



DXC REALTIME PROJECTS

AZ-900, DP-900, DP - 203



JUNE 12, 2022

DXC TECHNOLOGY PVT.LTD.

Name: Mohana Likhitha Thotakura

Reg No: DXC262AB1219

Project1 Name: Smart Vehicles

Date: 12th June,2022

Project 1 : Connected Vehicles

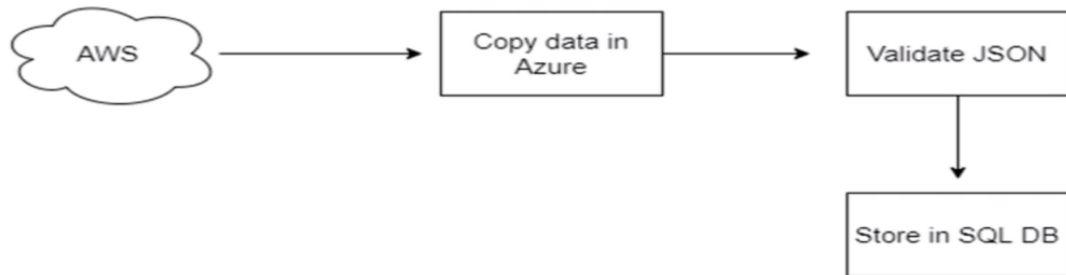
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



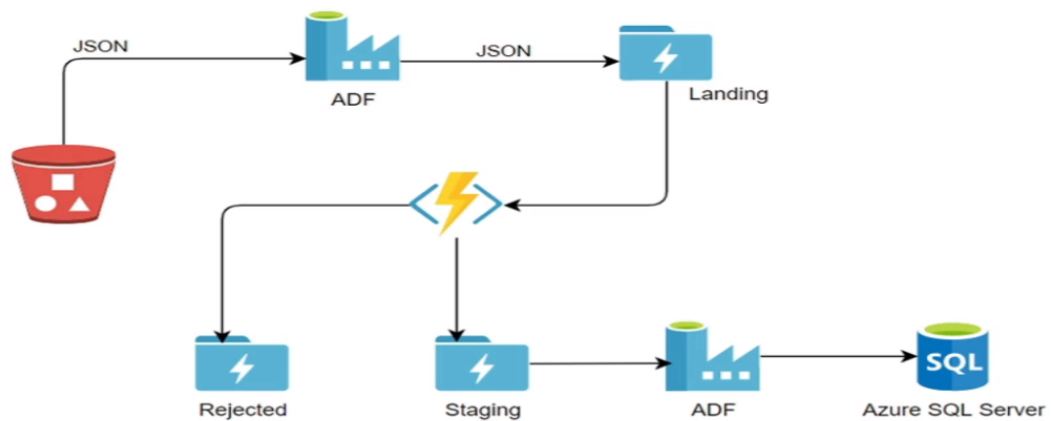
Project 1 : Connected Vehicles

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

Project 1 : Connected Vehicles



Project 1 : Connected Vehicles

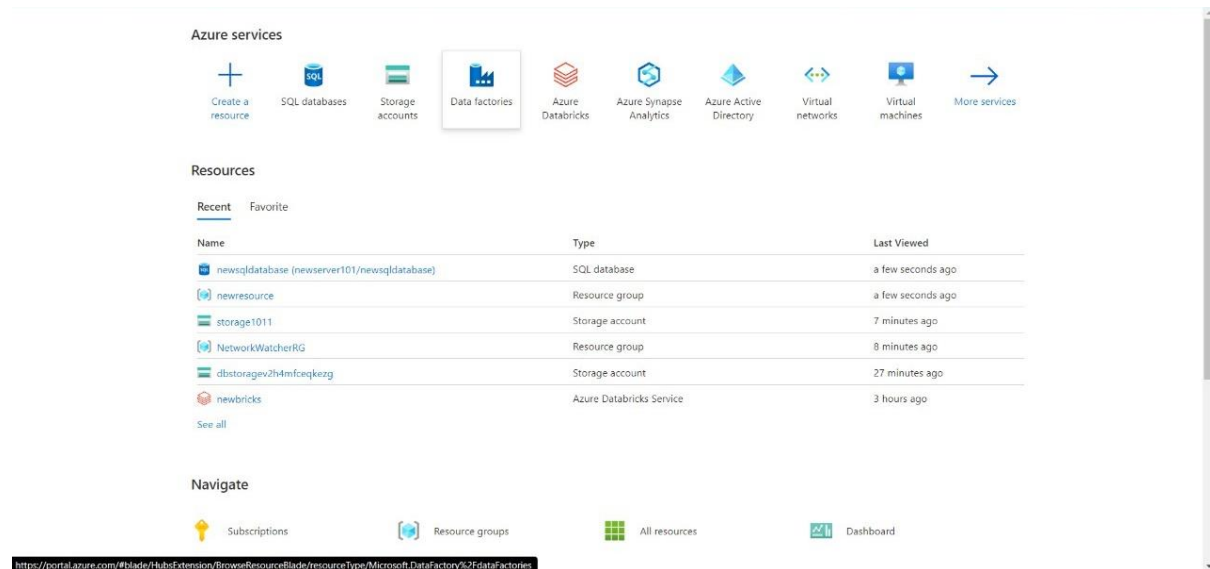


Architecture Diagram for Connected Vehicle Project

STEPS TO CREATE THE FOLLOWING ARCHITECTURE:

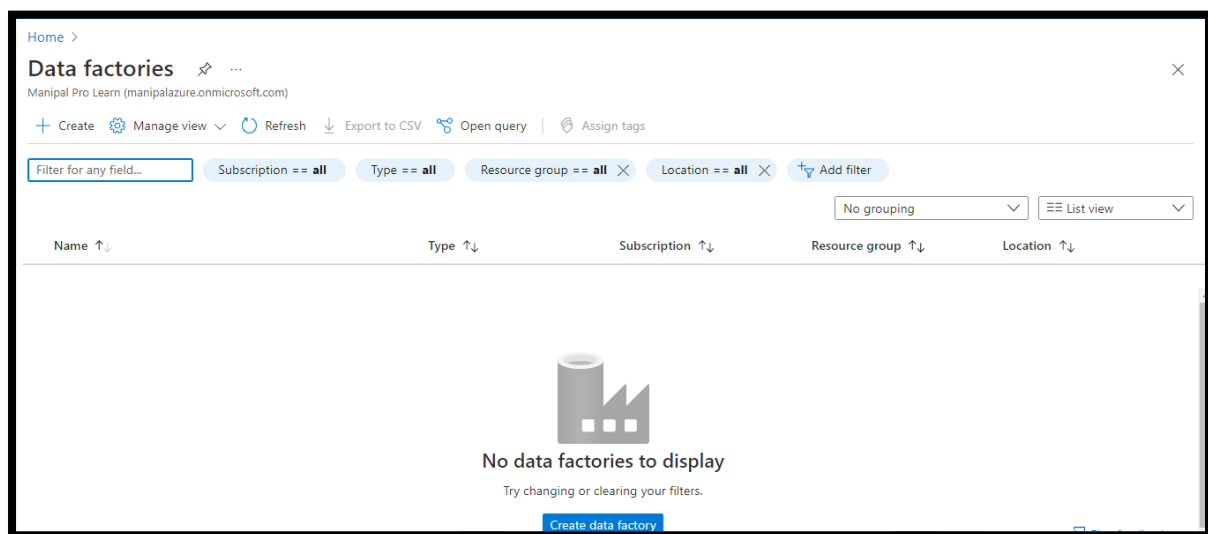
STEP-1

Open the Microsoft Azure and search for Azure data Factory:



STEP-2

Open the Azure Data Factory , Click on + Create option to create a new Data Factory account:



Enter the details for your Data factory account

Home > Data factories >

Create Data Factory

Basics | Git configuration | Networking | Advanced | Tags | Review + create

Project details

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

Subscription * ⓘ Azure-DXC262AB12Lab

Resource group * ⓘ newresource
[Create new](#)

Instance details

Name * ⓘ newdata-factory

Region * ⓘ East US

Version * ⓘ V2 (Recommended)

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

After entering the details move towards validating the fields and Click on Create when successfully validated.

Home > Data factories >

Create Data Factory

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

✓ Validation Passed

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

Basics

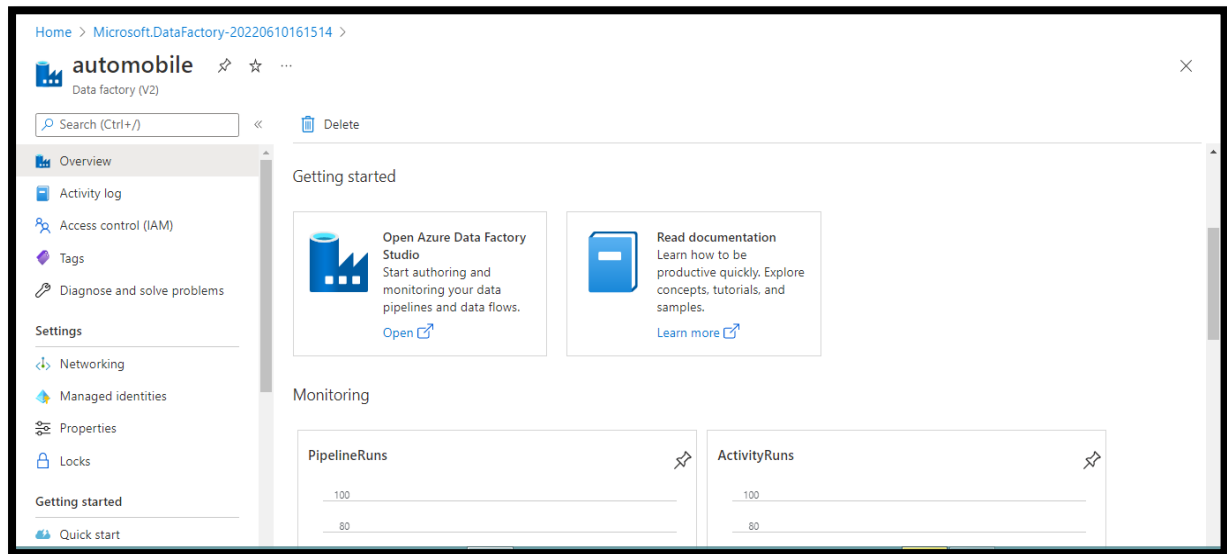
Subscription	Azure-DXC262AB12Lab
Resource group	DXCPROJ
Name	automobile
Region	East US
Version	V2 (Recommended)

Networking

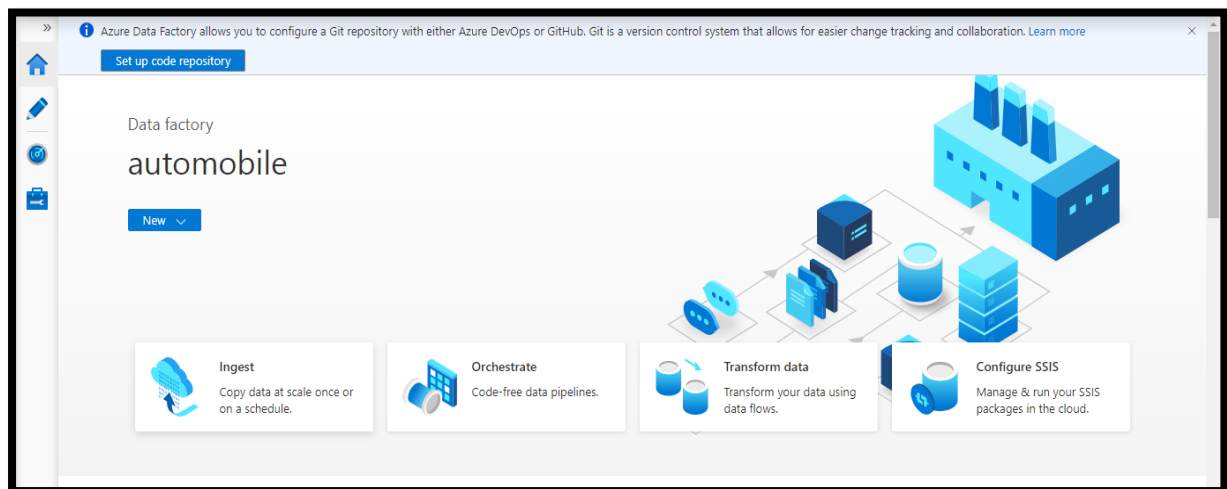
Connect via	Public endpoint
-------------	-----------------

STEP-3

After deployment is finished, click on **Go to Resource Group**.

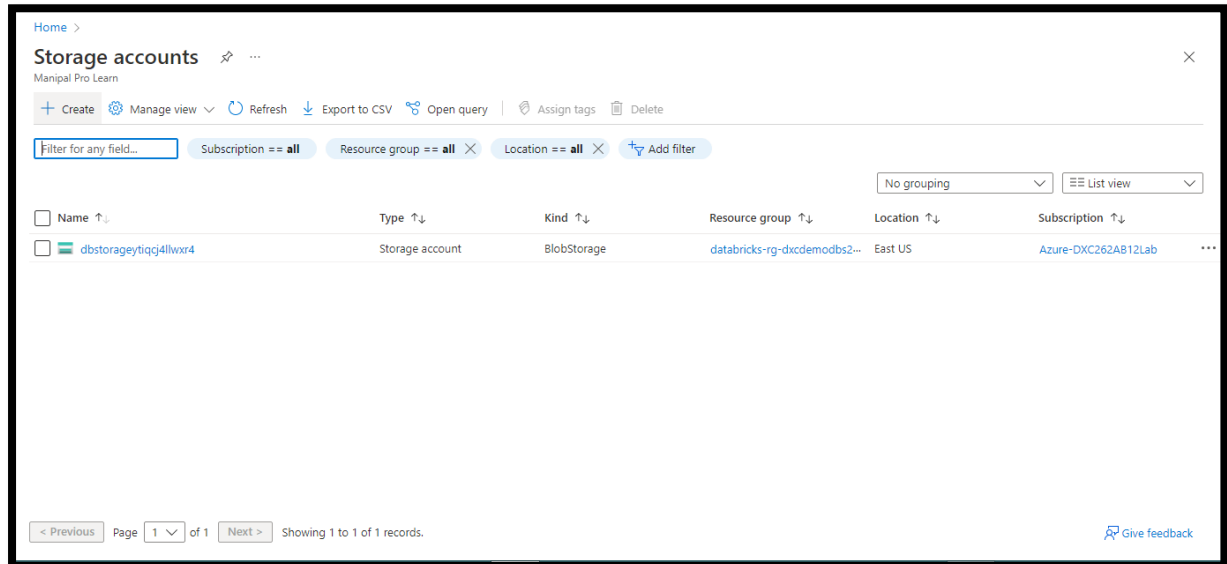


Click on Open Azure Data Factory.

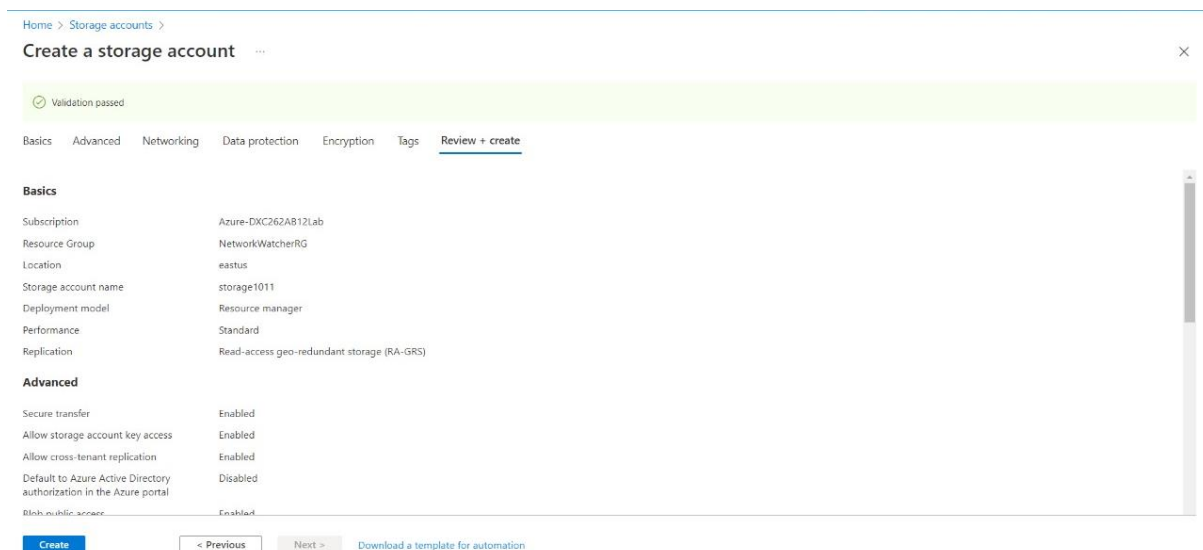


STEP-4

Go to Azure , search for storage. Click on Storage Account. Click on **+ Create**.



Fill up all the details and click on **Review+Create**. And, deploy it.



Once deployment is done click on Resource Group.

Home > storage1011_1654861967486 | Overview

Deployment

Search (Ctrl+/) << Delete Cancel Redeploy Refresh

We'd love your feedback →

✓ Your deployment is complete

Deployment name: storage1011_1654861967486 Start time: 6/10/2022, 5:23:21 PM
Subscription: Azure-DXC262AB12Lab Correlation ID: a0b808f9-9039-446b-81c2-976315b04c71
Resource group: NetworkWatcherRG

Deployment details (Download)

Next steps

Go to resource

Cost Management
Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >

Microsoft Defender for Cloud
Secure your apps and infrastructure
Go to Microsoft Defender for Cloud >

Free Microsoft tutorials
Start learning today >

Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
Find an Azure expert >

STEP-5

Navigate towards containers. Click on + Container to create a container.

Home > storage1011_1654861967486 > storage1011

storage1011 | Containers

Storage account

Search (Ctrl+/) << + Container Change access level Restore containers Refresh Delete

Search containers by prefix Show deleted containers

Name	Last modified	Public access level	Lease state	
\$logs	6/10/2022, 5:23:53 PM	Private	Available	...

Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser (preview)

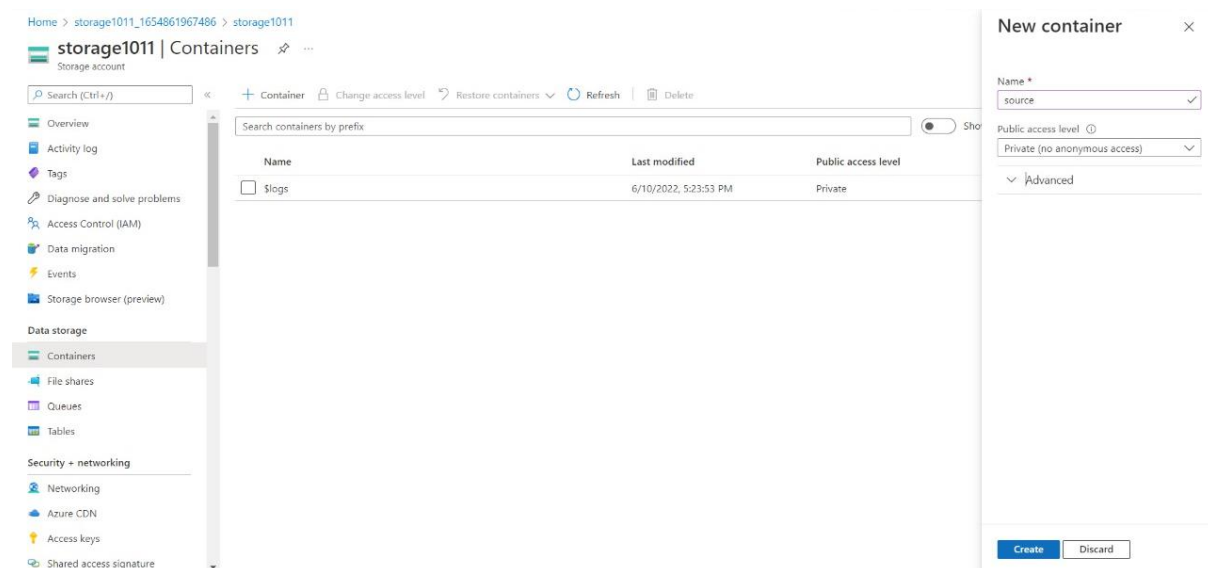
Data storage

Containers
File shares
Queues
Tables

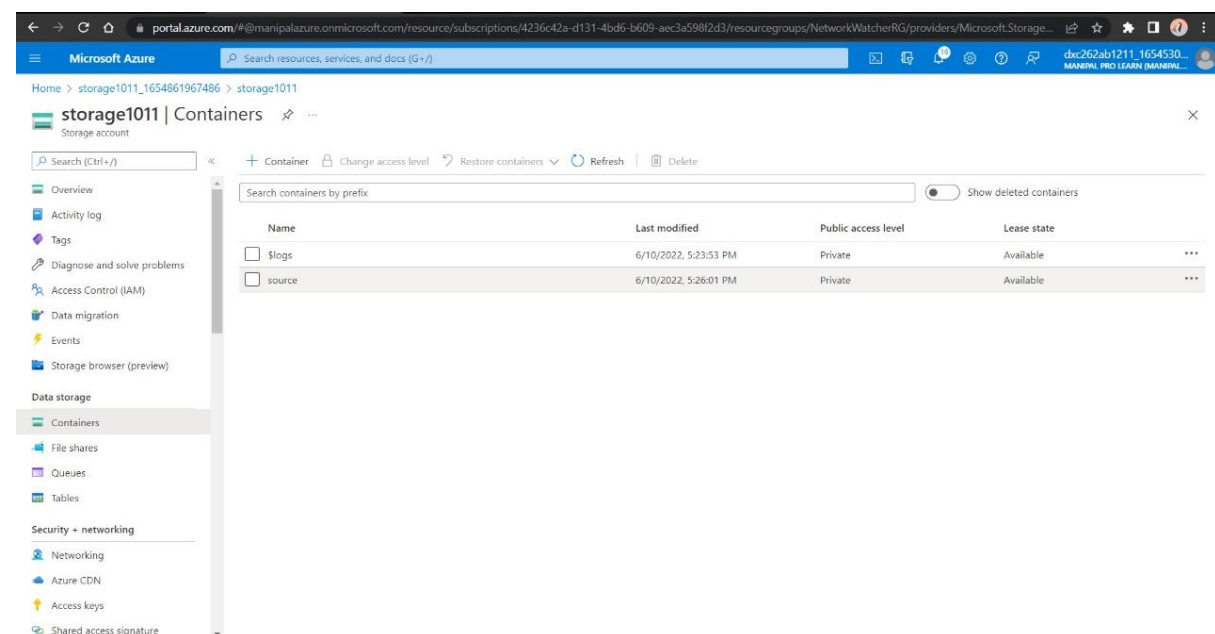
Security + networking

Networking
Azure CDN
Access keys
Shared access signature

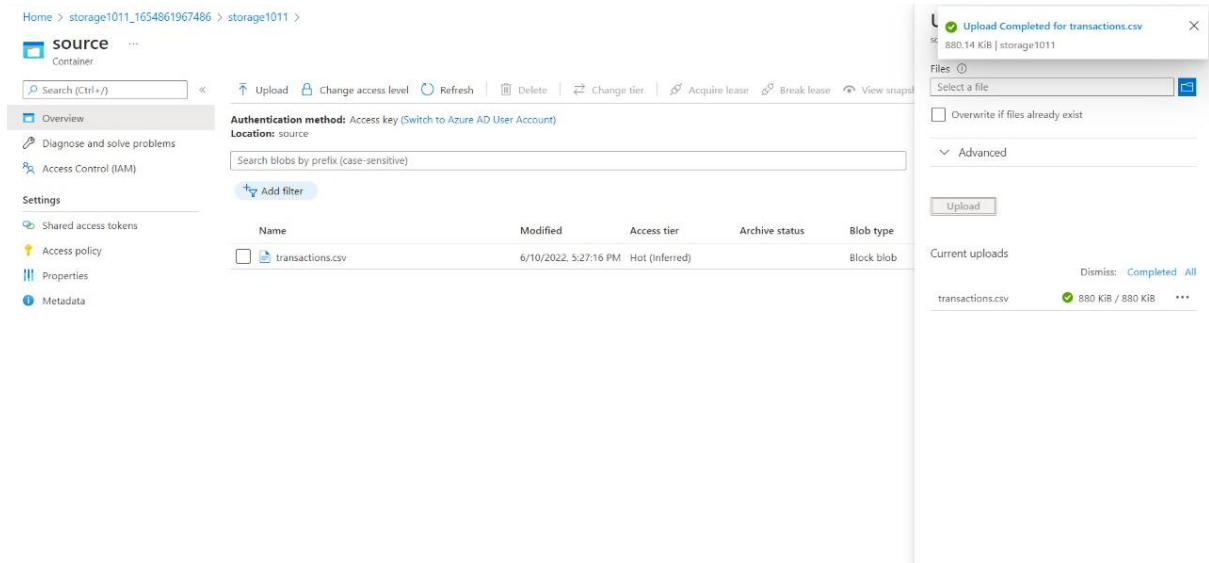
Container is used to store the data. Click create.



Give a name to the container Click on upload.



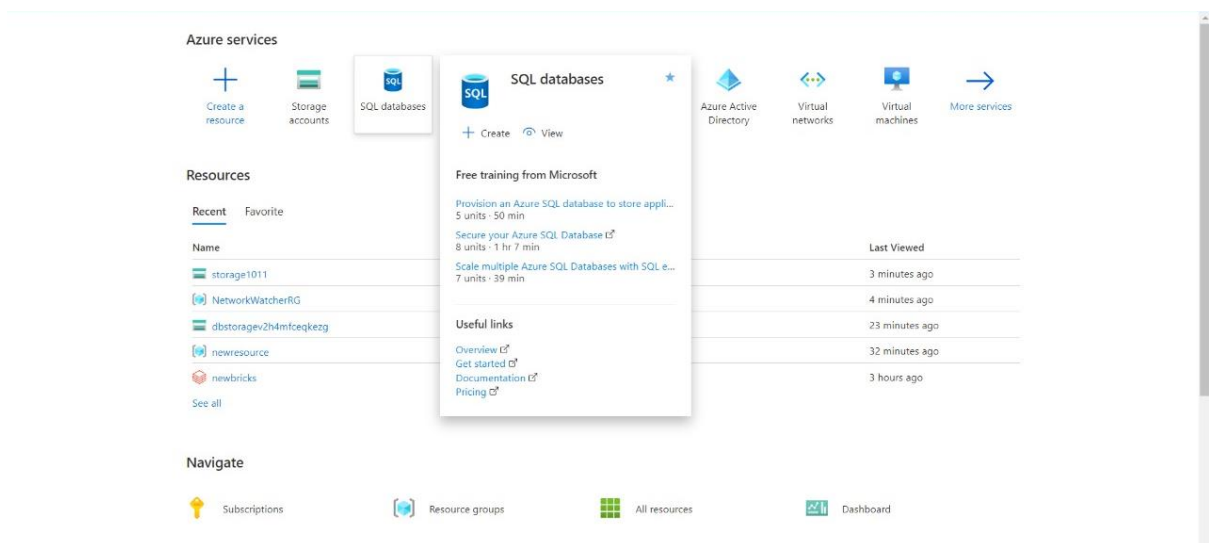
Since, we don't have the source of the data we will make this container as the source of our blob data and will fill in a sample file to paste into the SQL database.



STEP-6

Now we need a SQL database to send the data to. Go to **SQL DATABASES**.

Open it. Fill the necessary details and deploy it.



Home > SQL databases >

Create SQL Database

Microsoft

⚠ Changing Basic options may reset selections you have made. Review all options prior to creating the resource.

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 *

General Purpose
Gen5, 2 vCores, 32 GB storage, zone redundant disabled
[Configure database](#)

Backup storage redundancy

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

Make sure to enable the firewall for the current IP and for the same azure cloud otherwise connecting SQL database through Data Factory is not possible.

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 "newsrserver101" 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 * ☐ No ☒ Yes

Private endpoints

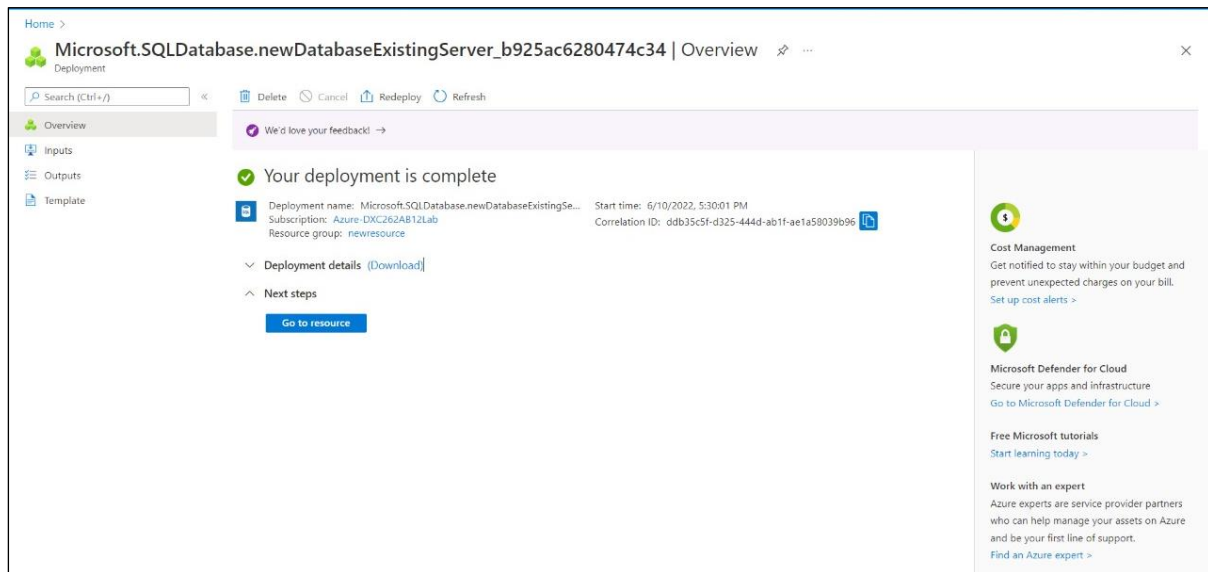
Private endpoint connections are associated with a private IP address within a Virtual Network. The list below shows all the private endpoint connections for this server. Note that private endpoint connections are defined at the server level and they provide access to all databases in the server. [Learn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

[+ Add private endpoint](#)

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

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

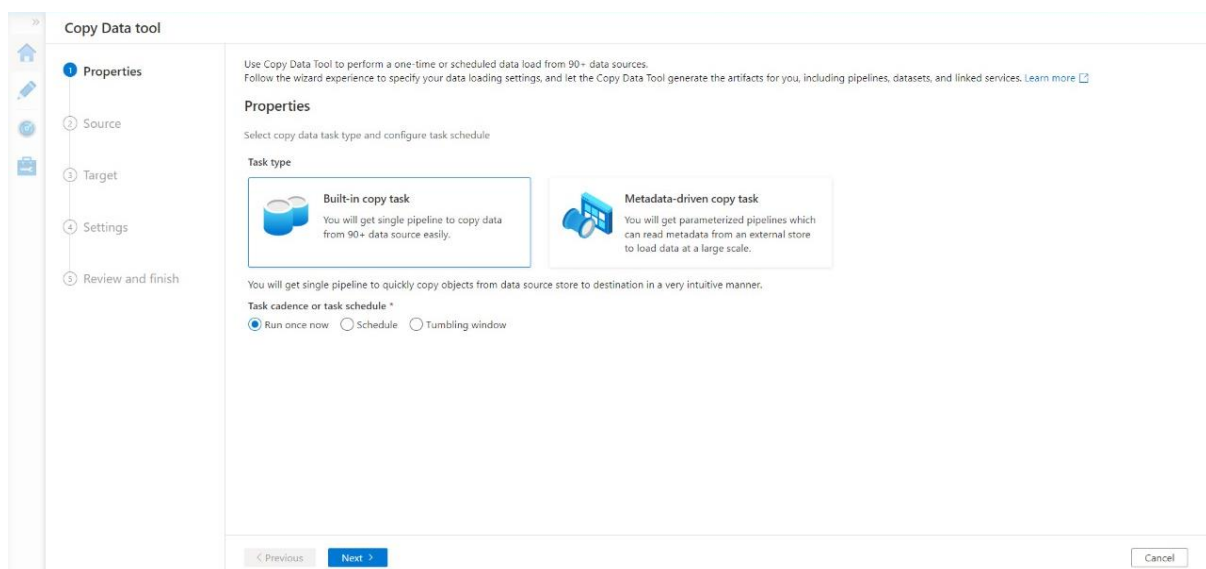
Go ahead to deploy the database



STEP - 7

After these steps come back to **Azure Data Factory Lab**. Here we need to create a pipeline that will take data from blob storage and feed into the SQL database.

Click on ingest



Enter the name of Azure blob storage linked file. Select the name of storage account name we created in this and click on Create.

The screenshot shows the 'New linked service' dialog in the 'Copy Data tool'. The 'Source data store' section on the left has 'Source type' set to 'Azure Blob Storage' and 'Connection *' set to 'select...'. The 'New linked service' section on the right has 'Name *' set to 'AzureBlobStorage1', 'Description' is empty, 'Connect via integration runtime *' is set to 'AutoResolveIntegrationRuntime', 'Authentication type' is 'Account key', 'Account selection method' is 'From Azure subscription', 'Azure subscription' is 'Azure-DXC262A812Lab (4236c42a-d131-4bd6-b609-aec3a598f2d3)', and 'Storage account name *' is 'storage1011'. There are 'Create' and 'Cancel' buttons at the bottom right, and a 'Test connection' link.

Select the source type as blob storage and connection name that we created in previous step. Click Recursively from Options. Click Next.

The screenshot shows the 'Source' step in the 'Copy Data tool'. The 'Source data store' section has 'Source type' set to 'Azure Blob Storage' and 'Connection *' set to 'AzureBlobStorage1'. The 'File or folder *' section has 'source/transactions.csv' entered. The 'Options' section has 'Binary copy' unchecked, 'Recursively' checked, and 'Enable partition discovery' unchecked. The 'Max concurrent connections' field is empty. The 'Filter by last modified' section has 'Start time (UTC)' and 'End time (UTC)' fields. There are 'Previous', 'Next', and 'Cancel' buttons at the bottom.

Choose the output format of the data

The screenshot shows the 'Copy Data tool' configuration window. On the left is a vertical navigation pane with steps: Properties, Source, Dataset, Configuration, Target, Settings, and Review and finish. The 'File format settings' tab is active. It contains the following fields: 'File format' (DelimitedText), 'Column delimiter' (Comma (,)), 'Row delimiter' (Default (\n,\r, or \r\n)), and a checked 'First row as header' option. There are also buttons for 'Detect text format' and 'Preview data'. Below these is an 'Advanced' section with 'Compression type' (None) and 'Additional columns' (+ New). At the bottom are 'Previous', 'Next', and 'Cancel' buttons.

Now, in the next step we will link the Data Factory with SQL database. Connect the database with new connection.

The screenshot shows the 'Copy Data tool' configuration window with two tabs visible. The 'Destination data store' tab is active, showing 'Target type' as 'Azure SQL Database' and a 'Connection' dropdown. The 'New linked service' tab is also visible, showing fields for 'Account selection method' (From Azure subscription), 'Azure subscription' (Azure-DXC262AB12Lab), 'Server name' (newsrver101), 'Database name' (newsqldatabase), 'Authentication type' (SQL authentication), 'User name' (user), and 'Password'. It also has 'Always encrypted' and 'Additional connection properties' sections. At the bottom are 'Create', 'Cancel', and 'Test connection' buttons.

Choose the file name you want to see in the SQL database. Click Next.

The screenshot shows the 'Copy Data tool' configuration interface. On the left, a vertical sidebar contains icons for home, settings, and a list of steps: Properties, Source, Target, Dataset, Configuration, Settings, and Review and finish. The 'Target' step is currently selected. The main panel is titled 'Destination data store' and includes the instruction: 'Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.' Below this, there are two dropdown menus: 'Target type' set to 'Azure SQL Database' and 'Connection *' set to 'AzureSqlDatabase1'. To the right of the connection dropdown are 'Edit' and '+ New connection' links. A visual mapping shows 'Source' as 'Azure Blob Storage file' and 'Target' as 'transaction' (with a 'Use existing table' link) pointing to 'Azure Blob Storage file (auto-create)'. At the bottom, there is a checkbox for 'Skip column mapping for all tables' and navigation buttons: '< Previous', 'Next >', and 'Cancel'.

Now we are the end stage of creating pipeline, click next.

The screenshot shows the 'Copy Data tool' configuration interface at the 'Settings' step. The left sidebar is the same as the previous screen, but now the 'Settings' step is selected. The main panel is titled 'Settings' with the instruction: 'Enter name and description for the copy data task, more options for data movement'. It contains a 'Task name *' field with the value 'CopyPipeline_xtw', a 'Task description' text area, and a 'Data consistency verification' radio button which is currently selected. Below these are 'Fault tolerance' (a dropdown menu), 'Enable logging' (checkbox), and 'Enable staging' (checkbox). An '> Advanced' link is at the bottom of the settings section. At the very bottom of the panel are navigation buttons: '< Previous', 'Next >', and 'Cancel'.

Validate the data, connection , details and move forwards with Next

Copy Data tool

Summary

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

Properties

Task name: CopyPipeline_zbw

Task description:

Source:

Connection name: AzureBlobStorage1

Dataset name: SourceDataset_zbw

Column delimiter: ;

Escape character: \

Quote char: "

First row as header: true

File name: transactions.csv

Container: source

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

Trigger the pipeline manually and navigate towards monitor , Here we can see our pipeline has successfully ran once.

Pipeline runs

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

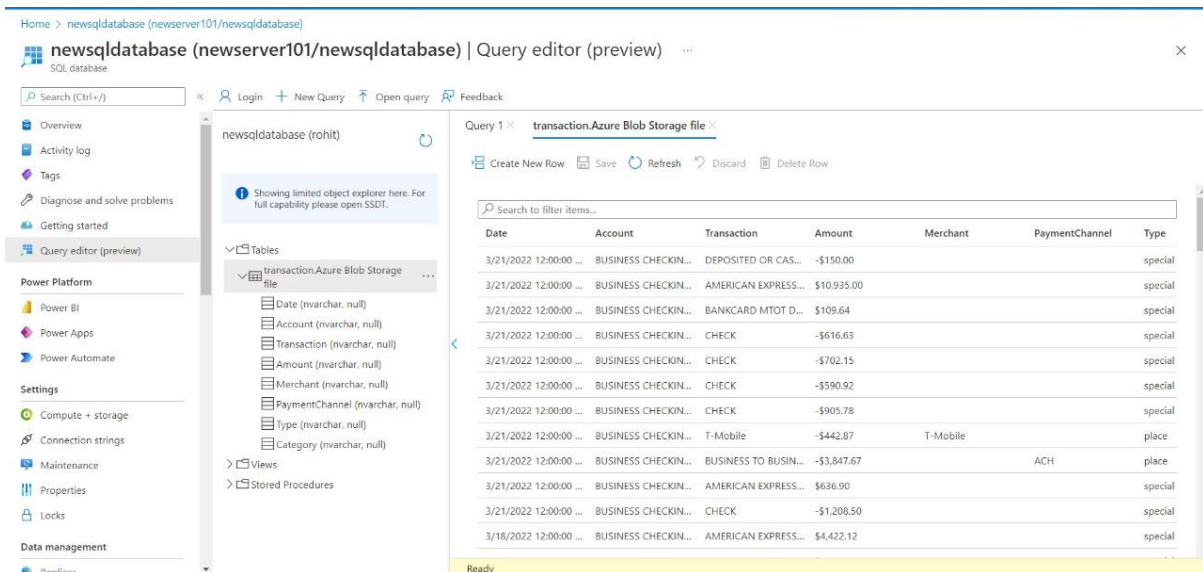
Filter by run ID or name Chennai, Kolkata, Mu... Last 24 hours Pipeline name: All Status: All Runs: Latest runs

Triggered by: All Add filter X

Showing 1 - 1 items Last refreshed 0 minutes ago

<input type="checkbox"/>	Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error	Run
<input type="checkbox"/>	CopyPipeline_zbw	Jun 10, 2022, 5:41:49 pm	Jun 10, 2022, 5:42:07 pm	00:00:17	Manual trigger	Succeeded		Original

Navigate towards SQL database to check the data.



Home > newsqldatabase (newserver101/newsqldatabase)

newsqldatabase (newserver101/newsqldatabase) | Query editor (preview)

Search (Ctrl+J)

newsql database

Search (Ctrl+J)

newsql database (rohit)

Showing limited object explorer here. For full capability please open SSDT.

Tables

- transaction.Azure Blob Storage file
 - Date (nvarchar, null)
 - Account (nvarchar, null)
 - Transaction (nvarchar, null)
 - Amount (nvarchar, null)
 - Merchant (nvarchar, null)
 - PaymentChannel (nvarchar, null)
 - Type (nvarchar, null)
 - Category (nvarchar, null)

Views

Stored Procedures

Query 1: transaction.Azure Blob Storage file

Create New Row Save Refresh Discard Delete Row

Search to filter items...

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

Ready

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

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

Name: Mohana Likhitha Thotakura

Reg No: DXC262AB1219

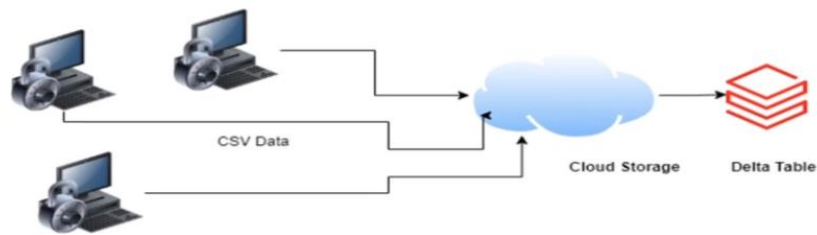
Project2 Name: AP Morgan Data Platform

Date: 12th June 2022

Project 2: AP Morgan Data Platform

Project 2 : AP Morgan

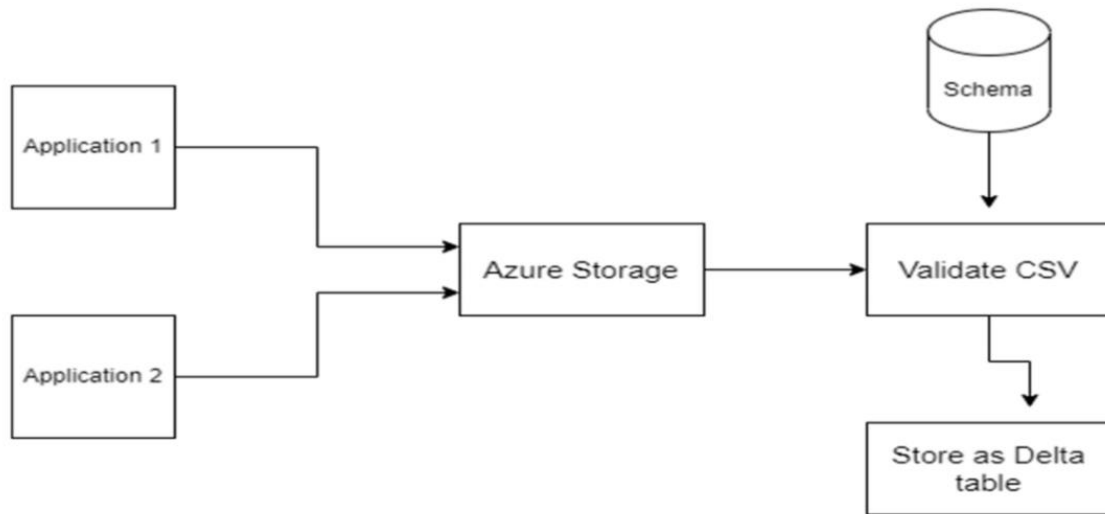
- Multiple Internal applications sends the data(huge size) in CSV format on daily basis in the cloud storage location. There are couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed data to be persisted as Delta table in Databricks for downstream system.



Project 2 : AP Morgan- High Level Detail

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

Project 2 : AP Morgan



Practical Lab: Create a **Databricks**

Practical Lab: Create **Cluster** in **Azure Databricks**

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

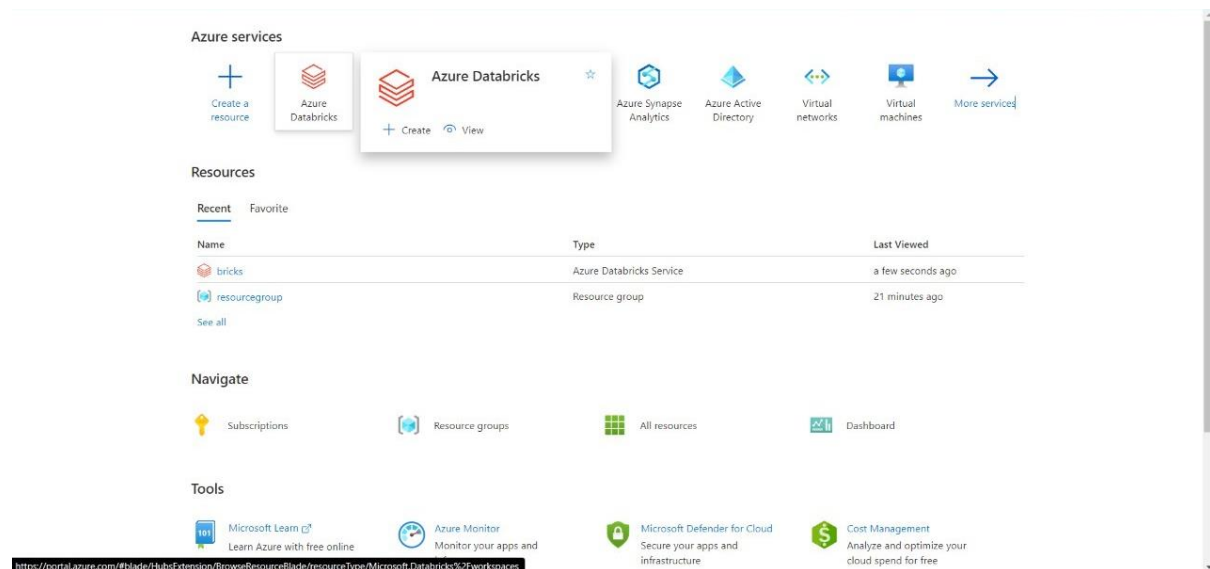
Practical Lab: **Azure Data Factory** for AP Morgan

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

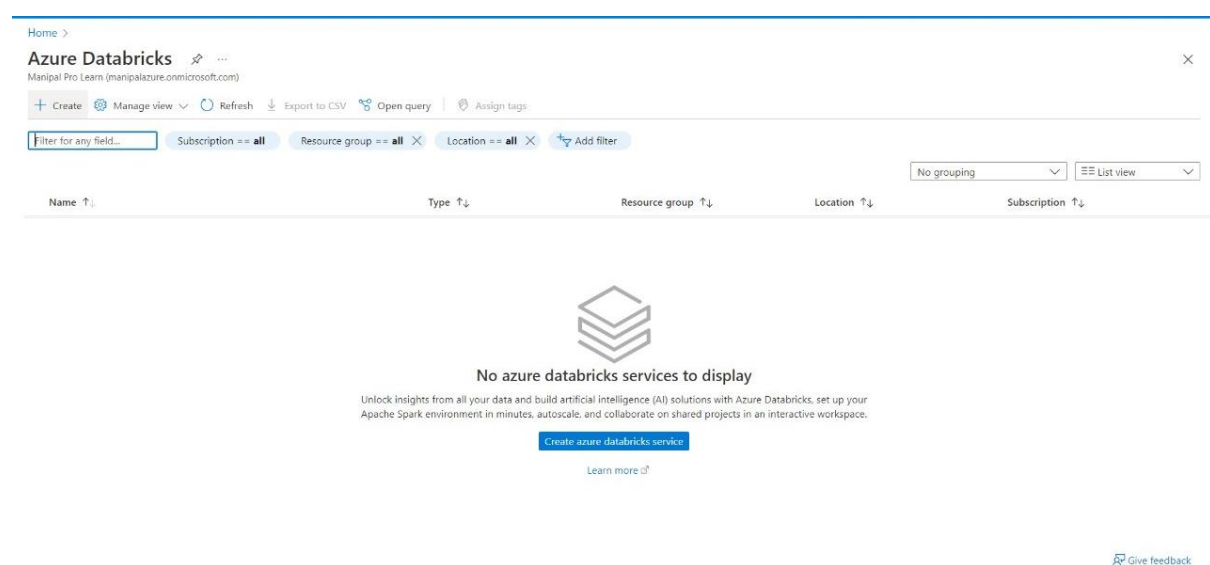
THESE ARE THE REQUIRES STEPS TO ACCOMPLISH THE ABOVE TASKS:

Step-1: Open Microsoft azure and create account for Data Factory and Azure Data Bricks

Search and open Azure Data Bricks



Step-2: Now create a new account into the Data Bricks



Step-3: Enter the account details

[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 *	<input type="text" value="Azure-DXC262AB12Lab"/>
Resource group *	<input type="text" value="resourcegroup"/>
	Create new

Instance Details

Workspace name *	<input type="text" value="bricks101"/>
Region *	<input type="text" value="East US"/>
Pricing Tier *	<input type="text" value="Standard (Apache Spark, Secure with Azure AD)"/>

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

Step-4: Make sure to give permission for Cluster connectivity

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

Create an Azure Databricks workspace ...

×

[Basics](#) [Networking](#) [Advanced](#) [Tags](#) [Review + create](#)

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP) [?](#)

☒ Yes ☐ No

Deploy Azure Databricks workspace in your own Virtual Network (VNet) [?](#)

☐ Yes ☒ No

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

Step-5: Validate and create a DataBricks account

Home > Azure Databricks >

Create an Azure Databricks workspace

Validation Succeeded

Basics Networking Advanced Tags Review + create

Summary

Basics

Workspace name	bricks101
Subscription	Azure-DXC262AB12Lab
Resource group	resourcegroup
Region	East US
Pricing Tier	standard

Networking

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

Advanced

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

Create < Previous Download a template for automation

*** Initializing deployment...
Initializing template deployment to resource group 'resourcegroup'.

Step-6: After the deployment is done Navigate towards “Go to Resource”

Home >

resourcegroup_bricks101 | Overview

Deployment

Search (Ctrl+J) Delete Cancel Redeploy Refresh

Overview Inputs Outputs Template

We'd love your feedback →

Your deployment is complete

Deployment name: resourcegroup_bricks101
Subscription: Azure-DXC262AB12Lab
Resource group: resourcegroup

Start time: 6/11/2022, 10:26:04 PM
Correlation ID: 89c0e208-0dcd-4273-9799-a00365c96492

Deployment details (Download)
Next steps
[Go to resource](#)

Deployment succeeded
Deployment 'resourcegroup_bricks101' to resource group 'resourcegroup' was successful.
[Go to resource](#) [Pin to dashboard](#)

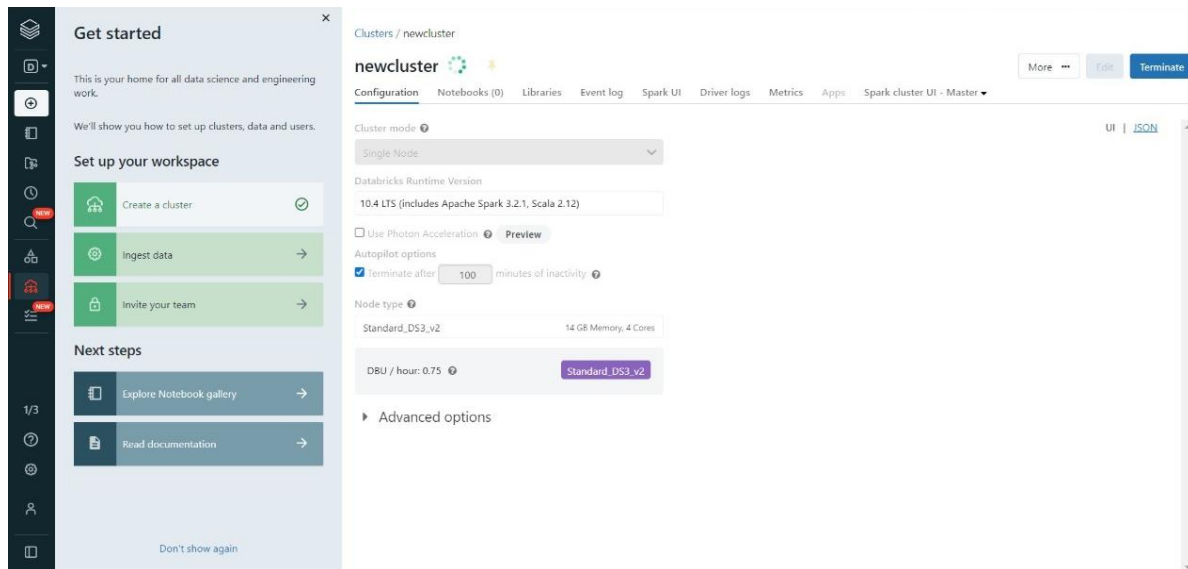
Cost Management
Get notified to stay within your budget and prevent unexpected charges on your bill.
[Set up cost alerts >](#)

Microsoft Defender for Cloud
Secure your apps and infrastructure
[Go to Microsoft Defender for Cloud >](#)

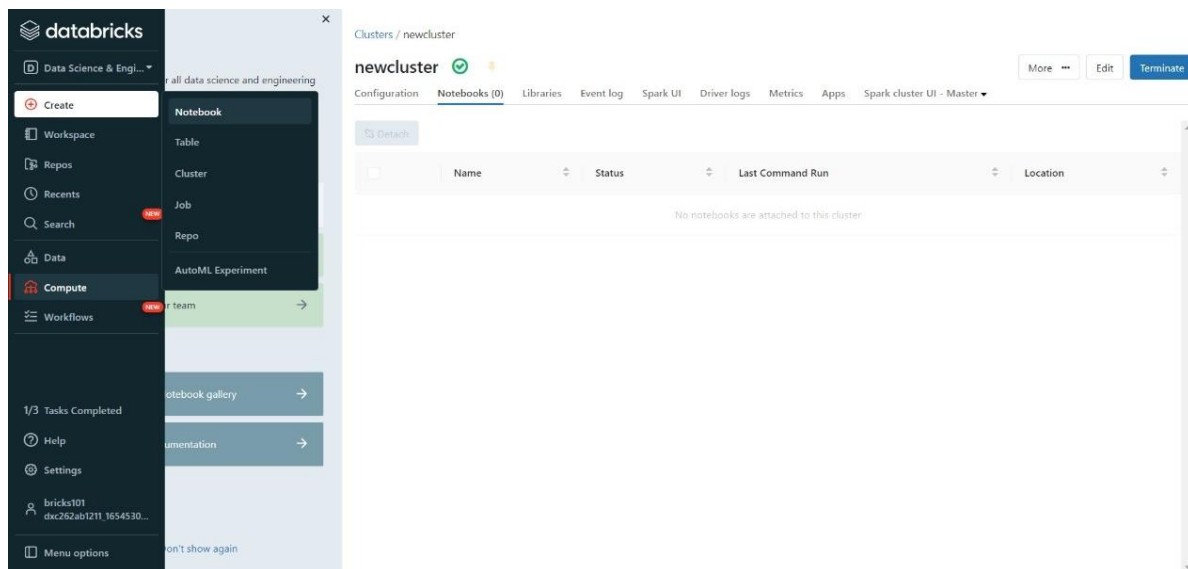
Free Microsoft tutorials
[Start learning today >](#)

Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
[Find an Azure expert >](#)

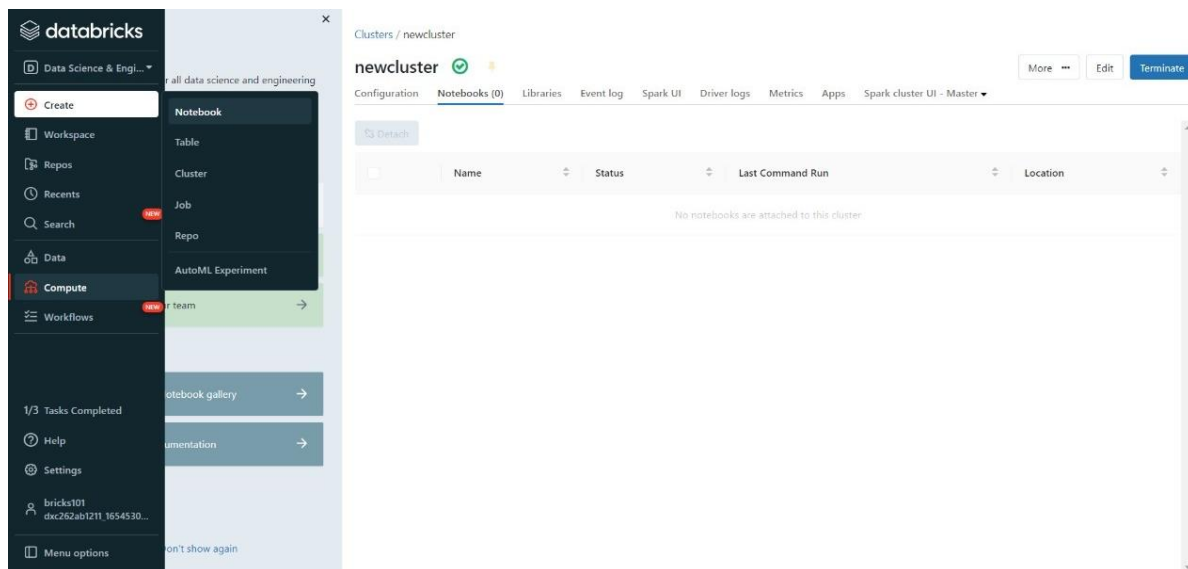
Step-8: Create a new cluster to work upon



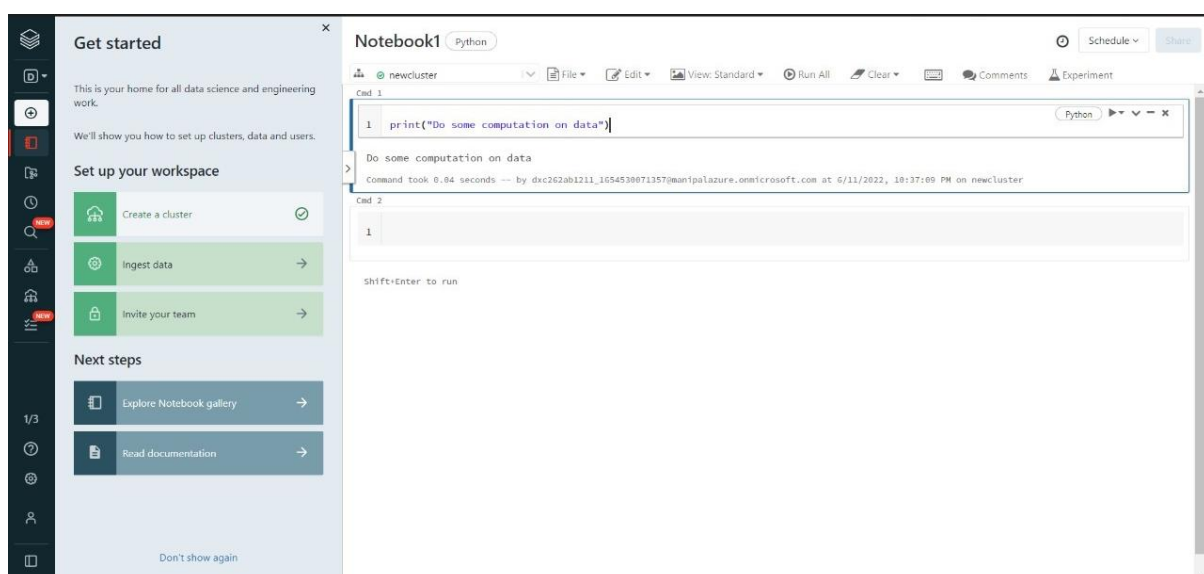
Step-9: Give the necessary credentials for the cluster and click on create



Step-10: After creating a new cluster, we need a notebook that will do some computation with the data present or incoming in databricks.

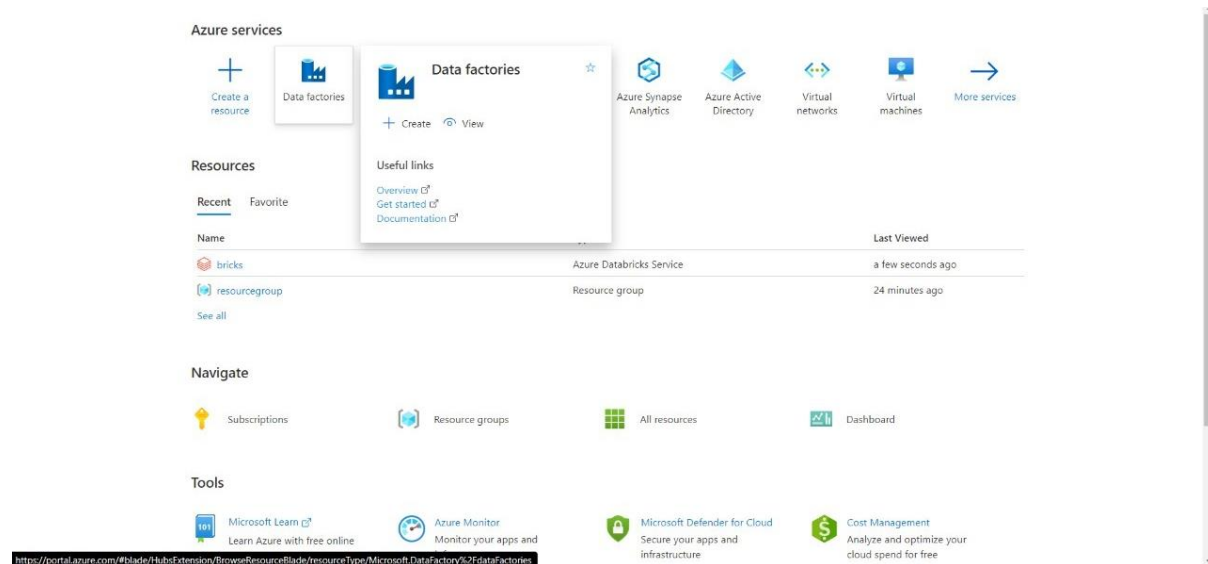


Step-11: Give some operation or do the logic building in the notebook

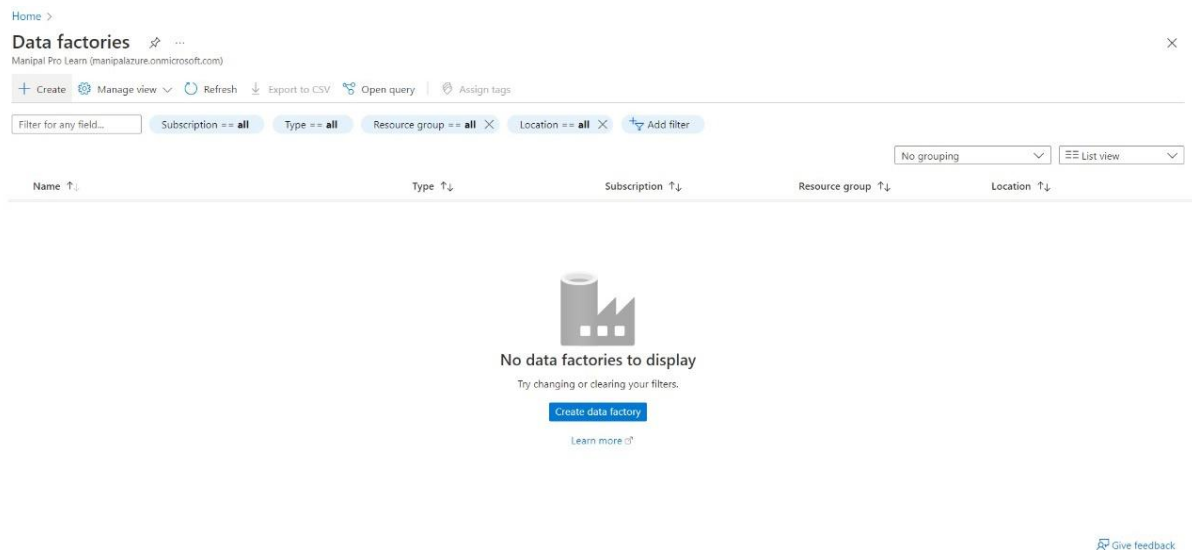


Now our notebook is ready to be linked and executed in Azure Data Factory

Go to azure home page and search for Data Factory



Create a new account on Azure Data Factory



Enter the details for the Azure Data Factory account

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

Create Data Factory

Basics

Git configuration

Networking

Advanced

Tags

Review + create

Project details

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

Subscription *

Azure-DXC262AB12Lab

Resource group *

resourcegroup

Create new

Instance details

Name *

AP-morgan

Region *

East US

Version *

V2 (Recommended)

Review + create

< Previous

Next : Git configuration >

Validate and deploy the Data Factory

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

Create Data Factory

*** Submitting deployment...
Submitting the deployment template for resource group 'resourcegroup'.

✓ 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

resourcegroup

Name

AP-morgan

Region

East US

Version

V2 (Recommended)

Networking

Connect via

Public endpoint

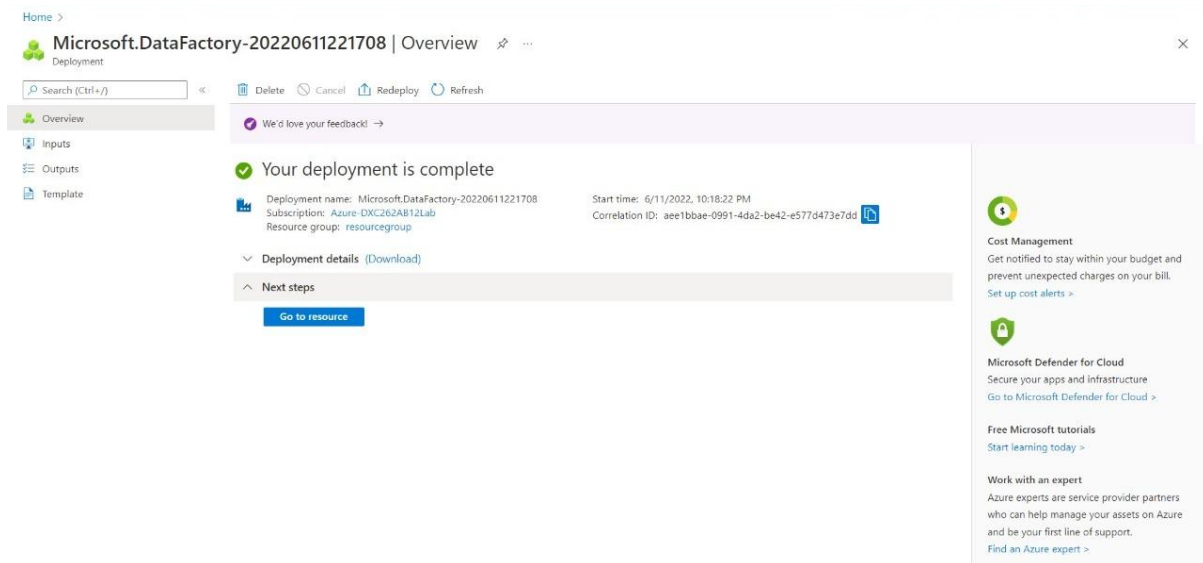
Create

< Previous

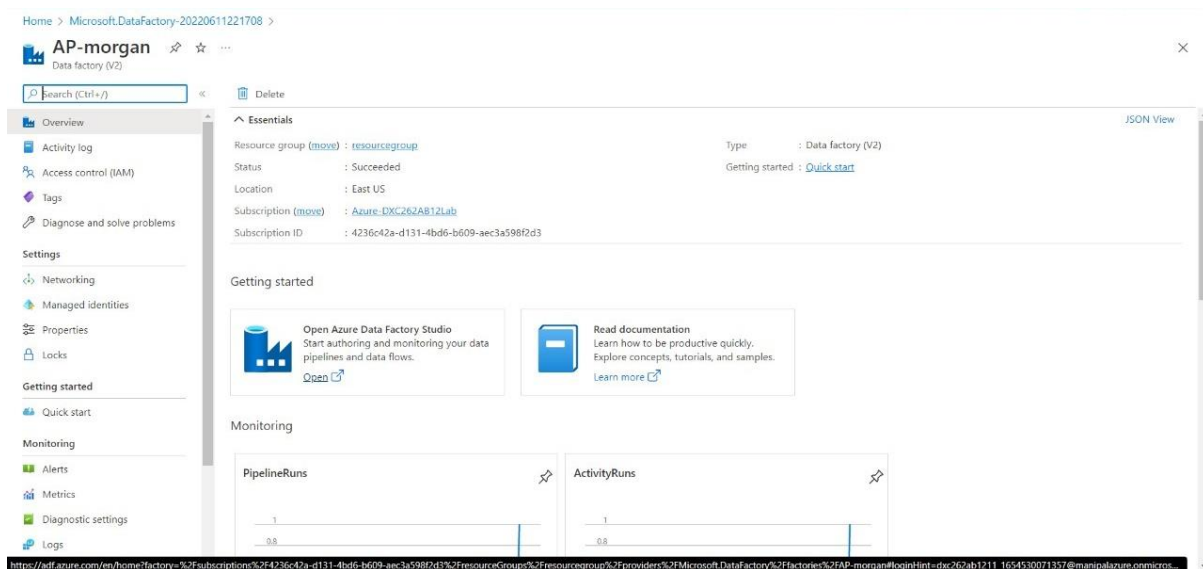
Next

Download a template for automation

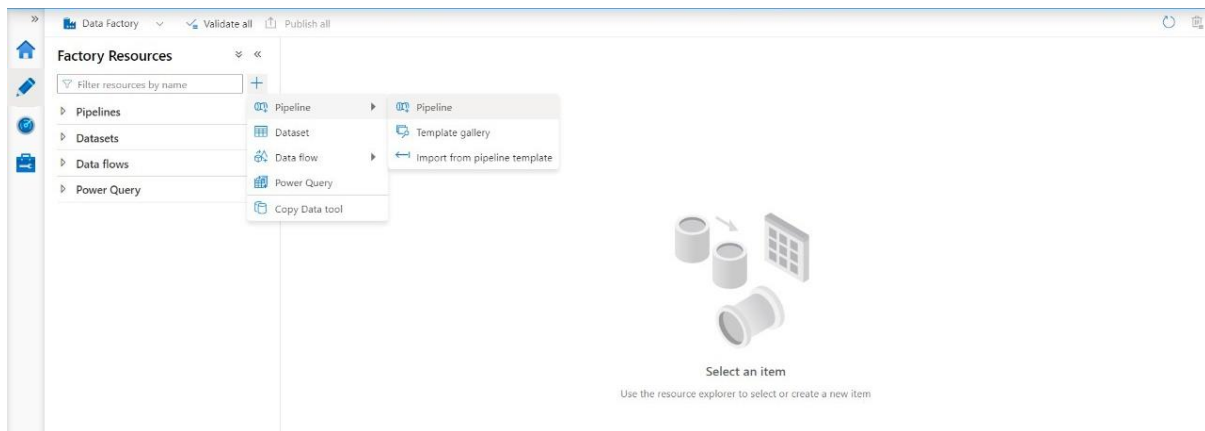
After the deployment is done, navigate to “Go to resource”



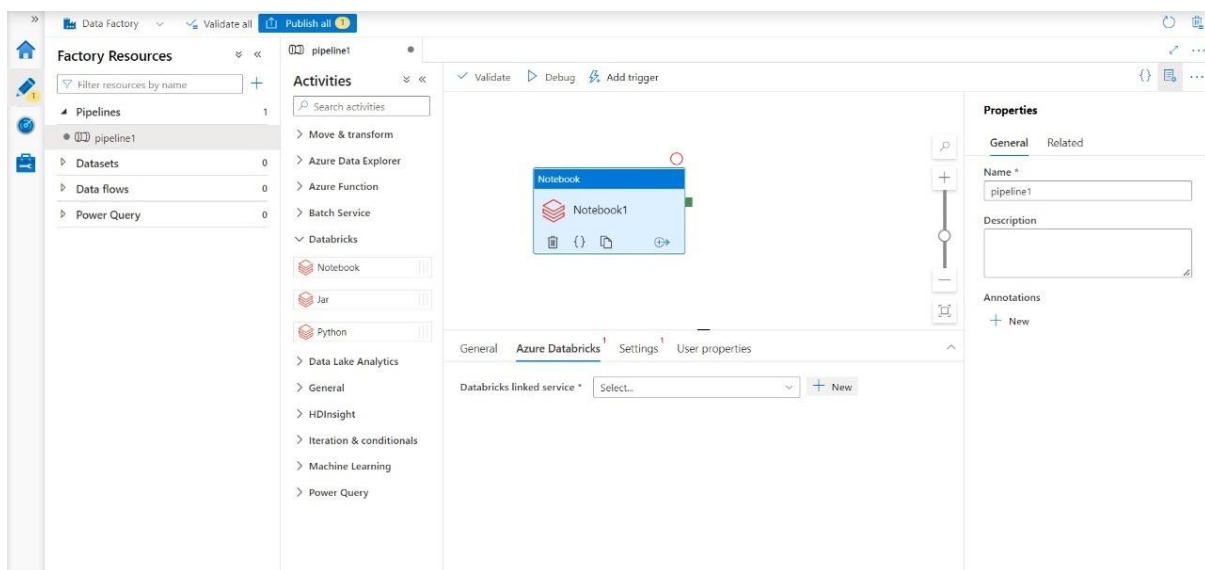
Now open the Data Factory studio



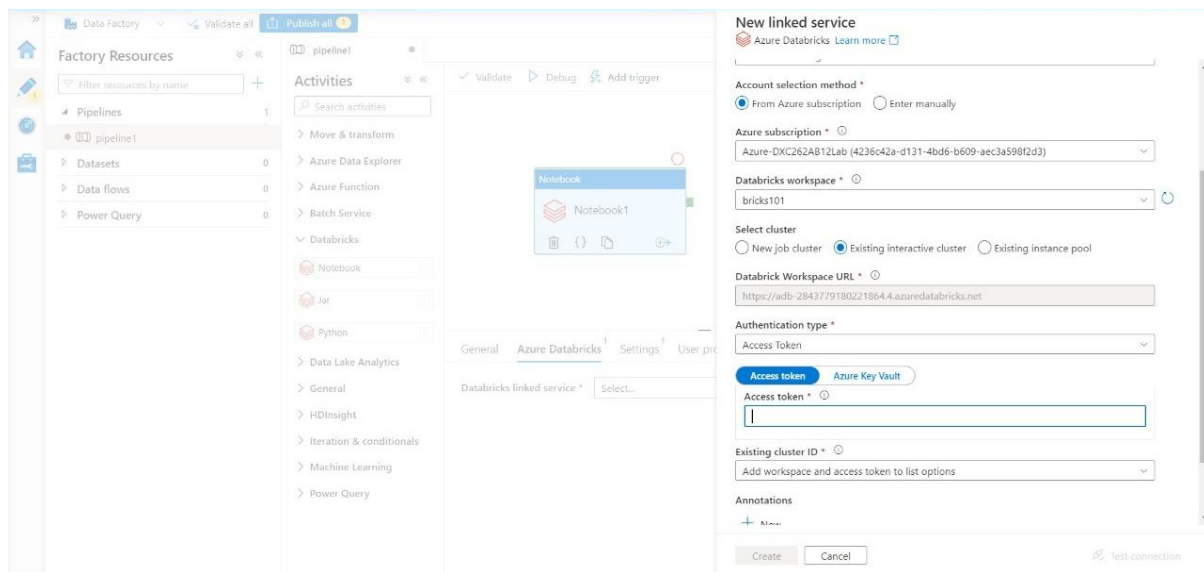
We need to create a pipeline that will connect Data Bricks notebook with Data Factory.



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

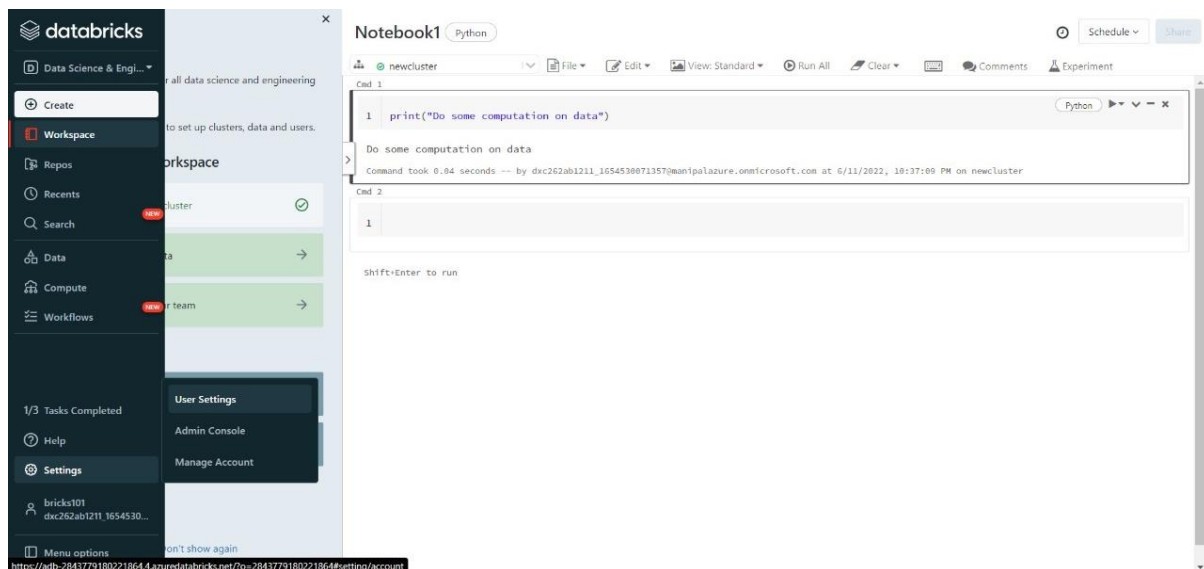


Create a new Linked service for databricks

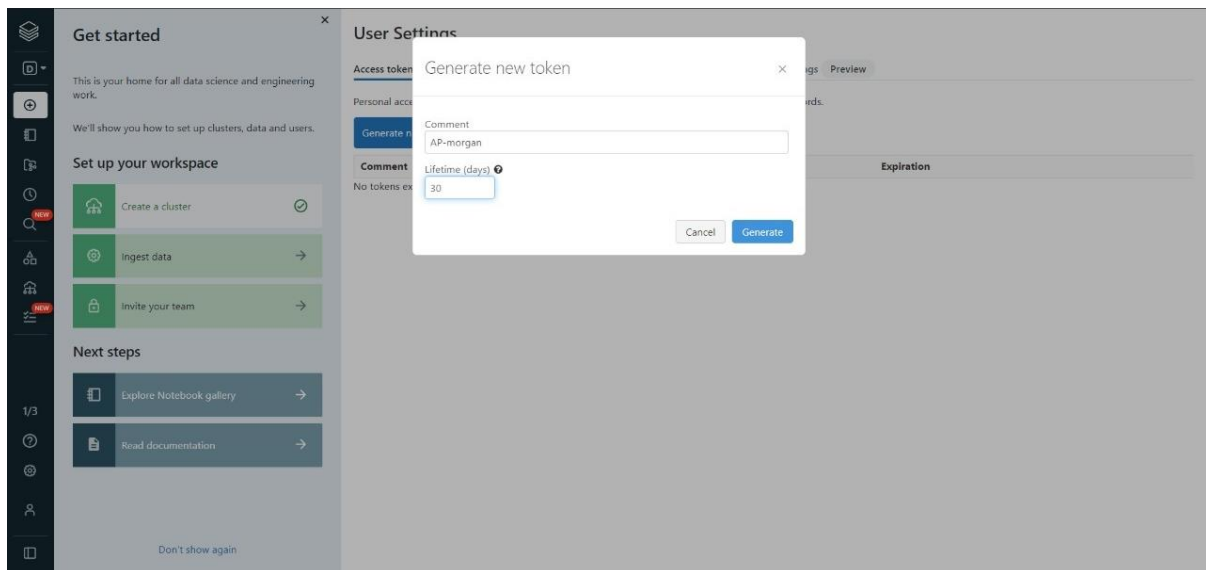


We need the access token of Data bricks account in order to access it.

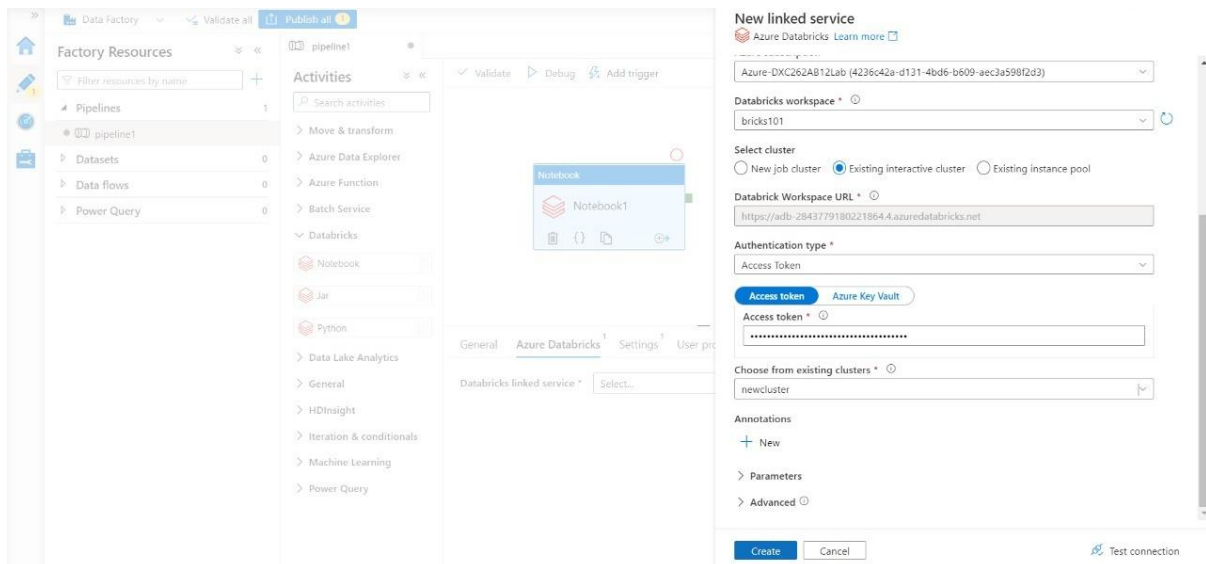
Navigate to Data bricks workspace and go to Users in settings



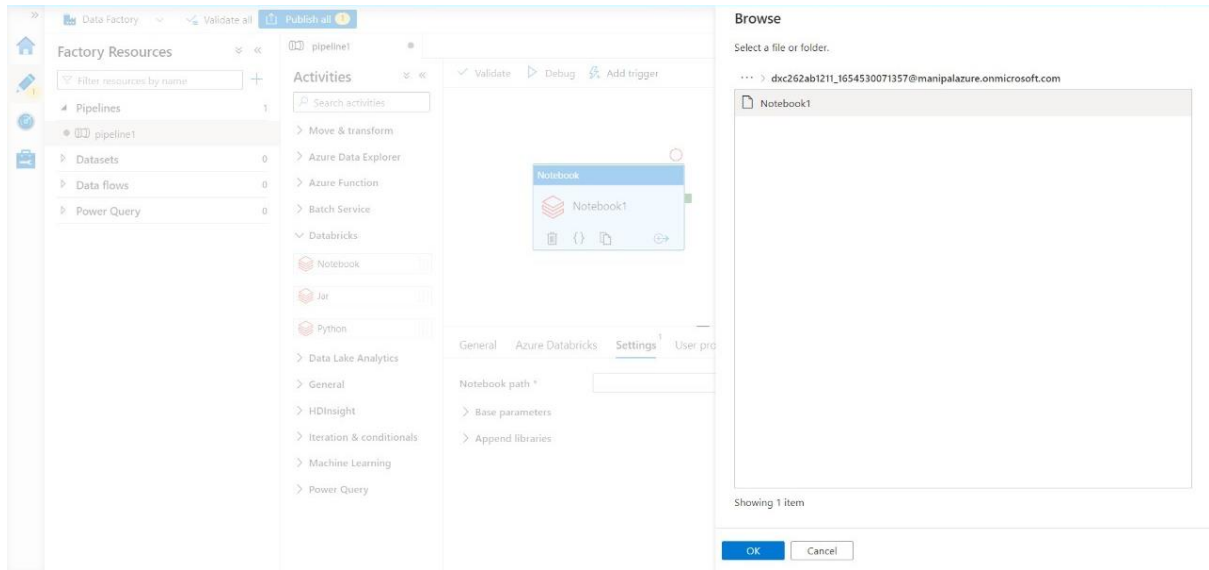
Click on “Generate token” and mention a small quick description name for the token



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

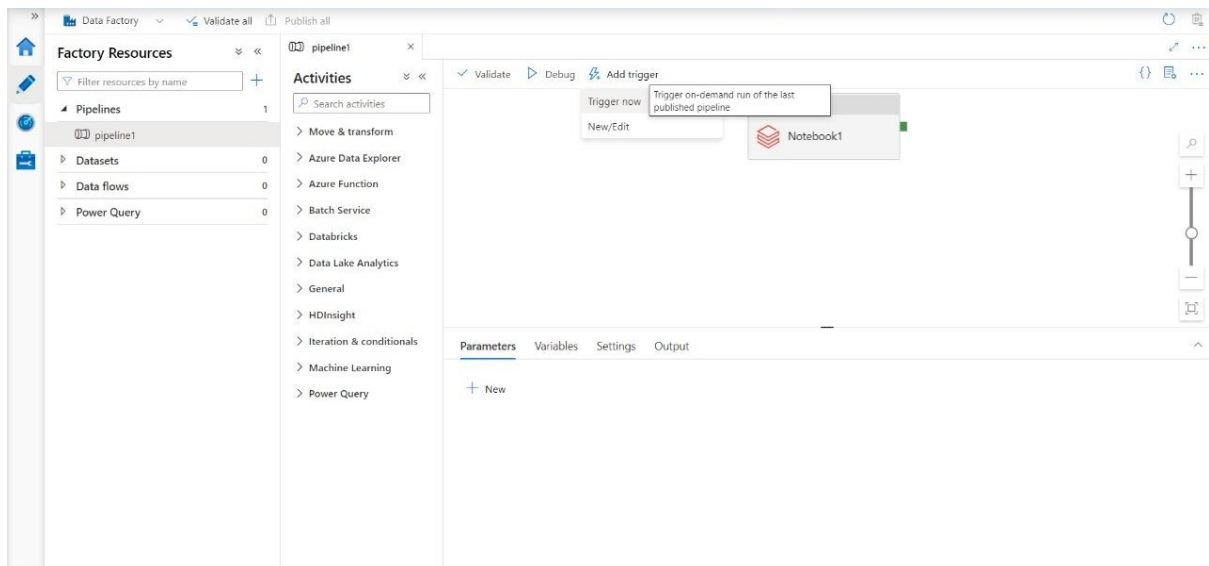


Choose the right cluster and click on create

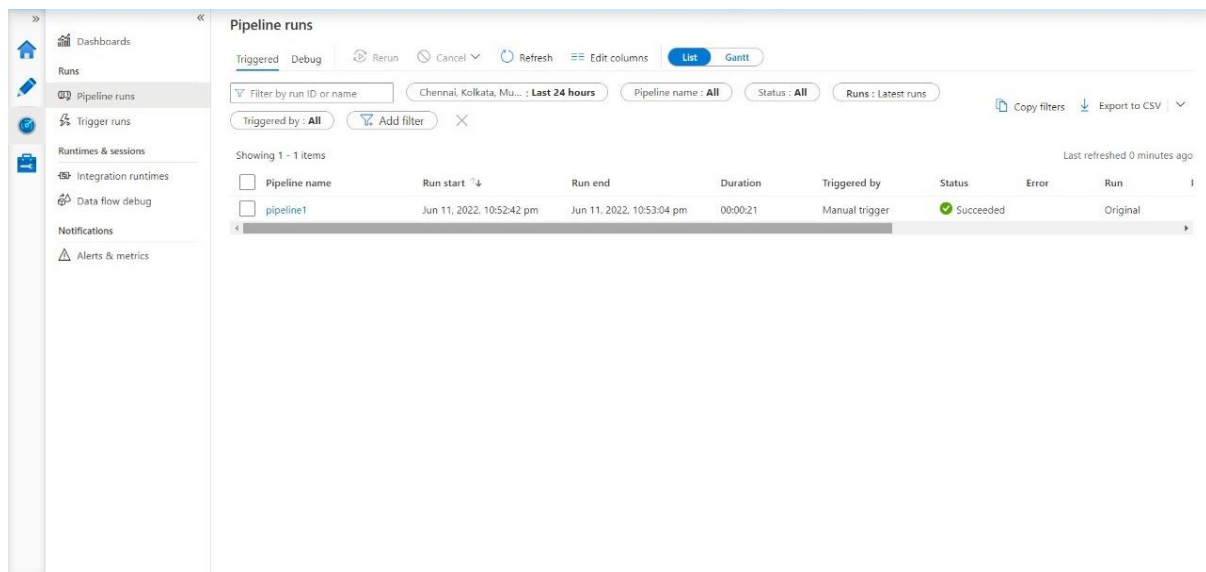


Now choose the notebook in the pipeline notebook tab setting to trigger it.

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



Go to Monitor window to check its execution



The screenshot shows the 'Pipeline runs' section in the Azure Data Factory Monitor. The left sidebar contains navigation links: 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. At the top, there are tabs for 'Triggered' and 'Debug', along with buttons for 'Rerun', 'Cancel', 'Refresh', 'Edit columns', 'List', and 'Gantt'. Below these are filters for 'Filter by run ID or name', 'Chennai, Kolkata, Mu...', 'Last 24 hours', 'Pipeline name: All', 'Status: All', and 'Runs: Latest runs'. There are also buttons for 'Copy filters', 'Export to CSV', and 'Add filter'. The table shows one run for 'pipeline1' with a status of 'Succeeded'.

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 have successfully triggered a linked notebook of Data bricks from Data Factory.

Result: Successfully able to link and trigger Azure DataBricks notebook using Data Factory.

Conclusion: Azure Data Factory linked with Azure Data bricks