



---

# DXC REALTIME PROJECTS

---

AZ-900, DP - 203



JUNE 10, 2022

DXC

Name: Nandi Vomkara Aditya Mohan

Reg No: DXCAB1203

**Project1 Name: Smart Vehicles**

Date: 10-06-2022

## Project 1 : Connected Vehicles

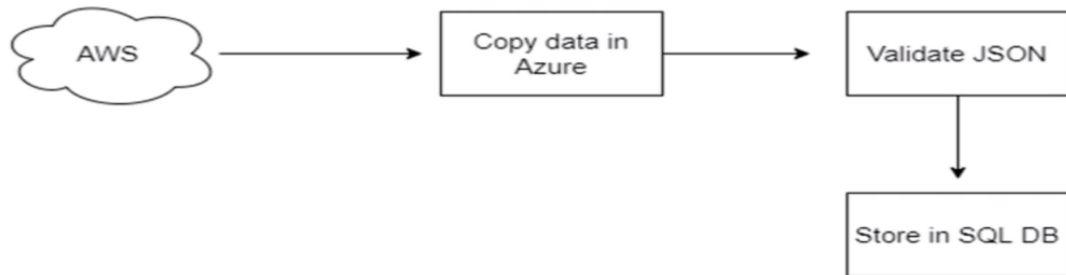
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



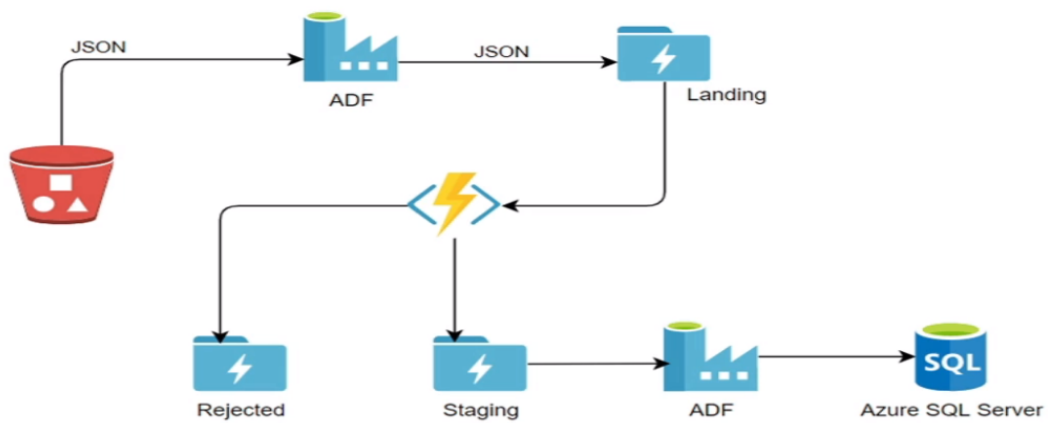
## Project 1 : Connected Vehicles

- 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



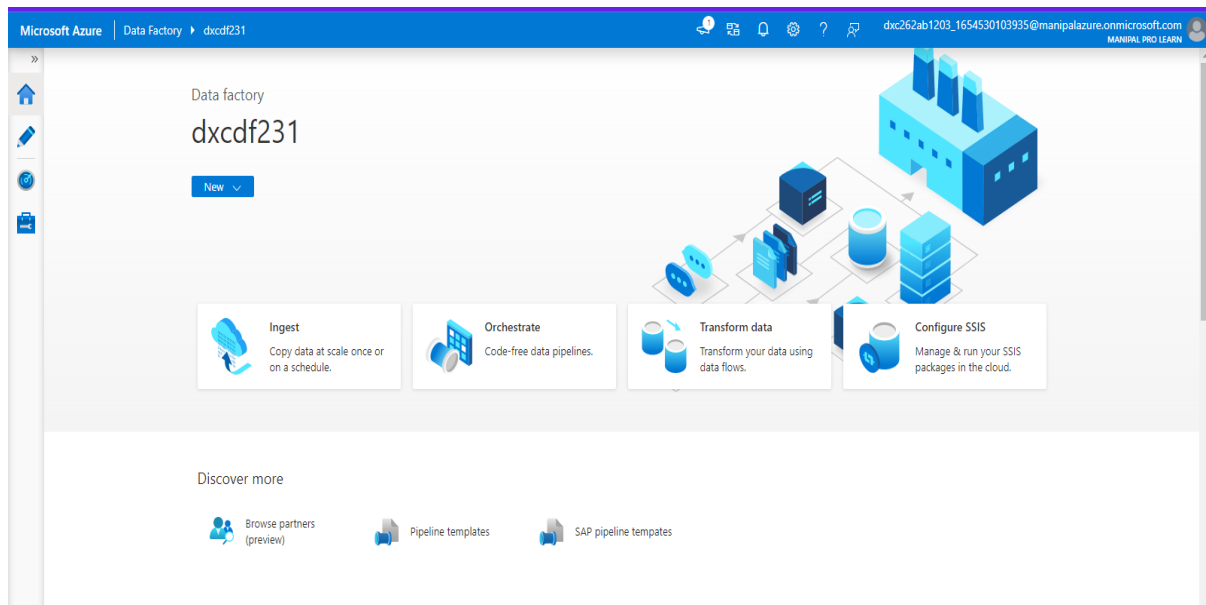
Architecture Diagram for Connected Vehicle Project

## Motivation:

Motivation of this project is to connect and migrate the data from IOT connected AWS cloud to our clients Azure Cloud. This action includes the linking of AWS data using azure data factory which provides insights of the data as well as migration in a same step with in no time and cost effective.

## Resource used:

Azure data factory



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

Azure data factory can be created by using the following steps

**Step-1:** login to the azure portal and search for azure data factory as shown in fig 1.1

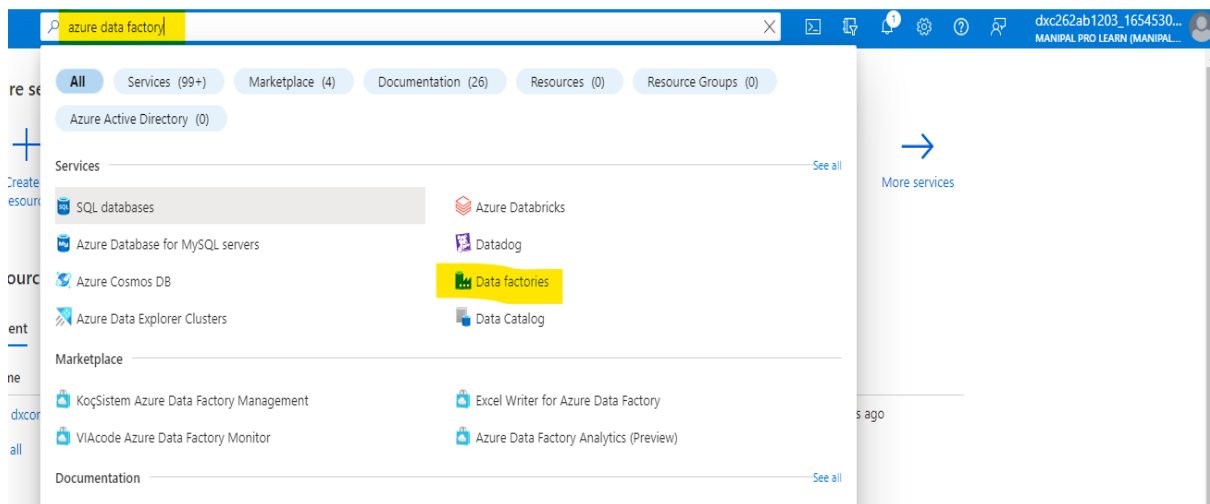


Fig 1.1 shows search for data factory in azure portal

**Step-2:** after navigating to the azure data factory page click on create and select the resource group, name, region, version after that click on next follow the fig 1.2

Fig-1.2 shows the creation procedure of data factory

**Step-3:** click on git configurations and checkbox it as configure git later as shown in fig 1.3

[Home](#) > [Data factories](#) >

## Create Data Factory ...

Basics

**Git configuration**

Networking

Advanced

Tags

Review + create

Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration.

[Learn more about Git integration in Azure Data Factory](#)

Configure Git later ⓘ



Fig-1.3 shows the git configuration during data factory creation

Step-4: Go through the next steps followed by successful completion of validation click on create as shown in fig-1.4

[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	dxccorg231
Name	dxcdf231
Region	East US
Version	V2 (Recommended)

### Networking

Connect via Public endpoint

**Create**

< Previous

Next

[Download a template for automation](#)

Fig-1.4 shows the steps in creation of data factory

**Step-5:** after clicking on create it takes some time for deployment after deployment you can see the following as shown in fig-1.5

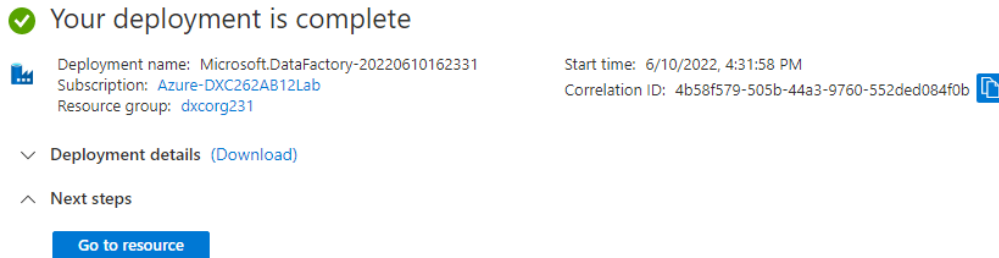


Fig1.5 shows the successful deployment of azure data factory

**Step-6:** click on go to resources after that click on open as shown in the figure 1.6

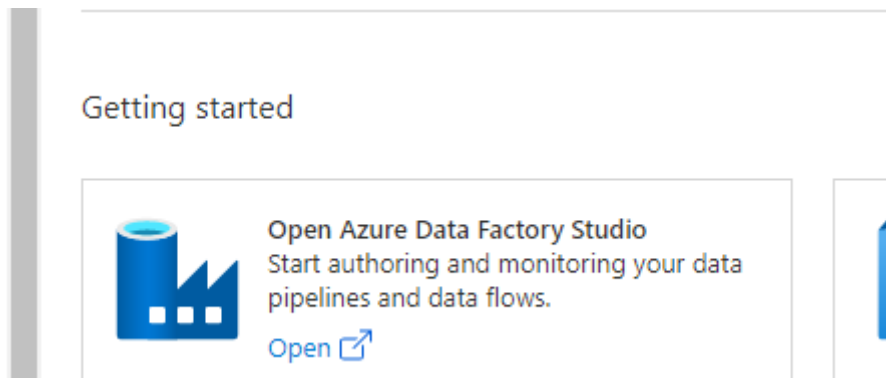


Fig-1.6 helps to navigate to the DF studio

**Step-7:** After clicking over that it will open azure data factory in new window as mentioned in fig 1.7 and we can use this creation of pipelines

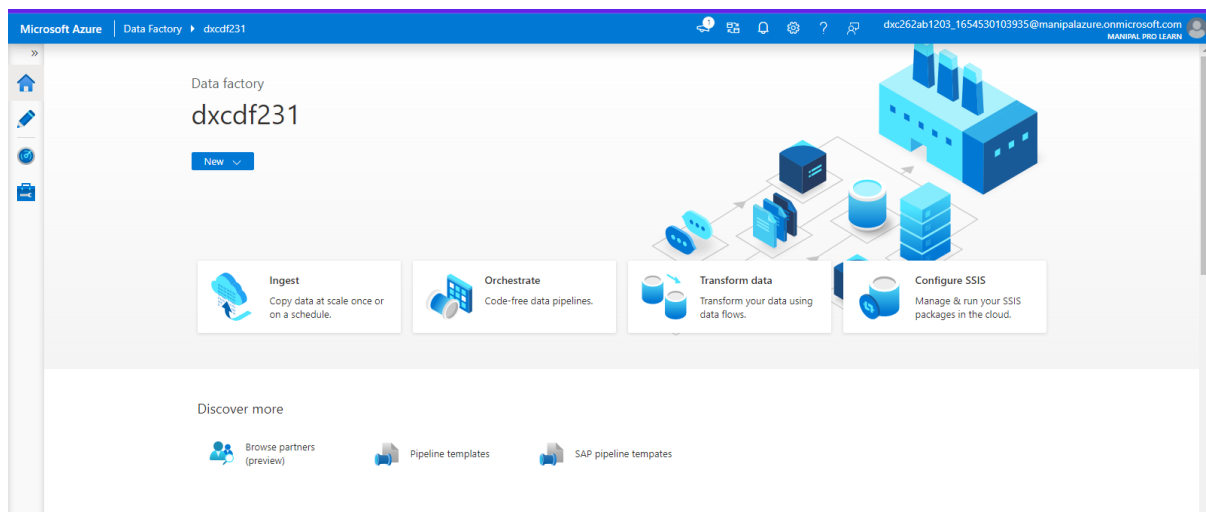


Fig-1.7 shows the main page of DATA FACTORY



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

In Azure data factory we use pipe lines for channeling the source to the correct destination which are having set of constraints. In order to create an end-to-end pipe line along with triggers enabled we have to follow the below mentioned steps

Step-1: As per our project the data is in json format we will get data from aws and we are storing data in file and we are moving this data through the source to the destination in our case it is



**Step-2 :** we are moving the data from the source to the destination using the copy data tool and creating a pipeline

The screenshot shows the 'Copy Data tool' configuration window in Azure Data Factory. The left sidebar contains a navigation pane with icons for Home, Edit, Monitor, and a vertical list of steps: Properties (checked), Source (checked), Target (checked), Settings (checked), Review and finish (selected), Review, and Deployment. The main area is titled 'Copy Data tool' and contains a 'Summary' section with the text 'You are running pipeline to copy data from Azure Blob Storage to Azure Blob Storage.' Below this is a visual representation of the data flow: 'Azure Blob Storage' (source icon) → 'Azure Blob Storage' (target icon). The 'Properties' section lists the following configuration details:

Property	Value	Action
Task name	pipeline1	Edit
Task description		
Source		Edit
Connection name	AzureBlobStorage1	
Dataset name	SourceDataset_v7x	
Column delimiter	,	
Row delimiter		
Escape character	\	
Quote char	"	
First row as header	true	
File name	1000 Companies.csv	

At the bottom of the configuration area are two buttons: '< Previous' and 'Next >'.

**Step-3:** The validation and deployment is done and pipeline is created successfully.

The screenshot shows the 'Copy Data tool' interface. On the left, a navigation pane lists steps: Properties, Source, Target, Settings, Review and finish (highlighted with a blue circle and checkmark), Review, and Deployment. The main area displays a diagram of data flow from 'Azure Blob Storage' to 'Azure Blob Storage'. Below this, the text 'Deployment complete' is shown. A table lists the deployment steps and their status:

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

Below the table, 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.' At the bottom, there are three buttons: 'Finish' (highlighted with a red checkmark), 'Edit pipeline', and 'Monitor'.

**Step-4:** The data is successfully copied from source to destination.

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

**Location:** source

Search blobs by prefix (case-sensitive)

☐ Show deleted blobs

+ Add filter

Name	Modified	Access tier
<input type="checkbox"/> 1000_Companies.csv	6/10/2022, 5:01:31 PM	Hot (Inferred)

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

**Location:** destination

Search blobs by prefix (case-sensitive)

☐ Show deleted blobs

+ Add filter

Name	Modified	Access tier
<input type="checkbox"/> 1000_Companies.txt	6/10/2022, 5:13:30 PM	Hot (Inferred)

## Practical Lab: Create Azure blob trigger logic

To create a azure blob trigger we have to have to follow the below mentioned steps. We have to keep an idea why we are triggering the data is the data from the device or source is to updated time to time. so in my point of view I used to update the information every minute that's why I set the trigger schedule for every minute.

**Copy Data tool**

Use Copy Data Tool to perform a one-time or scheduled data load from 90+ data sources. Follow the wizard experience to specify your data loading settings, and let the Copy Data Tool generate the artifacts for you, including pipelines, datasets, and linked services. [Learn more](#)

**Properties**

Select copy data task type and configure task schedule

**Task type**

**Built-in copy task**  
You will get single pipeline to copy data from 90+ data source easily.

**Metadata-driven copy task**  
You will get parameterized pipelines which can read metadata from an external store to load data at a large scale.

You will get single pipeline to quickly copy objects from data source store to destination in a very intuitive manner.

**Task cadence or task schedule \***

☐ Run once now ☒ **Schedule** ☐ Tumbling window

**Start Date (UTC) \***

**Recurrence \***

☒ **Specify an end date**

**End On (UTC) \***

[< Previous](#) [Next >](#) [Cancel](#)

After that we have to follow the steps for setting source and destination for the pipeline

**Copy Data tool**

**Summary**

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

**Azure Blob Storage** → **Azure Blob Storage**

**Properties** [Edit](#)

**Task name** pipeline1

**Task description**

**Source** [Edit](#)

**Connection name** AzureBlobStorage1

**Dataset name** SourceDataset\_v7x

**Column delimiter** ,

**Row delimiter**

**Escape character** \

**Quote char** "

**First row as header** true

**File name** 1000\_Campaigns.csv

[< Previous](#) [Next >](#)

After the pipeline is created wait for deployment and after deployment click on finish

Copy Data tool

Properties

Source

Target

Settings

Review and finish

Review

Deployment

Azure Blob Storage → Azure Blob Storage

### 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

And the pipeline starts data to trigger and below figure shows the triggered runs of the blob storage.

Pipeline runs

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

Filter by run ID or name Chennai, Kolkata, Mu... : Last 24 hours Pipeline name : All Status : All

Runs : Latest runs Triggered by : All Add filter

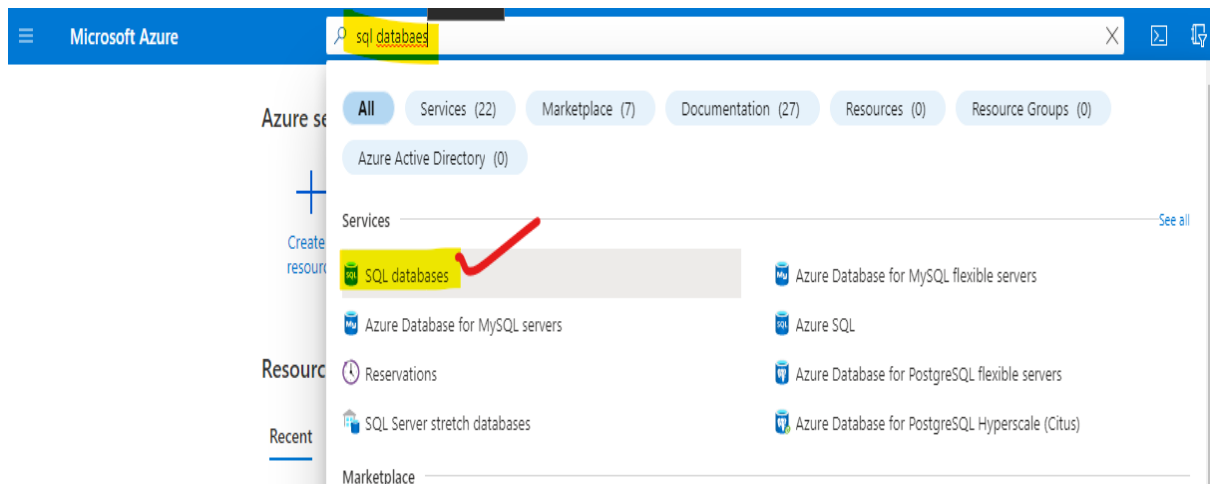
Showing 1 - 9 items Last refreshed 0 minutes ago

<input type="checkbox"/>	Pipeline name	Run start ↑↓	Run end	Duration	Triggered by	Status	Error
<input type="checkbox"/>	CopyPipeline_5mt	Jun 10, 2022, 5:47:00 pm	Jun 10, 2022, 5:47:09 pm	00:00:09	Trigger_5mt	✓ Succeeded	
<input type="checkbox"/>	CopyPipeline_5mt	Jun 10, 2022, 5:46:00 pm	Jun 10, 2022, 5:46:14 pm	00:00:14	Trigger_5mt	✓ Succeeded	
<input type="checkbox"/>	CopyPipeline_5mt	Jun 10, 2022, 5:45:00 pm	Jun 10, 2022, 5:45:22 pm	00:00:21	Trigger_5mt	✓ Succeeded	
<input type="checkbox"/>	CopyPipeline_5mt	Jun 10, 2022, 5:44:00 pm	Jun 10, 2022, 5:44:10 pm	00:00:10	Trigger_5mt	✓ Succeeded	
<input type="checkbox"/>	CopyPipeline_5mt	Jun 10, 2022, 5:43:01 pm	Jun 10, 2022, 5:43:11 pm	00:00:10	Trigger_5mt	✓ Succeeded	
<input type="checkbox"/>	CopyPipeline_5mt	Jun 10, 2022, 5:42:00 pm	Jun 10, 2022, 5:42:10 pm	00:00:09	Trigger_5mt	✓ Succeeded	
<input type="checkbox"/>	CopyPipeline_5mt	Jun 10, 2022, 5:41:00 pm	Jun 10, 2022, 5:41:13 pm	00:00:13	Trigger_5mt	✓ Succeeded	
<input type="checkbox"/>	pipeline1	Jun 10, 2022, 5:28:04 pm	Jun 10, 2022, 5:28:14 pm	00:00:10	Manual trigger	✓ Succeeded	
<input type="checkbox"/>	pipeline1	Jun 10, 2022, 5:13:21 pm	Jun 10, 2022, 5:13:32 pm	00:00:11	Manual trigger	✓ Succeeded	

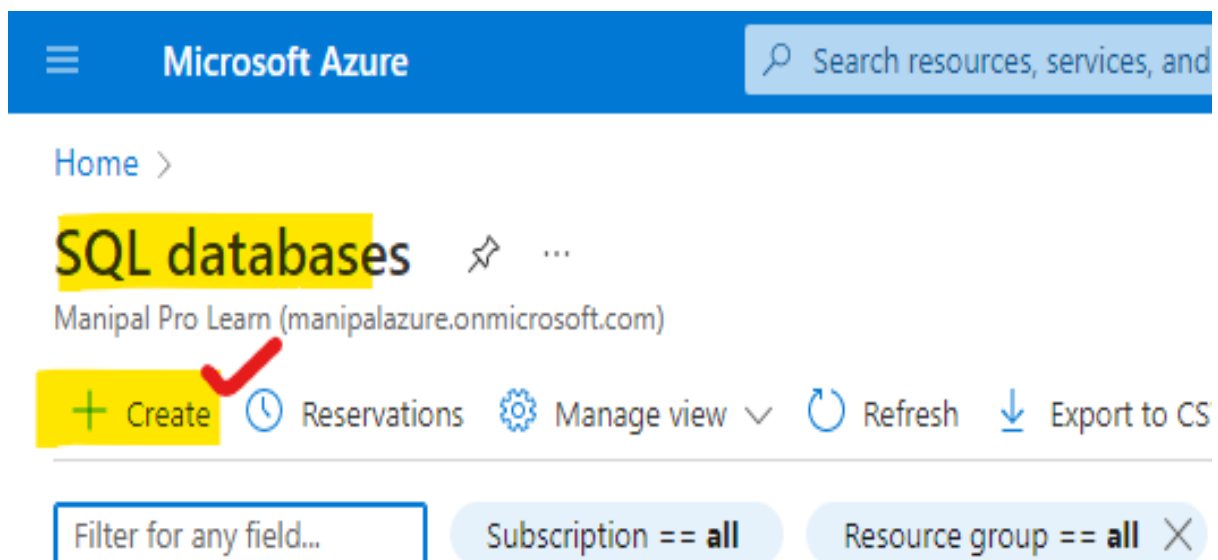
## Practical Lab: Create Azure SQL Server and Database

To create azure SQL SERVER & DATABASE, we have to follow the steps which are mentioned below

**Step-1:** search for azure sql database in azure portal and select it and click on create.



**Step-2:** After selecting the Sql databases. Click on create button as shown below



### Step-3: select all options as mentioned below

Microsoft Azure

Search resources, services, and docs (G+)

dxcc262ab1203\_16  
MANIPAL PRO LEARN (M)

Show portal menu

## Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

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

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

### Project details

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

Subscription \*

Resource group \*  [Create new](#)

### Database details

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

Database name \*

Server \*  [Create new](#)

Want to use SQL elastic pool? ☐ Yes ☒ No

Compute + storage \*    
Gen5, 1 vCore, 1 GB storage, zone redundant disabled   
[Configure database](#)

### Backup storage redundancy

[Review + create](#) [Next: Networking >](#)

<https://portal.azure.com/#>

**Step-4:** after completing this step navigate to the next menus without changing any settings and click on create

The screenshot shows the 'Create SQL Database' wizard in the Microsoft Azure portal, specifically the 'Review + create' step. The interface includes a top navigation bar with the Microsoft Azure logo, a search bar, and user information. The breadcrumb trail is 'Home > SQL databases >'. The main heading is 'Create SQL Database' with a Microsoft logo below it. The wizard has five tabs: 'Basics', 'Networking', 'Security', 'Additional settings', and 'Tags', with 'Review + create' being the active tab. Under 'Product details', it identifies the offering as 'SQL database by Microsoft' and provides links for 'Terms of use' and 'Privacy policy'. An 'Estimated cost' section shows 'Storage cost -- -- / month + Compute cost -- -- / vCore / second' with a link to 'View pricing details'. A 'Terms' section contains a legal disclaimer. The 'Basics' section lists configuration details: Subscription (Azure-DXC262AB12Lab), Resource group (dxc231), Region (eastus), Database name (dxcorg123), and Server (dxc231). Below this, it shows 'Compute + storage' as 'General Purpose - Serverless: Gen5, 1 vCore, 1 GB storage, zone redundant disabled' and 'Backup storage redundancy' as 'Geo-redundant backup storage'. The 'Networking' section shows 'Allow Azure services and resources to access this server' as 'No' and 'Private endpoint' as 'None'. The 'Security' section shows 'Service principal (preview)' as 'Off'. At the bottom, there is a red checkmark over the 'Create' button, a '< Previous' button, and a link to 'Download a template for automation'.

Microsoft Azure

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

Home > SQL databases >

## Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

### Product details

SQL database  
by Microsoft  
[Terms of use](#) | [Privacy policy](#)

**Estimated cost**  
Storage cost -- -- / month + Compute cost -- -- / vCore / second  
[View pricing details](#)

### Terms

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) 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. For additional details see [Azure Marketplace Terms](#).

### Basics

Subscription	Azure-DXC262AB12Lab
Resource group	dxc231
Region	eastus
Database name	dxcorg123
Server	dxc231

Compute + storage: General Purpose - Serverless: Gen5, 1 vCore, 1 GB storage, zone redundant disabled

Backup storage redundancy: Geo-redundant backup storage

### Networking

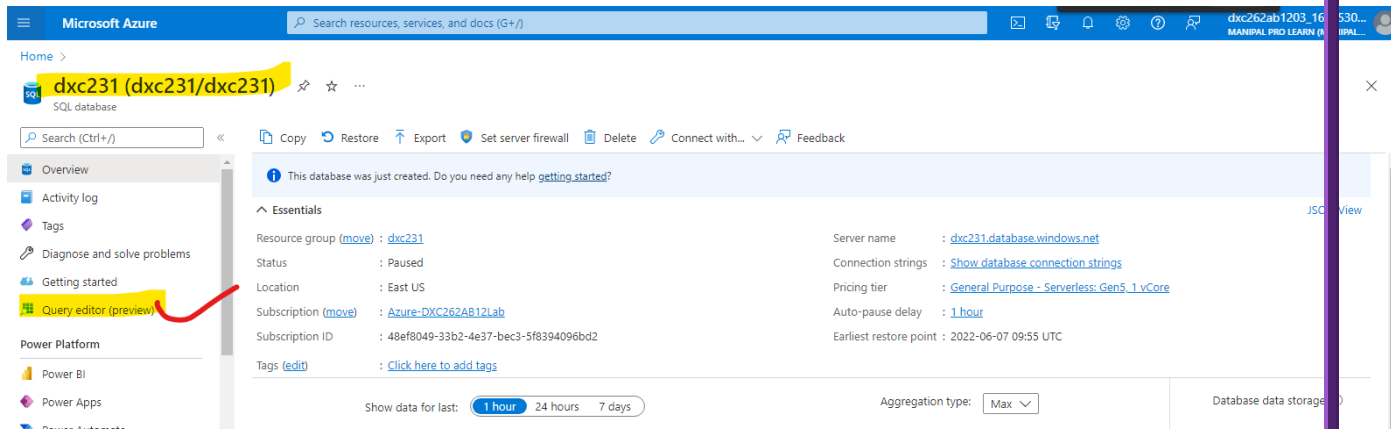
Allow Azure services and resources to access this server	No
Private endpoint	None

### Security

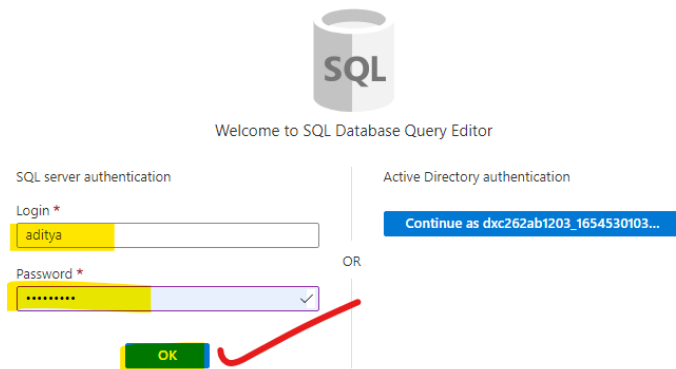
Service principal (preview)	Off
-----------------------------	-----

**Create** < Previous [Download a template for automation](#)

**Step-5:** after clicking on create it takes some time to deploy after that we can access the Database and click on query editor



**Step-6:** login with your login credentials and creation of Sql data base and server is done.





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

in this lab we are doing to perform the data from destination to the sql data base to perform this we have to link the blob trigger to the sql data base to perform that task we have to follow the below mentioned steps

Note: to perform this task we need to enable firewall for the current IP and for the same azure cloud otherwise connecting sql database through data factory is difficult.

Home > SQL databases >

### Create SQL Database

Microsoft

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

Configure network access and connectivity for your server. The configuration selected below will apply to the selected server 'newserv101' and all databases it manages. [Learn more](#)

#### Firewall rules

The settings displayed below are read-only. They can be modified from the "Firewalls and virtual networks" blade for the selected server after database creation. [Learn more](#)

Allow Azure services and resources to access this server

☐ No ☒ Yes

Add current client IP address \*

#### Private endpoints

Private endpoint connections are associated with a private IP address within a Virtual Network. The list below shows all the private endpoint connections for this server. Note that private endpoint connections are defined at the server level and they provide access to all databases in the server. [Learn more](#)

+ Add private endpoint

Name	Subscription	Resource group	Region	Subnet
Click on add to create private endpoint				

**Review + create**

< Previous

Next : Security >

**Step-1:** go to azure data factory and perform the action of creating a pipeline from blob storage to sql data base.

**Copy Data tool**

Use Copy Data Tool to perform a one-time or scheduled data load from 90+ data sources. Follow the wizard experience to specify your data loading settings, and let the Copy Data Tool create the pipeline for you.

**Properties**

Select copy data task type and configure task schedule

**Task type**

**Built-in copy task**  
You will get single pipeline to copy data from 90+ data source easily.

**Metac**  
You will get single pipeline to quickly copy objects from data source store to destination in

**Task cadence or task schedule \***

☒ Run once now ☐ Schedule ☐ Tumbling window

< Previous **Next** >

**Step-2:** now fill the details for the source of the pipeline that is blob storage account name and click on create button

The screenshot shows the 'New linked service' dialog for Azure Blob Storage. The 'Name' field is filled with 'AzureBlobStorage1'. The 'Description' field is empty. The 'Connect via integration runtime' dropdown is set to 'AutoResolveIntegrationRuntime'. The 'Authentication type' dropdown is set to 'Account key'. The 'Connection string' tab is selected. The 'Account selection method' is set to 'From Azure subscription'. The 'Azure subscription' dropdown is filled with 'Azure: OXC262AB12Lab (4236c42a-d131-4bd6-b609-aec3a598f2d3)'. The 'Storage account name' field is filled with 'storage1011'. The 'Additional connection properties' section is empty. The 'Test connection' button is disabled. The 'Create' button is highlighted with a red checkmark.

**Step-3:** After completing the above task we are successfully given the source location of the data and now we have to give the location of destination that is sql database.

The screenshot shows the 'Copy Data tool' interface. The 'Destination data store' section is active. The 'Target type' dropdown is set to 'Azure SQL Database'. The 'Connection' dropdown is set to 'Select...'. The 'New linked service' dialog is open on the right, showing the configuration for the 'Azure SQL Database' linked service. The 'Connection string' tab is selected. The 'Account selection method' is set to 'From Azure subscription'. The 'Azure subscription' dropdown is filled with 'Azure: OXC262AB12Lab (4236c42a-d131-4bd6-b609-aec3a598f2d3)'. The 'Server name' field is filled with 'newsserver101'. The 'Database name' field is filled with 'newsdatabase'. The 'Authentication type' dropdown is set to 'SQL authentication'. The 'User name' field is filled with 'user'. The 'Password' field is filled with '\*\*\*\*\*'. The 'Always encrypted' checkbox is unchecked. The 'Additional connection properties' section is empty. The 'Test connection' button is disabled. The 'Create' button is highlighted.

**Step-4:** now, the destination is set and pipeline is ready

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: AzureSqlDatabase1

Source: Azure Blob Storage file

Target: transaction (Use existing table)

Azure Blob Storage file (auto create)

☐ Skip column mapping for tables

< Previous Next > Cancel

**Step-5:** enter the task name and description and click on next

Settings

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

Task name \*: CopyPipeline\_zbw

Task description

Data consistency verification ☐

Fault tolerance

Enable logging ☐

Enable staging ☐

> Advanced

< Previous Next >

**Step-6:** review all and go ahead by clicking the next button.

**Summary**  
You are running pipeline to copy data from Azure Blob Storage to Azure SQL Database.

**Properties**

Task name: CopyPipeline\_rbw

Task description:

**Source**

Connection name: AzureBlobStorage1

Dataset name: SourceDataset\_L2bw

Column delimiter: ,

Escape character: \

Quote char: "

First row as header: True

File name: transactions.csv

Container: source

< Previous **Next** >

**Step-8:** now trigger the pipeline and the pipe line will be triggered

**Pipeline runs**

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

Filter by run ID or name: Chennai, Kolkata, Ma... Last 24 hours Pipeline name: All Status: All Runs: Latest runs

Triggered by: All Add filter

Showing 1 - 1 items

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error	Run
CopyPipeline_rbw	Jun 10, 2022, 5:41:49 pm	Jun 10, 2022, 5:42:07 pm	00:00:17	Manual trigger	Succeeded		Original

Last refreshed 0 minutes ago

**Step-9:** The data was moved successfully to the sql database. Due to limited access and time constraint the lab is closed so, screen shot is not available.

## **Result:**

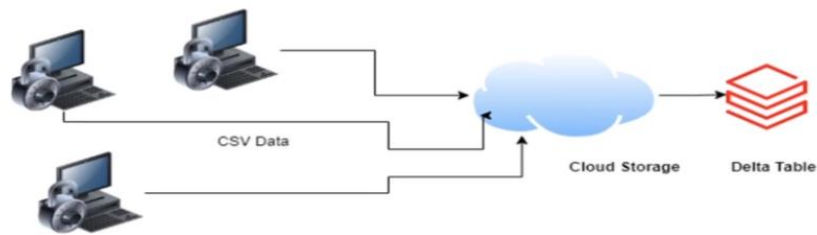
In this project we are successfully created a blob storage account, SQL database, azure data factory and used data to move from source that is AWS to Azure and the data has been moved successfully.

**Conclusion:** By using the resources from azure we migrated the data from AWS to azure successfully

## Project 2: AP Morgan Data Platform

## Project 2 : AP Morgan

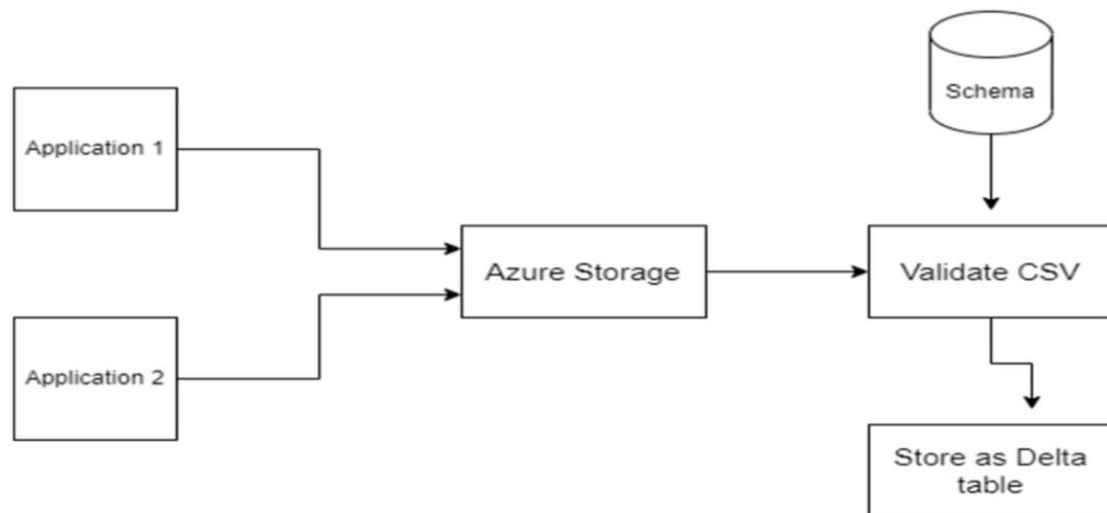
- Multiple Internal applications send the data (huge size) in CSV format on a daily basis in the cloud storage location. There are a couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed, data to be persisted as Delta table in Databricks for downstream 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: to create data bricks we have to search for data bricks and click on create button.

The screenshot shows the Microsoft Azure portal interface. At the top, there is a search bar with the text "Search resources, services, and docs (G+/)". Below the search bar, the "Home" link is visible. The main heading is "Azure Databricks", followed by a link to "Manipal Pro Learn (manipalazure.onmicrosoft.com)". Below this, there are several action buttons: "Create", "Manage view", "Refresh", "Export to CSV", "Open query", and "Assign tags". A filter bar is present with a search input "Filter for any field..." and several filter buttons: "Subscription == all", "Resource group == all", "Location == all", and "Add". At the bottom, there are sorting options for "Name" and "Type".



## Step-2: we have to create an azure data bricks workspace

[Home](#) > [Azure Databricks](#) >

### Create an Azure Databricks workspace ...

×

Basics Networking Advanced Tags Review + create

#### Project Details

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

Subscription \* ⓘ

Resource group \* ⓘ  [Create new](#)

#### Instance Details

Workspace name \*

Region \*

Pricing Tier \* ⓘ

[Review + create](#) < Previous Next: Networking >

## Step-3: after validation and click on create.

[Home](#) > [Azure Databricks](#) >

### Create an Azure Databricks workspace ...

×

✓ Validation Succeeded

Basics Networking Advanced Tags Review + create

#### Summary

##### Basics

Workspace name	databricksdemo
Subscription	Azure-DXC262AB12Lab
Resource group	dxcorg231
Region	East US 2
Pricing Tier	trial

##### Networking

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

##### Advanced

Enable Infrastructure Encryption	No
----------------------------------	----

Create

< Previous

[Download a template for automation](#)

Step-4: after clicking on create wait for deployment & deployment has been done, now click on go to resource button

✓ **Deployment succeeded**

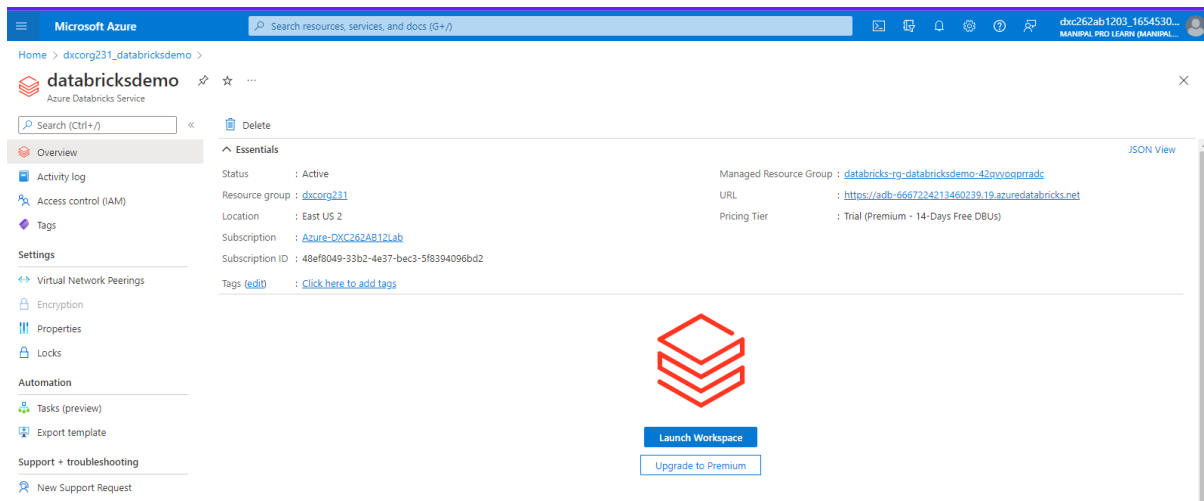
Deployment '**dxcoreg231\_databricksdemo**' to resource group '**dxcoreg231**' was successful.

[Go to resource](#)

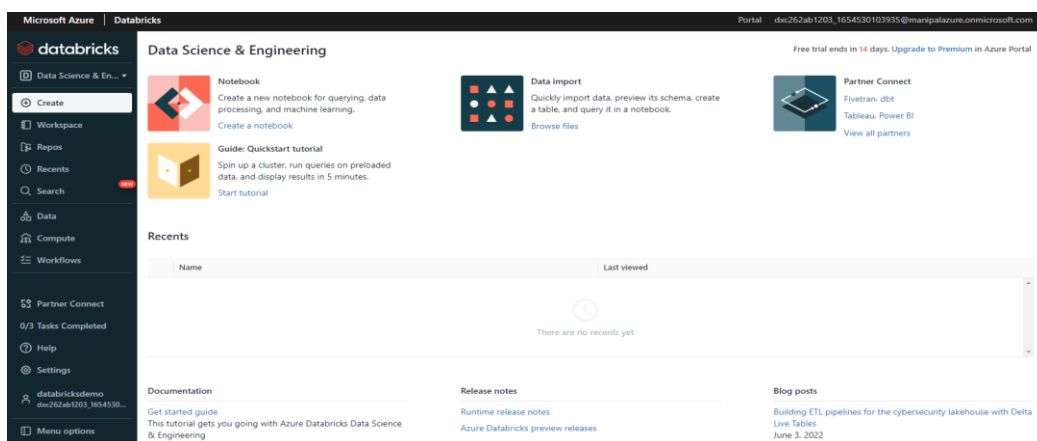
[Pin to dashboard](#)

3 minutes ago

Step-4: this page will appear and now click on launch workspace



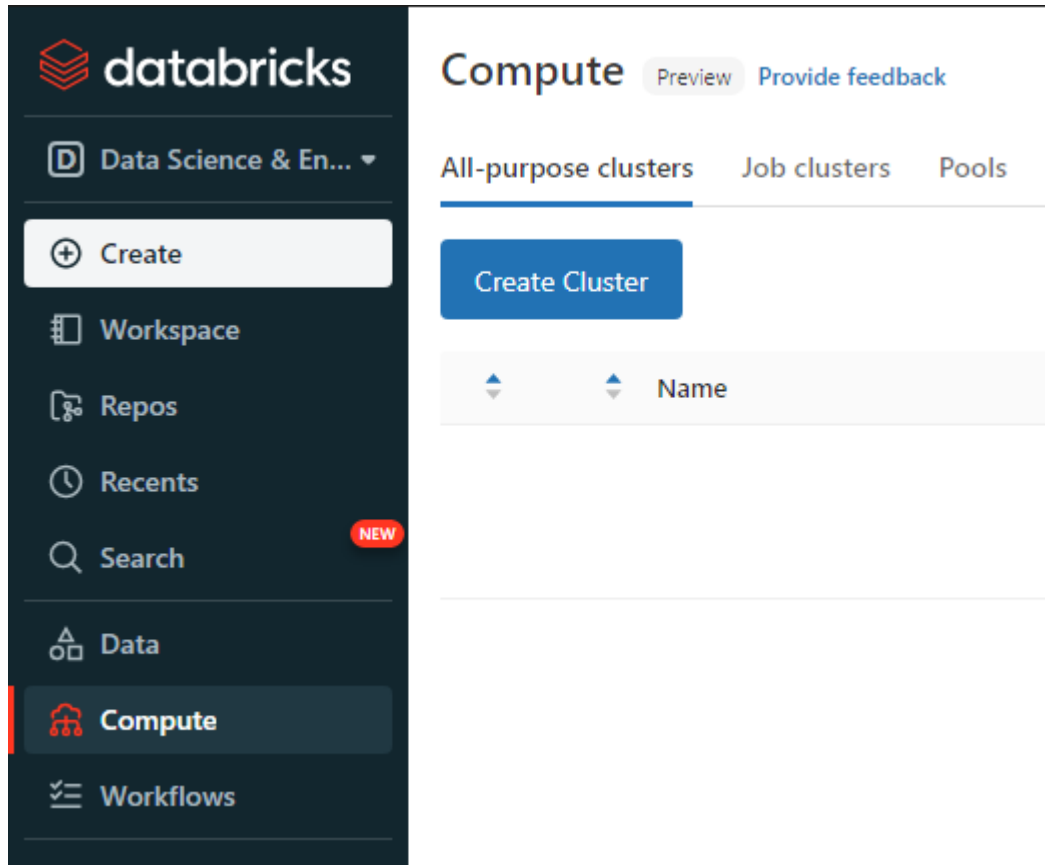
Step-5: after launching the work space we are navigated to the new tab



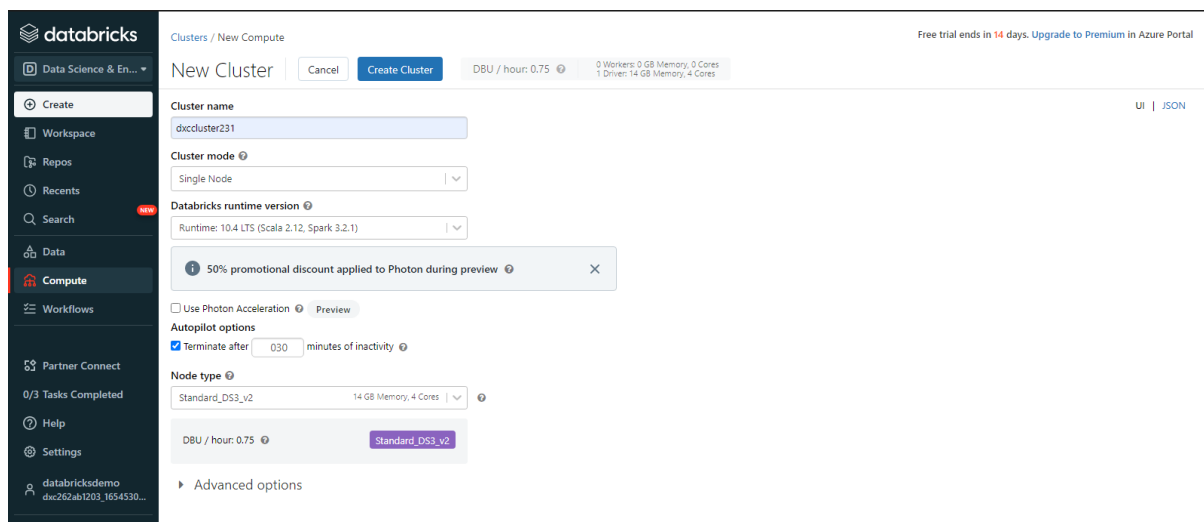
## Practical Lab: Create Cluster in Azure Databricks

To create a cluster in the data bricks we have to follow the below mentioned steps

Step-1: after navigating to the home page of azure data bricks click on compute the you can find create cluster button



Step-2: name the cluster and select cluster mode and click create



## Step-3: after that the cluster is created successfully

The screenshot displays the Databricks Clusters management interface. On the left is a dark sidebar with navigation options: Create, Workspace, Repos, Recents, Search, Data, Compute (highlighted), Workflows, Partner Connect, 1/3 Tasks Completed, Help, Settings, and a user profile section. The main content area is titled 'Clusters / dxcluster231' and includes a 'Free trial ends in 14 days. Upgrade to Premium in /' notification. Below the cluster name are tabs for Configuration, Notebooks, Libraries, Event log, Spark UI, Driver logs, Metrics, Apps, and Spark cluster UI - Master. The Configuration tab is active, showing settings for 'dxcluster231'. The settings include: Policy (Unrestricted), Cluster mode (Single Node), Databricks Runtime Version (10.4 LTS), Autopilot options (Terminate after 30 minutes), and Node type (Standard\_DS3\_v2, 14 GB Memory, 4 Cores). A DBU usage summary shows 0.75 DBU per hour. An 'Advanced options' section is partially visible at the bottom.

**databricks**

Clusters / dxcluster231

Free trial ends in 14 days. Upgrade to Premium in /

**dxcluster231**

Configuration | Notebooks | Libraries | Event log | Spark UI | Driver logs | Metrics | Apps | Spark cluster UI - Master

Policy: Unrestricted

Cluster mode: Single Node

Databricks Runtime Version: 10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)

Use Photon Acceleration: ☐ Preview

Autopilot options: ☒ Terminate after 30 minutes of inactivity

Node type: Standard\_DS3\_v2 (14 GB Memory, 4 Cores)

DBU / hour: 0.75

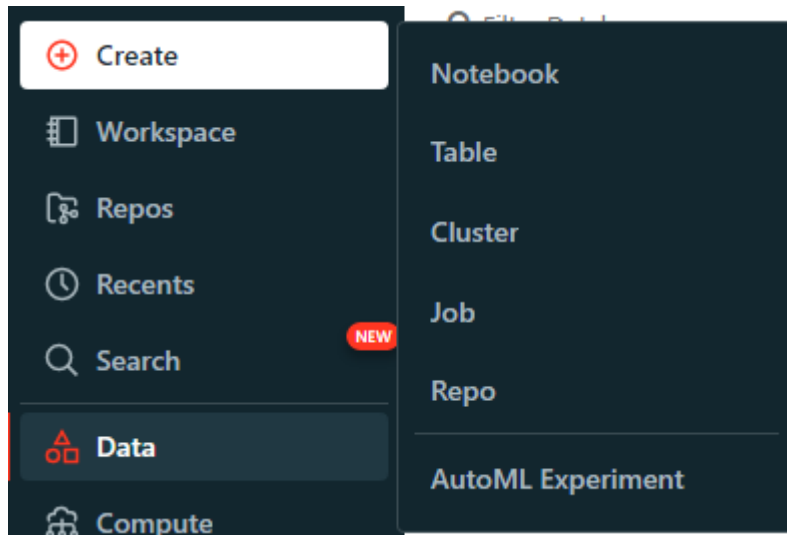
Standard\_DS3\_v2

Advanced options

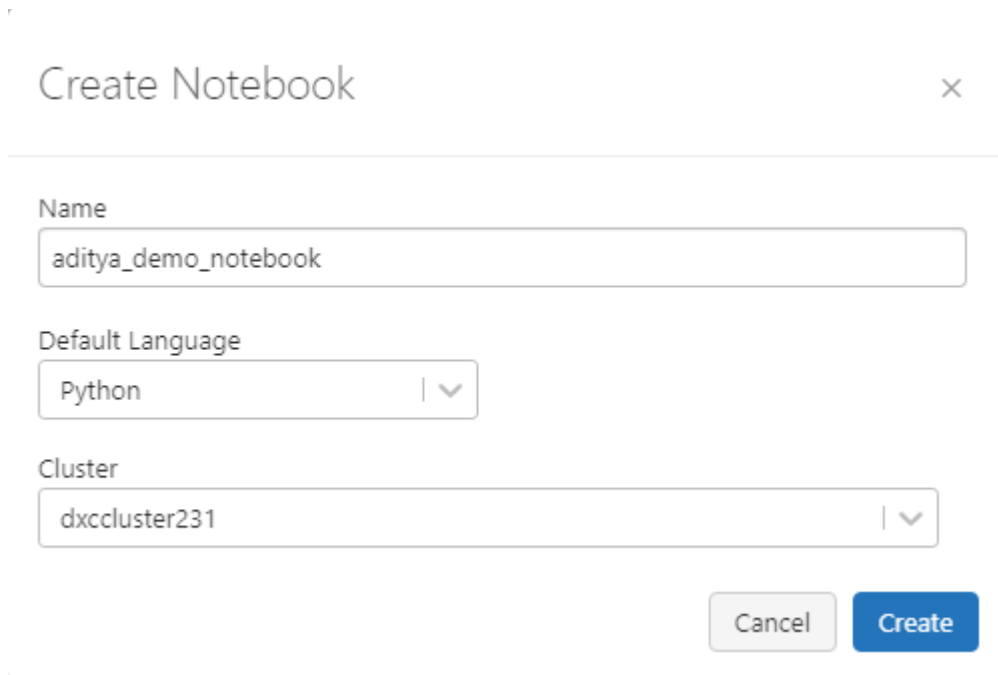
## Practical Lab: Add notebook in Databricks and Implement the Business Logic

To add notebook in data bricks we have to follow the below mentioned steps

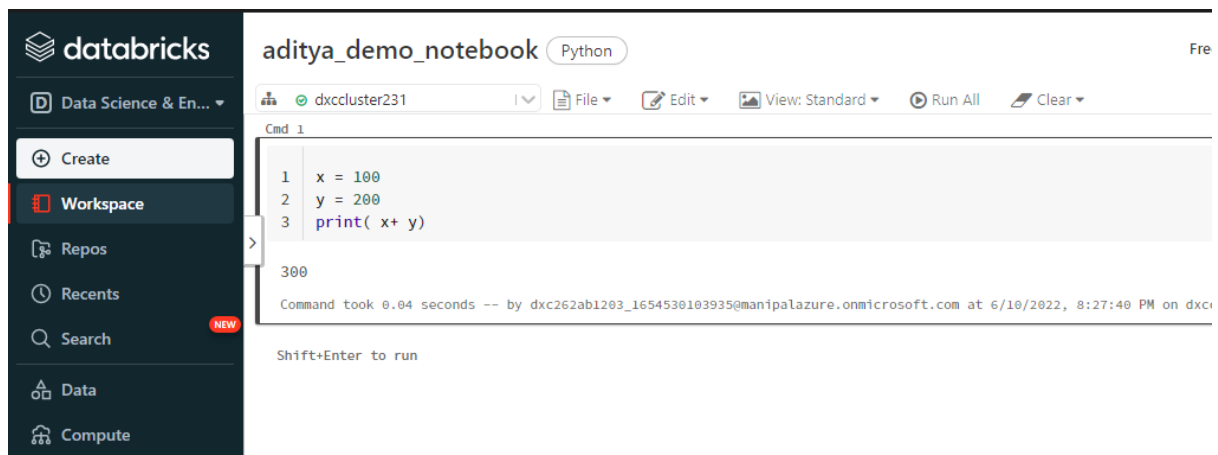
Step-1:click on create and select notebook



Step-2: name the note book, default language and cluster and click on create

A screenshot of the 'Create Notebook' dialog box in Databricks. The dialog has a title bar 'Create Notebook' with a close button (X). It contains three input fields: 'Name' with the value 'aditya\_demo\_notebook', 'Default Language' with a dropdown menu showing 'Python', and 'Cluster' with a dropdown menu showing 'dxcccluster231'. At the bottom right, there are two buttons: 'Cancel' and 'Create'.

Step-3: After clicking on create the note book is created

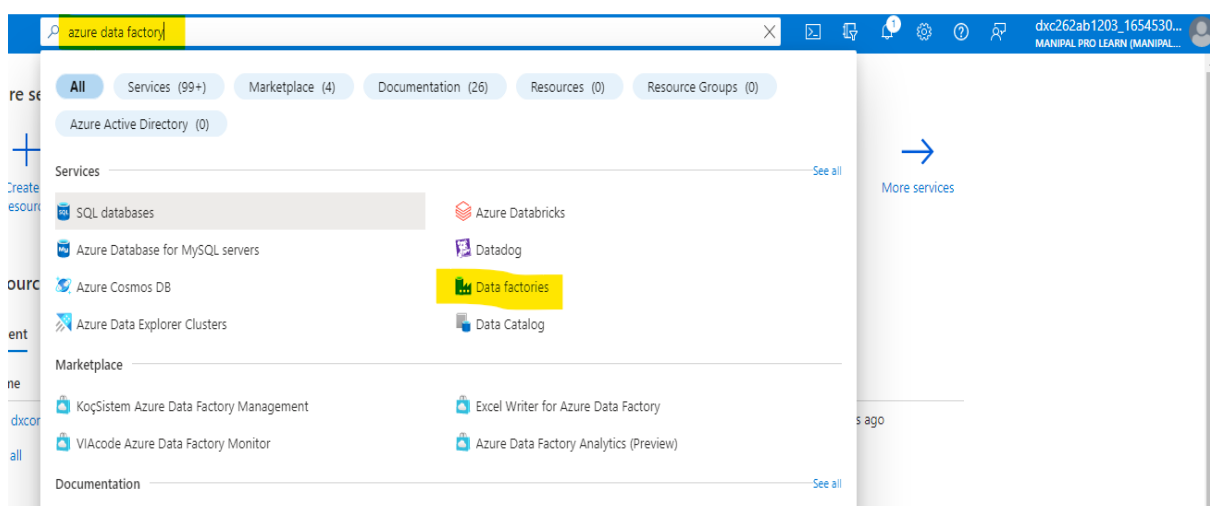


Step-4: And to implement the business logic we need to go to the partner connect and select the power bi (Business intelligence tool by Microsoft) and

After clicking on Power BI, we have to choose the cluster and download the file to connect using desktop power bi after that go to user settings and generate token and click the file downloaded and enter the access token. And we are done with connecting azure data bricks with the power BI.

## Practical Lab: Azure Data Factory for AP Morgan

**Step-1:** login to the azure portal and search for azure data factory as shown in fig



**Step-2:** after navigating to the azure data factory page click on create and select the resource group, name, region, version after that click on next follow the fig

**Basics**

Git configuration

Networking

Advanced

Tags

Review + create

**Project details**

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

Subscription \* ⓘ

Azure-DXC262AB12Lab



Resource group \* ⓘ

dxccorg231

[Create new](#)**Instance details**

Name \* ⓘ

dxccdf231

Region \* ⓘ

East US

Version \* ⓘ

V2 (Recommended)

[Review + create](#)[< Previous](#)[Next : Git configuration >](#)

**Step-3:** click on git configurations and checkbox it as configure git later as shown in fig

[Home](#) > [Data factories](#) >

## Create Data Factory ...

Basics

**Git configuration**

Networking

Advanced

Tags

Review + create

Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration.

[Learn more about Git integration in Azure Data Factory](#)

Configure Git later ⓘ



**Step-4:** Go through the next steps followed by successful completion of validation click on create as shown in

[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	dxccorg231
Name	dxcdf231
Region	East US
Version	V2 (Recommended)

### Networking

Connect via Public endpoint

 **Create**

< Previous

Next >

[Download a template for automation](#)


**Step-5:** after clicking on create it takes some time for deployment after deployment you can see the following as shown in fig

✓ Your deployment is complete



Deployment name: Microsoft.DataFactory-20220610162331  
Subscription: [Azure-DXC262AB12Lab](#)  
Resource group: [dxccorg231](#)

Start time: 6/10/2022, 4:31:58 PM

Correlation ID: 4b58f579-505b-44a3-9760-552ded084f0b 

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

∧ Next steps

[Go to resource](#)


**Step-6:** click on go to resources after that click on open as shown in the figure



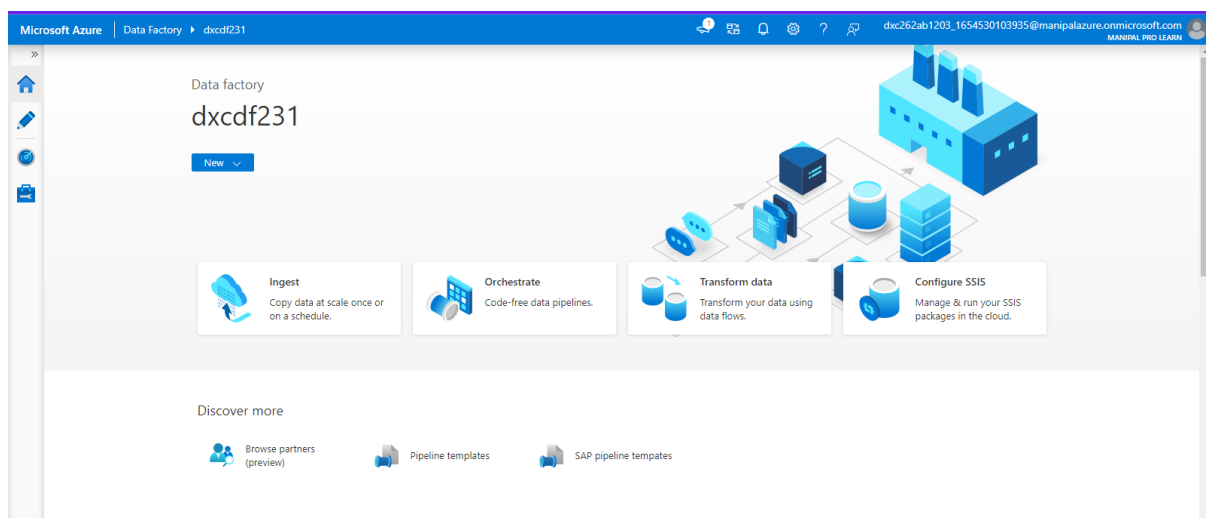
## Getting started



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

Open 

Step-7: After clicking over that it will open azure data factory in new window as mentioned in fig



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

To Create azure data bricks linked service in azure data factory we have to follow the below mentioned steps

Step-1: select integration and then select data factory

Step-2: now create azure data factory

Step-3: Now create linked services and click on new

Step-4: select compute and azure data bricks and select continue

**New linked service (Azure Databricks)**

Name \*  
AzureDatabricks\_LinkedService

Description

Connect via integration runtime \* ⓘ  
AutoResolveIntegrationRuntime

Account selection method \*  
From Azure subscription

Azure subscription \* ⓘ

Databricks workspace \* ⓘ

Select cluster  
☒ New job cluster ☐ Existing interactive cluster ☐ Existing instance pool

Databricks Workspace URL ⓘ

Authentication type \*  
Access Token

☒ Access token ☐ Azure Key Vault

Access token \* ⓘ  
\*\*\*\*\*

Cluster version \* ⓘ

Cluster node type \* ⓘ  
Standard\_D3\_v2

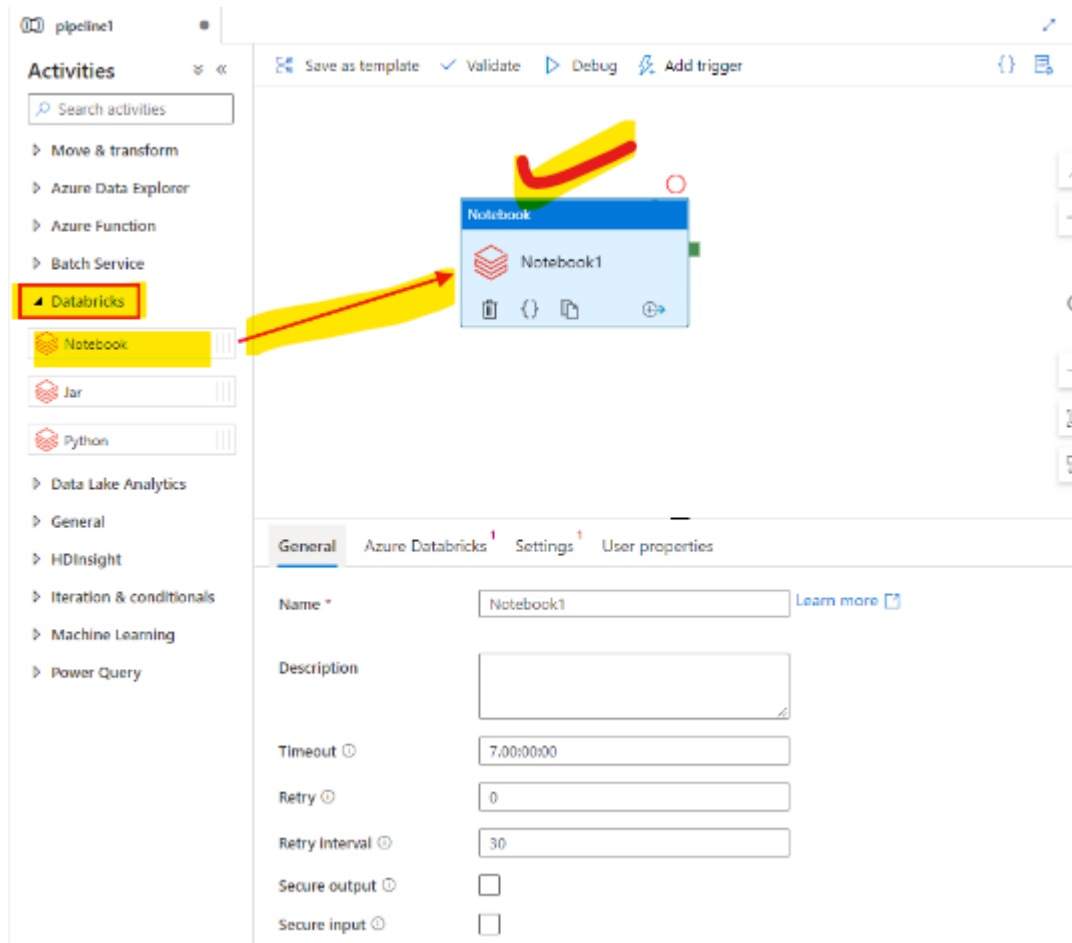
Python Version \*  
2

Worker options ⓘ  
☒ Fixed ☐ Autoscaling

Workers \*  
2

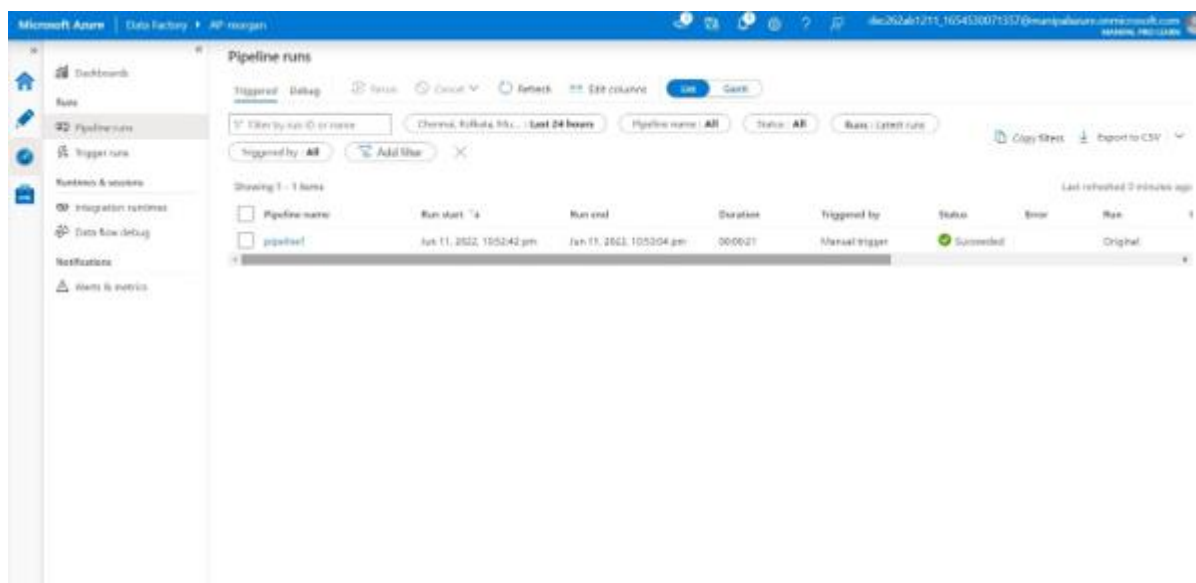
Connection successful

Step-5: create a pipe line and integrate it with the note book



Step-6: now open data bricks and click on workspace and click on create and create a note book

Step-7: now do any action in the note book and perform trigger.



**Result:** The data from the different application are stored in azure storage and then into the data bricks and we can now access the tables and the data is stored in the form of delta tables.

**Conclusion:** By using the resources like Data bricks, Clusters, Notebooks, And created azure data factory for Ap Morgan and linked Azure data bricks and azure data factory

## References:

- <https://azure.microsoft.com/en-in/blog/operationalize-azure-databricks-notebooks-using-data-factory/>
- <https://docs.microsoft.com/en-us/azure/data-factory/transform-data-using-databricks-notebook>