'. The 'Server name' is 'newserver101' and the 'Database name' is 'newsqldatabase'. 'Authentication type' is set to 'SQL authentication' with 'User name' 'user' and 'Password' '\*\*\*\*\*'. The 'Always encrypted' checkbox is unchecked. At the bottom are 'Previous' and 'Next >' buttons, and a 'Create' button.

## Step 23:

The screenshot shows the Microsoft Azure Copy Data tool interface. The left sidebar lists steps: Properties, Source, Target (which is selected), Dataset, Configuration, Settings, and Review and finish. The main panel is titled "Destination data store" and specifies the target type as "Azure SQL Database" and the connection as "AzureSqlDatabase1". It shows a mapping from "Azure Blob Storage file" to "transaction" in "Azure Blob Storage file (auto-create)". There is a checkbox for "Skip column mapping for all tables" and navigation buttons for "Previous", "Next", and "Cancel".

## Step 24:

The screenshot shows the Microsoft Azure Copy Data tool interface. The left sidebar lists steps: Properties, Source, Target (which is selected), Settings, and Review and finish. The main panel is titled "Settings" and includes fields for "Task name" (set to "CopyPipeline\_zbw"), "Task description", and several configuration options: "Data consistency verification" (radio button), "Fault tolerance" (dropdown menu), "Enable logging" (checkbox), and "Enable staging" (checkbox). A link for "Advanced" settings is also present. Navigation buttons for "Previous", "Next", and "Cancel" are at the bottom.

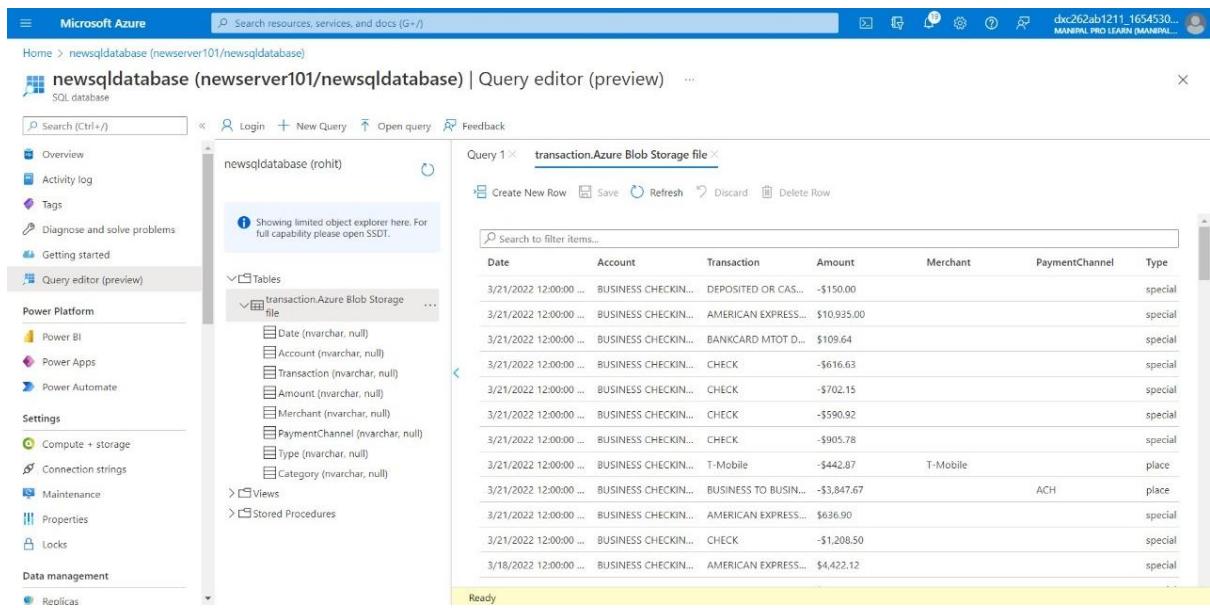
## Step 25:

The screenshot shows the Microsoft Azure Copy Data tool interface. On the left, a vertical navigation bar lists steps: Properties, Source, Target, Settings, Review and finish, Review, and Deployment. The 'Properties' step is selected. The main area is titled 'Summary' and shows a flow from 'Azure Blob Storage' to 'Azure SQL Database'. Below this, the 'Properties' section displays the task name 'CopyPipeline\_zbw' and various source settings: Connection name (AzureBlobStorage1), Dataset name (SourceDataset\_zbw), Column delimiter (,), Escape character (\), Quote char ("), First row as header (true), File name (transactions.csv), and Container (source). At the bottom are 'Previous' and 'Next' buttons, and a 'Cancel' button.

## Step 26:

The screenshot shows the Microsoft Azure Pipeline runs page. The left sidebar includes options like Dashboards, Runs, Pipeline runs (selected), Trigger runs, Runtimes & sessions, Integration runtimes, Data flow debug, Notifications, and Alerts & metrics. The main area is titled 'Pipeline runs' and shows a table of runs. The table has columns: Pipeline name, Run start, Run end, Duration, Triggered by, Status, Error, and Run. One run is listed: 'CopyPipeline\_zbw' with a run start of 'Jun 10, 2022, 5:41:49 pm', run end of 'Jun 10, 2022, 5:42:07 pm', duration of '00:00:17', triggered by 'Manual trigger', status 'Succeeded', and error 'Original'. The table also includes a 'Last refreshed 0 minutes ago' timestamp and buttons for 'List' and 'Gantt' views, as well as filters and export options.

## Step 27:



The screenshot shows the Microsoft Azure portal interface. The left sidebar has 'Query editor (preview)' selected under 'Data management'. The main area displays a table titled 'transaction.Azure Blob Storage file' with the following data:

Date	Account	Transaction	Amount	Merchant	PaymentChannel	Type
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	DEPOSITED OR CAS...	-\$150.00			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	AMERICAN EXPRESS...	\$10,935.00			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	BANKCARD MTOT D...	\$109.64			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$616.63			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$702.15			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$590.92			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$905.78			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	T-Mobile	\$442.87	T-Mobile	place	
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	BUSINESS TO BUSIN...	-\$3,847.67		ACH	place
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	AMERICAN EXPRESS...	\$636.90			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$1,208.50			special
3/18/2022 12:00:00 ...	BUSINESS CHECKIN...	AMERICAN EXPRESS...	\$4,422.12			special

**Result:** In project 1 we successfully created a pipeline that will validate and copy the blob data into the SQL database using Azure Data Factory

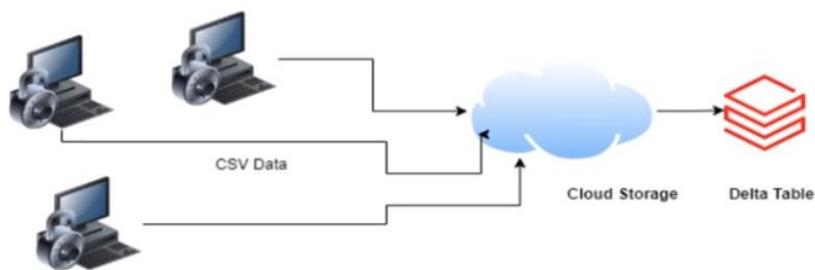
**Conclusion:** Azure Pipelines supports continuous integration (CI) and continuous delivery (CD) to continuously test, build, and deploy the code.

In this project we successfully created a pipeline that validated and copied the blob data into the SQL database using Azure Data Factory

Continuous Integration (CI) is the practice used by development teams of automating, merging and testing code. Implementing CI helps to catch bugs early in the development cycle, which makes them less expensive to fix.

## 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

## Practical Lab: Create a Databricks

## Practical Lab: Create Cluster in Azure Databricks

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

## Practical Lab: Azure Data Factory For AP Morgan

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

### Step 1:

The screenshot shows the Microsoft Azure portal's main dashboard. At the top, there's a search bar and a navigation bar with various icons. Below the search bar, the 'Azure services' section is visible, featuring a 'Create a resource' button and several service tiles. The 'Azure Databricks' tile is highlighted with a callout box. To the right of the tiles are links for 'Azure Synapse Analytics', 'Azure Active Directory', 'Virtual networks', 'Virtual machines', and 'More services'. Below the tiles, the 'Resources' section displays a table of recent resources. The first item is 'bricks' (Type: Azure Databricks Service), which was last viewed 'a few seconds ago'. The second item is 'resourcegroup' (Type: Resource group), last viewed '21 minutes ago'. There are also 'Recent' and 'Favorite' tabs, and a 'See all' link. The 'Navigate' section includes links for 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'. The 'Tools' section at the bottom includes links for 'Microsoft Learn' (Learn Azure with free online), 'Azure Monitor' (Monitor your apps and), 'Microsoft Defender for Cloud' (Secure your apps and infrastructure), and 'Cost Management' (Analyze and optimize your cloud spend for free). A URL at the bottom left of the dashboard area is: <https://portal.azure.com/#blade/HubsExtension/BrowseResourceBlade/resourceType/Microsoft.Databricks%2Fworkspaces>.

## Step 2:

The screenshot shows the Microsoft Azure portal interface. At the top, there's a navigation bar with 'Microsoft Azure' and a search bar. Below the navigation bar, the URL 'Azure Databricks' is visible, along with a 'MANIPAL PRO LEARN (manipalazure.onmicrosoft.com)' sign-in link. The main content area has a heading 'Azure Databricks' with a small icon and three dots. A sub-header 'Manipal Pro Learn (manipalazure.onmicrosoft.com)' follows. Below this, there are several filter options: '+ Create', 'Manage view', 'Refresh', 'Export to CSV', 'Open query', and 'Assign tags'. There are also dropdown filters for 'Subscription == all', 'Resource group == all', 'Location == all', and a 'Filter' button. The main body of the page says 'No azure databricks services to display' and features a large icon of three stacked cubes. Below the icon, a sub-headline reads '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.' A prominent blue button labeled 'Create azure databricks service' is centered. At the bottom right, there's a 'Give feedback' link.

## Step 3:

The screenshot shows the Microsoft Azure portal interface for creating a new Azure Databricks workspace. At the top, there's a navigation bar with 'Microsoft Azure' and a search bar. Below the navigation bar, the URL 'Create an Azure Databricks workspace' is visible, along with a 'MANIPAL PRO LEARN (manipalazure.onmicrosoft.com)' sign-in link. The main content area has a heading 'Create an Azure Databricks workspace' with three dots. Below it, there are tabs for 'Basics', 'Networking', 'Advanced', 'Tags', and 'Review + create'. The 'Basics' tab is currently selected. Under 'Project Details', there's a note about selecting a subscription and resource group. The 'Subscription' dropdown is set to 'Azure-DXC262AB12Lab' and the 'Resource group' dropdown is set to 'resourcegroup'. Under 'Instance Details', there are fields for 'Workspace name' (set to 'bricks10'), 'Region' (set to 'East US'), and 'Pricing Tier' (set to 'Standard (Apache Spark, Secure with Azure AD)'). At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next : Networking >'.

## Step 4:

The screenshot shows the Microsoft Azure portal with the title "Create an Azure Databricks workspace". The "Networking" tab is selected. Under "Secure Cluster Connectivity (No Public IP)", the "Yes" radio button is selected. Under "Deploy Azure Databricks workspace in your own Virtual Network (VNet)", the "No" radio button is selected. At the bottom, there are buttons for "Review + create", "< Previous", and "Next : Advanced >".

## Step 5:

The screenshot shows the Microsoft Azure portal with the title "Create an Azure Databricks workspace". The "Review + create" tab is selected. A green bar at the top indicates "Validation Succeeded". A message box in the top right says "... Initializing deployment... Initializing template deployment to resource group 'resourcegroup'." The "Summary" section shows the workspace name is "bricks101", located in "Azure-DXC262AB12Lab" subscription, "resourcegroup" resource group, "East US" region, and "standard" pricing tier. The "Networking" section shows "Secure Cluster Connectivity (No Public IP)" is set to "Yes" and "Deploy Azure Databricks workspace in your own Virtual Network (VNet)" is set to "No". The "Advanced" section shows "Enable Infrastructure Encryption" is set to "No". At the bottom, there are buttons for "Create", "< Previous", and "Download a template for automation".

## Step 6:

The screenshot shows the Microsoft Azure Resource Group Overview page for 'resourcegroup\_bricks101'. The main content area displays a green checkmark indicating 'Your deployment is complete' with the message: 'Deployment name: resourcegroup.bricks101 Subscription: Azure-DXC262AB12Lab Resource group: resourcegroup'. Below this, there are sections for 'Deployment details' (with a download link) and 'Next steps' (with a 'Go to resource' button). A sidebar on the right provides links to 'Cost Management', 'Microsoft Defender for Cloud', 'Free Microsoft tutorials', and 'Work with an expert'. The top navigation bar includes a search bar, account information, and various Azure service icons.

## Step 7:

The screenshot shows the Microsoft Azure Databricks Service Overview page for 'bricks101'. The left sidebar lists navigation options: Overview, Activity log, Access control (IAM), Tags, Settings (Virtual Network Peering, Encryption, Properties, Locks, Automation, Tasks (preview), Export template), Support + troubleshooting (New Support Request), Documentation, Getting Started, Import Data from File, and Import Data from Azure Storage. The main content area displays the 'Essentials' section with details: Status: Active, Resource group: resourcegroup, Location: East US, Subscription: Azure-DXC262AB12Lab, Subscription ID: 4236c42a-d131-4bd6-b609-aec3a598f2d3, and Tags (edit): Click here to add tags. To the right, there is a summary table with columns Managed Resource Group, URL, Pricing Tier, and Enable No Public IP. Below the table is a large red Databricks logo with the text 'Launch Workspace' underneath. At the bottom, there are four cards: Documentation, Getting Started, Import Data from File, and Import Data from Azure Storage.

## Step 8:

The screenshot shows the Microsoft Azure Databricks Data Science & Engineering interface. On the left, there is a sidebar with various navigation options: Data Science & Engineering (selected), Create, Workspace, Repos, Recents, Search, Data, Compute, Workflows, Help, Documentation, and Settings. Below the sidebar, it says '0/3 Tasks Completed'. In the center, there's a 'Data Science & Engineering' section with a 'Notebook' icon, a 'Cluster' icon, and a 'Job' icon. Below these are sections for 'Recent' notebooks, 'Documentation' (Get started guide, Best practices, Data guide), 'Release notes' (Runtime release notes, Platform release notes, More release notes), and 'Blog posts' (Building ETL pipelines for the cybersecurity lakehouse with Delta Live Tables, Streaming Windows Event Logs into the Cybersecurity Lakehouse, Speed Up Streaming Queries With Asynchronous State Checkpointing). At the bottom of the page, there's a URL: <https://adb-28437791802218644.azuredatabricks.net/o=28437791802218644#create/cluster>.

## Step 9:

The screenshot shows the Microsoft Azure Databricks Get started interface. On the left, there's a sidebar with 'Get started', 'Set up your workspace' (Create a cluster, Ingest data, Invite your team), and 'Next steps' (Explore Notebook gallery, Read documentation). Below the sidebar, it says '1/3' and 'Don't show again'. On the right, there's a 'Clusters / newcluster' configuration panel. It shows 'Configuration' tab selected. Under 'Cluster mode', it's set to 'Single Node'. Under 'Databricks Runtime Version', it's set to '10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)'. Under 'Autopilot options', there's a checked checkbox for 'Terminate after [100] minutes of inactivity'. Under 'Node type', it's set to 'Standard\_DS3\_v2' (14 GB Memory, 4 Cores). There's also a 'DBU / hour: 0.75' field and a 'Standard\_DS3\_v2' button. At the bottom of the configuration panel, there's a 'Advanced options' section.

## Step 10:

The screenshot shows the Microsoft Azure Databricks interface. On the left, there is a sidebar with various options: Data Science & Engineering, Create, Workspace, Repos, Recents, Search, Data, Compute (which is selected), Workflows, Tasks Completed (1/3), Help, Settings, and a user profile. The main area is titled 'Clusters / newcluster' and shows a cluster named 'newcluster' with a green checkmark icon. Below the cluster name are tabs for Configuration, Notebooks (0), Libraries, Event log, Spark UI, Driver logs, Metrics, Apps, and Spark cluster UI - Master. A 'More' button, an 'Edit' button, and a 'Terminate' button are at the top right. The 'Notebooks' tab is active, showing a table with columns Name, Status, Last Command Run, and Location. The table body contains the message 'No notebooks are attached to this cluster'.

## Step 11:

This screenshot is identical to the one in Step 10, showing the Microsoft Azure Databricks interface with the 'Compute' sidebar selected. It displays a cluster named 'newcluster' with a green checkmark icon. The 'Notebooks' tab is active, showing a table with the message 'No notebooks are attached to this cluster'.

## Step 12:

The screenshot shows the Microsoft Azure Databricks interface. On the left, the 'Get started' sidebar provides links for creating a cluster, ingest data, and invite your team. It also includes a 'Next steps' section with links to explore the Notebook gallery and read documentation. The main area is titled 'Notebook1 (Python)' and contains a code cell with the Python command: `print("Do some computation on data")`. The output of the command is displayed below, showing the execution took 0.84 seconds and was run at 6/11/2022, 10:37:09 PM on a new cluster.

## Step 13:

The screenshot shows the Microsoft Azure portal with the 'Data factories' service selected. The top navigation bar includes a search bar and user information. The main content area displays the 'Data factories' blade with options to 'Create' or 'View'. Below this, there are sections for 'Resources' (Recent and Favorite), 'Useful links' (Overview, Get started, Documentation), and 'Last Viewed' resources (Azure Databricks Service, Resource group). At the bottom, there are 'Navigate' and 'Tools' sections, along with a URL for the current page: <https://portal.azure.com/#blade/HubsExtension/BrowseResourceBlade/resourceType/Microsoft.DataFactory%2fdatafactories>.

## Step 14:

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and various navigation icons. The main title is "Data factories". Below the title, there are filter options: "Subscription == all", "Type == all", "Resource group == all", and "Location == all". There are also buttons for "+ Create", "Manage view", "Refresh", "Export to CSV", "Open query", and "Assign tags". 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. At the bottom right, there is a "Give feedback" button.

## Step 15:

The screenshot shows the "Create Data Factory" wizard in the Microsoft Azure portal. The title bar says "Create Data Factory". The "Basics" tab is selected. The "Project details" section asks to select a subscription and resource group. The "Subscription" dropdown is set to "Azure-DXC262AB12Lab" and the "Resource group" dropdown is set to "resourcegroup". The "Instance details" section allows setting the data factory name ("AP-morgan"), region ("East US"), and version ("V2 (Recommended)"). Navigation buttons at the bottom include "Review + create", "< Previous", and "Next : Git configuration >".

## Step 16:

The screenshot shows the Microsoft Azure 'Create Data Factory' wizard. The top navigation bar includes 'Microsoft Azure', a search bar, and user information 'dx262ab1211\_1654530... MANHAI PRO LEARN (MANHAI...)'. The main title is 'Create Data Factory' with a '...' button. A green banner at the top says 'Validation Passed'. Below it, tabs include 'Basics', 'Git configuration', 'Networking', 'Advanced', 'Tags', and 'Review + create' (which is underlined). A 'TERMS' section contains legal text about agreeing to terms and privacy statements. Under 'Basics', fields show 'Subscription: Azure-DXC262AB12Lab', 'Resource group: resourcegroup', 'Name: AP-morgan', 'Region: East US', and 'Version: V2 (Recommended)'. Under 'Networking', 'Connect via' is set to 'Public endpoint'. At the bottom are buttons for 'Create', '< Previous', 'Next >', and 'Download a template for automation'.

## Step 17:

The screenshot shows the Microsoft Data Factory Overview page for 'Microsoft.DataFactory-20220611221708'. The top navigation bar includes 'Microsoft Azure', a search bar, and user information 'dx262ab1211\_1654530... MANHAI PRO LEARN (MANHAI...)'. The main title is 'Microsoft.DataFactory-20220611221708 | Overview' with a '...' button. On the left, a sidebar has 'Overview' selected, along with 'Inputs', 'Outputs', and 'Template' buttons. The main content area shows a green checkmark icon and the message 'Your deployment is complete'. It lists deployment details: 'Deployment name: Microsoft.DataFactory-20220611221708', 'Subscription: Azure-DXC262AB12Lab', and 'Resource group: resourcegroup'. To the right, a 'Cost Management' section encourages staying within budget, with a 'Set up cost alerts' link. Other sections include 'Microsoft Defender for Cloud' (with a lock icon), 'Free Microsoft tutorials' (with a play icon), and 'Work with an expert' (with a person icon).

## Step 18:

The screenshot shows the Microsoft Azure portal interface for a Data Factory V2 resource named 'AP-morgan'. The top navigation bar includes the Microsoft Azure logo, a search bar, and various navigation icons. The main content area displays the 'Overview' tab for the resource group 'resourcegroup'. Key details shown include:

- Resource group (move):** resourcegroup
- Status:** Succeeded
- Location:** East US
- Subscription (move):** Azure-DXC262AB12Lab
- Subscription ID:** 4236c42a-d131-4bd6-b609-aec3a598f2d3

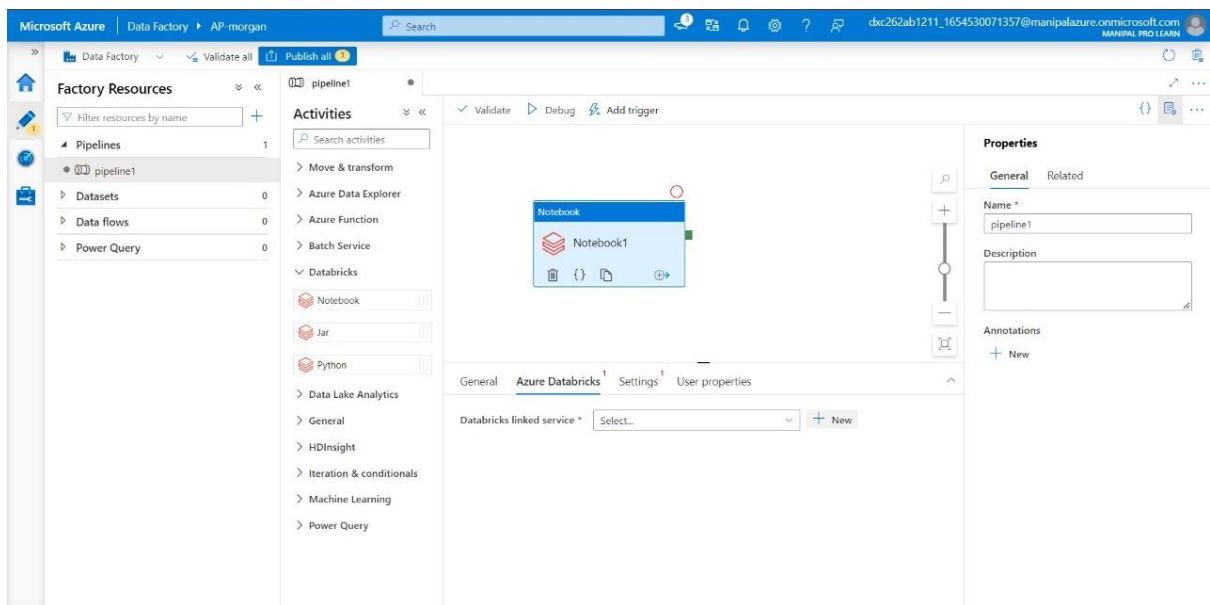
The left sidebar contains navigation links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (Networking, Managed identities, Properties, Locks), Getting started (Quick start), Monitoring (Alerts, Metrics, Diagnostic settings, Logs), and a Delete button.

The right side features a 'Getting started' section with links to 'Open Azure Data Factory Studio' and 'Read documentation', and a 'Monitoring' section showing PipelineRuns and ActivityRuns with one item each.

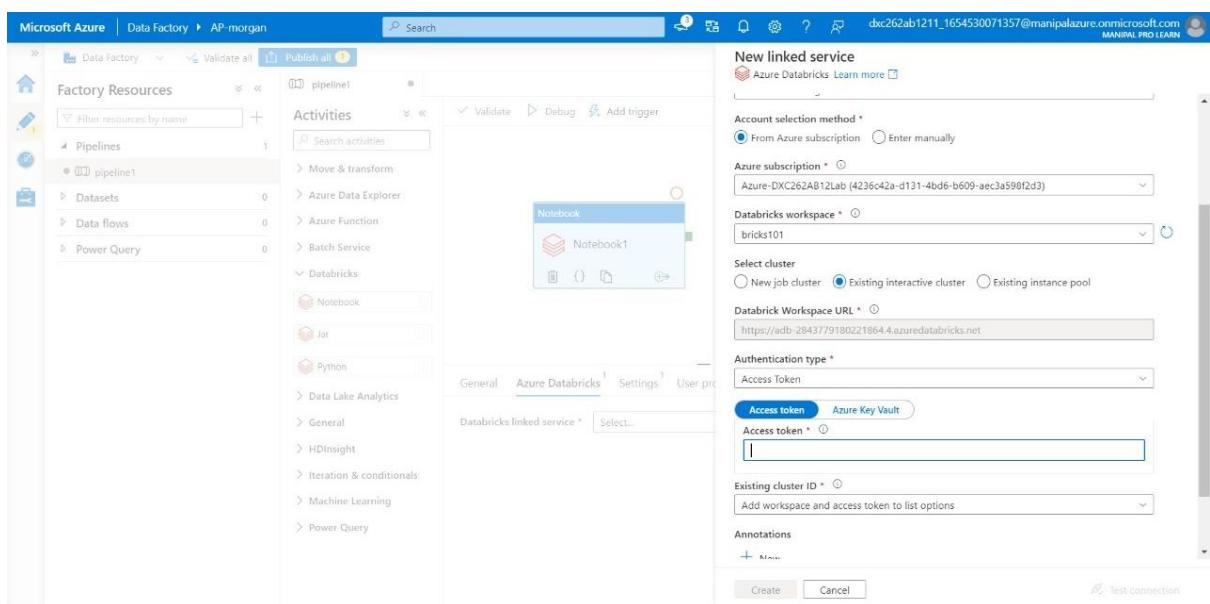
## Step 19:

The screenshot shows the Microsoft Azure portal interface for the 'AP-morgan' Data Factory. The top navigation bar includes the Microsoft Azure logo, a search bar, and various navigation icons. The main content area displays the 'Factory Resources' section for the 'Data Factory' category. The left sidebar lists 'Factory Resources' such as Pipelines, Datasets, Data flows, and Power Query. A central panel shows a resource creation dialog with options for 'Pipeline', 'Dataset', 'Data flow', 'Template gallery', and 'Import from pipeline template'. Below the dialog is a decorative icon of pipes and a calendar, and a callout text 'Select an item' with the sub-instruction 'Use the resource explorer to select or create a new item'.

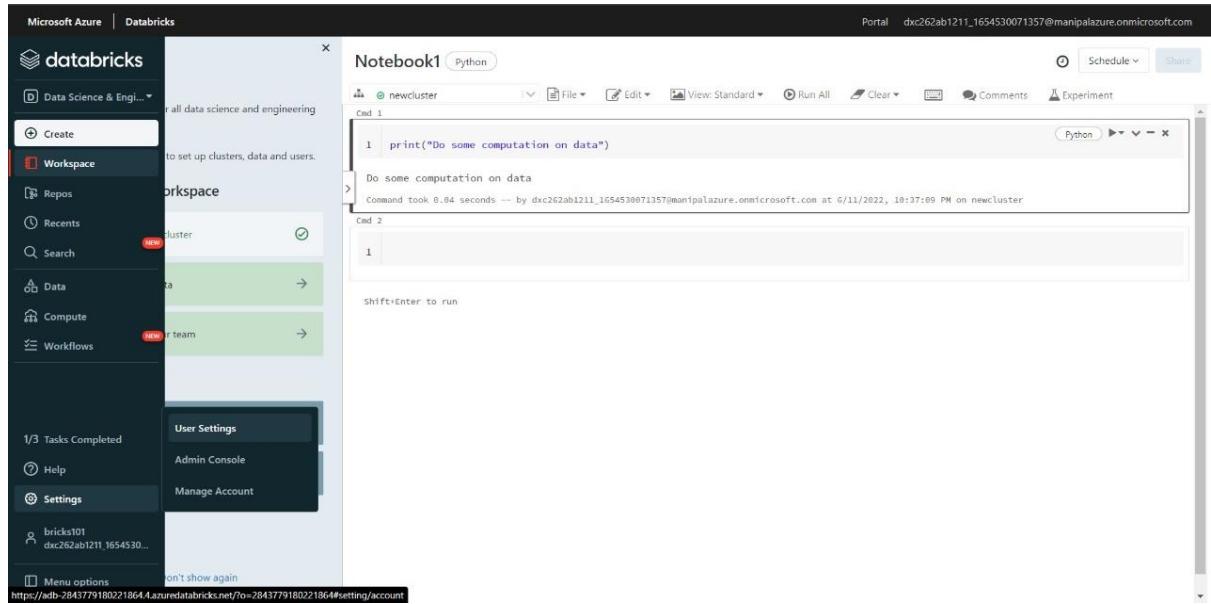
## Step 20:



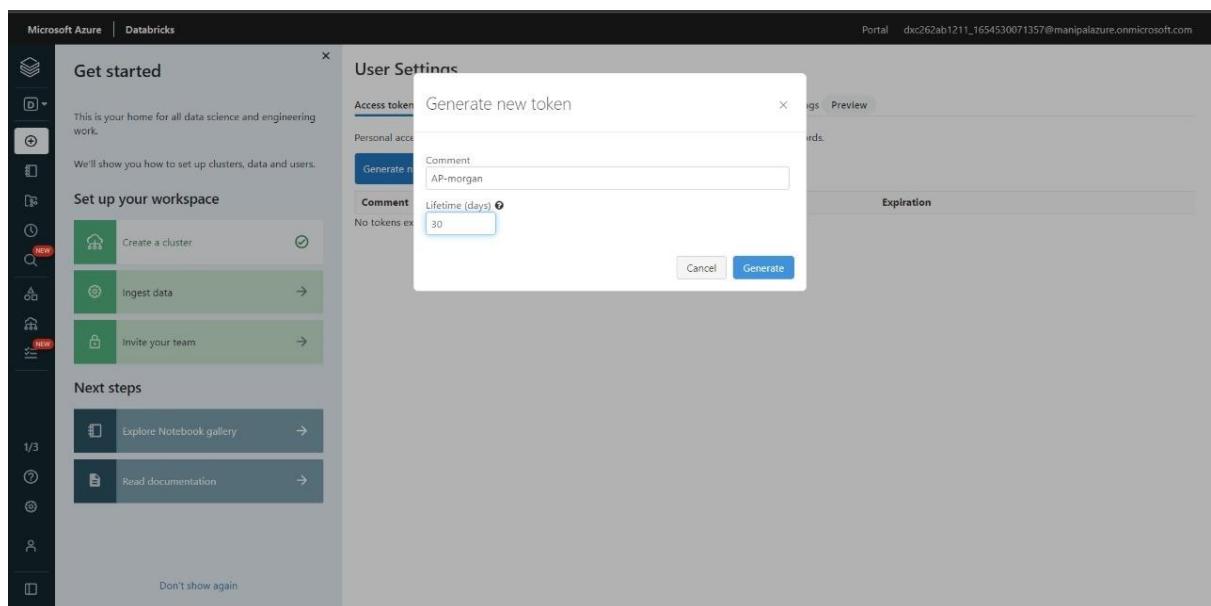
## Step 21:



## Step 22:



## Step 23:



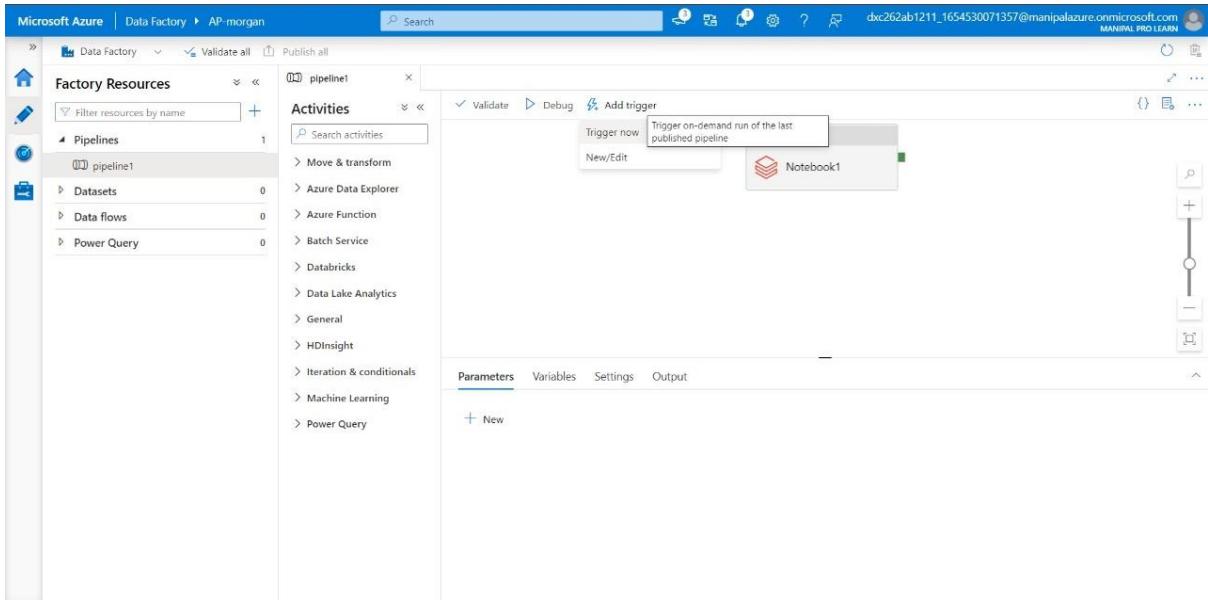
## Step 24:

The screenshot shows the Microsoft Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. Under Pipelines, 'pipeline1' is selected. In the main workspace, a 'Notebook' activity is being configured. The 'Activities' pane on the right shows various options like Move & transform, Azure Data Explorer, and Databricks. The 'Databricks' section is expanded, showing 'Notebook' selected. The 'Azure Databricks' tab is active. The configuration pane on the right is titled 'New linked service' and 'Azure Databricks'. It includes fields for 'Databricks workspace' (set to 'Azure-DX262AB12Lab'), 'Databricks linked service' (set to 'bricks101'), 'Select cluster' (radio button for 'Existing interactive cluster' selected), 'Databrick Workspace URL' (set to 'https://adb-2843779180221864.azuredatabricks.net'), 'Authentication type' (radio button for 'Access Token' selected), and 'Access token' (a redacted token). Below these are sections for 'Annotations', 'Parameters', and 'Advanced'. At the bottom are 'Create' and 'Cancel' buttons, and a 'Test connection' link.

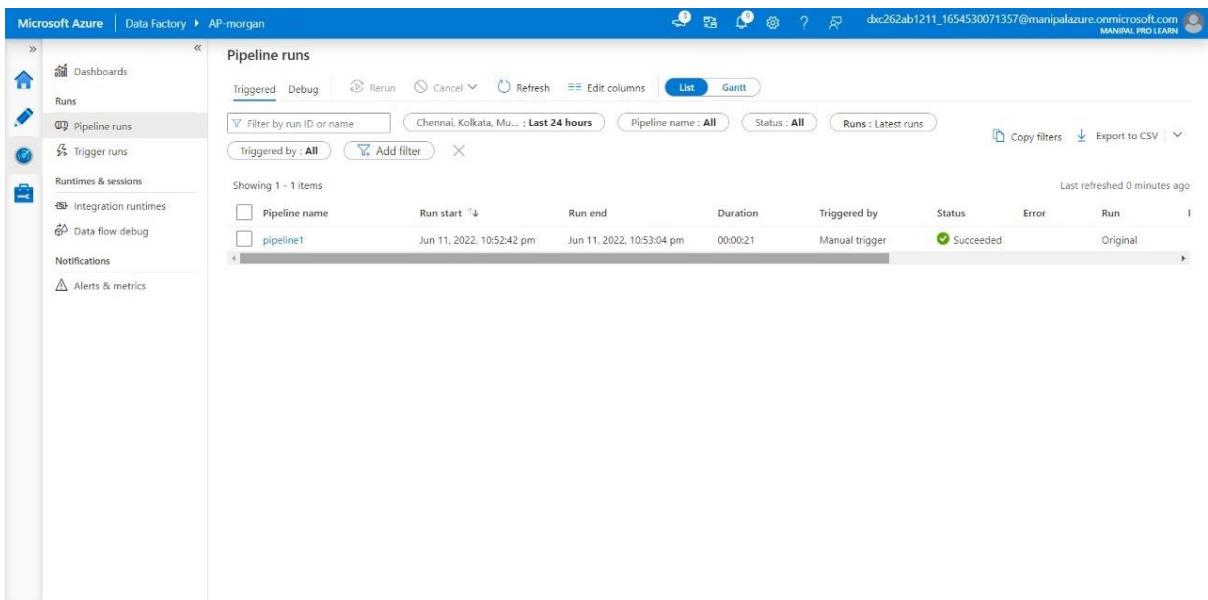
## Step 25:

This screenshot shows the same Microsoft Azure Data Factory pipeline editor interface as Step 24. The 'Notebook' activity is still being configured. The 'Azure Databricks' tab is active in the configuration pane. The 'Notebook path' field is now populated with 'Notebook1'. The 'Browse' button is visible above the path input. The configuration pane also shows sections for 'Base parameters' and 'Append libraries'. At the bottom are 'OK' and 'Cancel' buttons.

## Step 26:



## Step 27:



**Result:** Successfully linked and triggered azure DataBricks notebook using Data Factory.

## **Conclusion:**

- Our motive was to link Azure Databricks notebook using Data factory.
- By orchestrating Databricks notebooks through Azure Data Factory, we get the best of both worlds from each tool. The native connectivity, workflow management and trigger functionality built into Azure Data Factory, and at the same time we get the limitless flexibility to code within Databricks.
- In this project we were successfully able to link and trigger azure Data Bricks notebook using Data Factory.