

---

# DXC REAL TIME PROJECTS

---

AZ-900, DP - 203



[10-06-2022]  
[DXC TECHNOLOGY]  
[Bangalore]

**Name:** Jyothi Chandana Voleti

**Reg No:** DXC262AB1218

**Project1 Name:** Smart Vehicles

**Date:** 10-06-2022

## Project 1 : Connected Vehicles

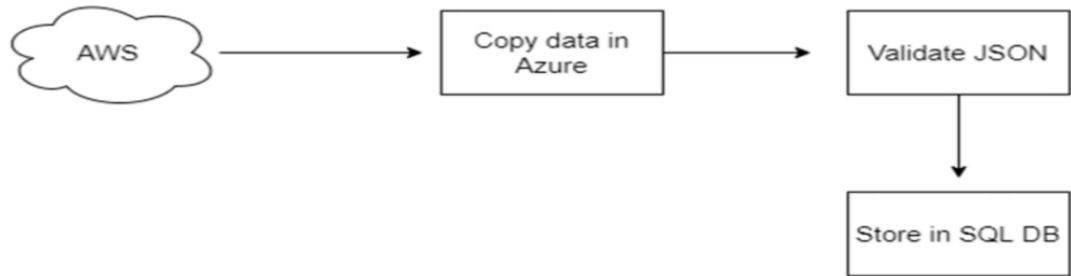
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



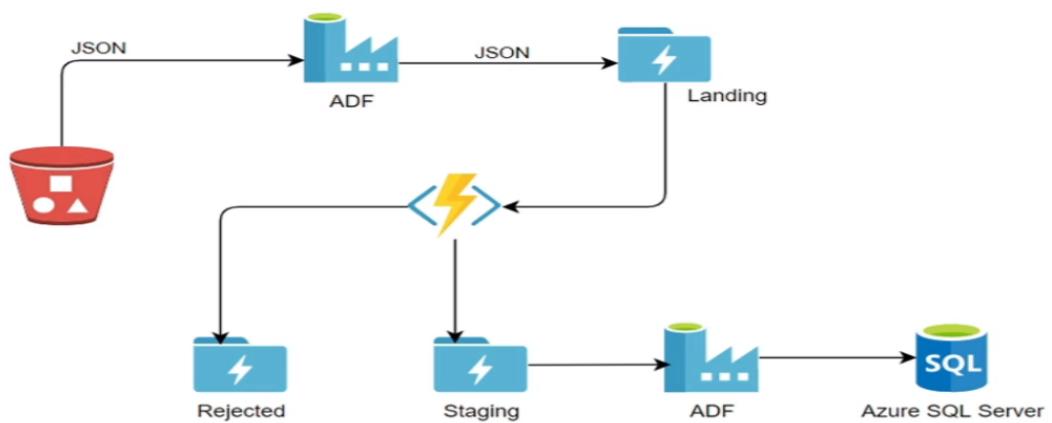
## Project 1 : Connected Vehicles

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

## Project 1 : Connected Vehicles



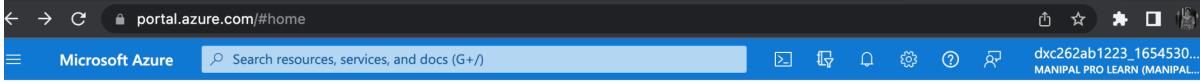
## Project 1 : Connected Vehicles



Architecture Diagram for Connected Vehicle Project

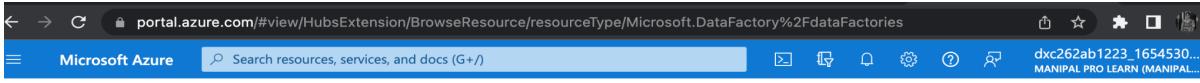
# Practical Lab: Create Azure Data Factory Account For Data pipelines

**Step 1:** Login into “Microsoft Azure Portal” and search for “Data Factories”.



The screenshot shows the Microsoft Azure Portal home page. In the 'Azure services' section, there is a row of icons for various services, including 'Create a resource', 'Data factories', 'Azure Databricks', 'Azure Synapse Analytics', 'Virtual machines', 'Azure Cosmos DB', 'SQL databases', 'Quickstart Center', and 'App Services'. Below this row is a large blue arrow pointing right, with the text 'More services' underneath it. In the 'Resources' section, there are tabs for 'Recent' and 'Favorite', followed by a table with columns for 'Name', 'Type', and 'Last Viewed'. A large gray cube icon is centered in the middle of the page, with the text 'No resources have been viewed recently' below it and a 'View all resources' button underneath.

**Step 2:** Now click on “Data Factories” and then on “+create” to create a Azure Data Factory.



The screenshot shows the 'Data factories' page in the Microsoft Azure Portal. At the top, there is a header with a 'Create' button, a 'Manage view' dropdown, a 'Refresh' button, an 'Export to CSV' button, an 'Open query' button, and an 'Assign tags' button. Below the header is a search bar and a filter bar with options for 'Subscription == all', 'Type == all', 'Resource group == all', 'Location == all', and an 'Add filter' button. The main area has sorting and grouping buttons for 'Name ↑↓', 'Type ↑↓', 'Subscription ↑↓', 'Resource group ↑↓', and 'Location ↑↓'. In the center, there is a factory icon and the text 'No data factories to display'. Below this text is a message 'Try changing or clearing your filters.' and a 'Create data factory' button.

## Step 3: Now will the required fields.

The screenshot shows the 'Create Data Factory' wizard in the Azure portal. The 'Basics' tab is selected. In the 'Project details' section, the subscription is set to 'Azure-DXC262AB12Lab' and the resource group is set to '(New) dxc1218'. In the 'Instance details' section, the name is 'dxcadf1218', the region is 'East US', and the version is 'V2 (Recommended)'.

## Step 4: After filling the required fields click on “Review + create” to validate the fields.

The screenshot shows the 'Create Data Factory' wizard in the Azure portal, with the 'Review + create' tab selected. A green banner at the top indicates 'Validation Passed'. The page displays the validation status and the summary of the selected basic settings:

Setting	Value
Subscription	Azure-DXC262AB12Lab
Resource group	dxc1218
Name	dxcadf1218
Region	East US
Version	V2 (Recommended)

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

**Step 5:** Now click on “create” at the left bottom to deploy the Data Factory.

Microsoft.DataFactory-20220610161601 | Overview

Deployment

Search (Cmd +/)

Overview

Inputs

Outputs

Template

We'd love your feedback! →

Deployment is in progress

Deployment name: Microsoft.DataFactory-20220610161601  
Subscription: Azure-DXC262AB12Lab  
Resource group: dxc1218

Start time: 6/10/2022, 4:36:22 PM  
Correlation ID: f396e62d-962c-46e5-8ef9-850c58b0ec63

Deployment details (Download)

Resource	Type	Status	Operation details
No results.			

**Step 6:** Deployment is completed.

Microsoft.DataFactory-20220610161601 | Overview

Deployment

Search (Cmd +/)

Overview

Inputs

Outputs

Template

We'd love your feedback! →

Your deployment is complete

Deployment name: Microsoft.DataFactory-20220610161601...  
Subscription: Azure-DXC262AB12Lab  
Resource group: dxc1218

Start time: 6/10/2022, 4:36:22 PM  
Correlation ID: f396e62d-962c-46e5-8ef9-850c58b0ec63

Deployment details (Download)

Next steps

Go to resource

Cost Management  
Get notified to sl prevent unexp Set up cost alert:

Microsoft Defense

## Step 7: Now click on “Go to Resource” and then open Azure Data Factory Studio.

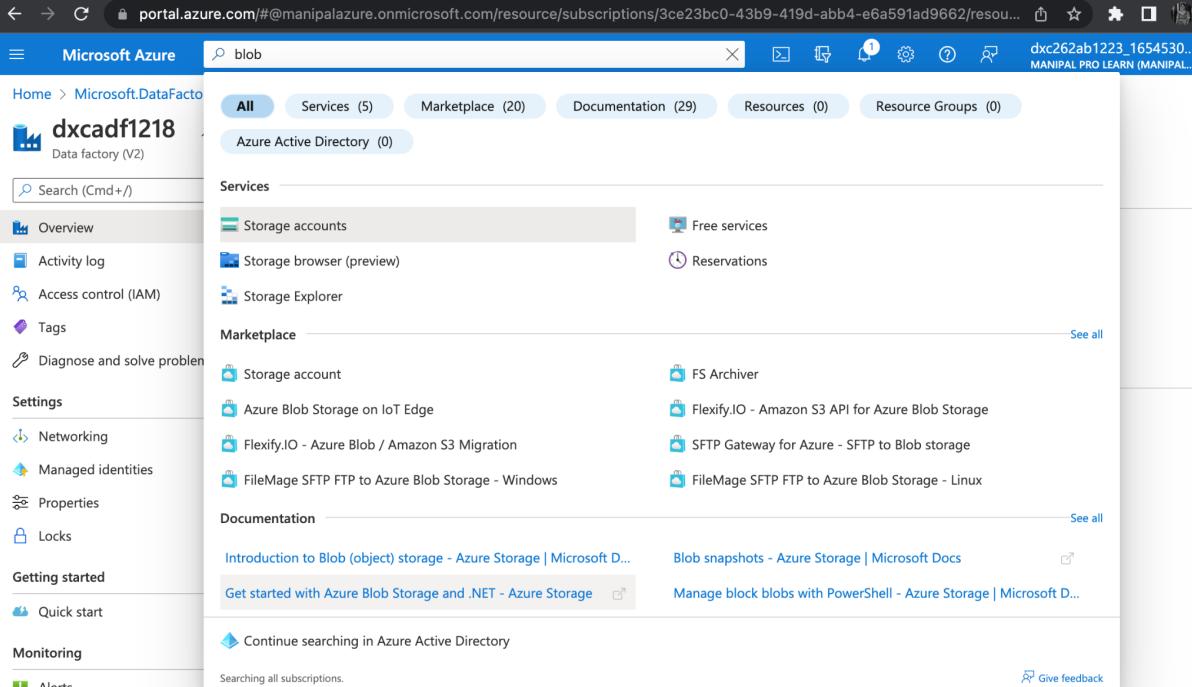
The screenshot shows the Microsoft Azure portal interface. At the top, the URL is <https://portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/3ce23bc0-43b9-419d-abb4-e6a591ad9662/resourceGroups/dxcadf1218/providers/Microsoft.DataFactory/factories/dxcadf1218>. The page title is "dxcadf1218" and the sub-title is "Data factory (V2)". The left sidebar shows navigation options like 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 displays the factory's status as "Succeeded", location as "East US", and subscription information. It includes a "Getting started" section with links to "Open Azure Data Factory Studio" and "Read documentation".

## Step 8: Creation of “Azure Data Factory” is done.

The screenshot shows the Azure Data Factory (ADF) studio home page. The URL is <https://adf.azure.com/en/home?factory=%2Fsubscriptions%2F3ce23bc0-43b9-419d-abb4-e6a591ad9662%2FresourceGroups%2Fd...>. The page title is "dxcadf1218" and the sub-title is "Data factory". The left sidebar has icons for Home, Set up code repository, Data factory, and New. The main content area features a large 3D diagram of a factory building with pipes and containers. Below the diagram are four service 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).

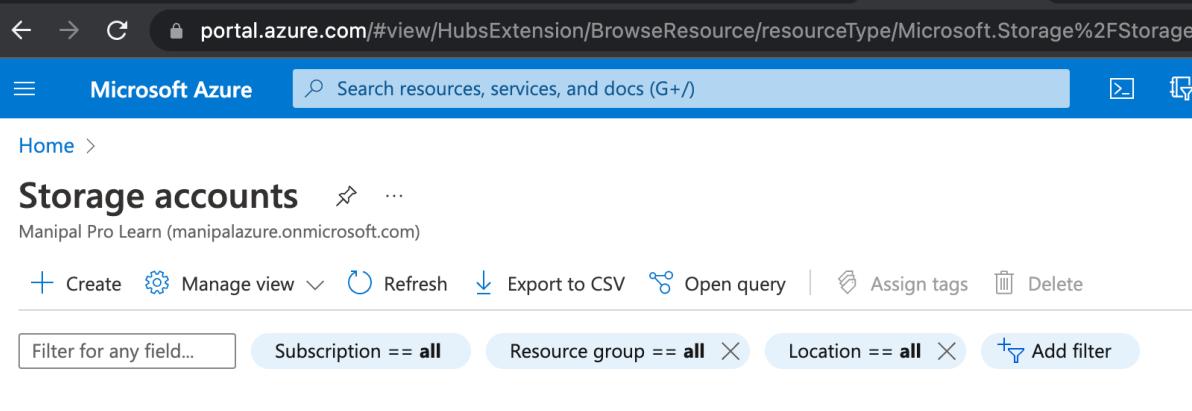
# Practical Lab: Create ADF Pipeline End to end pipeline with triggers enabled.

**Step 1:** To create a pipeline first you need to create a storage account. For that search for storage accounts in azure portal.



The screenshot shows the Microsoft Azure portal interface. The search bar at the top contains the text "blob". Below the search bar, there are several navigation tabs: All, Services (5), Marketplace (20), Documentation (29), Resources (0), and Resource Groups (0). The "All" tab is selected. Under the "Services" section, the "Storage accounts" option is highlighted and selected. Other visible service options include Storage browser (preview), Storage Explorer, and FS Archiver. The "Marketplace" section lists various storage-related services like Storage account, Flexify.IO - Amazon S3 API for Azure Blob Storage, and SFTP Gateway for Azure - SFTP to Blob storage. The "Documentation" section provides links to introductory articles for Blob storage. The left sidebar shows the current context is "dxcadf1218 Data factory (V2)".

**Step 2:** Open Storage accounts and click on “+create”.



The screenshot shows the "Storage accounts" page in the Azure portal. At the top, there is a search bar with the placeholder "Search resources, services, and docs (G+/)". Below the search bar, there are several navigation buttons: "Create", "Manage view", "Refresh", "Export to CSV", "Open query", "Assign tags", and "Delete". There are also filters for "Subscription", "Resource group", and "Location", along with a "Add filter" button. The main area displays a table with columns for Name, Type, Status, and Actions. The first row in the table is for a storage account named "dxcadfd1218".

## Step 3: Now fill all the required fields.

The screenshot shows the 'Create a storage account' wizard on the 'Basics' tab. It includes fields for Subscription (selected: Azure-DXC262AB12Lab), Resource group (selected: dxc1218), and Instance details (Storage account name: dxc1218, Region: (US) East US, Performance: Standard). Buttons at the bottom include 'Review + create' (highlighted in blue), '< Previous', and 'Next : Advanced >'.

## Step 4: After filling the required fields click on “Review + create” to validate the fields.

The screenshot shows the 'Create a storage account' wizard on the 'Review + create' tab. A green banner at the top indicates 'Validation passed'. The page displays the summary of filled fields under 'Basics' and 'Advanced' sections. Under 'Basics', fields include Subscription (Azure-DXC262AB12Lab), Resource Group (dxc1218), Location (eastus), Storage account name (dxc1218), Deployment model (Resource manager), Performance (Standard), and Replication (Read-access geo-redundant storage (RA-GRS)). Under 'Advanced', fields include Secure transfer (Enabled) and Allow storage account key access (Enabled). Buttons at the bottom include 'Create' (highlighted in blue), '< Previous', and 'Download a template for automation'.

## Step 5: Now click on “create” to deploy the storage account.

The screenshot shows the Azure portal's deployment details blade for a deployment named 'dxc1218\_1654863952463'. A green checkmark indicates 'Deployment succeeded' with the message: 'Deployment 'dxc1218\_1654863952463' to resource group 'dxc1218' was successful.' Below this, there are links to 'Go to resource' and 'Pin to dashboard'. On the left, a navigation menu includes 'Overview', 'Inputs', 'Outputs', and 'Template'. In the center, a summary box says 'Your deployment is complete' with deployment details: Deployment name: dxc1218\_1654863952..., Start time: 6/10/2022, 5:56:06 PM, Subscription: Azure-DXC26AB12Lab, Correlation ID: 0e84f8fd-692e-4314-854b-4a84b1873d1b. There are sections for 'Deployment details' (with a download link) and 'Next steps' (with a 'Go to resource' button). To the right, there are promotional cards for 'Cost Management', 'Microsoft Defender for Cloud', 'Free Microsoft tutorials', and 'Work with an expert'.

## Step 6: Now click on “Go to Resource” and create 2 new containers one for source and the other for destination.

The screenshot shows the 'Containers' blade for a storage account named 'dxc1218'. The left sidebar has sections for 'Overview', 'Activity log', 'Tags', 'Diagnose and solve problems', 'Access Control (IAM)', 'Data migration', 'Events', and 'Storage browser (preview)'. Under 'Data storage', 'Containers' is selected. The main area shows a table of existing containers: '\$logs' (Last modified: 6/10/2022, 5:56:39 PM, Public access: Private). A 'New container' button is located at the bottom right of this table. On the right, a 'New container' dialog box is open, prompting for a 'Name' (marked with a red asterisk) and a 'Public access level' (set to 'Private (no anonymous access)'). At the bottom right of the dialog are 'Create' and 'Discard' buttons.

## Step 7: Fill the required fields and click on create.

The screenshot shows the Microsoft Azure portal interface. On the left, there's a sidebar with various navigation options like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, and Storage browser (preview). The main area is titled 'dxc1218 | Containers'. A search bar at the top has 'Search resources, services, and docs (G+/-)' and a 'Container' button. Below the search bar, there's a table with columns 'Name' and 'Last modified'. One row shows '\$logs' last modified on 6/10/2022, 5:56:39. A modal window titled 'New container' is open on the right. It has a 'Name' field containing 'dxcblob1218', a 'Public access level' dropdown set to 'Blob (anonymous read access for blobs only)', and an 'Advanced' section. At the bottom of the modal are 'Create' and 'Discard' buttons.

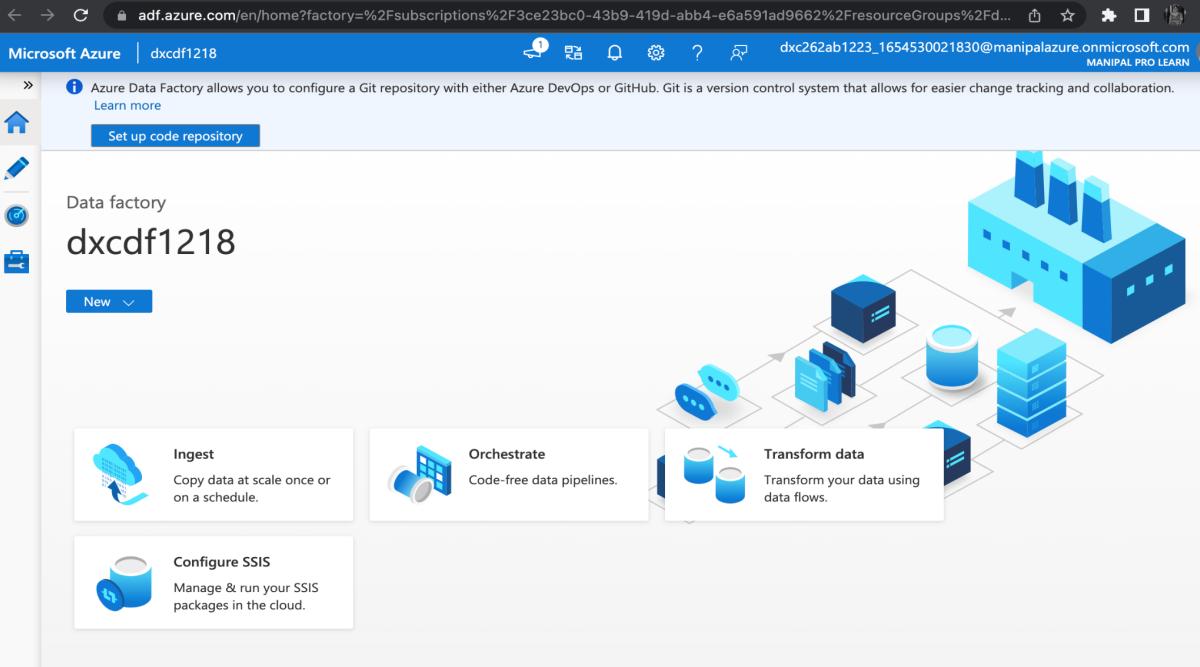
Source Storage container has been created.

The screenshot shows the Microsoft Azure portal interface. The left sidebar includes 'Overview', 'Diagnose and solve problems', 'Access Control (IAM)', 'Settings' (with Shared access tokens, Access policy, Properties, and Metadata), and 'Storage browser (preview)'. The main area is titled 'dxcblob1218'. It shows an 'Upload' button, 'Change access level', 'Refresh', 'Delete', 'Change tier', 'Acquire lease', and 'Break lease' buttons. A search bar says 'Search blobs by prefix (case-sensitive)' and a 'Show deleted blobs' toggle is off. Below is a table with columns 'Name', 'Modified', 'Access tier', 'Archive status', and 'Blob type'. The table shows 'No results'.

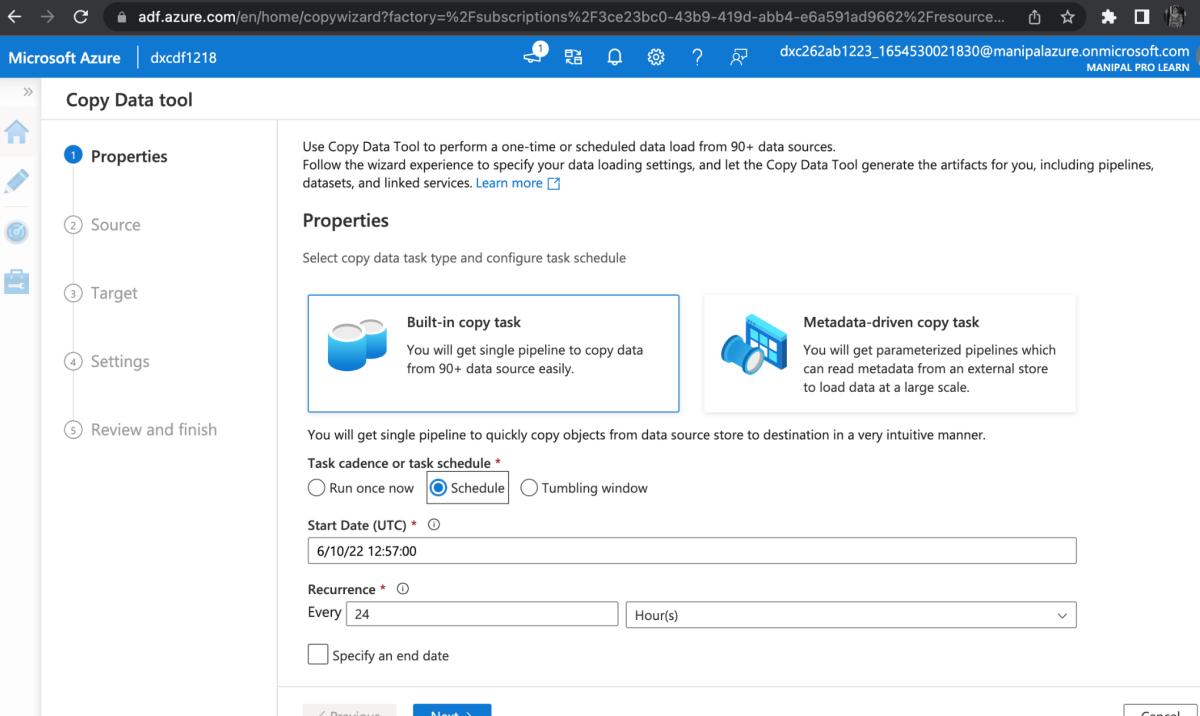
## Step 8: Now create a container for the destination container.

The screenshot shows the Microsoft Azure portal interface. The left sidebar is identical to the previous screenshot. The main area is titled 'dxc1218 | Containers'. It shows a table with columns 'Name', 'Last modified', 'Public access level', and 'Lease state'. The table lists '\$logs' (last modified 6/10/2022, 6:08:57 PM, Private, Available), 'dxcblob1218' (last modified 6/10/2022, 6:09:33 PM, Blob, Available), and 'dxcdestination12' (last modified 6/10/2022, 6:20:46 PM, Blob, Available). A 'Show deleted containers' toggle is off.

## Step 9: Now to create a pipeline click “ingest” in the data factory.



## Step 10: Now schedule the pipeline.



After filling the required fields in the copy data tool review the fields and create the pipeline and trigger.

<adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F3ce23bc0-43b9-419d-abb4-e6a591ad9662%2Fresource...>

Microsoft Azure | dxcdf1218

Copy Data tool

Properties

Source

Target

Settings

Review and finish

Review

Deployment

**Summary**

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

**Dataset name** DestinationDataset\_2r8

**Column delimiter**,

**Escape character**\

**Quote char**"

**Copy settings**

**Timeout** 7.00:00:00

**Retry** 0

**Retry interval (sec)** 30

**Secure output** false

**Secure input** false

**Trigger**

**Name** Trigger\_2r8

**Trigger type** ScheduleTrigger

[Edit](#)

[Edit](#)

< Previous **Next >** Cancel

<adf.azure.com/en/home/copywizard?factory=%2Fsubscriptions%2F3ce23bc0-43b9-419d-abb4-e6a591ad9662%2Fresource...>

Microsoft Azure | dxcdf1218

Copy Data tool

Properties

**Source**

Dataset

Configuration

Target

Settings

Review and finish

**Source data store**

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

**Source type** Azure Blob Storage

**Connection \*** AzureBlobStorage1 [Edit](#) [New connection](#)

**File or folder \***

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

dxcblob1218/1000\_Companies.csv [Browse](#)

**Options**

**File loading behavior** Load all files

Binary copy

Recursively

Enable partition discovery

**Max concurrent connections**

< Previous **Next >** Cancel

## Step 11: Pipeline has been created, click on finish.

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 (which is highlighted), Review, and Deployment. The main area displays a flow from 'Azure Blob Storage' to 'Azure Blob Storage'. Below this, a message says 'Deployment complete'. A table titled 'Deployment step' shows the status of various tasks: 'Validating copy runtime environment' (Succeeded), 'Creating datasets' (Succeeded), 'Creating pipelines' (Succeeded), 'Creating triggers' (Succeeded), and 'Starting triggers' (Succeeded). At the bottom, a message states: 'Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.' Three buttons are at the bottom: 'Finish' (highlighted in blue), 'Edit pipeline', and 'Monitor'.

## Step 12: Enabling pipeline trigger.

The screenshot shows the Microsoft Azure Pipeline runs page. The left sidebar includes options like Dashboards, Runs (which is selected and highlighted in blue), Pipeline runs, 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, and Triggered by. One row is visible: 'CopyPipeline\_2r8' with a run starting on Jun 10, 2022, at 6:54:09 pm, ending at 6:54:22 pm, duration 0:00:12, and triggered by 'Manual trigger'. The table header includes filters for Pipeline name, Status, and Runs, along with 'List' and 'Gantt' buttons. The status bar at the bottom right indicates 'Last refreshed 0 minutes ago'.

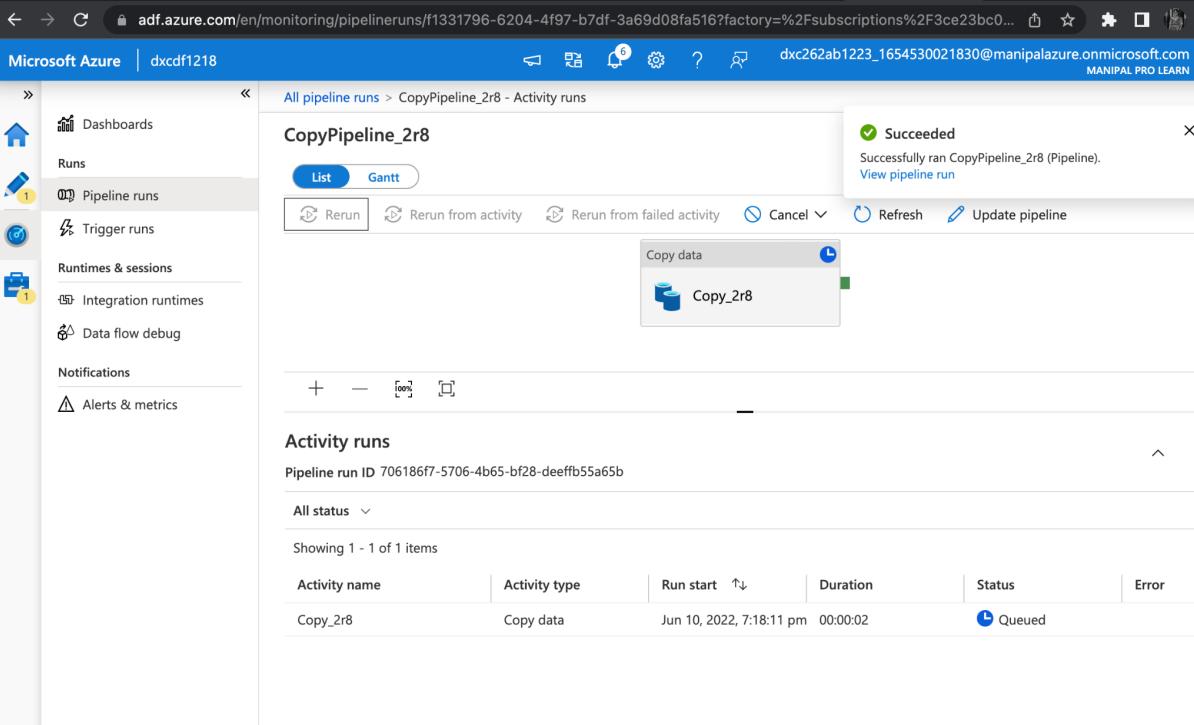
Activity name	Activity type	Run start	Duration	Status	Error
Copy_2r8	Copy data	Jun 10, 2022, 6:54:11 pm	00:00:10	Succeeded	

## Practical Lab: Create Azure blob trigger logic

**Step 1:** Create a Blob Storage Container as shown in the above steps.

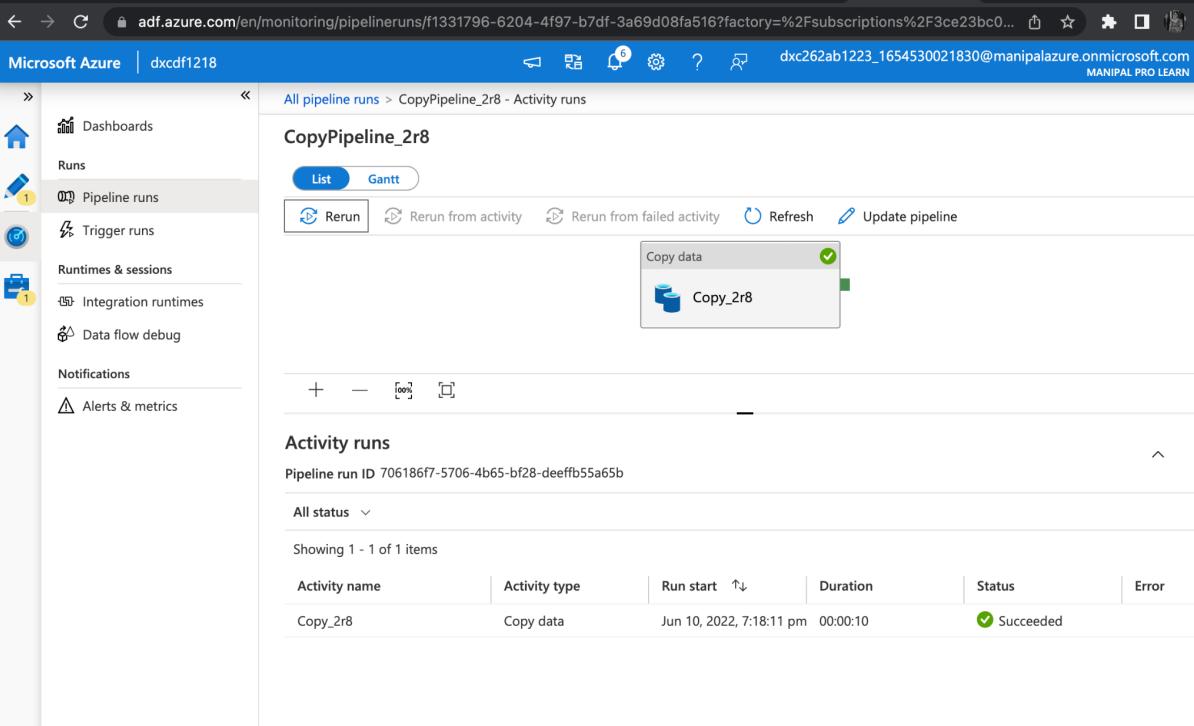
Name	Last modified	Public access level	Lease state
\$logs	6/10/2022, 6:08:57 PM	Private	Available
dxcblob1218	6/10/2022, 6:09:33 PM	Blob	Available
dxcdestination12	6/10/2022, 6:20:46 PM	Blob	Available

## Step 2: Enabling blob trigger.



The screenshot shows the Microsoft Azure portal interface for monitoring pipeline runs. The left sidebar navigation includes: Home, Dashboards, Runs, Pipeline runs (selected), Trigger runs, Runtimes & sessions, Integration runtimes, Data flow debug, Notifications, and Alerts & metrics. The main content area displays the 'All pipeline runs > CopyPipeline\_2r8 - Activity runs' section. A summary box indicates a 'Succeeded' status with the message 'Successfully ran CopyPipeline\_2r8 (Pipeline). View pipeline run'. Below this, a 'Copy data' activity named 'Copy\_2r8' is shown. The 'Activity runs' table lists one item:

Activity name	Activity type	Run start	Duration	Status	Error
Copy_2r8	Copy data	Jun 10, 2022, 7:18:11 pm	00:00:02	Queued	



This screenshot shows the same Microsoft Azure portal interface and pipeline run details as the first one, but with a key difference: the status of the 'Copy\_2r8' activity has changed from 'Queued' to 'Succeeded', indicated by a green checkmark icon in the status column of the activity table. The rest of the interface and data remain identical to the first screenshot.

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 displays a table of pipeline runs for 'CopyPipeline\_2r8' over the last 24 hours. The table columns are Pipeline name, Run start, Run end, Duration, and Triggered by. All runs are marked as Succeeded.

Pipeline name	Run start	Run end	Duration	Triggered by
CopyPipeline_2r8	Jun 10, 2022, 6:54:09 pm	Jun 10, 2022, 6:54:22 pm	00:00:12	Manual trigger
CopyPipeline_2r8	Jun 10, 2022, 7:16:59 pm	Jun 10, 2022, 7:17:14 pm	00:00:14	Manual trigger
CopyPipeline_2r8	Jun 10, 2022, 7:17:56 pm	Jun 10, 2022, 7:18:07 pm	00:00:11	Manual trigger
CopyPipeline_2r8	Jun 10, 2022, 7:18:10 pm	Jun 10, 2022, 7:18:22 pm	00:00:12	Manual trigger
CopyPipeline_2r8	Jun 10, 2022, 7:19:21 pm	Jun 10, 2022, 7:19:32 pm	00:00:11	Manual trigger
CopyPipeline_2r8	Jun 10, 2022, 7:19:41 pm	Jun 10, 2022, 7:19:52 pm	00:00:11	Manual trigger
CopyPipeline_2r8	Jun 10, 2022, 7:19:51 pm	Jun 10, 2022, 7:20:00 pm	00:00:09	Manual trigger
CopyPipeline_2r8	Jun 10, 2022, 7:20:03 pm	Jun 10, 2022, 7:20:15 pm	00:00:12	Manual trigger

## Practical Lab: Create Azure SQL Server and Database

### CREATION OF SQL SERVER:

Step 1: Login to Azure portal and search for sql server.

The screenshot shows the Microsoft Azure search results for 'sql'. The search bar at the top has 'sql' typed in. The main content area shows the 'Services' section with 'SQL servers' selected. Other options include 'SQL databases', 'SQL elastic pools', 'SQL managed instances', 'SQL Server registries', 'SQL virtual machines', 'SQL Server stretch databases', and 'Dedicated SQL pools (formerly SQL DW)'. Below the services, there are sections for 'Marketplace', 'Documentation', and 'See all' links. On the left sidebar, under 'Azure services', there are sections for 'Create a resource', 'More services', 'Resources', 'Recent', 'Favorite', and 'Documentation'. Under 'Recent', 'dxc1218' is listed. Under 'Documentation', links to 'Getting started with SQL queries in Azure Cosmos DB' and 'SSMS: Connect and query data - Azure SQL Database & SQL Mana...' are shown.

## Step 2: Open sql server and click on “+create”.

The screenshot shows the Azure portal's "SQL servers" blade. At the top, there are navigation links for "Home", "Microsoft Azure", and a search bar. Below the search bar are buttons for "+ Create", "Manage view", "Refresh", "Export to CSV", "Open query", and "Assign tags". There are also filters for "Subscription == all", "Resource group == all", "Location == all", and a "No grouping" dropdown. The main area displays a large "SQL" logo with a gear icon. Below it, the text "No SQL servers to display" is centered, followed by the instruction "Try changing or clearing your filters." A blue "Create SQL server" button is prominently displayed, along with a "Learn more" link and a "Give feedback" button at the bottom right.

## Step 3: Fill all the required fields.

The screenshot shows the "Create SQL Database Server" wizard. The top navigation bar includes "Home", "SQL servers", and the current step "Create SQL Database Server". The page title is "Create SQL Database Server". The "Basics" tab is selected, with other tabs for "Networking", "Additional settings", "Tags", and "Review + create". A note states: "SQL database server is a logical container for managing databases and elastic pools. Complete the Basic tab, then go to Review + Create to provision with smart defaults, or visit each tab to customize." Below this, the "Project details" section asks to select a subscription and resource group. The "Subscription" dropdown is set to "Azure-DXC26AB12Lab" and the "Resource group" dropdown is set to "dxc1218", with a "Create new" link. The "Server details" section requires entering a server name and location. The "Server name" field is filled with "dxcserver1218" and the "Location" field is set to "(US) East US". At the bottom, there are "Review + create" and "Next : Networking >" buttons.

## Step 4: Now click on “create” at the left bottom to start the deployment.

The screenshot shows the Azure portal interface for creating a SQL Database Server. The URL is [portal.azure.com/#create/Microsoft.SQLServer](https://portal.azure.com/#create/Microsoft.SQLServer). The page title is "Create SQL Database Server". The top navigation bar includes "Microsoft Azure", a search bar, and various icons. Below the title, it says "Home > SQL servers > Create SQL Database Server". The main content area has tabs: Basics, Networking, Additional settings, Tags, and Review + create (which is underlined). Under "Product details", it shows "SQL Database Server by Microsoft" and "Estimated cost per month: No additional charges". There are links for "Terms of use" and "Privacy policy". The "Terms" section contains legal text about agreeing to terms and conditions. The "Basics" section lists configuration details: Subscription (Azure-DXC262AB12Lab), Resource group (dxc1218), Server name (dxcserver1218), Authentication method (SQL authentication), and Server admin login (dxc1218). At the bottom, there are buttons for "Create", "< Previous", and "Download a template for automation".

## Step 5: Deployment completed.

The screenshot shows the Azure portal interface for a completed deployment named "Microsoft.SQLServer.createServer\_ade2174027e24e1e8c376901232c98b". The URL is [portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/-/overview/id/%2Fsubscriptions%2F3ce23bc0-43b9-41...](https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/-/overview/id/%2Fsubscriptions%2F3ce23bc0-43b9-41...). The page title is "Microsoft.SQLServer.createServer\_ade2174027e24e1e8c376901232c98b | Overview". The top navigation bar includes "Microsoft Azure", a search bar, and various icons. Below the title, it says "Home > Microsoft.SQLServer.createServer\_ade2174027e24e1e8c376901232c98b | Overview". The main content area shows a message "Your deployment is complete" with a checkmark icon. It provides deployment details: Deployment name: Microsoft.SQLServer.createServer\_ade2174027e..., Start time: 6/10/2022, 7:29:38 PM, Subscription: Azure-DXC262AB12Lab, Correlation ID: 0b96678f-ea9b-433b-ba05-8f3e8c996498, and Resource group: dxc1218. There are sections for "Deployment details" (with a "Download" link) and "Next steps". A "Go to resource" button is at the bottom. On the right side, there are sidebar links for "Cost Management", "Get notified", "Prevent", "Set up compliance", "Secure", "Go to Microsoft Secure", "Free Microsoft 365", "Start learning", and "Work with Azure".

## Step 6: Now create a sql database by clicking “+create database” in “Go to Resource”.

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and various navigation icons. Below the header, the URL is `portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/3ce23bc0-43b9-419d-abb4-e6a591ad9662/r...`. The main content area is titled "dxcserver1218" and "SQL server". On the left, there's a sidebar with links like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Quick start, Settings (with sub-links for Azure Active Directory, SQL databases, SQL elastic pools, DTU quota, Properties, Locks), Data management, and Backups. The main panel has tabs for Overview, Essentials, and Features (which is selected). Under Features, there are sections for Active Directory admin (NOT CONFIGURED) and Microsoft Defender for SQL (NOT CONFIGURED). At the top of the main panel, there are buttons for Create database, New elastic pool, New dedicated SQL pool (formerly SQL DW), Import database, and Reset password.

## Step 7: Fill all the required fields that are needed to create a sql database.

The screenshot shows the "Create SQL Database" wizard in the Microsoft Azure portal. The top navigation bar is identical to the previous screenshot. The main content area is titled "Create SQL Database" and shows a step-by-step process. Step 1: Project details. It asks to select a subscription and resource group. Step 2: Database details. It asks for the database name (dxcdb12) and to select a server (Select a server dropdown). Step 3: Review + create (button at the bottom left) and Next : Networking > (button at the bottom right). A note at the top says: "Did you know that new users in Azure can create a free Azure SQL Database and use it for 12 months using Azure free account? Learn more" with a link.

## Step 8: Click on Review+create to validate the details.

The screenshot shows the Azure portal interface for creating a SQL database. The URL is [portal.azure.com/#create/Microsoft.SQLDatabase](https://portal.azure.com/#create/Microsoft.SQLDatabase). The page title is "Create SQL Database". Below it, there are tabs: Basics, Networking, Security, Additional settings, Tags, and **Review + create**, which is underlined to indicate it's the active step. On the left, there's a sidebar with "Product details" showing "SQL database by Microsoft" and links to "Terms of use" and "Privacy policy". The main content area has a section titled "Estimated cost per month" with a "View pricing details" link. Below this is a "Terms" section with a detailed legal agreement. Under "Basics", the configuration is listed as follows:

Subscription	Azure-DXC262AB12Lab
Resource group	dxc1218
Region	East US
Database name	dxcdb12

At the bottom, there are three buttons: "Create" (in blue), "< Previous", and "Download a template for automation".

## Step 9: Now click on create to deploy the database.

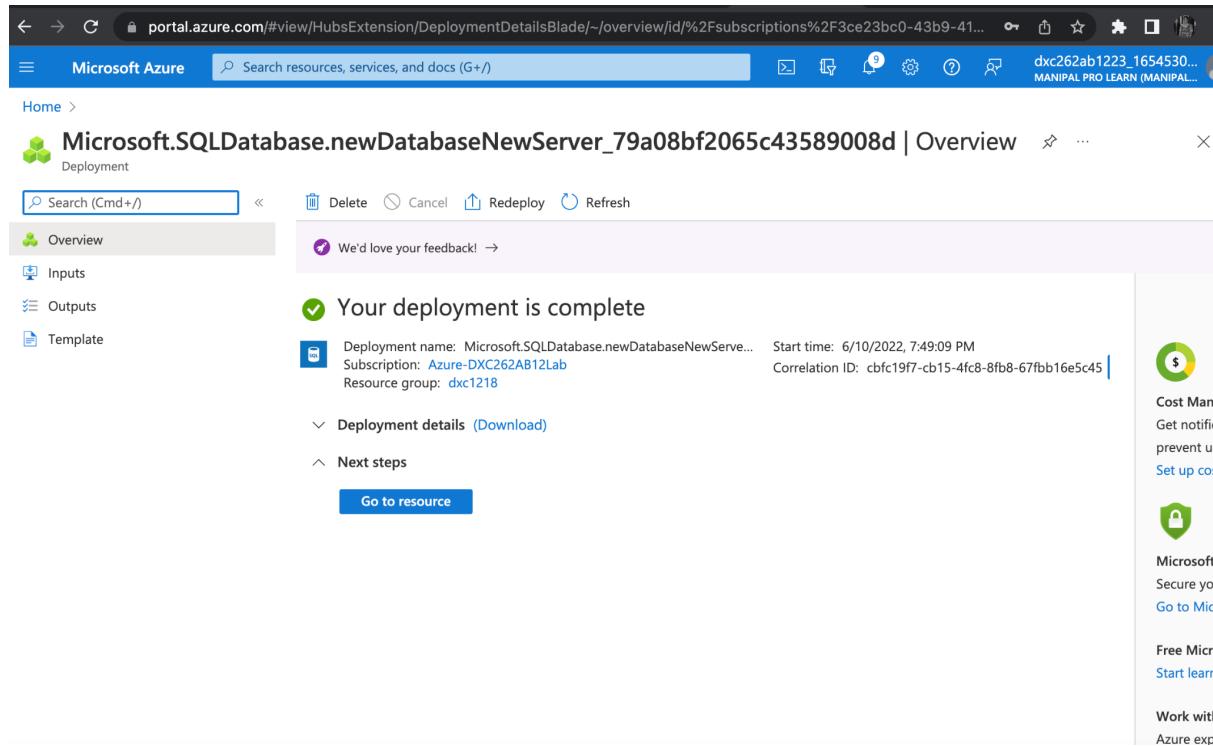
The screenshot shows the Azure portal interface for viewing deployment details. The URL is [portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F3ce23bc0-43b9-41...](https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F3ce23bc0-43b9-41...). The page title is "Microsoft.SQLDatabase.newDatabaseNewServer\_79a08bf2065c43589008d | Overview". On the left, there's a sidebar with "Overview", "Inputs", "Outputs", and "Template". The main content area shows a message "Deployment is in progress". Below it, deployment details are listed:

Deployment name:	Microsoft.SQLDatabase.newDatabaseNewServer_79a08bf2065c43589008d
Subscription:	Azure-DXC262AB12Lab
Resource group:	dxc1218

Below the deployment details, there's a section titled "Deployment details (Download)" with a table:

Resource	Type	Status	Operation details
No results.			

## Step 10: Deployment completed.



The screenshot shows the Microsoft Azure portal interface. At the top, there's a navigation bar with icons for back, forward, search, and notifications. Below it, the main header reads "Microsoft Azure" and "Search resources, services, and docs (G/)" followed by a search bar. The URL in the address bar is "portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/-/overview/id/%2Fsubscriptions%2F3ce23bc0-43b9-41...". On the right side of the header, there are several small icons and the text "MANIPAL PRO LEARN (MANIPAL...)".

The main content area has a title "Microsoft.SQLDatabase.newDatabaseNewServer\_79a08bf2065c43589008d | Overview". Below the title, there are several buttons: "Search (Cmd+/)", "Delete", "Cancel", "Redeploy", and "Refresh".

A sidebar on the left lists navigation options: "Overview" (which is selected and highlighted in blue), "Inputs", "Outputs", and "Template".

The main content area displays a message: "We'd love your feedback! →". Below this, a green checkmark icon indicates "Your deployment is complete".

Deployment details are shown in a table:

Deployment name:	Microsoft.SQLDatabase.newDatabaseNewServer_79a08bf2065c43589008d
Subscription:	Azure-DXC262AB1Lab
Resource group:	dxc1218
Start time:	6/10/2022, 7:49:09 PM
Correlation ID:	cbfc19f7-cb15-4fc8-8fb8-67fbb16e5c45

Below the deployment details, there are two collapsed sections: "Deployment details (Download)" and "Next steps". A blue button labeled "Go to resource" is located at the bottom of this section.

On the right side of the page, there's a vertical sidebar with several cards:

- Cost Management**: Get notifications, prevent usage, Set up cost alerts.
- Microsoft Secure**: Go to Microsoft Secure.
- Free Microsoft**: Start learning.
- Work with Azure**: Azure exp.

## SQL SERVER AND DATABASE CREATION IS COMPLETED.

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

To solve this case we need to deploy a storage account, sql database and data factory as shown in the above steps.

Then we have to create a source container in the storage account to store the data.

To move the data to the SQL DB we need to use a pipeline.

So create a pipeline in the data factory using a copy data tool.

The screenshot shows the Microsoft Azure Copy Data tool wizard on the 'Properties' step. The left sidebar lists steps 1 through 5: Properties, Source, Target, Settings, and Review and finish. The main area is titled 'Properties' and describes the tool for performing one-time or scheduled data loads from 90+ data sources. It includes sections for 'Task type' (Built-in copy task and Metadata-driven copy task), 'Task cadence or task schedule' (Run once now selected), and navigation buttons ('Next >', 'Cancel').

Fill all the required fields.

The screenshot shows the Microsoft Azure Copy Data tool wizard on the 'Source' step. The left sidebar shows steps 1 (Properties) completed with a checkmark, and steps 2 through 5 listed. The main area is titled 'Source data store' and specifies the source data store as 'Azure Blob Storage' connected via 'AzureBlobStorage1'. It includes options for 'File or folder' (source1218/1000\_Companies.csv), 'Options' (Binary copy, Recursively selected, Enable partition discovery), 'Max concurrent connections' (empty input field), and 'Filter by last modified' (empty input fields for Start time (UTC) and End time (UTC)). Navigation buttons ('Next >', 'Cancel') are at the bottom.

## Select the target destination as SQL DB.

The screenshot shows the Microsoft Azure Copy Data tool wizard. The left sidebar lists steps: Properties, Source, Target (selected), Dataset, Configuration, Settings, and Review and finish. The main area is titled 'Destination data store' and shows 'Target type' set to 'Azure SQL'. A 'Connection' dropdown is labeled 'Select...'. To the right, a 'New linked service' section is displayed, titled 'Azure SQL Database'. It includes fields for 'Name' (set to 'AzureSqlDatabase1'), 'Description', 'Connect via integration runtime' (set to 'AutoResolveIntegrationRuntime'), 'Account selection method' (radio button selected for 'From Azure subscription'), 'Azure subscription' (dropdown set to 'Select all'), 'Server name' (set to 'dxcsr12'), 'Database name' (set to 'From Azure subscription cdb1218'), and 'Authentication type' (dropdown). Buttons at the bottom include 'Create', 'Cancel', and 'Test connection'.

The screenshot shows the Microsoft Azure Copy Data tool wizard. The left sidebar lists steps: Properties, Source, Target (selected), Dataset, Configuration, Settings, and Review and finish. The main area is titled 'Destination data store' and shows 'Target type' set to 'Azure SQL Database'. A 'Connection' dropdown is set to 'AzureSqlDatabase3'. Below, a table maps 'Source' (Azure Blob Storage file) to 'Target' (Azure Blob Storage file). The 'Source' row has an arrow pointing to the 'Target' row, which contains a dropdown set to '(auto-create)' and a link 'Use existing table'. A checkbox 'Skip column mapping for all tables' is at the bottom. Navigation buttons include '< Previous', 'Next >', and 'Cancel'.

Check the summary and click on next to create a pipeline to move the data from blob to sql db.

**Copy Data tool**

**Properties**

**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\_r9y

Task description:

**Source**

Connection name: AzureBlobStorage1

Dataset name: SourceDataset\_r9y

Column delimiter: ,

Escape character: \

Quote char: "

... -

< Previous Next > Cancel

Pipeline has been created.

**Copy Data tool**

**Properties**

**Source**

**Target**

**Settings**

**Review and finish**

**Review**

**Deployment**

Azure Blob Storage → Azure SQL Database

**Deployment complete**

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

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

Finish Edit pipeline Monitor

Now run the pipeline to move the data.

The screenshot shows the Microsoft Azure Pipeline runs page. The left sidebar has a 'Runs' section selected. The main area displays a table of pipeline runs. One run is listed: 'CopyPipeline\_r9y' (Run start: Jun 12, 2022, 12:24:04 pm; Run end: Jun 12, 2022, 12:24:17 pm; Duration: 00:00:13; Triggered by: Manual trigger). The status bar at the bottom right indicates 'Last refreshed 0 minutes ago'.

Now navigate towards the SQL database in order to check the data.

The screenshot shows the Microsoft Azure portal's Query editor (preview) for the 'dxcdb2' database. The left sidebar has 'Query editor (preview)' selected. The main area shows a query editor with two tabs: 'Query 1' and 'Query 2'. 'Query 2' is active, containing the SQL command: 'SELECT TOP (1000) \* FROM [dbo].[Azure Blob Storage file]'. Below the query, the results pane shows a table with four columns: Column1, Column2, Column3, and Column4. The data is as follows:

Column1	Column2	Column3	Column4
R&D Spend	Administration	Marketing Spend	State
165349.2	136897.8	471784.1	New York
162597.7	151377.59	443898.53	California

The data is successfully moved from storage account to SQL DB and can be accessed through querying.

**Result:** Using Azure Data Factory, we were able to create a pipeline that will validate and copy the blob data into the SQL database in this project.

**Conclusion:** The Blob data has been validated and put in the SQL database.

**Name:** Jyothi Chandana Voleti

**Reg No:** DXC262AB1218

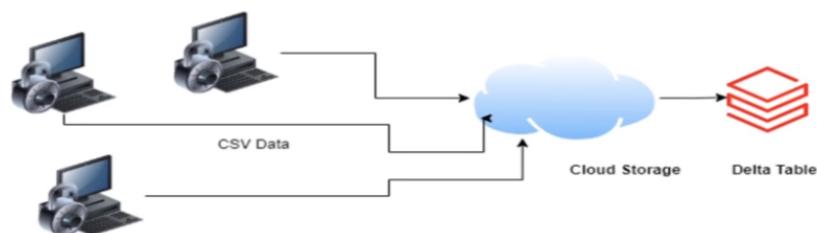
**Project2 Name:** AP Morgan Data Platform

**Date:** 10-06-2022

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

## Project 2 : AP Morgan

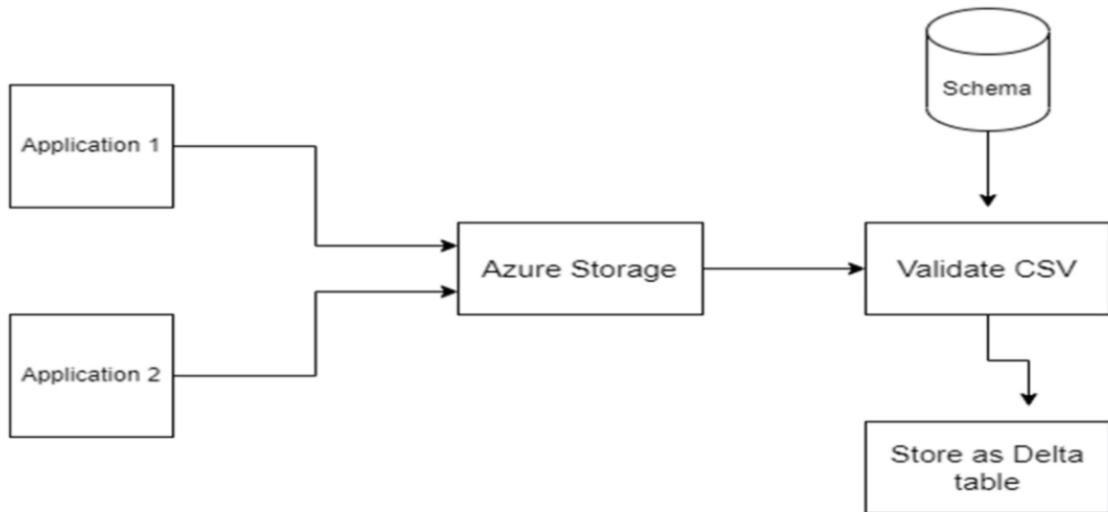
- Multiple Internal applications sends the data(huge size) in CSV format on daily basis in the cloud storage location. There are couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed data to be persisted as Delta table in Databricks for downstream system.



## Project 2 : AP Morgan- High Level Detail

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

## Project 2 : AP Morgan



## Practical Lab: Create a Databricks

Step 1: Login to Azure portal and search for Databricks.

The screenshot shows the Microsoft Azure portal search results for 'databr'. The search bar at the top has 'databr' typed into it. The left sidebar shows 'Azure services' with a 'Create a resource' button and a 'More services' button. The main search results area shows 'Services' with 'Azure Databricks' highlighted in grey. Below 'Services' are sections for 'Marketplace' and 'Documentation'. At the bottom of the page, there is a note about searching all subscriptions, a 'Storage account' section, and a timestamp '47 minutes ago'.

## Step 2: Click on Azure Databricks and click on “+create”.

The screenshot shows the Azure portal interface for managing Azure Databricks services. The top navigation bar includes the Microsoft Azure logo, a search bar, and various navigation icons. Below the header, there's a breadcrumb trail: Home > Azure Databricks. The main content area is titled 'Azure Databricks' and displays a message: 'No azure databricks services to display'. Below this message is a call-to-action button labeled 'Create azure databricks service'. There are also links for 'Learn more' and 'Give feedback'.

## Step 3: Fill all the required fields.

The screenshot shows the 'Create an Azure Databricks workspace' wizard. The 'Basics' tab is active. In the 'Project Details' section, the subscription is set to 'Azure-DXC262AB12Lab' and the resource group is 'dxc1218'. In the 'Instance Details' section, the workspace name is 'dxcdab1', the region is 'East US', and the pricing tier is 'Standard (Apache Spark, Secure with Azure AD)'. At the bottom, there are navigation buttons: 'Review + create', '< Previous', and 'Next : Networking >'.

**Step 4:** Now click on “Review+create” to validate and then click on create at the left bottom to deploy the databricks.

The screenshot shows the 'Create an Azure Databricks workspace' page. At the top, there's a green banner with a checkmark icon and the text 'Validation Succeeded'. Below it, the 'Review + create' tab is selected. The page is divided into sections: Basics, Networking, Advanced, Tags, and Review + create. Under Basics, the workspace name is 'dxcdbab1', subscription is 'Azure-DXC262AB12Lab', resource group is 'dxc1218', region is 'East US', and pricing tier is 'standard'. Under Networking, it shows 'Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)' set to 'No' and 'Deploy Azure Databricks workspace in your own Virtual Network (VNet)' also set to 'No'. At the bottom, there are three buttons: 'Create' (highlighted in blue), '< Previous', and 'Download a template for automation'.

**Step 5:** Deployment completed.

The screenshot shows the 'Deployment Details Blade' for deployment 'dxc1218\_dxcdb1'. The main message says 'Your deployment is complete'. It provides deployment details: Deployment name: dxc1218\_dxcdb1, Start time: 6/12/2022, 12:42:48 PM, Subscription: Azure-DXC262AB12..., Correlation ID: 1030b4a1-fa56-47f0-b20f-42107e3337ab, and Resource group: dxc1218. There are sections for 'Deployment details' (with a download link) and 'Next steps' (with a 'Go to resource' button). On the right, there are promotional cards for 'Cost Management', 'Microsoft Defender for Cloud', 'Free Microsoft tutorials', and 'Work with an expert'.

**Step 6:** Now click on “Go to Resource” and click on “launch workspace”.

The screenshot shows the Microsoft Azure portal interface for the 'dxc1218\_dxcdab1' Azure Databricks Service. The left sidebar includes sections for Overview, Activity log, Access control (IAM), Tags, Settings (Virtual Network Peering, Encryption, Properties, Locks), Automation (Tasks (preview), Export template), and Help (New Support Request). The main content area displays the 'Essentials' section with the following details:

Setting	Value
Status	Active
Resource group	dxc1218
Location	East US
Subscription	Azure-DXC262AB1Lab
Subscription ID	3ce23bc0-43b9-419d-abb4-e6a591ad9662
Tags	(edit) Click here to add tags

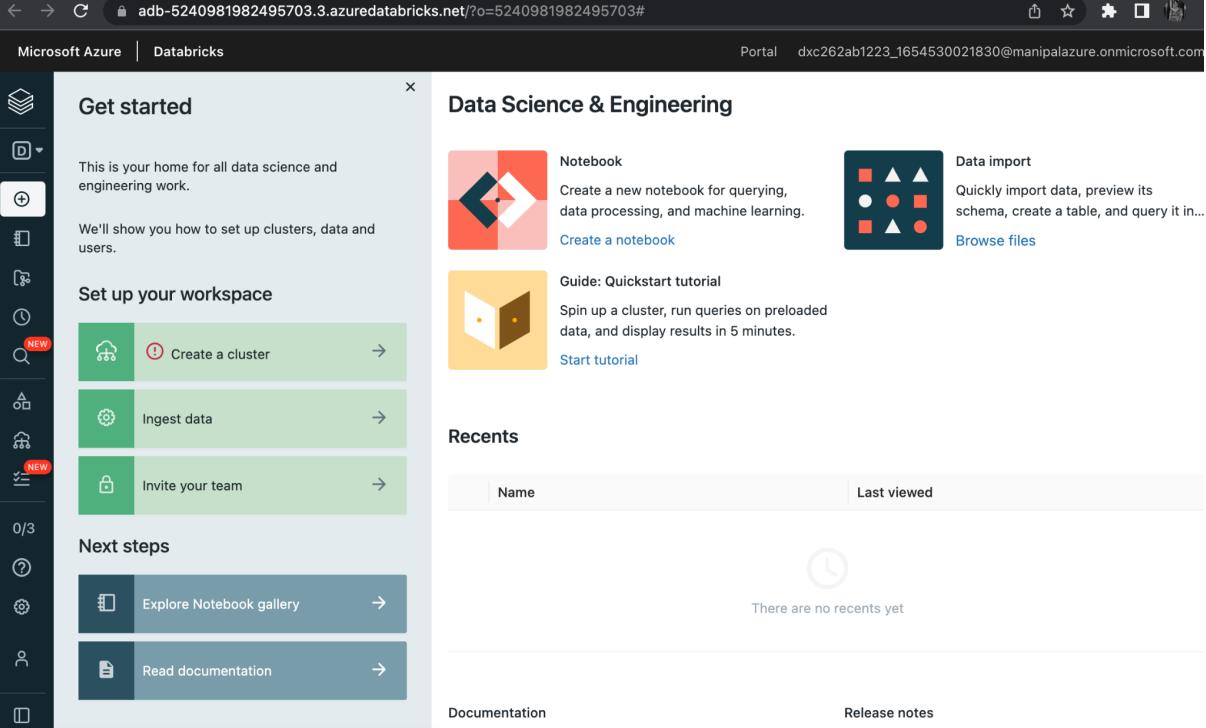
On the right, there are links for Managed Resource Group ([databricks-rg-dxcdab1-appdomphuphrc](#)), URL (<https://adb-5240981982495703.3.azuredatabricks.net>), and Pricing Tier (standard). A large red icon of three stacked cubes is centered below the details. At the bottom right is a blue 'Launch Workspace' button.

**Step 7:** You will sign into Azure Databricks and the Azure Databricks is created.

The screenshot shows the Azure Databricks workspace home page at [adb-5240981982495703.3.azuredatabricks.net/](https://adb-5240981982495703.3.azuredatabricks.net/). The left sidebar has sections for Get started, Set up your workspace (Create a cluster, Ingest data, Invite your team), and Next steps (Explore Notebook gallery, Read documentation). The main content area is titled 'Data Science & Engineering' and includes sections for Notebook (Create a new notebook for querying, data processing, and machine learning. [Create a notebook](#)), Data import (Quickly import data, preview its schema, create a table, and query it in... [Browse files](#)), Guide: Quickstart tutorial (Spin up a cluster, run queries on preloaded data, and display results in 5 minutes. [Start tutorial](#)), and Recents (a table showing recent items with columns for Name and Last viewed. It notes 'There are no recents yet'). At the bottom are links for Documentation and Release notes.

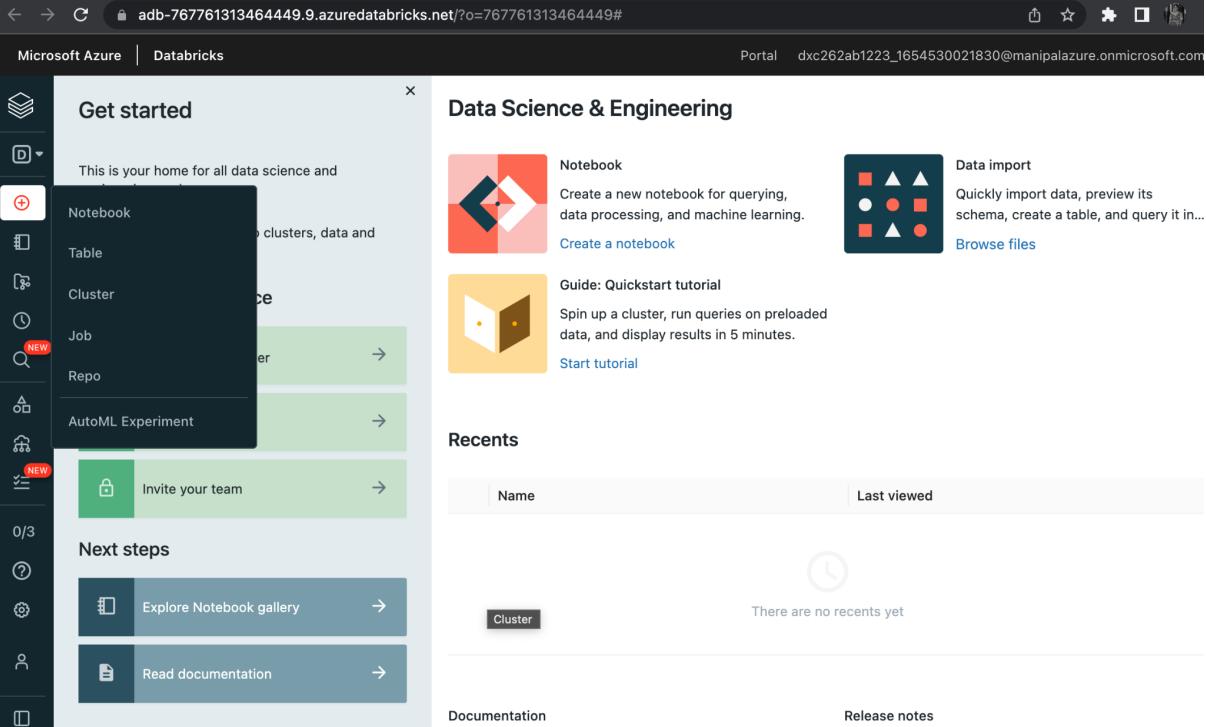
# Practical Lab: Create Cluster in Azure Databricks

## Step 1: Open Databricks workspace.



The screenshot shows the Microsoft Azure Databricks workspace interface. On the left, there's a sidebar with various icons for navigation. The main area is titled "Data Science & Engineering". It features several cards: "Notebook" (Create a new notebook for querying, data processing, and machine learning. [Create a notebook](#)), "Guide: Quickstart tutorial" (Spin up a cluster, run queries on preloaded data, and display results in 5 minutes. [Start tutorial](#)), "Data import" (Quickly import data, preview its schema, create a table, and query it in... [Browse files](#)), "Recents" (a table showing no recent items), and "Next steps" (links to "Explore Notebook gallery" and "Read documentation").

## Step 2: Click on “+” cluster to create a new cluster.



This screenshot is similar to the first one, but the "Cluster" icon in the sidebar is highlighted with a red box. This indicates that the user has clicked on it to proceed with creating a new cluster. The rest of the interface remains the same, showing the "Data Science & Engineering" dashboard with its various components and links.

## Step 3: Give the necessary credentials for the cluster and click on create.

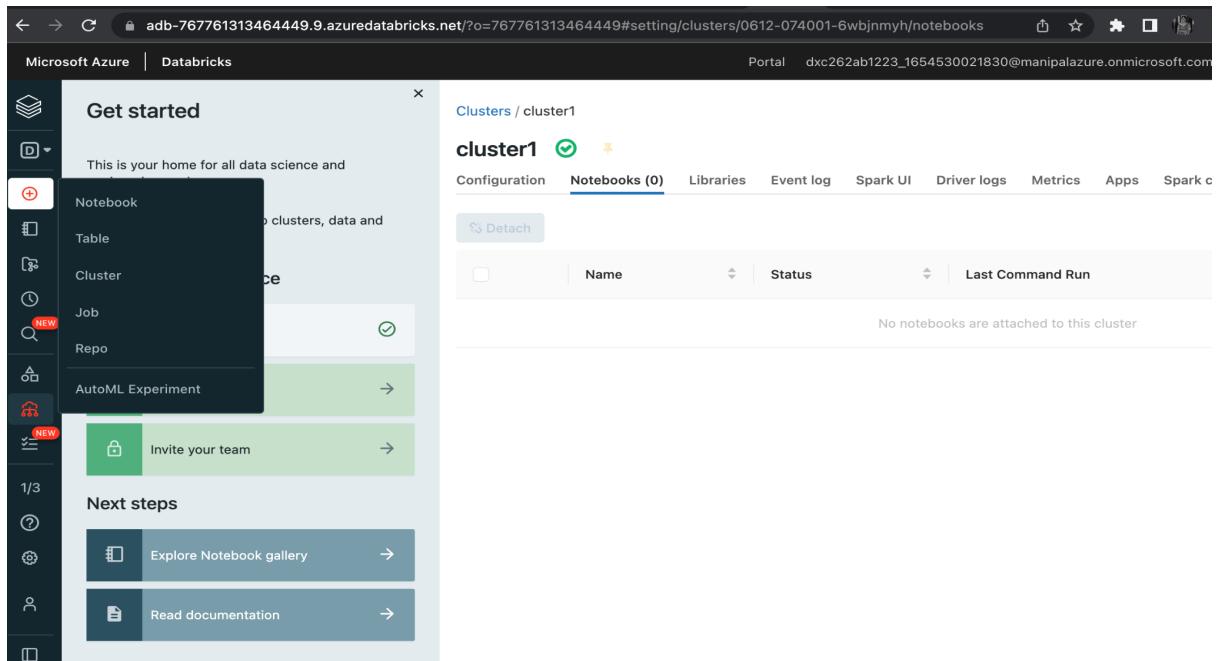
The screenshot shows the 'Clusters / New Compute' section of the Databricks interface. On the left, there's a sidebar with various icons and a 'Get started' section. The main area is titled 'New Cluster' with a 'Create Cluster' button. It shows configuration details: 'Standard' runtime, 'Runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1)', and a promotional discount message. Under 'Autopilot options', 'Enable autoscaling' is checked. 'Worker type' is set to 'Standard\_DS3\_v2' with '14 GB Memory, 4 Cores'. 'Min workers' is 2 and 'Max workers' is 8. 'Driver type' is 'Same as worker' with '14 GB Memory, 4 Cores'. A DBU range of '2.25 - 6.75' is shown next to a 'Standard\_DS3\_v2' button. At the bottom, there's an 'Advanced options' link.

## Step 4: Cluster is created.

The screenshot shows the 'Clusters / cluster1' page. The sidebar on the left is identical to the previous screenshot. The main area shows a cluster named 'cluster1' with a green status icon. Below it, there are tabs for Configuration, Notebooks (0), Libraries, Event log, Spark UI, Driver logs, Metrics, Apps, and Spark c. The 'Notebooks (0)' tab is selected. A 'Detach' button is visible. A table below shows no attached notebooks. A note at the bottom states 'No notebooks are attached to this cluster'.

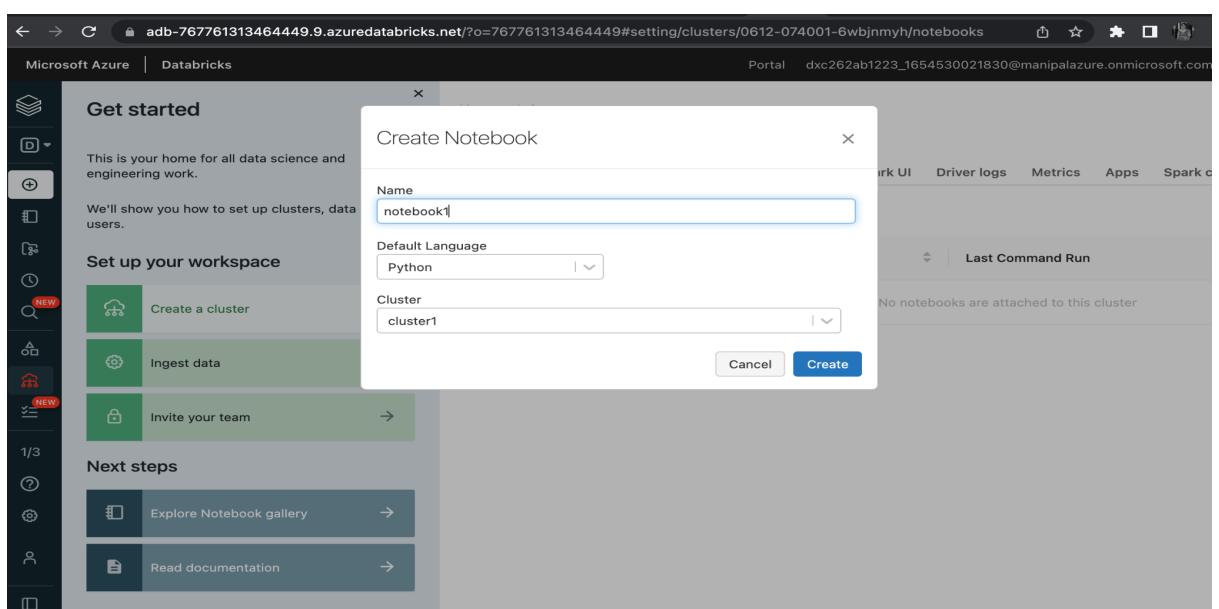
# Practical Lab: Add notebook in Databricks and Implement the Business Logic.

**Step 1:** We require a notebook that will perform calculations with the data that is now available or that is arriving in data bricks. So click on “+” notebook in the data bricks workspace to create a notebook.



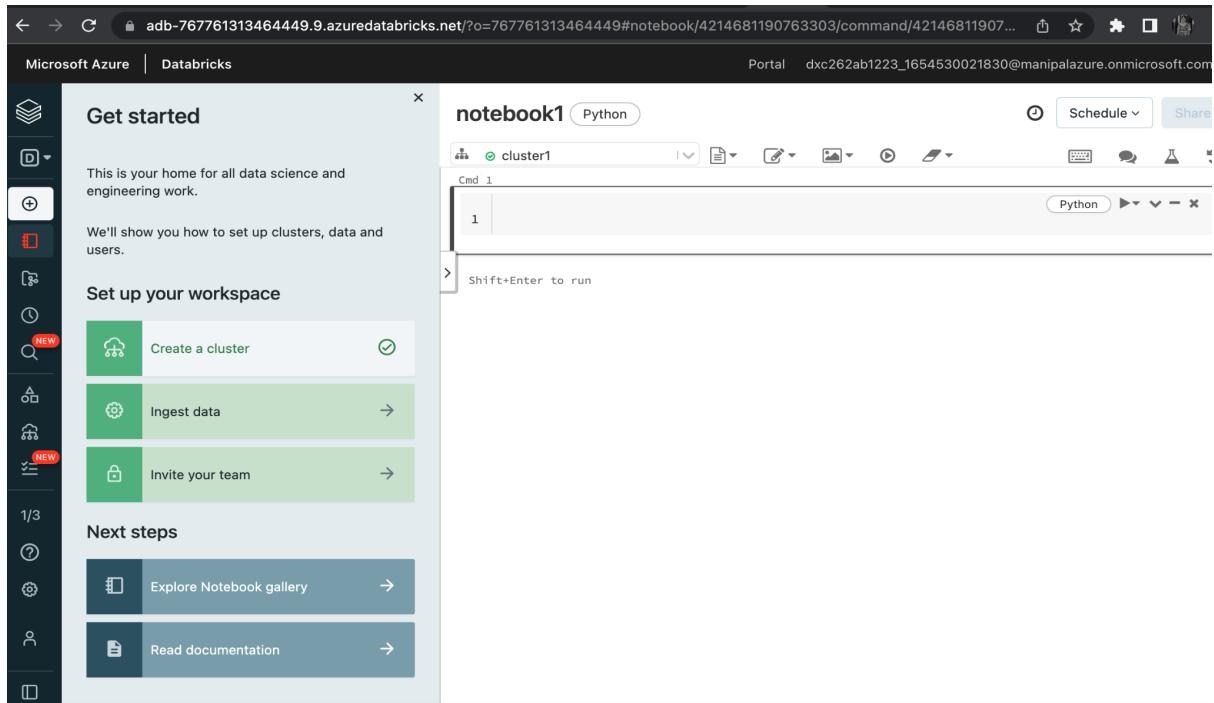
The screenshot shows the Microsoft Azure Databricks 'Get started' interface. On the left, there's a sidebar with various icons: a plus sign for 'New', a cluster icon, a table icon, a cluster icon with a plus sign, a job icon, a repo icon, an AutoML experiment icon, and a lock icon for 'Invite your team'. Below these are sections for 'Next steps': 'Explore Notebook gallery' and 'Read documentation'. The main area is titled 'Clusters / cluster1' and shows tabs for Configuration, Notebooks (0), Libraries, Event log, Spark UI, Driver logs, Metrics, Apps, and Spark c. The 'Notebooks' tab is selected. A message says 'No notebooks are attached to this cluster'. There's also a 'Detach' button.

**Step 2:** Give the necessary credentials for the notebook and click on create.

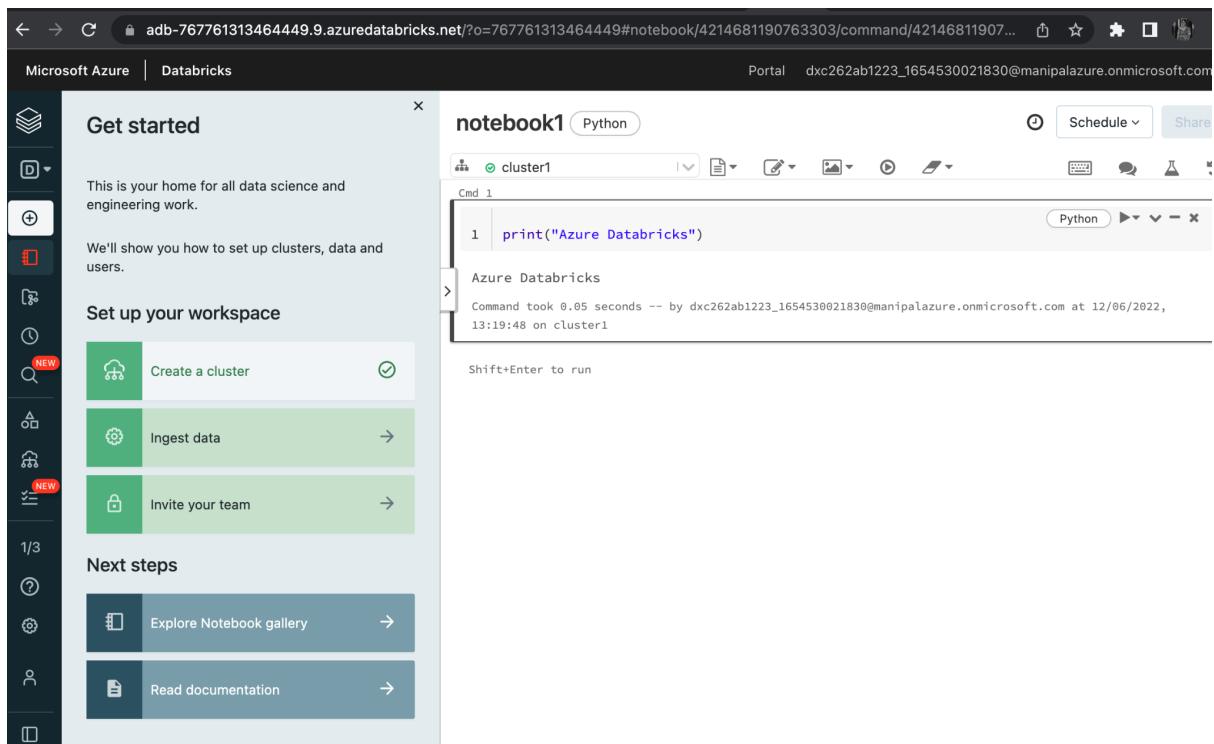


The screenshot shows the 'Create Notebook' dialog box over the 'Get started' interface. The dialog has fields for 'Name' (set to 'notebook1'), 'Default Language' (set to 'Python'), and 'Cluster' (set to 'cluster1'). It includes 'Cancel' and 'Create' buttons. The background shows the same 'Get started' sidebar and cluster details as the previous screenshot.

## Step 3: Notebook is created.



## Step 4: In the notebook, perform some operations or develop logic.



**Our notebook is now ready to be attached to Azure Data Factory and executed.**

# Practical Lab: Azure Data Factory For AP Morgan

**Step 1:** Login into “Microsoft Azure Portal” and search for “Data Factories”.

The screenshot shows the Microsoft Azure Portal homepage. In the 'Azure services' section, there are icons for Create a resource, Data factories, Azure Databricks, Azure Synapse Analytics, Virtual machines, Azure Cosmos DB, SQL databases, Quickstart Center, and App Services. Below this is a 'More services' button. In the 'Resources' section, tabs for Recent and Favorite are shown, along with columns for Name, Type, and Last Viewed. A message indicates 'No resources have been viewed recently' with a 'View all resources' button.

**Step 2:** Now click on “Data Factories” and then on “+create” to create a Azure Data Factory.

The screenshot shows the 'Data factories' list page in the Microsoft Azure Portal. The top navigation bar includes 'Home >', 'Data factories', and a 'More' button. The main area displays a search bar and filter options for Subscription, Type, Resource group, and Location. A message states 'No data factories to display' with a note to 'Try changing or clearing your filters.' and a prominent 'Create data factory' button.

## Step 3: Now will the required fields.

The screenshot shows the Azure portal's 'Create Data Factory' interface. The 'Basics' tab is selected. In the 'Project details' section, the 'Subscription' dropdown is set to 'Azure-DXC262AB12Lab'. Under 'Resource group', it shows '(New) dxc1218' with a 'Create new' link. The 'Instance details' section has the 'Name' field filled with 'dxcadf1218'. The 'Region' dropdown is set to 'East US' and 'Version' is set to 'V2 (Recommended)'.

## Step 4: After filling the required fields click on “Review + create” to validate the fields.

The screenshot shows the 'Create Data Factory' interface in the 'Review + create' step. A green banner at the top indicates 'Validation Passed'. The 'Review + create' tab is selected. The 'TERMS' section contains a paragraph about agreeing to legal terms and privacy statements. Below are two sections: 'Basics' and 'Networking'. The 'Basics' section shows the same configuration as the previous screenshot: Subscription 'Azure-DXC262AB12Lab', Resource group 'dxc1218', Name 'dxcadf1218', Region 'East US', and Version 'V2 (Recommended)'. The 'Networking' section is partially visible.

**Step 5:** Now click on “create” at the left bottom to deploy the Data Factory.

Microsoft.DataFactory-20220610161601 | Overview

Deployment

Search (Cmd+ /)

Delete Cancel Redeploy Refresh

We'd love your feedback! →

... Deployment is in progress

Deployment name: Microsoft.DataFactory-20220610161601  
Subscription: Azure-DXC262AB12Lab  
Resource group: dxc1218

Start time: 6/10/2022, 4:36:22 PM  
Correlation ID: f396e62d-962c-46e5-8ef9-850c58b0ec63

Deployment details (Download)

Resource	Type	Status	Operation details
No results.			

**Step 6:** Deployment completed.

Microsoft.DataFactory-20220610161601 | Overview

Deployment

Search (Cmd+ /)

Delete Cancel Redeploy Refresh

We'd love your feedback! →

✓ Your deployment is complete

Deployment name: Microsoft.DataFactory-20220610161601... Start time: 6/10/2022, 4:36:22 PM  
Subscription: Azure-DXC262AB12Lab Correlation ID: f396e62d-962c-46e5-8ef9-850c58b0ec63

Deployment details (Download)

Next steps

Go to resource

Cost Management  
Get notified to prevent unexpected costs  
Set up cost alert:

Microsoft Defense

**Step 7:** Now click on “Go to Resource” and then open Azure Data Factory Studio.

The screenshot shows the Microsoft Azure portal interface. At the top, the URL is <https://portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/3ce23bc0-43b9-419d-abb4-e6a591ad9662/resourceGroups/dxcadf1218/providers/Microsoft.DataFactory/factories/dxcadf1218>. The page title is "dxcadf1218" under the "Data factory (V2)" category. The main content area displays the following details:

- Overview:** Status Succeeded, Location East US, Subscription (move) Azure-DXC262AB12Lab, Subscription ID 3ce23bc0-43b9-419d-abb4-e6a591ad9662.
- Getting started:** Quick start link.
- Settings:** Options include Networking, Managed identities, Properties, Locks, and Getting started.
- Getting started:** Two cards:
  - Open Azure Data Factory Studio:** Start authoring and monitoring your data pipelines and data flows. [Open](#)
  - Read documentation:** Learn how to be productive quickly. Explore concepts, tutorials, and samples. [Learn more](#)

**Step 8:** Creation of “Azure Data Factory” is done.

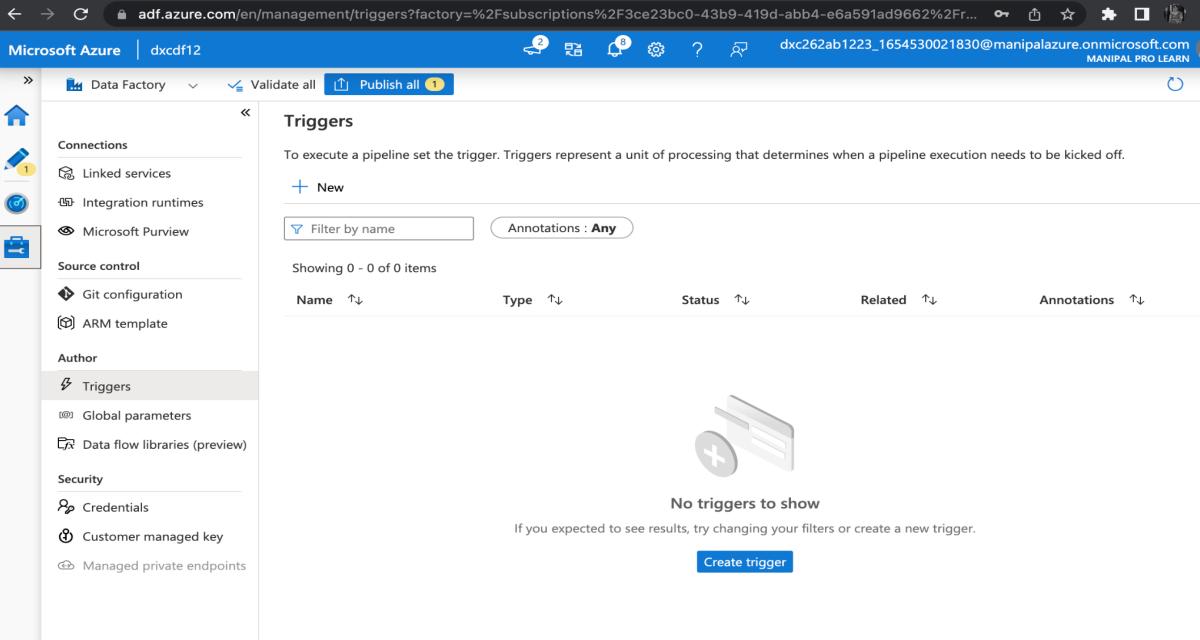
The screenshot shows the Azure Data Factory (ADF) studio home page for the data factory "dxcadf1218". The top navigation bar includes links for Home, Set up code repository, and a help section. The main content area features a large 3D diagram illustrating the data flow process, showing data being ingested, orchestrated, and transformed. Below the diagram, four key features are highlighted:

- Ingest:** Copy data at scale once or on a schedule.
- Orchestrate:** Code-free data pipelines.
- Transform data:** Transform your data using data flows.
- Configure SSIS:** Manage & run your SSIS packages in the cloud.

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

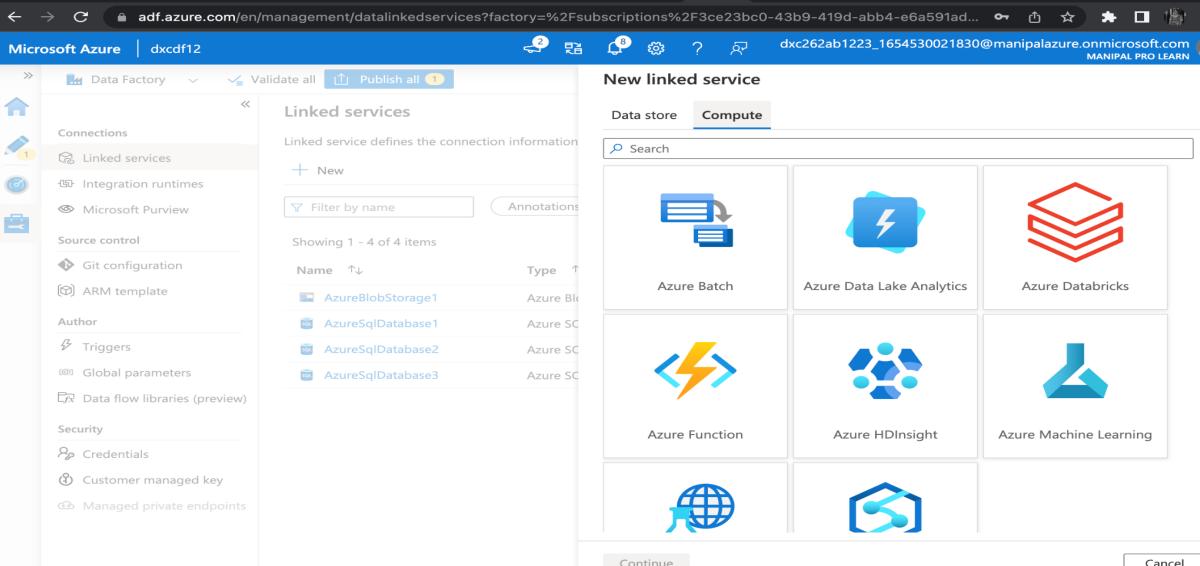
We'll need to build a pipeline to connect the Data Bricks notebook to Data Factory.

**Step 1:** Click Linked services under Connections, then select Add New from the Manage option on the left side.



The screenshot shows the Microsoft Azure Data Factory management interface. The left sidebar is titled 'Data Factory' and includes sections for Connections, Source control, Author, Security, and Compute. Under 'Author', 'Triggers' is selected. The main content area is titled 'Triggers' and contains a message: 'To execute a pipeline set the trigger. Triggers represent a unit of processing that determines when a pipeline execution needs to be kicked off.' Below this is a '+ New' button, a 'Filter by name' input field, and an 'Annotations : Any' dropdown. A table header row shows columns for Name, Type, Status, Related, and Annotations. The body of the table is empty, displaying 'Showing 0 - 0 of 0 items'. In the center, there's a large icon of a document with a plus sign and the text 'No triggers to show'. Below it, a message says 'If you expected to see results, try changing your filters or create a new trigger.' A blue 'Create trigger' button is at the bottom.

**Step 2:** Select Compute > Azure Databricks and then Continue in the New connected service box.



The screenshot shows the Microsoft Azure Data Factory management interface. The left sidebar is identical to the previous screenshot. The main content area is titled 'Linked services' and has tabs for 'Data store' and 'Compute'. The 'Compute' tab is selected. A search bar is above a grid of service icons. The grid contains six items: 'Azure Batch' (blue square with arrows), 'Azure Data Lake Analytics' (blue cube with a lightning bolt), 'Azure Databricks' (red stack of cubes), 'Azure Function' (orange lightning bolt), 'Azure HDInsight' (blue hexagon), and 'Azure Machine Learning' (blue triangle). At the bottom of the dialog are 'Continue' and 'Cancel' buttons.

## Step 3: Fill all the required fields to link a new databricks service.

New linked service

Azure Databricks [Learn more](#)

Name \*

Description

Connect via integration runtime \*

Account selection method \*  From Azure subscription  Enter manually

Azure subscription \*

Databricks workspace \*

Select cluster  New job cluster  Existing interactive cluster  Existing instance pool

Databrick Workspace URL

Create Back Test connection Cancel

Here to generate Access token use Azure Data Bricks.

In Azure DataBricks settings click on “User Settings”.

notebook1 Python

cluster1

Cmd 1

```
1 print("Azure Databricks")
```

Azure Databricks

Command took 0.05 seconds -- by dxc262ab1223\_1654530021830@manipalazure.onmicrosoft.com at 12/06/2022, 13:19:48 on cluster1

Shift+Enter to run

User Settings

Get started

This is your home for all data science and engineering work.

We'll show you how to set up clusters, data and users.

Set up your workspace

Create a cluster

Ingest data

Invite your team

User Settings

Admin Console

Manage Account

Read documentation

# Click on Generate New Token.

This is your home for all data science and engineering work.

We'll show you how to set up clusters, data and users.

### Set up your workspace

- Create a cluster
- Ingest data
- Invite your team

### Next steps

- Explore Notebook gallery
- Read documentation

Click on generate after filling the required fields.

This is your home for all data science and engineering work.

We'll show you how to set up clusters, data and users.

### Set up your workspace

- Create a cluster
- Ingest data
- Invite your team

### Next steps

- Explore Notebook gallery
- Read documentation

# Access Token is generated.

The screenshot shows the Microsoft Azure Databricks User Settings page. On the left, there's a sidebar with 'Get started' and 'Set up your workspace' sections, along with 'Next steps'. The main area is titled 'User Settings' and has tabs for 'Access tokens', 'Git integration', 'Notebook settings', 'Email preferences', 'Language settings', and 'Preview'. Under the 'Access tokens' tab, it says 'Personal access tokens can be used for secure authentication to the [Databricks API](#) instead of passwords.' A button labeled 'Generate new token' is visible. Below it is a table with columns 'Comment', 'Creation', and 'Expiration'. One row is shown: 'Comment' is empty, 'Creation' is '2022-06-12 13:46:31 IST', and 'Expiration' is '2022-09-10 13:46:31 IST'.

## Step 4: Click on create.

The screenshot shows the Azure Data Factory 'New linked service' creation dialog. The left sidebar includes options like 'Connections', 'Linked services', 'Integration runtimes', 'Microsoft Purview', 'Source control', 'Author', 'Triggers', 'Global parameters', 'Data flow libraries (preview)', 'Security', 'Credentials', 'Customer managed key', and 'Managed private endpoints'. The main area is titled 'New linked service' and shows 'Azure Databricks' as the selected service. It includes fields for 'Cluster version' (set to '10.1 (includes Apache Spark 3.2.0, Scala 2.12)'), 'Cluster node type' ('Standard\_DS3\_v2'), 'Python Version' ('2'), 'Worker options' (radio button for 'Fixed' selected), and 'Workers' ('1'). There are also sections for 'Additional cluster settings', 'Annotations', 'Parameters', and 'Advanced'. At the bottom are 'Create', 'Back', 'Test connection', and 'Cancel' buttons.

## Step 5: Azure Data Bricks is successfully linked.

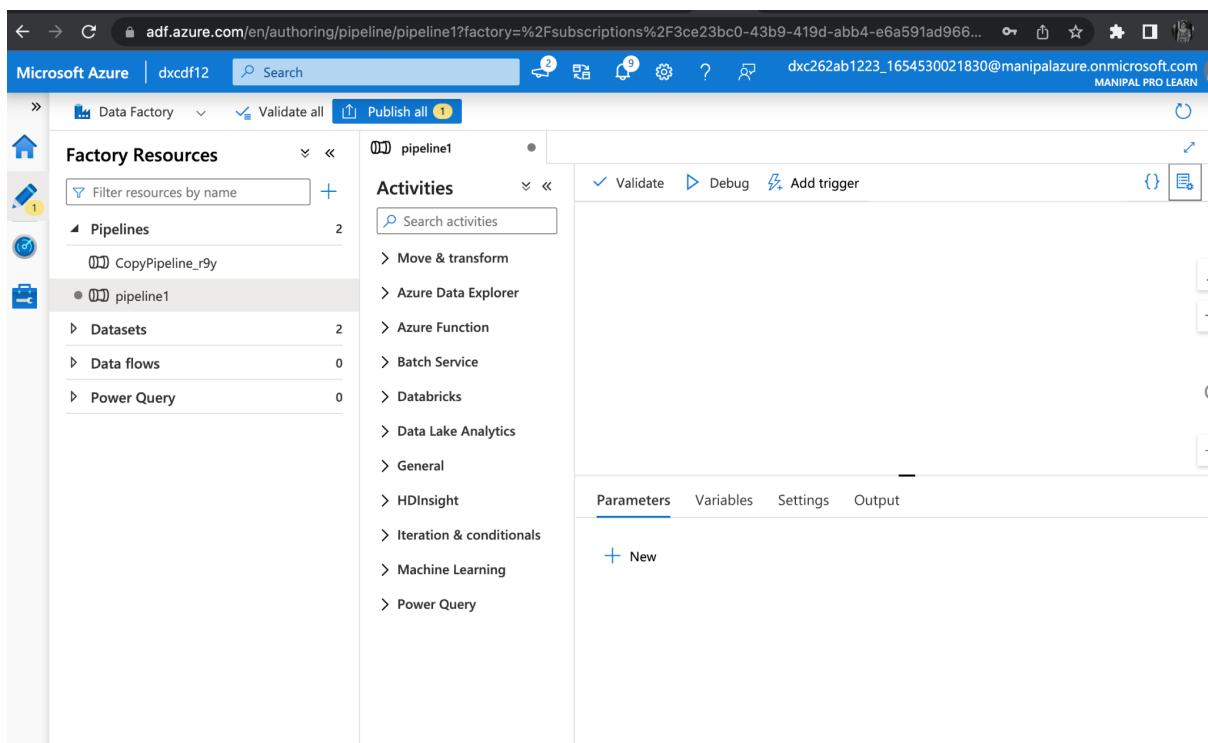
The screenshot shows the 'Linked services' section of the Azure Data Factory interface. On the left, there's a sidebar with various options like 'Connections', 'Source control', 'Author', and 'Security'. The main area displays a table with five rows of linked services:

Name	Type	Related	Annotations
AzureBlobStorage1	Azure Blob Storage	1	
AzureDatabricks1	Azure Databricks	0	
AzureSqlDatabase1	Azure SQL Database	0	
AzureSqlDatabase2	Azure SQL Database	0	
AzureSqlDatabase3	Azure SQL Database	1	

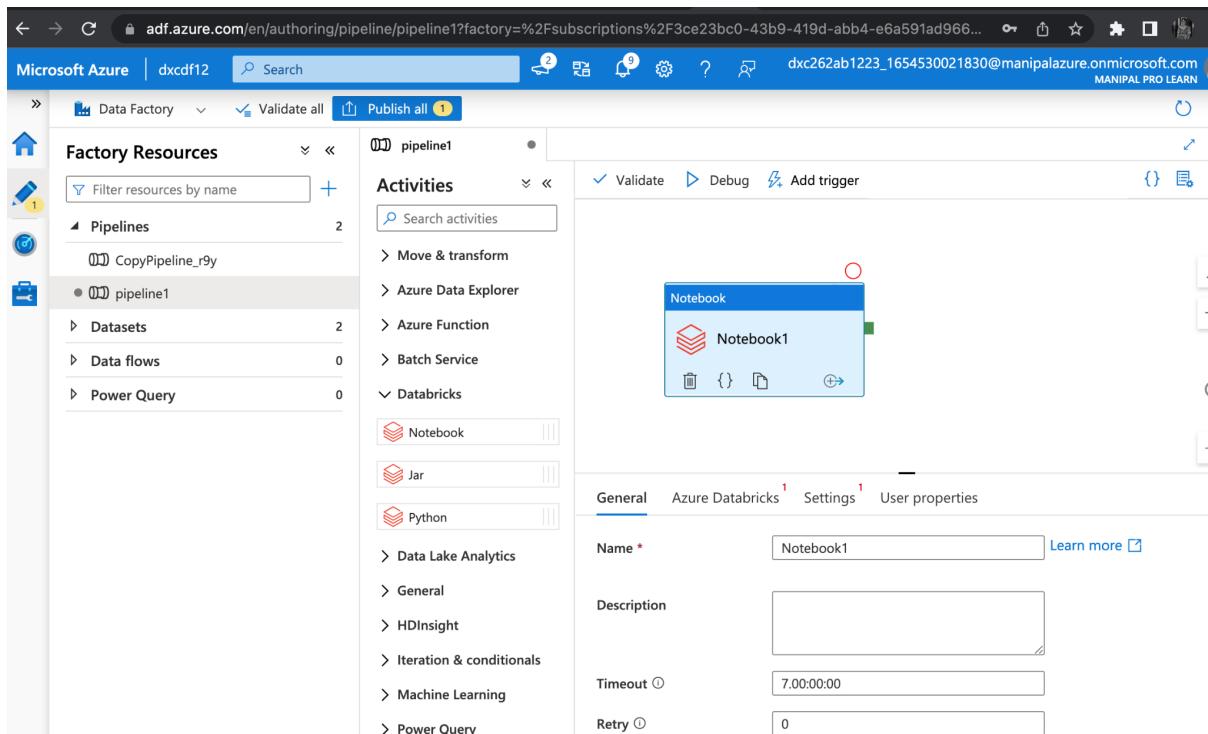
## Step 6: Now create a pipeline.

Click on Author and then on Add resource to create a pipeline.

The screenshot shows the 'Factory Resources' section of the Azure Data Factory interface. On the left, there's a sidebar with 'Factory Resources' expanded, showing 'Pipelines', 'Datasets', 'Data flows', and 'Power Query'. A context menu is open over the 'Pipelines' item, listing 'Pipeline', 'Dataset', 'Data flow', 'Power Query', and 'Copy Data tool'. Below the menu, there's a placeholder message 'Select an item' and a note 'Use the resource explorer to select or create a new item'.



**Step 7:** Drag the notebook tab from the Azure Databricks dropdown onto the workspace of the pipeline.



## Step 8: Fill all the details that are required and validate the pipeline.

The screenshot shows the Microsoft Azure Data Factory pipeline validation output. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (CopyPipeline\_r9y, pipeline1), 'Datasets' (2), 'Data flows' (0), and 'Power Query' (0). The main area displays the validation results for 'pipeline1'. A large green checkmark icon indicates success. Below it, a message says 'Your pipeline has been validated.' and 'No errors were found.' The pipeline structure is shown with a 'Notebook' activity containing a single 'Notebook1' step. The 'Settings' pane shows a 'Base parameters' section with a 'Name' field set to 'p1'. A 'Close' button is at the bottom right.

## Step 9: Click on publish all to publish the pipeline and then we can trigger and run the pipeline.

The screenshot shows the 'Publishing' dialog for the pipeline. The left sidebar is identical to the previous screenshot. The right panel is titled 'Publish all' and contains a message: 'You are about to publish all pending changes to the live environment.' with a 'Learn more' link. Below this is a 'Pending changes (1)' section. A table shows the change for 'pipeline1':

NAME	CHANGE	EXISTING
pipeline1	(New)	-

At the bottom are 'Publish' and 'Cancel' buttons.

The screenshot shows the Microsoft Azure Data Factory publishing interface. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (CopyPipeline\_r9y, pipeline1), 'Datasets', 'Data flows', and 'Power Query'. The main area shows the 'Activities' pane for 'pipeline1', which includes 'Move & transform', 'Azure Data Explorer', 'Azure Function', 'Batch Service', and 'Databricks' sections. Under 'Databricks', there is a 'Notebook' activity named 'Notebook1'. A message at the top right says 'Publishing completed' and 'Successfully published'. The 'Settings' tab is selected, showing a parameter 'p1' with the value '@pipeline().parameters.p1'.

## Step 10: Trigger the pipeline and monitor the pipeline run.

The screenshot shows the Microsoft Azure Data Factory 'Pipeline runs' monitoring interface. The left sidebar includes 'Dashboards', 'Runs', 'Pipeline runs' (selected), 'Trigger runs', 'Runtimes & sessions' (Integration runtimes, Data flow debug), 'Notifications', and 'Alerts & metrics'. The main area displays a table of pipeline runs:

Pipeline name	Run start	Run end	Duration	Triggered by
CopyPipeline_r9y	Jun 12, 2022, 12:24:04 pm	Jun 12, 2022, 12:24:17 pm	00:00:13	Manual trigger
pipeline1	Jun 12, 2022, 2:05:31 pm	Jun 12, 2022, 2:05:45 pm	00:00:14	Manual trigger

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has a 'Pipeline runs' section selected. 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 row is visible: pipeline1, Jun 11, 2022, 10:52:42 pm, Jun 11, 2022, 10:53:04 pm, 00:00:21, Manual trigger, Succeeded, and Original.

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error	Run
pipeline1	Jun 11, 2022, 10:52:42 pm	Jun 11, 2022, 10:53:04 pm	00:00:21	Manual trigger	Succeeded		Original

We can observe that Data Factory has successfully triggered a linked notebook of Data bricks.

**Result:** Using Data Factory, I was able to successfully link and trigger an Azure DataBricks notebook.

**Conclusion:** Azure Data Factory and Azure Data Bricks are related.

## References:

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