

DXC REALTIME PROJECTS

AZ-900, DP - 203



[DATE]
[COMPANY NAME]
[Company address]

Name: Shaik mohammed Raihan

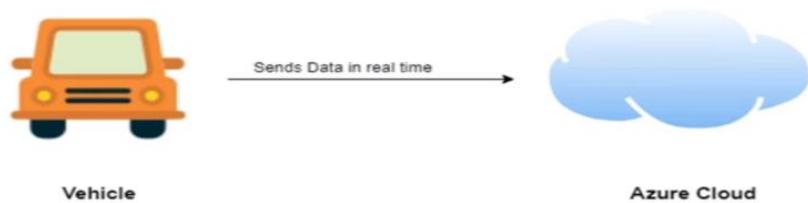
Reg No: DXC262AB12036

Project1 Name: Smart Vehicles

Date:12/6/2022

Project 1 : Connected Vehicles

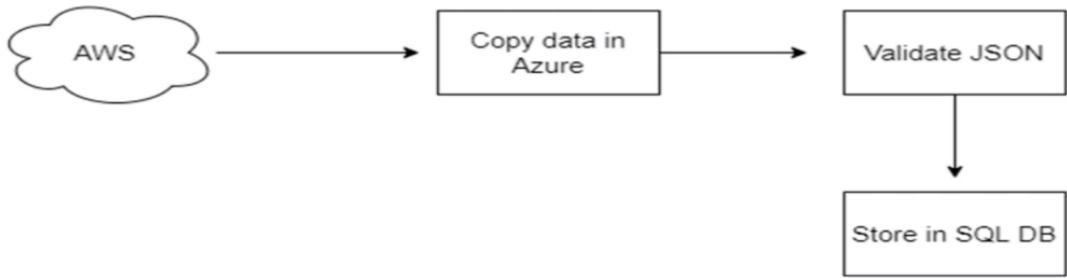
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



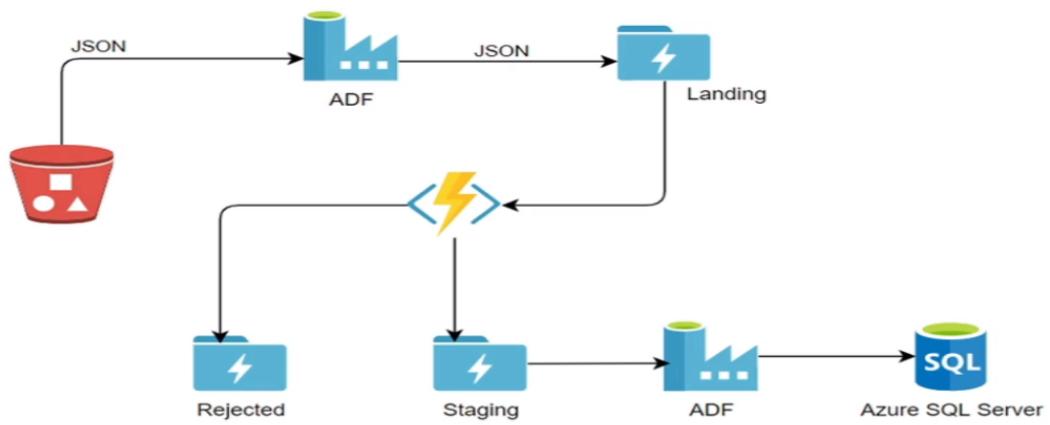
Project 1 : Connected Vehicles

- 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 enabled

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

Create datafactory

The screenshot shows the Microsoft Azure portal interface. At the top, there are three tabs: "Subscription Details | Nuvepro", "Home - Microsoft Azure", and "Power BI". The main search bar contains the text "data fa". Below the search bar, the "Microsoft Azure" navigation bar is visible with options like "Create resource", "Recent", and "Name". The "Services" section is selected, showing categories such as "Data factories", "Reservations", "SQL databases", and "Azure Database for MySQL servers". A search bar at the bottom of this section shows "Searching all subscriptions." and a "type" dropdown. To the right, there's a "More services" button and a "Give feedback" link. The "Recent" section lists "dxcdb231" and "Azure Databricks Service" with a timestamp of "2 minutes ago". Below this, there are "Navigate" and "Tools" sections with links to "Subscriptions", "Resource groups", "All resources", and "Dashboard". The status bar at the bottom shows "ENG IN 4:00 PM 6/10/2022".

The main content area displays the "Data factories" page under the "Data factories" category. The URL is <https://portal.azure.com/#view/HubsExtension/BrowseResource/resourceType/Microsoft.DataFactory%2FdataFactories>. The page title is "Data factories". It includes a toolbar with "Create", "Manage view", "Refresh", "Export to CSV", "Open query", and "Assign tags". There are filter buttons for "Subscription == all", "Type == all", "Resource group == all", and "Location == all". Below the filters, there are dropdowns for "No grouping" and "List view". A "Name" column header is present. In the center, there is a large icon of a factory building and the text "No data factories to display. Try changing or clearing your filters." A blue "Create data factory" button is available. The status bar at the bottom shows "ENG IN 4:00 PM 6/10/2022".

After opening datafactory click on create then give required

The screenshot shows the 'Create Data Factory' wizard on the Microsoft Azure portal. The current step is 'Project details'. It asks for a subscription and a resource group. The subscription dropdown shows 'Azure-DXC262AB12Lab' and the resource group dropdown shows 'dxcrg231'. Below this, there are fields for 'Name' (set to 'dxcpro'), 'Region' (set to 'East US'), and 'Version' (set to 'V2 (Recommended)'). At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next : Git configuration >'.

The screenshot shows the 'Create Data Factory' wizard on the Microsoft Azure portal. The current step is 'Git configuration'. It allows configuring a Git repository with either Azure DevOps or GitHub. A note 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.' There is a link to 'Learn more about Git integration in Azure Data Factory'. A checkbox labeled 'Configure Git later' is checked. At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next : Networking >'.

The screenshot shows the 'Create Data Factory' wizard on the Microsoft Azure portal. The current step is 'Networking'. This step is currently empty, indicated by a large white area. At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next : Networking >'.

Subscription Details | Nuvepro

Create Data Factory - Microsoft

Power BI

https://portal.azure.com/#create/Microsoft.DataFactory

Microsoft Azure

Search resources, services, and docs (G+)

Home > Data factories >

Create Data Factory

Validation Passed

Basics Git configuration Networking Advanced Tags Review + create

TERMS

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

Basics

Subscription: Azure-DXC262AB12Lab
Resource group: dxcrg231
Name: dxcproj
Region: East US
Version: V2 (Recommended)

Create < Previous Next Download a template for automation

Subscription Details | Nuvepro

Microsoft.DataFactory-20220610160032

Power BI

https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F3d94c073-3bce-4ba8-a171-6976f9...

Microsoft Azure

Search resources, services, and docs (G+)

Home >

Microsoft.DataFactory-20220610160032 | Overview

Deployment

Search (Ctrl+)

Delete Cancel Redeploy Refresh

We'd love your feedback! →

Deployment is in progress

Deployment name: Microsoft.DataFactory-20220610160032
Subscription: Azure-DXC262AB12Lab
Resource group: dxcrg231

Start time: 6/10/2022, 4:09:34 PM
Correlation ID: c15c4755-a2bd-403a-86c7-bef616c5ebbe

Deployment details (Download)

Resource	Type	Status	Operation details
No results.			



The screenshot shows the Microsoft Azure Data Factory Overview page for a deployment named "Microsoft.DataFactory-20220610160032". The main message is "Your deployment is complete". Deployment details include: Deployment name: Microsoft.DataFactory-20220610160032, Start time: 6/10/2022, 4:09:34 PM, Subscription: Azure-DXC262AB12Lab, Correlation ID: c15c4755-a2bd-403a-86c7-bef616c5ebee, Resource group: dxrcrg231. A "Go to resource" button is present. On the right, there are promotional cards for Cost Management, Microsoft Defender for Cloud, and Free Microsoft tutorials.

After deployment click on open azure data factory studio

The screenshot shows the Microsoft Azure Data Factory studio for the "dxproj" resource. The left sidebar lists navigation options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Networking, Managed identities, Properties, Locks, Getting started, Quick start, Monitoring, and Alerts. The main area displays the "Essentials" section with details: Resource group (move) : dxrcrg231, Status : Succeeded, Location : East US, Subscription (move) : Azure-DXC262AB12Lab, Subscription ID : 3d94c073-3bce-4ba8-a171-6976f92e1322. It also includes sections for Getting started (with links to Open Azure Data Factory Studio and Read documentation) and Monitoring (PipelineRuns and ActivityRuns).

Click on ingest

The screenshot shows the Microsoft Azure Data Factory portal. At the top, there are three tabs: "Subscription Details | Nuvepro", "dxcpoj - Microsoft Azure", and "dxcpoj - Azure Data Factory". The current view is under the "Data Factory" tab, specifically for the "dxcpoj" project. A banner at the top 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. Learn more". Below the banner, there's a "Set up code repository" button. The main area is titled "Data factory" and "dxcpoj". It features a large 3D-style diagram of industrial buildings and data flows. 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). Further down, there's a "Discover more" section with links to "Browse partners", "Pipeline templates", and "SAP pipeline templates". The bottom of the screen shows a Windows taskbar with various pinned icons and the date/time as 6/10/2022, 4:12 PM.

Now open storage account and create it

The screenshot shows the Microsoft Azure portal home page. The search bar at the top contains the text "blob". Below the search bar, the "Services" section is expanded, showing "Storage accounts" as the selected item. Other options include "Storage browser (preview)" and "Storage Explorer". To the right, there are sections for "Free services" (FS Archiver, SFTP Gateway for Azure - SFTP to Blob storage, Flexify.IO - Amazon S3 API for Azure Blob Storage, FileMage SFTP FTP to Azure Blob Storage - Linux) and "Reservations". The left sidebar includes sections for "Recent" (with items like "dxcproj" and "dxcproj"), "Create resource", "Azure Active Directory", and "Marketplace". The bottom of the screen shows a Windows taskbar with various pinned icons and the date/time as 6/10/2022, 4:32 PM.

Screenshot of the Microsoft Azure Storage accounts page showing no storage accounts displayed.

No storage accounts to display

Create a storage account

Validation passed

Basics Advanced Networking Data protection Encryption Tags Review + create

Enable SFTP (preview) Disabled
Large file shares Disabled

Networking

Network connectivity Public endpoint (all networks)
Default routing tier Microsoft network routing

Data protection

Point-in-time restore Disabled
Blob soft delete Enabled
Blob retention period in days 7
Container soft delete Enabled

Create < Previous Next > Download a template for automation

The screenshot shows the Microsoft Azure portal interface. The main title bar displays "Subscription Details | Nuvepro" and "dxctr_1654858999168 - Microsoft Azure Data Factory". The URL in the address bar is "https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F3d94c073-3bce-4ba8-a171-6976f9...". A search bar at the top says "Search resources, services, and docs (G+)". On the left, there's a navigation menu with "Home", "Overview", "Inputs", "Outputs", and "Template". The main content area has a heading "Deployment is in progress" with sub-sections for "Deployment name: dxctr_1654858999168", "Subscription: Azure-DXC262AB1Lab", and "Resource group: dxcrg231". It also shows "Start time: 6/10/2022, 4:33:28 PM", "Correlation ID: 885f7191-4a25-48d1-8663-7b129bd04e03", and a download link. A tooltip message "Deployment to resource group 'dxcrg231' is in progress." is visible. Below this, a section titled "Deployment details" with a download link is shown, followed by a table with columns "Resource", "Type", "Status", and "Operation details". The table body contains the message "No results."

This screenshot is nearly identical to the one above, showing the same Azure portal interface and deployment details. The main difference is in the status message: instead of "Deployment is in progress", it now says "Your deployment is complete". The deployment information remains the same: "Deployment name: dxctr_1654858999168", "Subscription: Azure-DXC262AB1Lab", and "Resource group: dxcrg231". The start time is still "6/10/2022, 4:33:28 PM" and the correlation ID is "885f7191-4a25-48d1-8663-7b129bd04e03". The deployment details table still shows "No results.". To the right of the main content, there are several promotional cards: "Cost Management" (Get notified to stay within your budget and prevent unexpected charges on your bill. Set up cost alerts >), "Microsoft Defender for Cloud" (Secure your apps and infrastructure. Go to Microsoft Defender for Cloud >), "Free Microsoft tutorials" (Start learning today >), and "Work with an expert".

Go to containers and create source container and upload a file

The screenshot shows the Microsoft Azure Storage account overview for 'dxctr'. Key details include:

- Resource group: dxrg231
- Location: East US
- Primary/Secondary Location: Primary: East US, Secondary: West US
- Subscription: Azure-DXC262AB12Lab
- Subscription ID: 3d94c073-3bce-4ba8-a171-6976f92e1322
- Disk state: Primary: Available, Secondary: Available
- Tags: Click here to add tags

The 'Properties' tab is selected, showing Blob service settings like Hierarchical namespace (Disabled), Default access tier (Hot), Blob public access (Enabled), and Container soft delete (Enabled (7 days)).

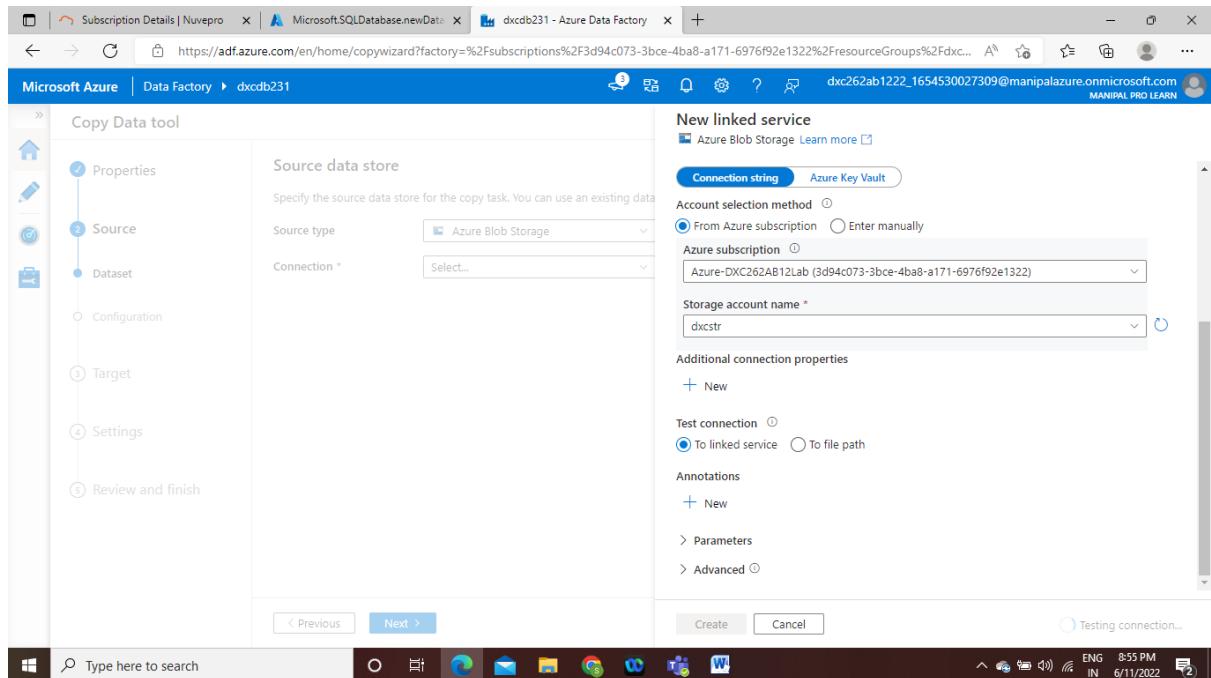
The screenshot shows the 'New container' dialog for creating a new blob container named 'sourceblob'. The 'Public access level' is set to 'Blob (anonymous read access for blobs only)'. A warning message states: 'Blobs within the container can be read by anonymous request, but container data is not available. Anonymous clients cannot enumerate the blobs within the container.' The 'Create' button is visible at the bottom.

The screenshot shows the Microsoft Azure Storage Container 'surceblob'. On the right, an 'Upload blob' dialog is open, prompting for a file to upload. The container details show an 'Authentication method' of 'Access key' and a 'Location' of 'surceblob'. The main pane displays a table with columns: Name, Modified, Access tier, Archive status, Blob type, and Size. A message indicates 'No results'.

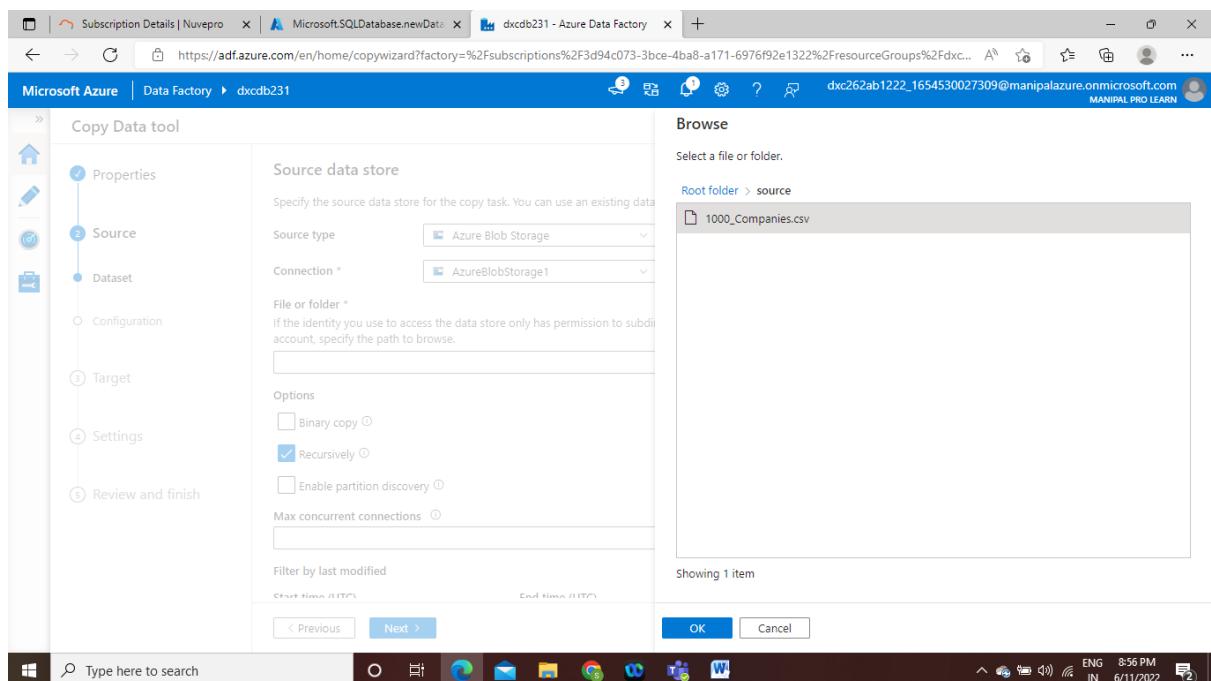
The screenshot shows the Microsoft Azure Storage Container 'surceblob'. The table now lists a single blob named '1000_Companies.csv'. The blob details are: Name - 1000_Companies.csv, Modified - 6/10/2022, 4:37:13 PM, Access tier - Hot (Inferred), Blob type - Block blob, Size - 50 KiB, Lease state - Available.

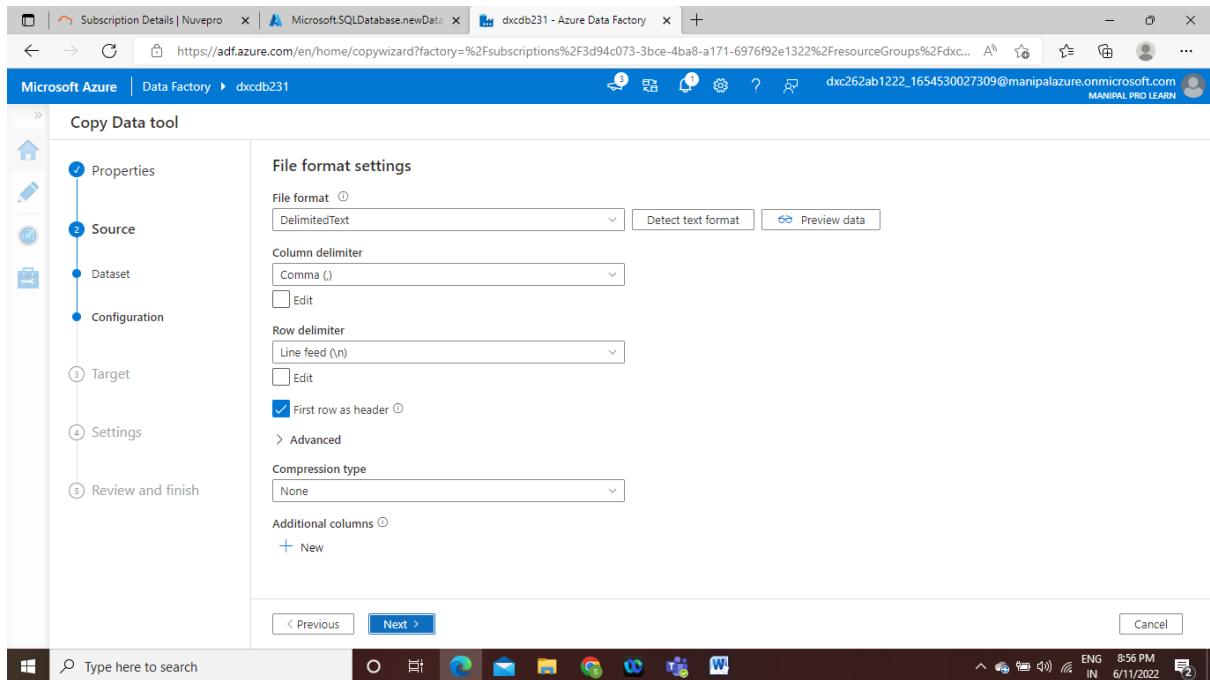


Then to link we connect data storage and sql data base

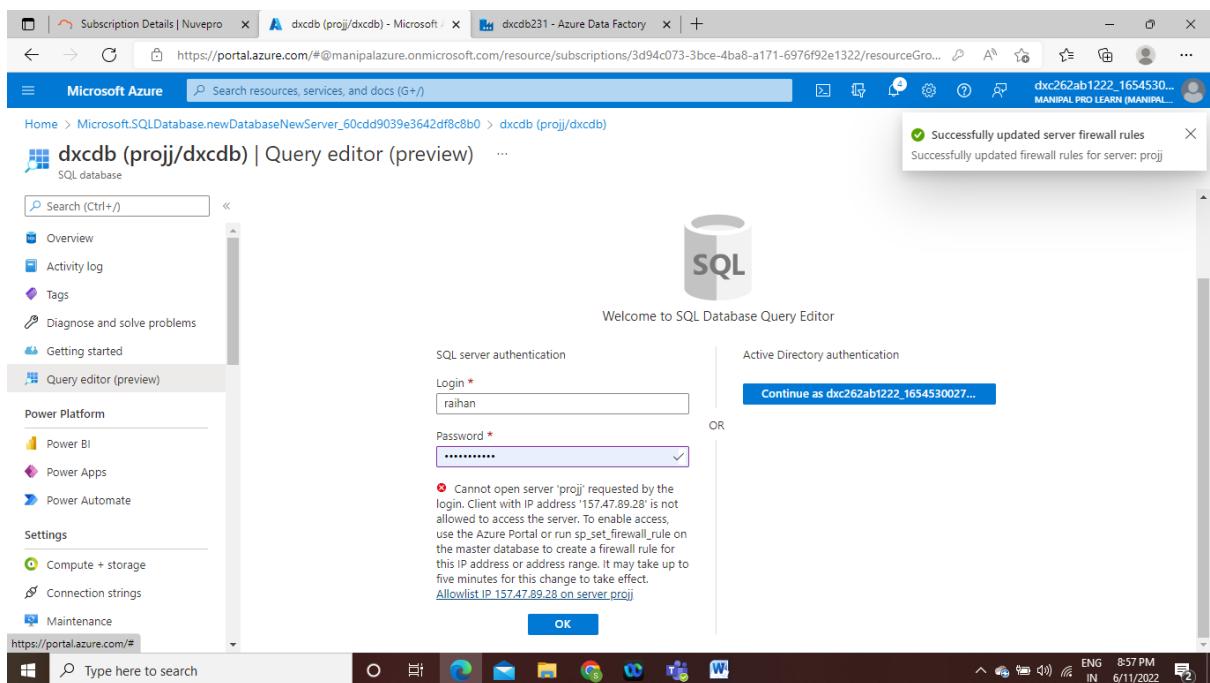


Now select source data





Now create sql data to link storage accounts



The screenshot shows two windows side-by-side. The left window is titled 'Copy Data tool' and is part of the 'dxedb (proj/dxedb) - Microsoft' project in 'dxedb231 - Azure Data Factory'. It displays the 'Destination data store' configuration step, showing a tree view with 'Properties', 'Source', 'Target' (selected), 'Dataset', 'Configuration', 'Settings', and 'Review and finish'. Under 'Target', 'Target type' is set to 'Azure SQL Database' and 'Connection' is set to 'Select...'. To the right, a 'New linked service' dialog is open for an 'Azure SQL Database'. It includes fields for 'Database name' (set to 'dxedb'), 'Authentication type' (set to 'SQL authentication'), 'User name' (set to 'raihan'), 'Password' (set to '*****'), and 'Always encrypted' (unchecked). Below these are 'Additional connection properties' and 'Annotations' sections. At the bottom are 'Create' and 'Cancel' buttons, and a 'Test connection' link. The status bar at the bottom of the browser window shows 'ENG IN 8:59 PM 6/11/2022'. The right window is titled 'Create SQL Database - Microsoft' and is part of the 'Subscription Details | Nuvepro' project. It shows the 'Networking' tab of the 'Create SQL Database' wizard. The 'Basics' tab is also visible. The 'Networking' tab has 'Configure network access and connectivity for your server. The configuration selected below will apply to the selected server 'projjj' and all databases it manages.' A note says 'Learn more'. Under 'Network connectivity', 'Public endpoint' is selected. Under 'Connectivity method', 'Allow Azure services and resources to access this server' is set to 'Yes'. At the bottom are 'Review + create', '< Previous', and 'Next : Security >' buttons. The status bar at the bottom of this window shows 'ENG IN 9:18 PM 6/11/2022'.

Subscription Details | Nuvepro **Networking - Microsoft Azure** **dxcdb231 - Azure Data Factory**

Microsoft Azure | Data Factory > dxcdb231

Copy Data tool

Destination data store

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

Target type: Azure SQL Database

Connection: AzureSqlDatabase2

Source: Azure Blob Storage file → **Target**: Loading...

Skip column mapping for all tables

< Previous **Next >** Cancel

Successfully created
Successfully created AzureSqlDatabase2 (Linked service).

Type here to search

Subscription Details | Nuvepro **Networking - Microsoft Azure** **dxcdb231 - Azure Data Factory**

Microsoft Azure | Data Factory > dxcdb231

Copy Data tool

Properties

Source

Target

Settings

Review and finish

Settings

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

Task name *: CopyPipeline_yk0

Task description:

Data consistency verification:

Fault tolerance:

Enable logging:

Enable staging:

> Advanced

< Previous **Next >** Cancel

Subscription Details | Nuvepro | Networking - Microsoft Azure | dxcdb231 - Azure Data Factory | https://adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F3d94c073-3bce-4ba8-a171-6976f92e1322%2FresourceGroups%2B... | Microsoft Azure | Data Factory > dxcdb231 | dxc262ab1222_1654530027309@manipalazure.onmicrosoft.com | MANIPAL PRO LEARN

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 SQL Database.

Azure Blob Storage → Azure SQL Database

Properties

Task name: CopyPipeline_yk0

Task description:

Source

Connection name: AzureBlobStorage3

Dataset name: SourceDataset_yk0

Column delimiter: ,

Row delimiter: \

Escape character: \

< Previous | Next > | Cancel

The screenshot shows the Microsoft Azure Copy Data tool interface. The left sidebar lists steps: Properties (checked), Source (checked), Target (checked), Settings (checked), Review and finish (checked), Review (unchecked), and Deployment (unchecked). The main area is titled 'Summary' and shows a flow from 'Azure Blob Storage' to 'Azure SQL Database'. Below this, the 'Properties' section details a task named 'CopyPipeline_yk0'. Under 'Source', it specifies 'Connection name: AzureBlobStorage3' and 'Dataset name: SourceDataset_yk0'. Other source settings include column delimiter (,), row delimiter (\), and escape character (\). Navigation buttons at the bottom include '< Previous' and 'Next >'.

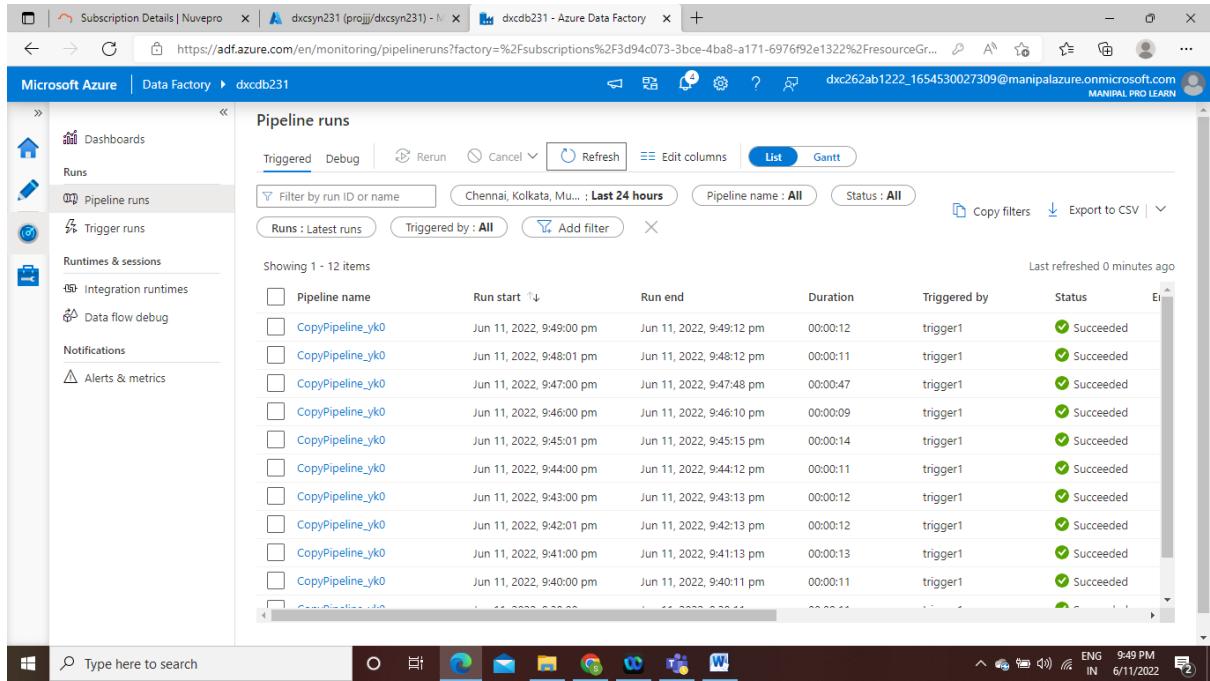
See the file uploaded in storage account is visible here

The screenshot shows the Microsoft Azure portal interface. At the top, there are three tabs: 'Subscription Details | Nuvepro', 'dxcsyn231 (projjj/dxcsyn231) - A', and 'dxdb231 - Azure Data Factory'. The main browser window displays the 'Query editor (preview)' for the 'dxcsyn231 (projjj/dxcsyn231)' database. The left sidebar includes links for Overview, Activity log, Tags, Diagnose and solve problems, Getting started, and a prominent 'Query editor (preview)' section. Below these are sections for Power Platform, Settings, and Properties. The central pane shows a tree view under 'dxcsyn231 (raihan)': 'Tables' (selected), '1000_companies.Azure Blob Storage file', 'Views', and 'Stored Procedures'. A tooltip indicates: 'Showing limited object explorer here. For full capability please open SSDT.' In the 'Query 1' editor, the query 'select * from [1000_companies].[Azure Blob Storage file]' is entered. The results pane shows a table with the following data:

R&D Spend	Administration	Marketing Spend	State	Profit
165349.2	136897.8	471784.1	New York	192261.83
162597.7	151377.59	443898.53	California	191792.06
153441.51	101145.55	407934.54	Florida	191050.39

The status bar at the bottom right shows 'ENG 9:36 PM IN 6/11/2022'.

We gave trigger to run those files



The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has a 'Pipeline runs' section selected. The main area displays a table of 'Pipeline runs' with the following columns: Pipeline name, Run start, Run end, Duration, Triggered by, and Status. There are 12 rows, each representing a successful run ('Succeeded') triggered by 'trigger1'. The table includes filters for 'Run ID or name' (Chennai, Kolkata, Mu...), 'Last 24 hours', 'Pipeline name : All', and 'Status : All'. The status bar at the bottom indicates 'Last refreshed 0 minutes ago'.

Pipeline name	Run start	Run end	Duration	Triggered by	Status
CopyPipeline_yk0	Jun 11, 2022, 9:49:00 pm	Jun 11, 2022, 9:49:12 pm	00:00:12	trigger1	Succeeded
CopyPipeline_yk0	Jun 11, 2022, 9:48:01 pm	Jun 11, 2022, 9:48:12 pm	00:00:11	trigger1	Succeeded
CopyPipeline_yk0	Jun 11, 2022, 9:47:00 pm	Jun 11, 2022, 9:47:48 pm	00:00:47	trigger1	Succeeded
CopyPipeline_yk0	Jun 11, 2022, 9:46:00 pm	Jun 11, 2022, 9:46:10 pm	00:00:09	trigger1	Succeeded
CopyPipeline_yk0	Jun 11, 2022, 9:45:01 pm	Jun 11, 2022, 9:45:15 pm	00:00:14	trigger1	Succeeded
CopyPipeline_yk0	Jun 11, 2022, 9:44:00 pm	Jun 11, 2022, 9:44:12 pm	00:00:11	trigger1	Succeeded
CopyPipeline_yk0	Jun 11, 2022, 9:43:00 pm	Jun 11, 2022, 9:43:13 pm	00:00:12	trigger1	Succeeded
CopyPipeline_yk0	Jun 11, 2022, 9:42:01 pm	Jun 11, 2022, 9:42:13 pm	00:00:12	trigger1	Succeeded
CopyPipeline_yk0	Jun 11, 2022, 9:41:00 pm	Jun 11, 2022, 9:41:13 pm	00:00:13	trigger1	Succeeded
CopyPipeline_yk0	Jun 11, 2022, 9:40:00 pm	Jun 11, 2022, 9:40:11 pm	00:00:11	trigger1	Succeeded
CopyPipeline_yk0	Jun 11, 2022, 9:39:00 pm	Jun 11, 2022, 9:39:11 pm	00:00:11	trigger1	Succeeded

Result: In this project we were successful in creating a pipeline that will validate and copy the blob data into the SQL database using Azure Data Factory

Conclusion: The Blob data is being successfully validated and stored into SQL database.

Name:Shaik mohammed Raihan

Reg No: DXC262AB12036

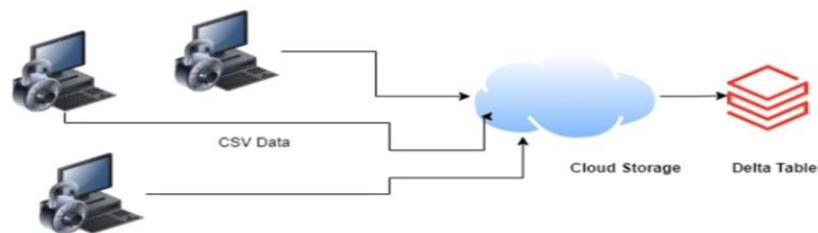
Project2 Name: AP Morgan Data Platform

Date:12/6/2022

Project 2: AP Morgan Data Platform

Project 2 : AP Morgan

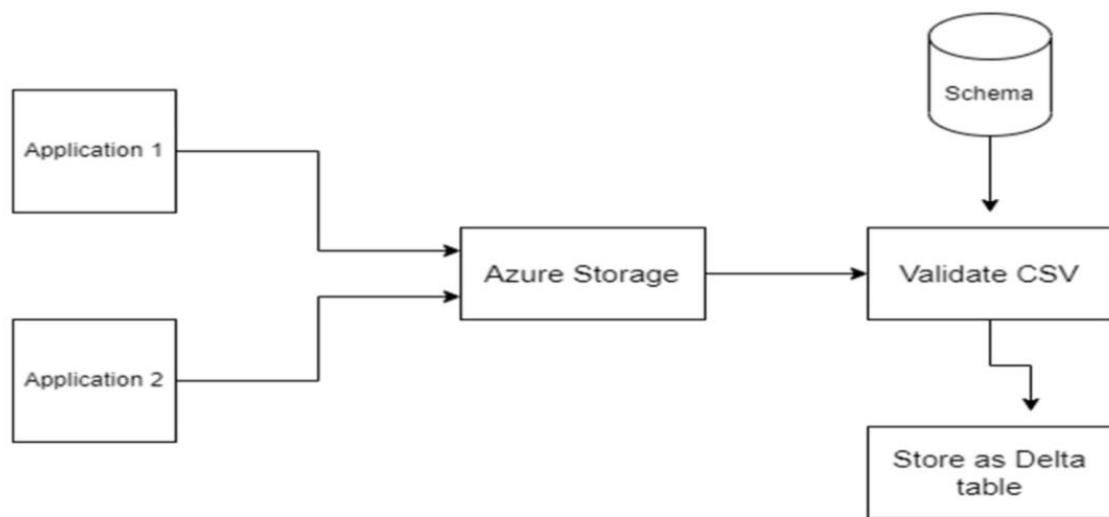
- Multiple Internal applications sends the data(huge size) in CSV format on daily basis in the cloud storage location. There are couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed data to be persisted as Delta table in Databricks for downstream system.



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**

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

Create data bricks

The screenshot shows the Microsoft Azure portal with the URL <https://portal.azure.com/#create/Microsoft.Databricks>. The page title is "Create an Azure Databricks workspace". The top navigation bar includes "Microsoft Azure", a search bar, and user information. Below the title, the breadcrumb navigation shows "Home > Azure Databricks > Create an Azure Databricks workspace". The main content area has tabs for "Basics", "Networking", "Advanced", "Tags", and "Review + create". The "Basics" tab is selected. The "Project Details" section asks to select a subscription and resource group. A dropdown for "Subscription" shows "Azure-DXC262AB12Lab". A dropdown for "Resource group" shows "(New) dxproj" with "Create new" as an option. The "Instance Details" section includes fields for "Workspace name" (with placeholder "Enter name for Databricks workspace"), "Region" (set to "West 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 >". The taskbar at the bottom shows various pinned icons and the system status bar indicates "ENG IN 9:39 AM 6/12/2022".

Give required credentials to create data bricks

This screenshot is identical to the one above, but the "Resource group" dropdown now shows "(New) dxproj" instead of "Create new", indicating that a new resource group has been selected.

Subscription Details | Nuvepro | Create an Azure Databricks workspace | https://portal.azure.com/#create/Microsoft.Databricks

Microsoft Azure | Search resources, services, and docs (G+) | Home > Azure Databricks > Create an Azure Databricks workspace

Validation Succeeded

Basics Networking Advanced Tags Review + create

Summary

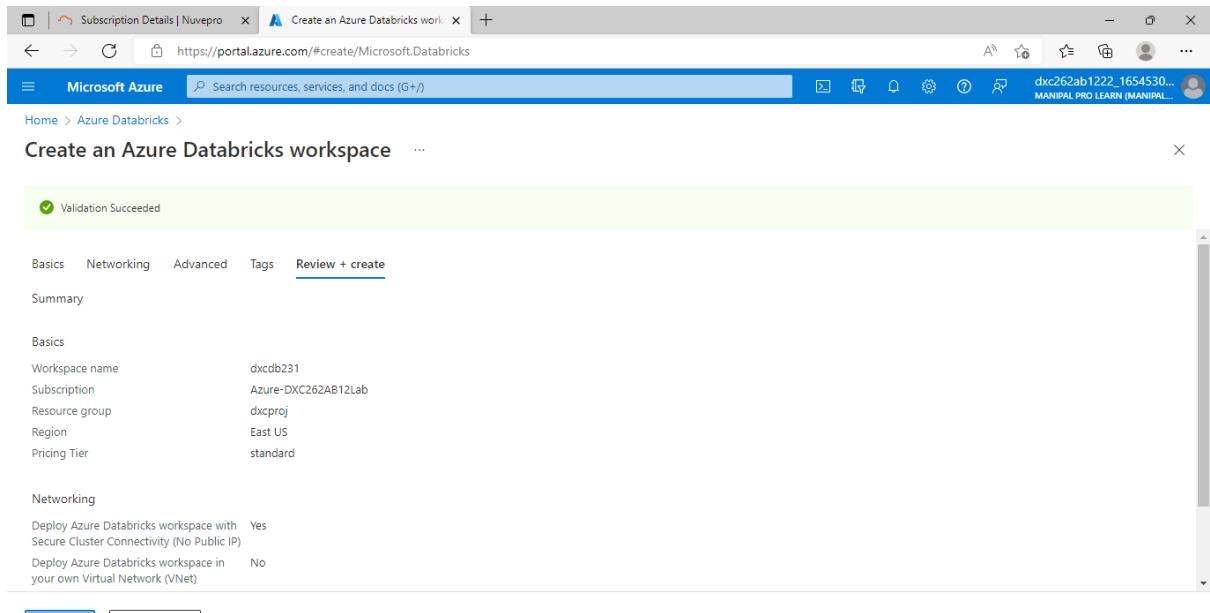
Basics

Workspace name	dxcdb231
Subscription	Azure-DXC262AB12Lab
Resource group	dxcproj
Region	East US
Pricing Tier	standard

Networking

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

Create < Previous Download a template for automation



Subscription Details | Nuvepro | dxcproj_dxcdb231 - Microsoft Azure | https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F3d94c073-3bce-4ba8-a171-6976f9... | ENG IN 9:41 AM 6/12/2022

Microsoft Azure | Search resources, services, and docs (G+) | Home > dxcproj_dxcdb231 | Overview

Deployment

Search (Ctrl+F) Overview Delete Cancel Redeploy Refresh

We'd love your feedback! →

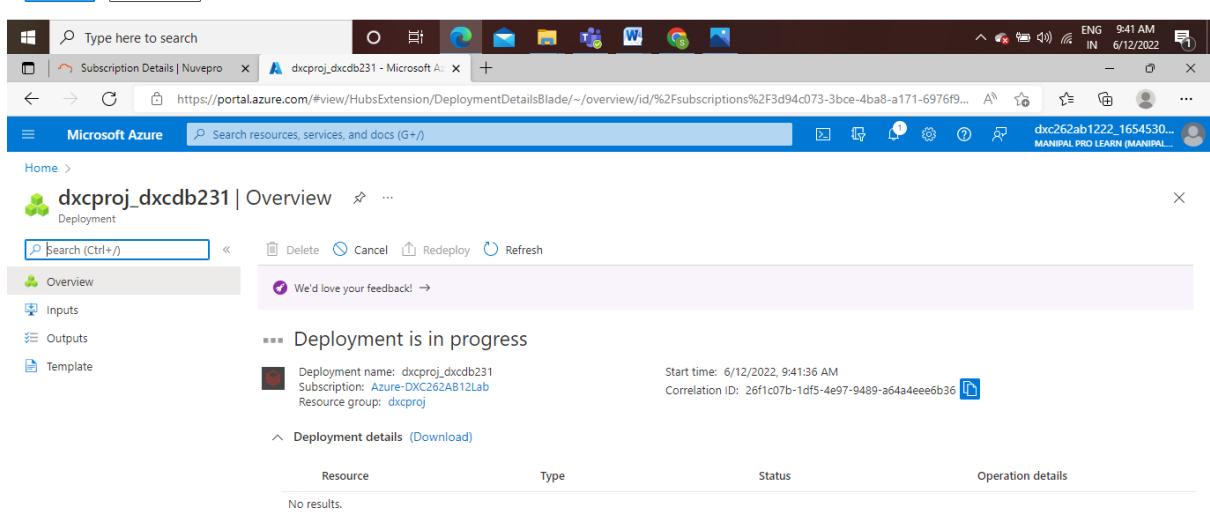
Deployment is in progress

Deployment name: dxcproj_dxcdb231
Subscription: Azure-DXC262AB12Lab
Resource group: dxcproj

Start time: 6/12/2022, 9:41:36 AM
Correlation ID: 26f1c07b-1df5-4e97-9489-a64a4eee6b36

Deployment details (Download)

Resource	Type	Status	Operation details
No results.			



The screenshot shows the Microsoft Azure portal interface. The main title bar reads "Subscription Details | Nuvepro" and "dxcproj_dxcd231 - Microsoft Azure". The search bar says "Search resources, services, and docs (G+)". The user is signed in as "MANIPAL PRO LEARN (MANIPAL...)". The main content area is titled "dxcproj_dxcd231 | Overview" and displays a message: "Your deployment is complete". It shows deployment details: Deployment name: dxcproj_dxcd231, Subscription: Azure-DXC262AB12Lab, Resource group: dxcproj. Below this are sections for "Deployment details" (with a download link) and "Next steps". A "Go to resource" button is present. On the right side, there are promotional cards for "Cost Management" (Get notified to stay within your budget and prevent unexpected charges on your bill. Set up cost alerts >) and "Microsoft Defender for Cloud" (Secure your apps and infrastructure. Go to Microsoft Defender for Cloud >). Below these are links for "Free Microsoft tutorials" (Start learning today >) and "Work with an expert". The taskbar at the bottom includes icons for File Explorer, Task View, Edge browser, Mail, OneDrive, Word, Excel, and Power BI, along with system status indicators like battery level, signal strength, and date/time (9:45 AM, 6/12/2022).

After deployment go to launch workspace

The screenshot shows the Microsoft Azure portal interface. The main title bar reads "Subscription Details | Nuvepro" and "dxcd231 - Microsoft Azure". The search bar says "Search resources, services, and docs (G+)". The user is signed in as "MANIPAL PRO LEARN (MANIPAL...)". The main content area is titled "dxcd231" and "Azure Databricks Service". It shows an "Overview" section with a "Delete" button. Below it is an "Essentials" table with the following data:

Status	: Active	Managed Resource Group : databricks-rg-dxcd231-gamwkcj7oxj56
Resource group	: dxcproj	URL : https://adb-1484508246371556.16.azuredatabricks...
Location	: East US	Pricing Tier : standard
Subscription	: Azure-DXC262AB12Lab	Enable No Public IP : Yes
Subscription ID	: 3d94c073-3bce-4ba8-a171-6976f92e1322	

Below the table is a red Databricks logo and a "Launch Workspace" button. At the bottom, there are navigation links for "Documentation", "Getting Started", "Import Data from File", and "Import Data from Azure Storage". The taskbar at the bottom includes icons for File Explorer, Task View, Edge browser, Mail, OneDrive, Word, Excel, and Power BI, along with system status indicators like battery level, signal strength, and date/time (9:45 AM, 6/12/2022).

This screenshot shows the Microsoft Azure Databricks portal. The left sidebar has a 'Get started' section with links to 'Create a cluster', 'Ingest data', and 'Invite your team'. Below that is a 'Next steps' section with 'Explore Notebook gallery' and 'Read documentation'. The main area is titled 'Data Science & Engineering' and contains sections for 'Notebook' (with a 'Create a notebook' link), 'Data import' (with a 'Browse files' link), and a 'Guide: Quickstart tutorial' (with a 'Start tutorial' link). A 'Recents' section shows a table with columns 'Name' and 'Last viewed', which is currently empty. The bottom status bar shows the date and time as 6/12/2022 9:46 AM.

Create cluster

This screenshot shows the 'Create a cluster' configuration page. The left sidebar has a 'Completed' message and a 'Next step' section with a 'Ingest data' link. The main area shows a cluster named 'apmorgan' with a green status icon. It includes tabs for 'Configuration' (selected), 'Notebooks (0)', 'Libraries', 'Event log', 'Spark UI', 'Driver logs', 'Metrics', 'Apps', and 'Spark cluster UI - Master'. Under 'Configuration', settings include 'Cluster mode' (Single Node), 'Databricks Runtime Version' (10.4 LTS), 'Use Photon Acceleration' (unchecked), 'Autopilot options' (checkbox checked, 'Terminate after 100 minutes of inactivity'), 'Node type' (Standard_DS3_v2, 14 GB Memory, 4 Cores), and 'DBU / hour' (0.75). There is also an 'Advanced options' button. The bottom status bar shows the date and time as 6/12/2022 9:48 AM.

The screenshot shows the 'Create a cluster' wizard in the Microsoft Azure Databricks portal. The 'Completed' step is highlighted in green. The configuration details are as follows:

- Cluster mode:** Single Node
- Databricks Runtime Version:** 10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)
- Autopilot options:** Use Photon Acceleration (checkbox checked), Preview, Autopilot options (checkbox checked), Terminate after 100 minutes of inactivity (checkbox checked).
- Node type:** Standard_DS3_v2 (14 GB Memory, 4 Cores)
- DBU / hour:** 0.75 (Standard_DS3_v2)

A 'Next step' button is visible at the bottom.

Create notebook to run pipeline

The screenshot shows a Microsoft Azure Databricks notebook titled 'apmorg - Python'. The notebook contains the following code in the first cell:

```
print("project 2 ap morgan")
```

The output of the command is displayed in the second cell:

```
project 2 ap morgan
Command took 0.05 seconds -- by dxc262ab1222_1654530027309@manipalazure.onmicrosoft.com at 6/12/2022, 9:52:27 AM on apmorg
```

The notebook interface includes a toolbar with various icons for file operations, a search bar, and a status bar at the bottom.

Create datafactory to run pipeline

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 *

Instance details

Name *

Region *

Version *

Review + create < Previous Next : Git configuration >

Validation Passed

Basics

Subscription: Azure-DXC262AB12Lab
Resource group: dxcproj
Name: dxcprj
Region: East US
Version: V2 (Recommended)

TERMS

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

Review + create

Create < Previous Next Download a template for automation

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

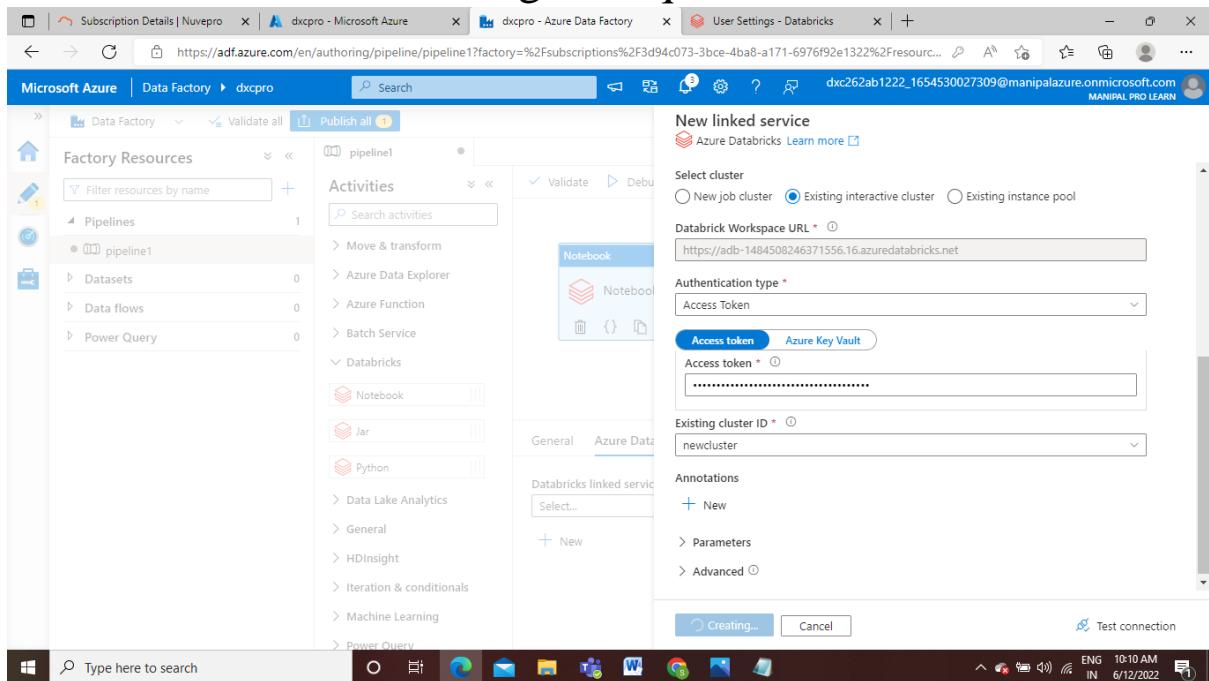
The screenshot shows the Microsoft Azure Data Factory Overview page for the deployment name "Microsoft.DataFactory-20220612095243". A modal window titled "... Deployment in progress..." is open, stating "Deployment to resource group 'dxcproj' is in progress." The deployment details section shows the deployment name, subscription, and resource group. The deployment status is "Deployment is in progress".



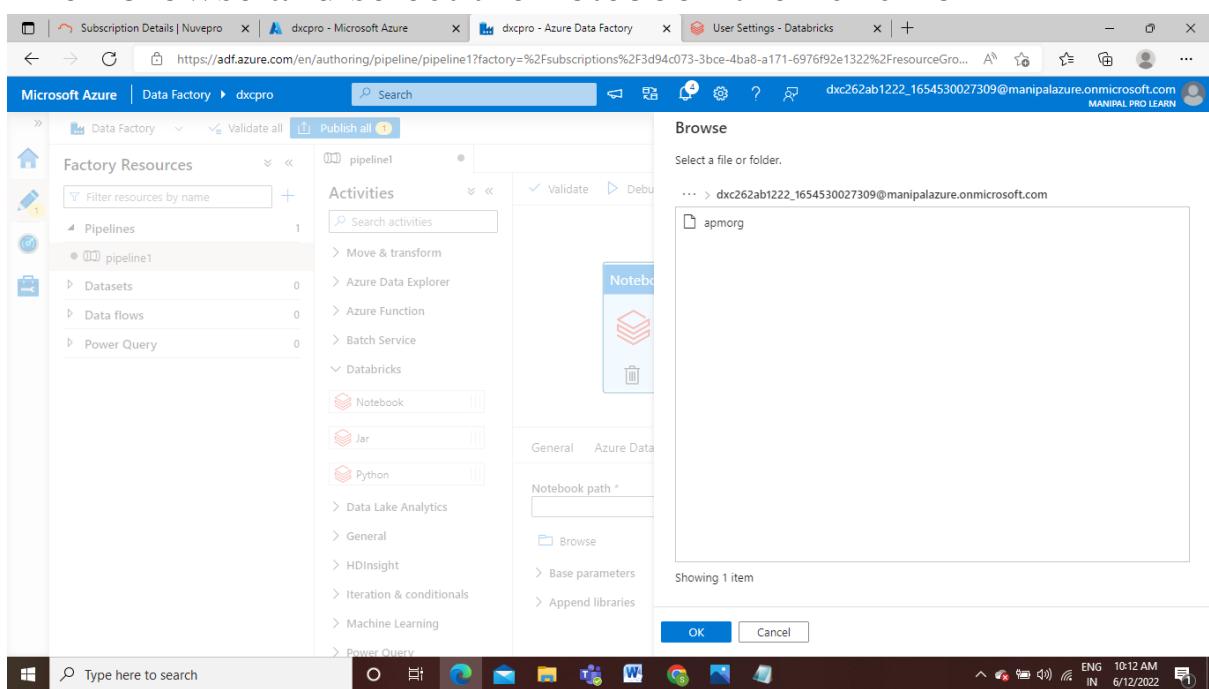
Generate token to link data bricks and data factory

The screenshot shows the Databricks portal interface. A modal window titled "Generate new token" is open, prompting for a comment ("apmorgan") and a lifetime in days (set to 10). The background shows the "Create a cluster" wizard with the "Completed" step highlighted.

Create linked service then give required credentials to link it



Then browse and select the notebook then click ok



Then run pipeline by giving trigger

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has 'Pipeline runs' selected. The main area displays a table titled 'Pipeline runs' with one item: 'pipeline1' (Run start: Jun 12, 2022, 10:16:48 am, Duration: 00:00:03, Triggered by: Manual trigger, Status: In progress). The top navigation bar shows the URL as https://adf.azure.com/en/monitoring/pipelineruns?factory=%2Fsubscriptions%2F3d94c073-3bce-4ba8-a171-6976f92e1322%2FresourceGroups%... and the user as dxc262ab1222_1654530027309@manipalazure.onmicrosoft.com.

Trigger and pipelines run successfully

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has 'Trigger runs' selected. The main area displays a table titled 'Trigger runs' with 1 - 26 items. All triggers are named 'trigger1' and are of type 'Schedule trigger'. They all have a status of 'Succeeded' with a green checkmark icon. The trigger times range from Jun 12, 2022, 10:41:0 to Jun 12, 2022, 10:31:5. The table includes columns for Trigger name, Trigger type, Trigger time, Status, Pipelines, Run, Message, and Properties. The top navigation bar shows the URL as https://adf.azure.com/en/monitoring/triggerruns?factory=%2Fsubscriptions%2F3d94c073-3bce-4ba8-a171-6976f92e1322%2FresourceGro... and the user as dxc262ab1222_1654530027309@manipalazure.onmicrosoft.com.

Result:

Conclusion:

References: