



DXC REALTIME PROJECTS

AZ-900, DP - 203



JUNE 10, 2022

DXC TECHNOLOGY

Name: Shaik Abdul Khadar Jilani

Reg No: DXC262AB12038

Project1 Name: Smart Vehicles

Date: 10-6-2022

Project 1 : Connected Vehicles

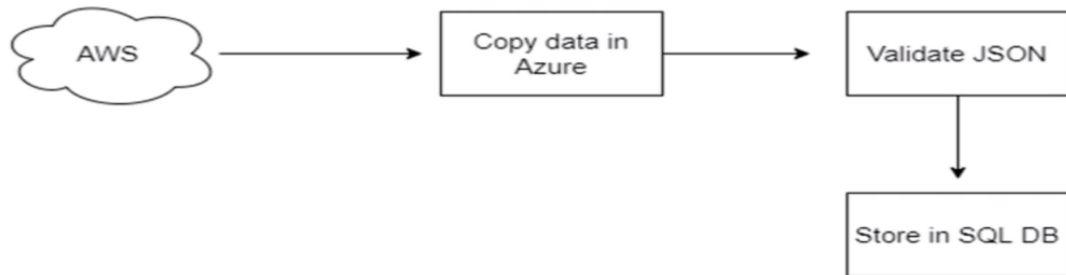
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



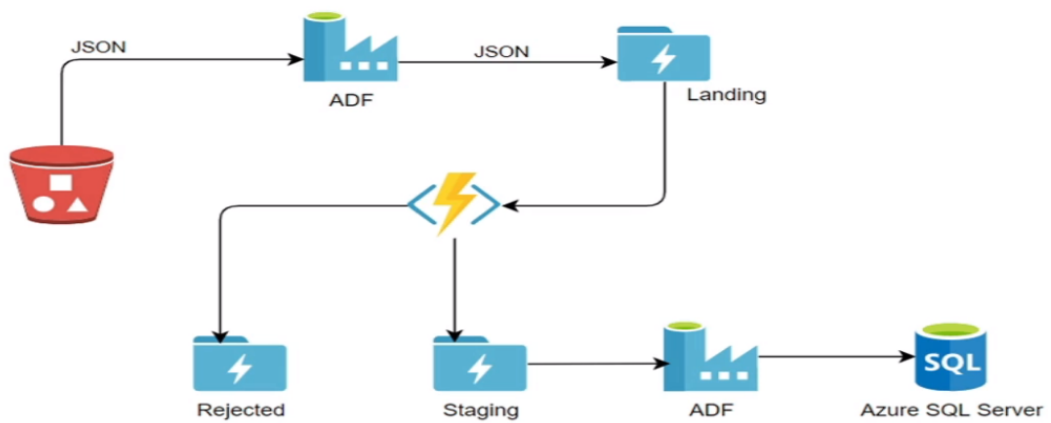
Project 1 : Connected Vehicles

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

Project 1 : Connected Vehicles



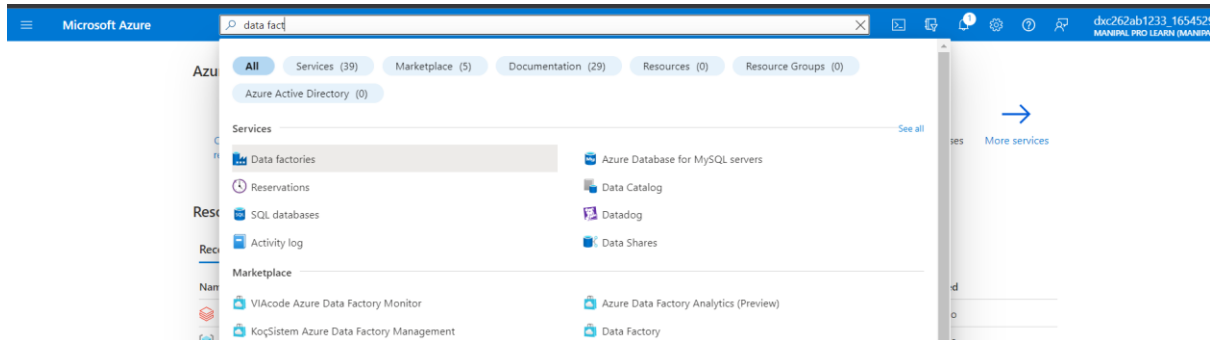
Project 1 : Connected Vehicles



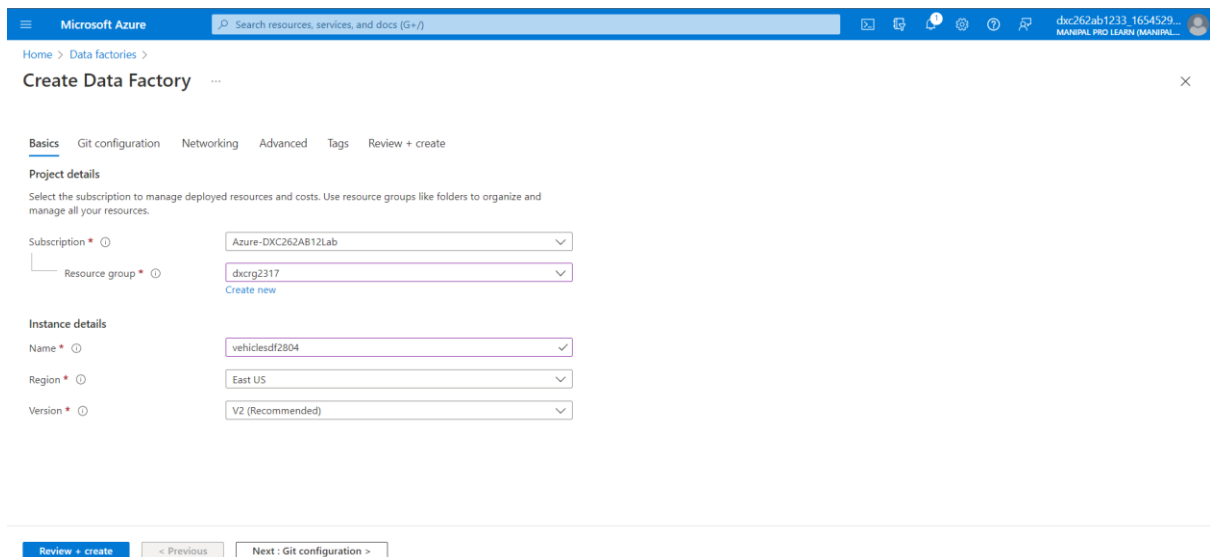
Architecture Diagram for Connected Vehicle Project

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

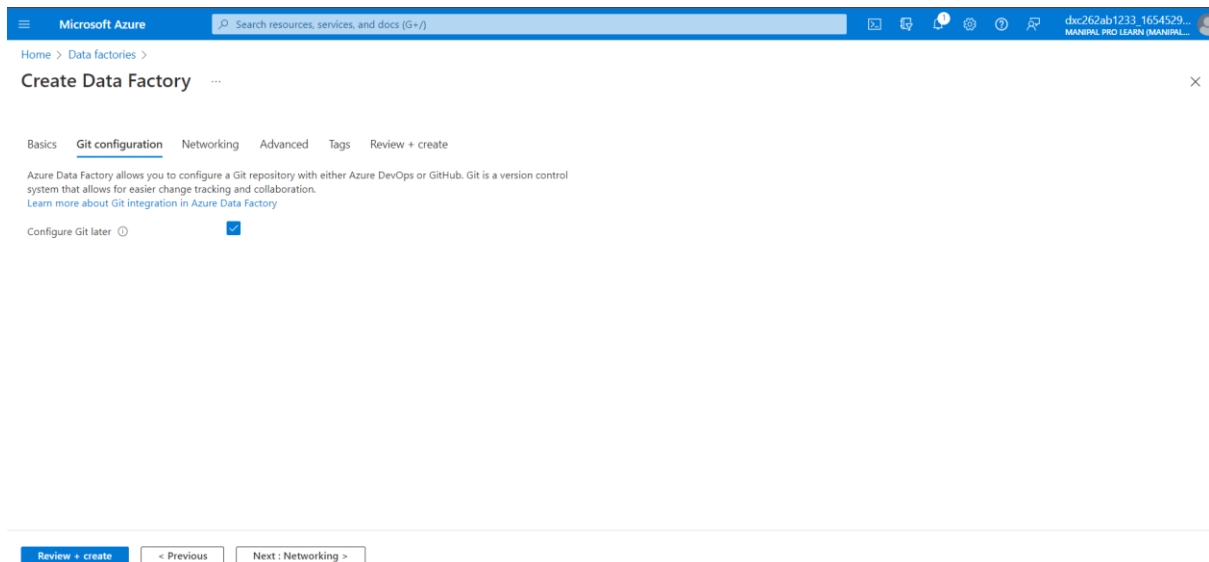
Step 1: First we need to open Azure and search for Data Factory and click on it



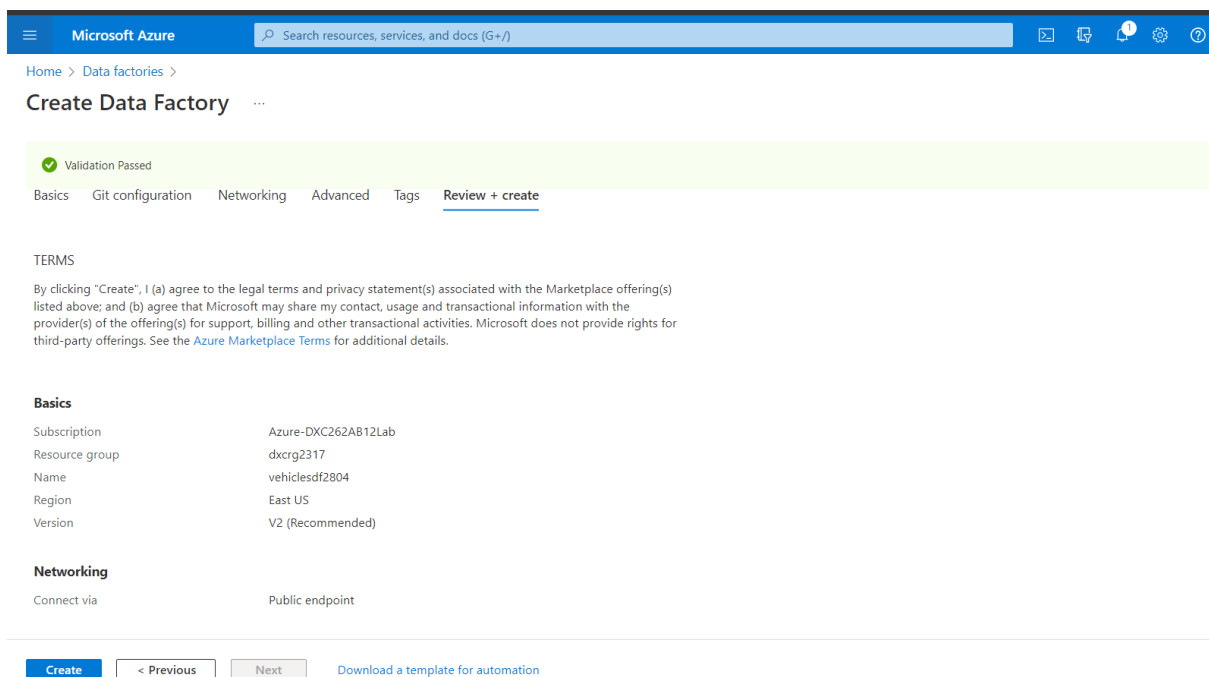
Step 2: Select data factory and click the + create and Give the Resource group name and Name vehiclesdf2804 and click next Git configuration



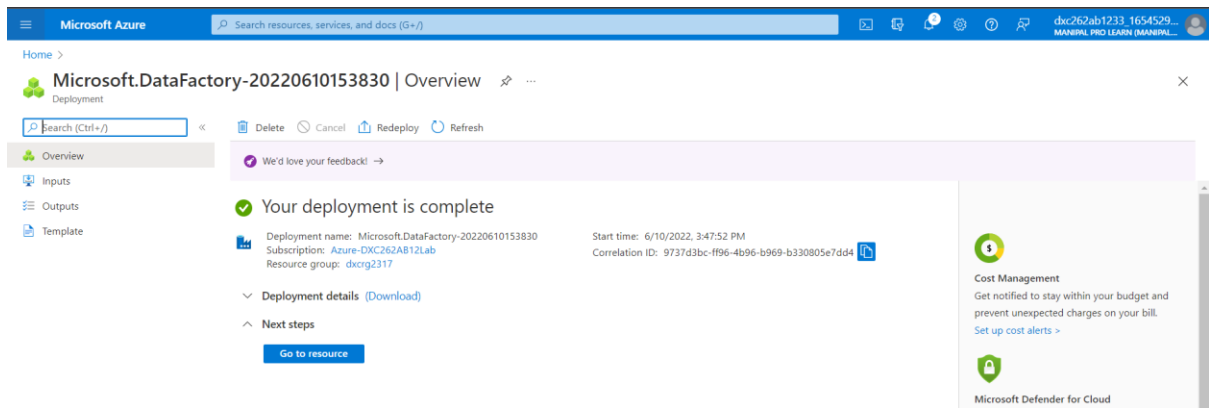
Step 3: Check box the git configuration later and all other are default and click Review + create



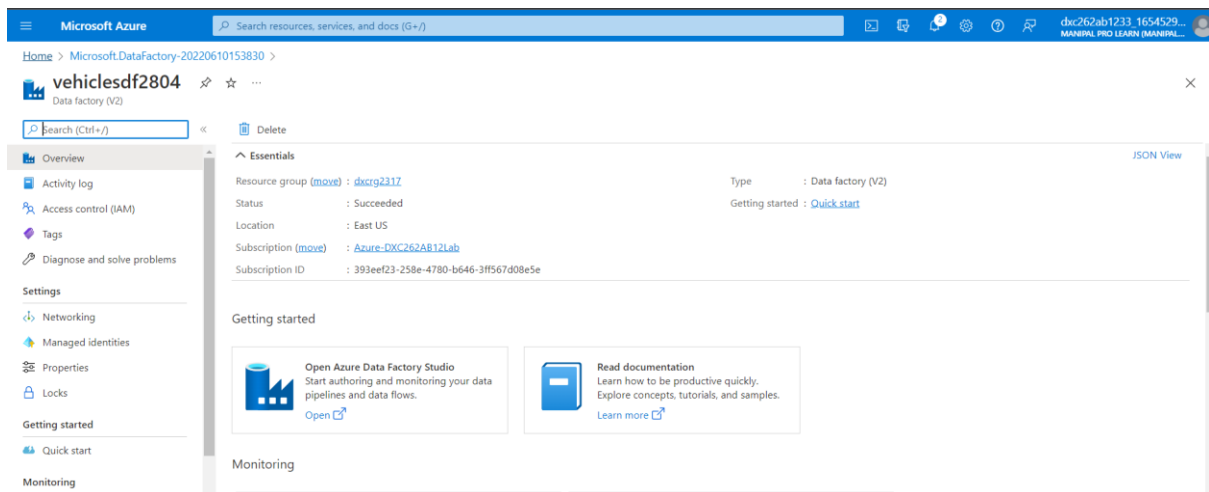
Step 4: After that we get the validation is passed and click on Create



Step 5: It will take some time and it will show Your Deployment is complete

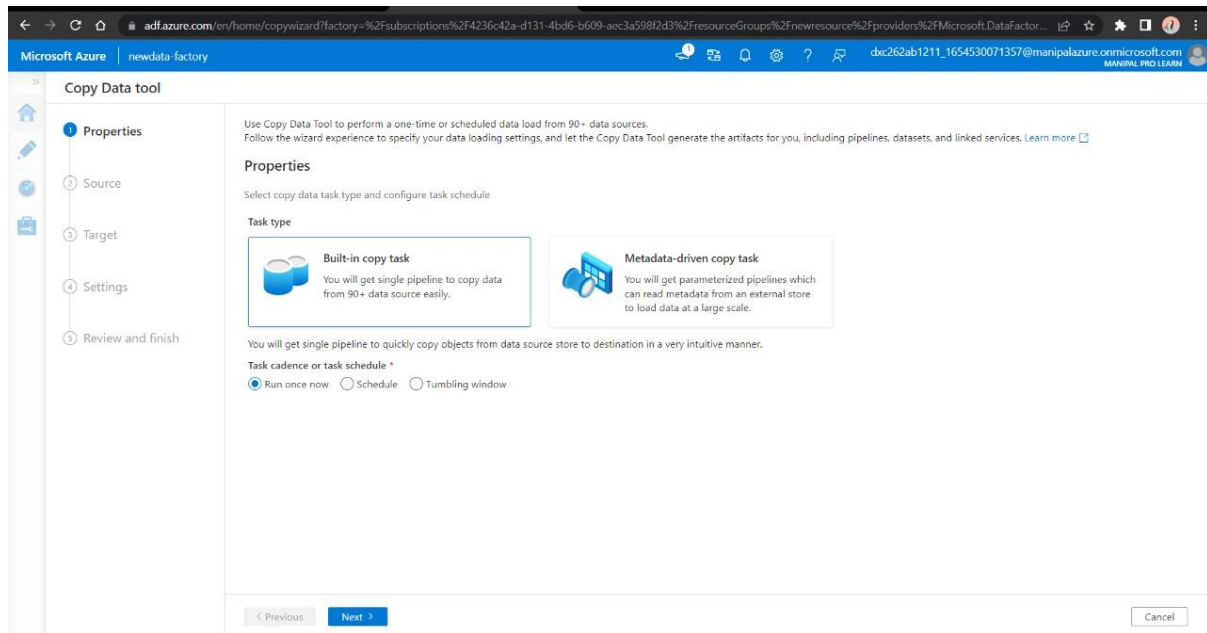


Step 6: click on Go to resource and in that click on open Azure data factory studio where data factory is created



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

Step 1: Here we need to create a pipeline that will take data from blob storage and feed into the SQL database. Click on ingest



Step 2: We need to fill the name of Azure blob storage linked file. Select the name of storage account name we created in this and click on Create.

Microsoft Azure | newdata-factory

Copy Data tool

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

New linked service

Name: AzureBlobStorage1

Description:

Connect via integration runtime: AutoResolveIntegrationRuntime

Authentication type: Account key

Account selection method: From Azure subscription

Azure subscription: Azure-DXC262AB12Lab (4236c42a-d131-4bd6-b609-aec3a59812d3)

Storage account name: storage1011

Additional connection properties: + New

Test connection: Test connection

Previous Next Create Cancel

Step 3: Select the source type as blob storage and connection name that we created in previous step. Click from Options. Click Next.

Microsoft Azure | newdata-factory

Copy Data tool

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

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.

File or folder: source/transactions.csv

Options

Binary copy: Binary copy

Recursively: Recursively

Enable partition discovery: Enable partition discovery

Max concurrent connections: Max concurrent connections

Filter by last modified

Start time (UTC): Start time (UTC)

End time (UTC): End time (UTC)

Previous Next Cancel

Choose the output format of the data

Copy Data tool

File format settings

File format: DelimitedText [Detect text format] [Preview data]

Column delimiter: Comma (,) [Edit]

Row delimiter: Default (\r\n, or \n\r\n) [Edit]

☒ First row as header

Advanced

Compression type: None

Additional columns: [New]

[Previous] [Next] [Cancel]

Step 4: Now, in the next step we will link the Data Factory with SQL database .Connect the database with new connection.

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 one.

Target type: Azure SQL Database

Connection: Select... [New connection]

New linked service

Azure SQL Database [Learn more]

Connection string: Azure Key Vault

Account selection method: ☒ From Azure subscription ☐ Enter manually

Azure subscription: Azure-DXC262AB12Lab (4236c42a-d131-4bd6-b609-aec3a598f2d3)

Server name: newserver101

Database name: newsql database

Authentication type: SQL authentication

User name: user

Password: [Password field]

Always encrypted: ☐

Additional connection properties: [New]

[Previous] [Next] [Create] [Cancel] [Test connection]

Step 6: Choose the file name you want to see in the SQL database.
Click Next.

The screenshot shows the 'Copy Data tool' configuration page in the Microsoft Azure portal. The left sidebar contains a navigation menu with the following items: Properties, Source, Target, Dataset, Configuration, Settings, and Review and finish. The 'Destination data store' section is active, showing the following configuration:

- Target type:** Azure SQL Database
- Connection:** AzureSqlDatabase1
- Source:** Azure Blob Storage file
- Target:** transaction (with a 'Use existing table' link)
- Destination:** Azure Blob Storage file (auto-create)

At the bottom, there is a checkbox for 'Skip column mapping for all tables' and navigation buttons for '< Previous', 'Next >', and 'Cancel'.

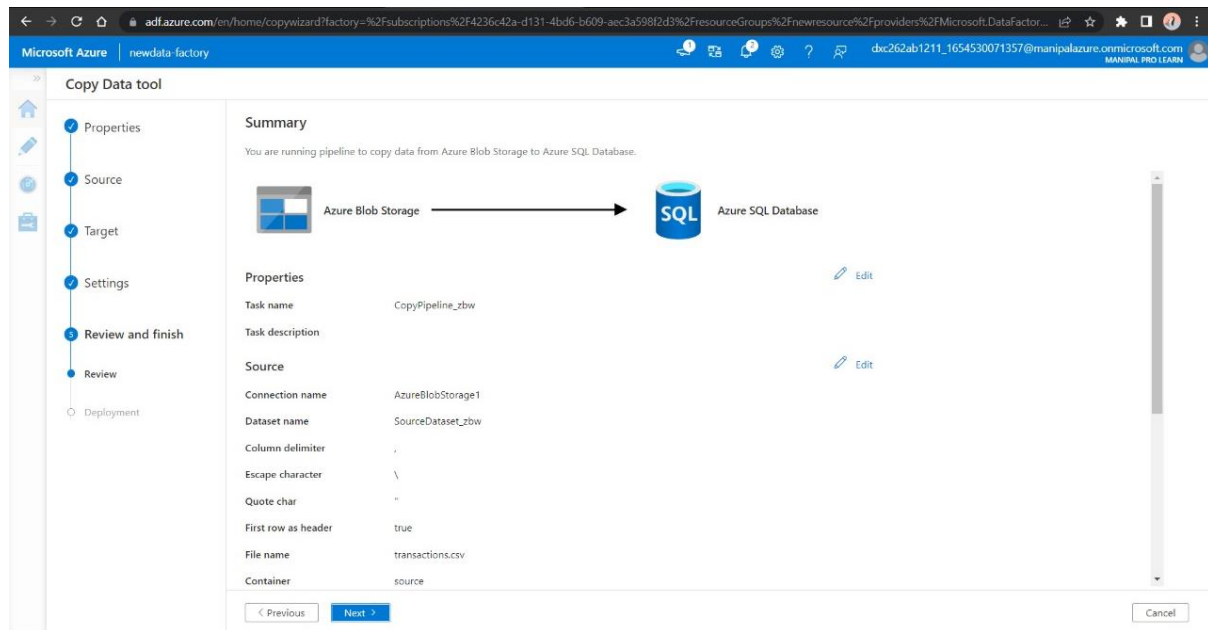
Step 7: Now we are the end stage of creating pipeline, click next.

The screenshot shows the 'Copy Data tool' configuration page in the Microsoft Azure portal, specifically the 'Settings' section. The left sidebar contains a navigation menu with the following items: Properties, Source, Target, Settings, and Review and finish. The 'Settings' section is active, showing the following configuration:

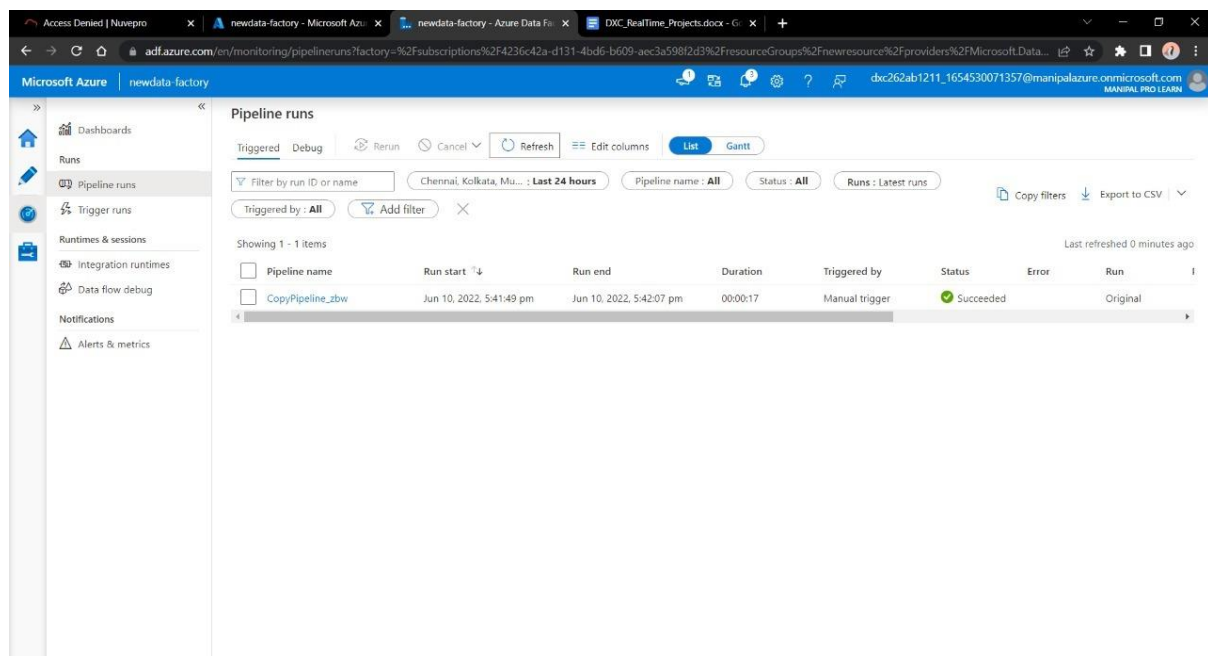
- Task name:** CopyPipeline_zbw
- Task description:** (empty text box)
- Data consistency verification:** ☐
- Fault tolerance:** (dropdown menu)
- Enable logging:** ☐
- Enable staging:** ☐
- Advanced:** (expandable section)

At the bottom, there are navigation buttons for '< Previous', 'Next >', and 'Cancel'.

Step 8: Validate the details and move forwards with Next



Step 9: Trigger the pipeline manually and go to monitor tab , Here we need to refresh see our pipeline has successfully



Step 10: Go to SQL database to know that the data is there

Access Denied | Nuvepro

newsqldatabase (newserver101/)

newdata-factory - Azure Data

DXC_RealTime_Projects.docx - G

portal.azure.com/@manipalazure.onmicrosoft.com/resource/subscriptions/4236c42a-d131-4bd6-b609-aec3a5982d3/resourceGroups/newresource/providers/Microsoft.Sql/servers/ne...

Microsoft Azure

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

dxm262ab1211.1654530...

MANIPAL PRO LEARN (MANIPAL...

Home > newsqldatabase (newserver101/newsqldatabase)

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

SQL database

Search (Ctrl+J)

Login

New Query

Open query

Feedback

Overview

Activity log

Tags

Diagnose and solve problems

Getting started

Query editor (preview)

Power Platform

Power BI

Power Apps

Power Automate

Settings

Compute + storage

Connection strings

Maintenance

Properties

Locks

Data management

Replicas

newsqldatabase (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 x transaction.Azure Blob Storage file x

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

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

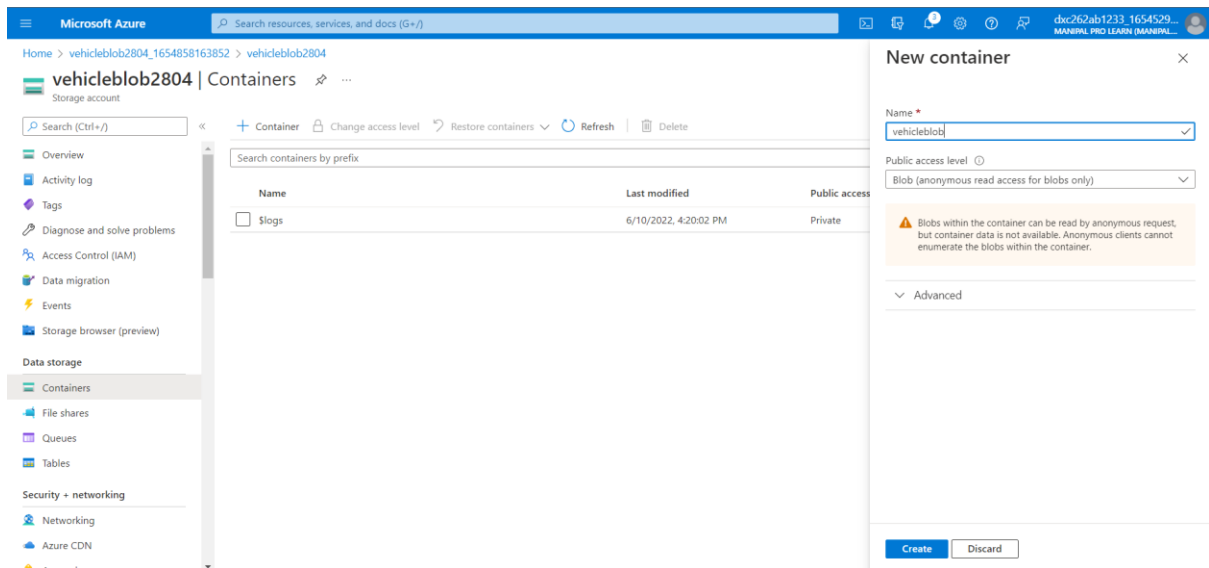
Step 1: Search for Storage account and click + create

The screenshot shows the 'Create a storage account' page in the Microsoft Azure portal. The page has a blue header with the Microsoft Azure logo and a search bar. Below the header, there's a breadcrumb trail: 'Home > Storage accounts >'. The main heading is 'Create a storage account' with a close button (X) on the right. There are tabs for 'Basics', 'Advanced', 'Networking', 'Data protection', 'Encryption', 'Tags', and 'Review + create'. The 'Basics' tab is selected. Under 'Instance details', there's a link: 'If you need to create a legacy storage account type, please click here.' The form fields are: 'Storage account name' (vehicleblob2804), 'Region' (US East US), 'Performance' (Standard: Recommended for most scenarios (general-purpose v2 account) is selected, Premium: Recommended for scenarios that require low latency. is unselected), and 'Redundancy' (Geo-redundant storage (GRS) is selected, Make read access to data available in the event of regional unavailability. is checked). At the bottom, there are buttons: 'Review + create', '< Previous', and 'Next : Advanced >'.

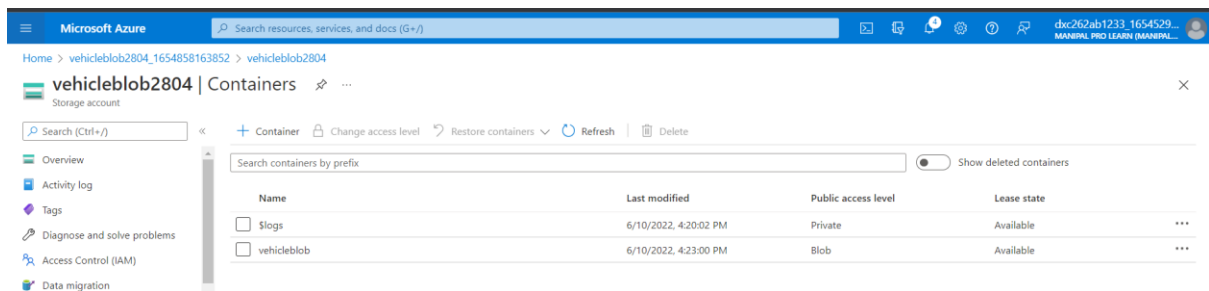
Step 2: Click on Review + create and create

The screenshot shows the 'Overview' page for a storage account named 'vehicleblob2804_1654858163852'. The page has a blue header with the Microsoft Azure logo and a search bar. Below the header, there's a breadcrumb trail: 'Home >'. The main heading is 'vehicleblob2804_1654858163852 | Overview' with a close button (X) on the right. There's a search bar (Search (Ctrl+J)) and buttons for 'Delete', 'Cancel', 'Redeploy', and 'Refresh'. On the left, there's a sidebar with 'Overview', 'Inputs', 'Outputs', and 'Template'. The main content area shows a green checkmark and the text 'Your deployment is complete'. Below this, there's a table with deployment details: 'Deployment name: vehicleblob2804_1654858163852', 'Subscription: Azure-DXC262AB12Lab', 'Resource group: dxcrig2317', 'Start time: 6/10/2022, 4:19:33 PM', and 'Correlation ID: 532c98e1-7d60-46a1-b9cc-dcd204460617'. There are links for 'Deployment details (Download)' and 'Next steps'. At the bottom, there's a button 'Go to resource'. On the right, there's a 'Cost Management' section with a green checkmark and the text 'Get notified to stay within your budget and prevent unexpected charges on your bill. Set up cost alerts >'. Below that, there's a 'Microsoft Defender for Cloud' section with a green checkmark and the text 'Secure your apps and infrastructure'.

Step 3: Go to resource and go to container option and click + container and give the name of it and select public access level Blob

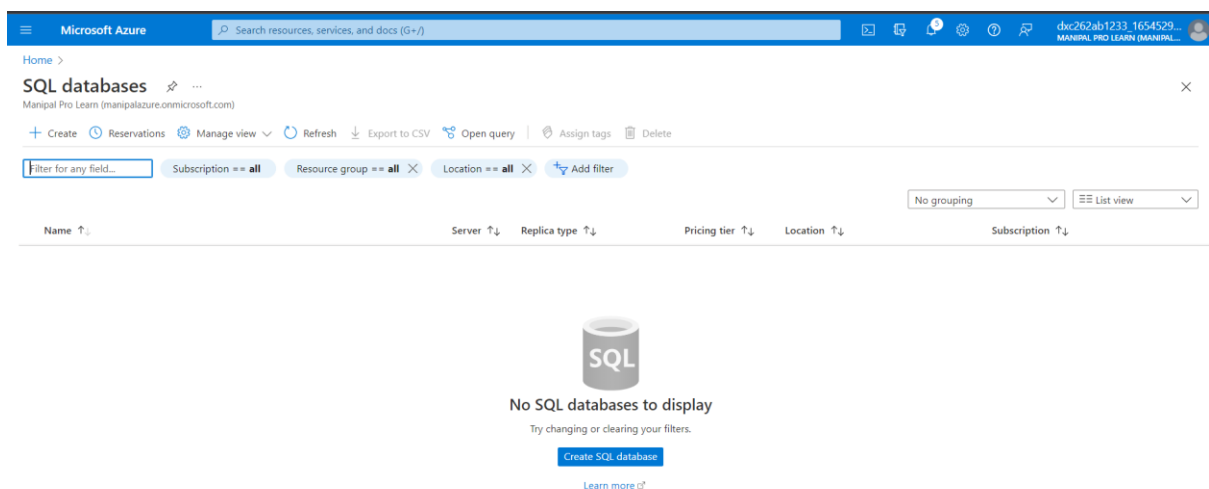


Step 4: Blob is created



Practical Lab: Create Azure SQL Server and Database

Step 1: search for sql data base and click + create



Step 2: Give the Resource name and Name of database and create new server give admin login and click ok

The screenshot shows the 'Create SQL Database Server' form in the Microsoft Azure portal. The form is titled 'Create SQL Database Server' and includes a search bar at the top. The form fields are as follows:

- Server name ***: A text input field containing 'vehicle2804'.
- Location ***: A dropdown menu showing '(US) East US'.
- Authentication**: A section with three radio buttons: 'Use SQL authentication' (selected), 'Use only Azure Active Directory (Azure AD) authentication', and 'Use both SQL and Azure AD authentication'.
- Server admin login ***: A text input field containing 'jilani'.
- Password ***: A text input field containing '*****'.
- Confirm password ***: A text input field containing '*****'.

A green checkmark icon and the text 'Password and confirm password must match.' are visible next to the 'Confirm password' field. At the bottom of the form, there is a blue 'OK' button.

Step 3: All other options remain same and click review + create after that validation is passed click on create

The screenshot shows the 'Review + create' step in the 'Create SQL Database' form in the Microsoft Azure portal. The form is titled 'Create SQL Database' and includes a search bar at the top. The form is divided into two main sections: 'Product details' and 'Terms'.

Product details

- SQL database by Microsoft**: A section with links for 'Terms of use' and 'Privacy policy'.
- Estimated cost per month**: A section with a 'View pricing details' link.

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	dxcrq2317
Region	East US
Database name	vehicledb2804
Server	(new) vehicle2804
Authentication method	SQL authentication

At the bottom of the form, there is a blue 'Create' button, a grey '< Previous' button, and a link 'Download a template for automation'.

Step 4: We will get Your deployment is complete

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information. The main content area displays the deployment status for a resource named 'Microsoft.SQLDatabase.newDatabaseNewServer_1896e009f147449a88731'. The deployment is marked as 'Complete' with a green checkmark. Below this, deployment details are listed: Deployment name, Subscription, Resource group, Start time, and Correlation ID. A 'Go to resource' button is visible. On the right, there are sections for 'Cost Management' and 'Microsoft Defender for Cloud'.

Microsoft Azure

Home > Microsoft.SQLDatabase.newDatabaseNewServer_1896e009f147449a88731 | Overview

Deployment

Search (Ctrl+/)

Delete Cancel Redeploy Refresh

We'd love your feedback! →

✓ Your deployment is complete

Deployment name: Microsoft.SQLDatabase.newDatabaseNewServer_1896e009f147449a88731 Start time: 6/10/2022, 4:35:53 PM
Subscription: Azure-DXC262AB12Lab Correlation ID: ec2ee825-52d7-4368-bf19-8c472c4e319b
Resource group: dxcrq2317

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 >

Step 5: Go to resource here we can see SQL database , server

The screenshot shows the Microsoft Azure portal interface for a specific SQL database resource named 'vehicledb2804 (vehicle2804/vehicledb2804)'. The left sidebar contains navigation options like Overview, Activity log, Tags, Diagnose and solve problems, Getting started, Query editor (preview), Power Platform, Power BI, Power Apps, Power Automate, and Settings. The main content area displays the database's 'Essentials' information, including Resource group, Status, Location, Subscription, Subscription ID, and Tags. It also shows 'Compute utilization' and 'Database data storage' sections. A message at the top indicates that the database was just created and provides a link to get started.

Microsoft Azure

Home > Microsoft.SQLDatabase.newDatabaseNewServer_1896e009f147449a88731 > vehicledb2804 (vehicle2804/vehicledb2804)

SQL database

Search (Ctrl+/)

Copy Restore Export Set server firewall Delete Connect with... Feedback

This database was just created. Do you need any help [getting started?](#)

Essentials

Resource group (move): dxcrq2317 Server name: vehicle2804.database.windows.net
Status: Online Elastic pool: No elastic pool
Location: East US Connection strings: Show database connection strings
Subscription (move): Azure-DXC262AB12Lab Pricing tier: General Purpose: Gen5, 2 vCores
Subscription ID: 393eef23-258e-4780-b646-3ff567d08e5e Earliest restore point: No restore point available
Tags (edit): Click here to add tags

Show data for last: 1 hour 24 hours 7 days Aggregation type: Max Database data storage

Compute utilization

Result: data factory, SQL database and SQL data server, pipeline

Conclusion: I got how to create a data factory and how to create a pipeline and also understand the way to create and also got to know how to create a SQL database and SQL server and blob data is successfully validated and stored into SQL database.

References:

Name: Shaik Abdul Khadar Jilani

Reg No: DXC262AB12038

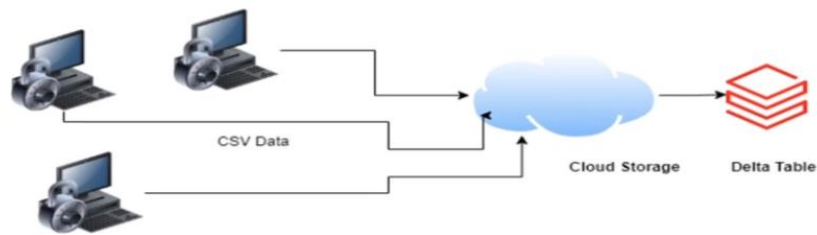
Project2 Name: AP Morgan Data Platform

Date: 10-6-2022

Project 2: AP Morgan Data Platform

Project 2 : AP Morgan

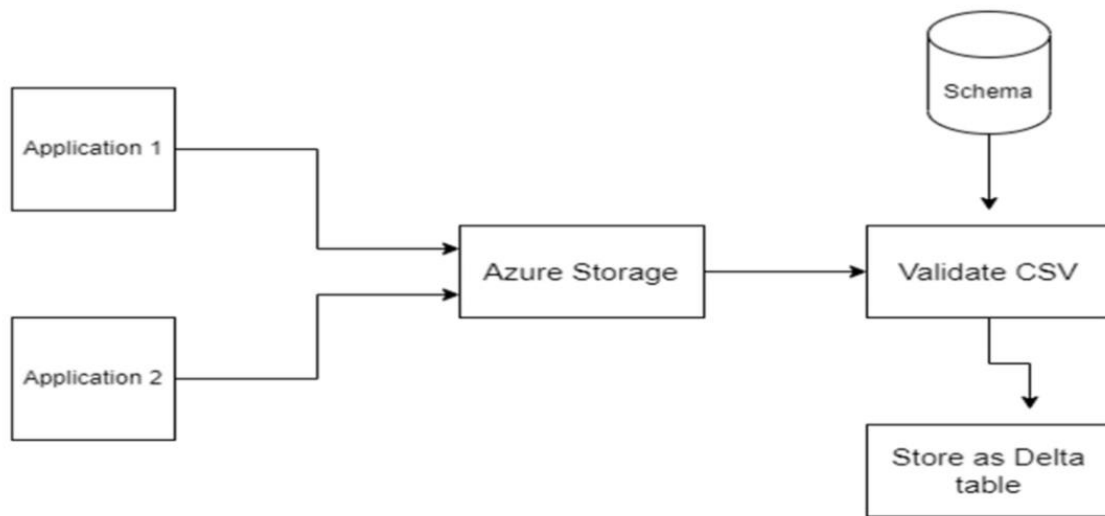
- Multiple Internal applications sends the data(huge size) in CSV format on daily basis in the cloud storage location. There are couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed data to be persisted as Delta table in Databricks for downstream system.



Project 2 : AP Morgan- High Level Detail

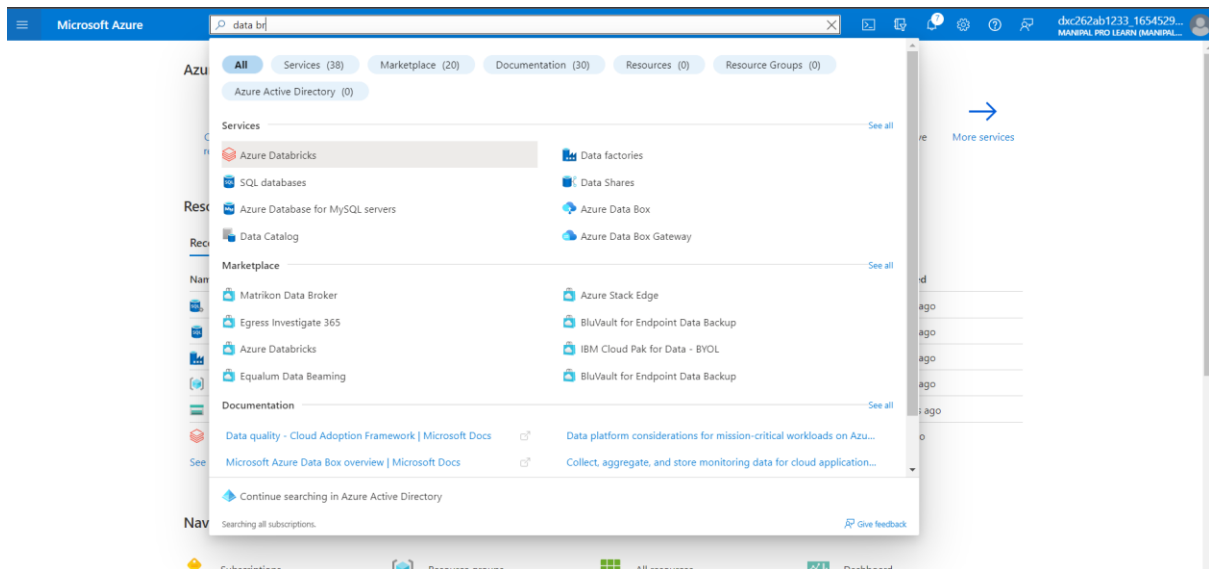
- Internal Application sends CSV file in Azure data lake storage.
- Validation needed to apply on this follows:
 - Check for duplicate rows. If it contains duplicate rows, file need to be rejected.
 - Need to validate the date format for all the date fields. Date column names and desired date format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

Project 2 : AP Morgan



Practical Lab: Create a **Databricks**




Step 1: Search for data bricks and select Azure data bricks



Step 2: Click on + create and fill the details in and pricing tier select Trial and all other are default and click review and create

Microsoft Azure

Search resources, services, and docs (G+)



[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 *

Azure-DXC262AB12Lab

Resource group *

dxcrq2317

[Create new](#)

Instance Details

Workspace name *

morgandb2804

Region *

East US

Pricing Tier *

Trial (Premium - 14-Days Free DBUs)

Review + create







< Previous

Next : Networking >

Step 3: After validation click on create

Microsoft Azure

Search resources, services, and docs (G+)



dxcr262ab1233.1654529...
MANUAL PRO LEARN (MANUAL...

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

Create an Azure Databricks workspace

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

Summary

Basics

Workspace name

morgandb2804

Subscription

Azure-DXC262AB12Lab

Resource group

dxcrq2317

Region

East US

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

Public Information Provisioning

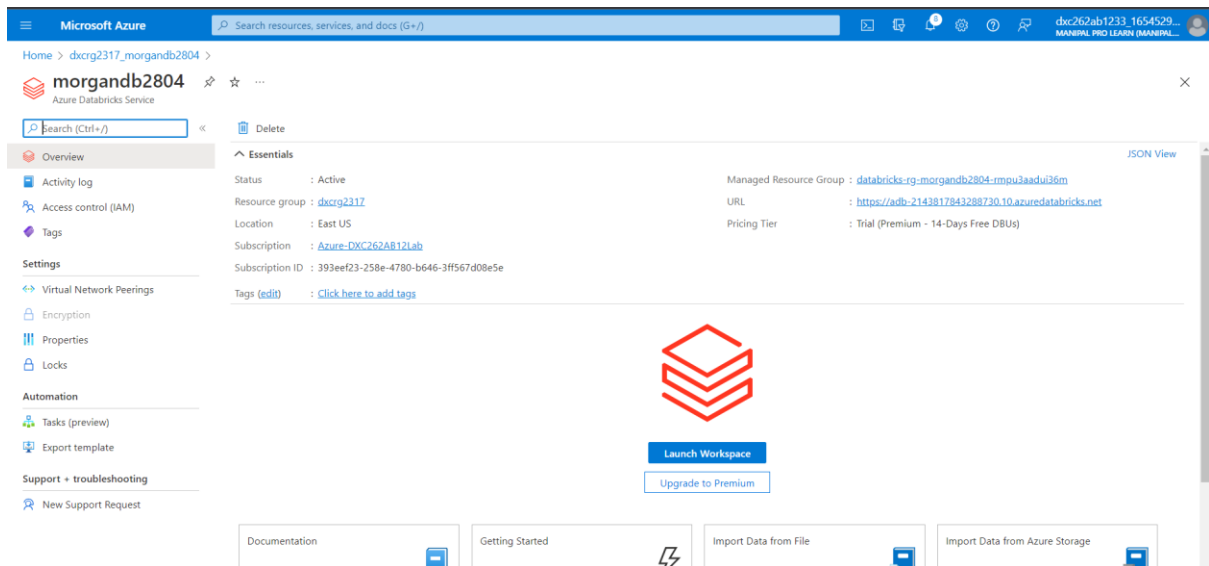
No

Create

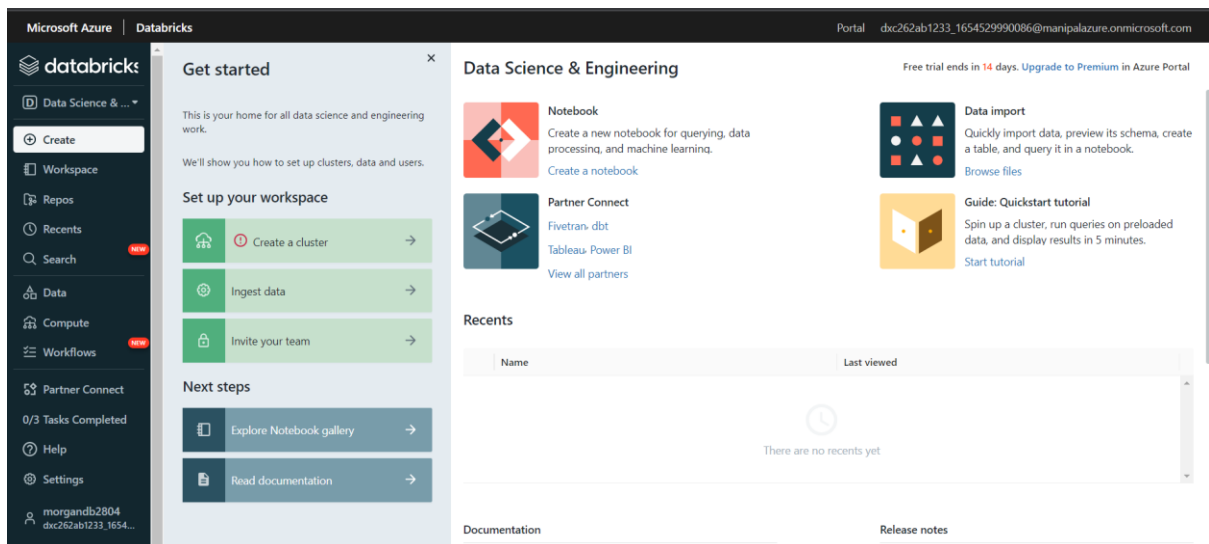
< Previous

[Download a template for automation](#)

Step 4: We get Your deployment is complete

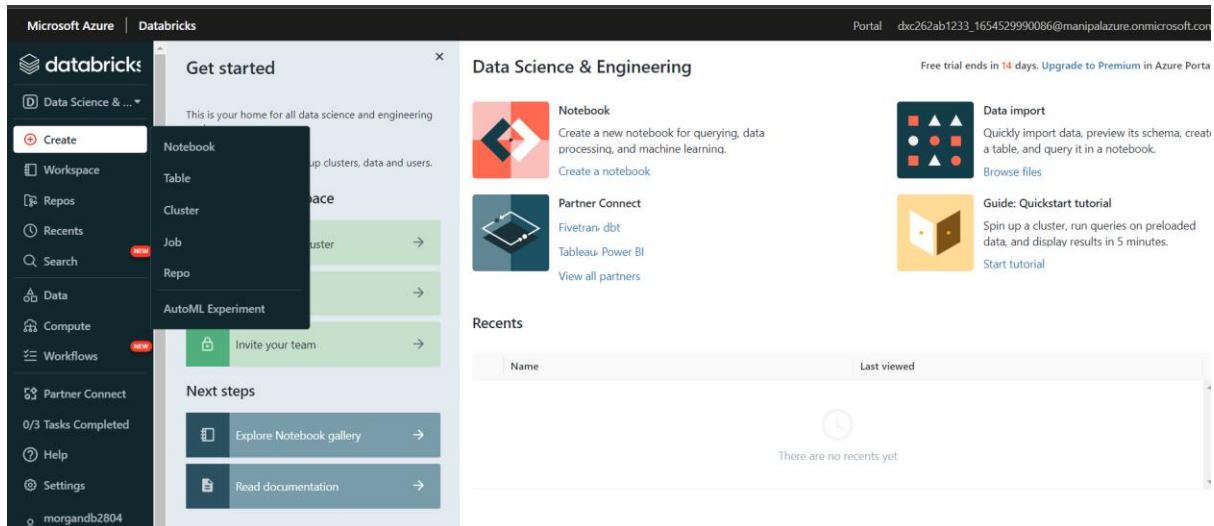


Step 5: Launch Workspace to go to data bricks that we created
This is the how page of data bricks we created

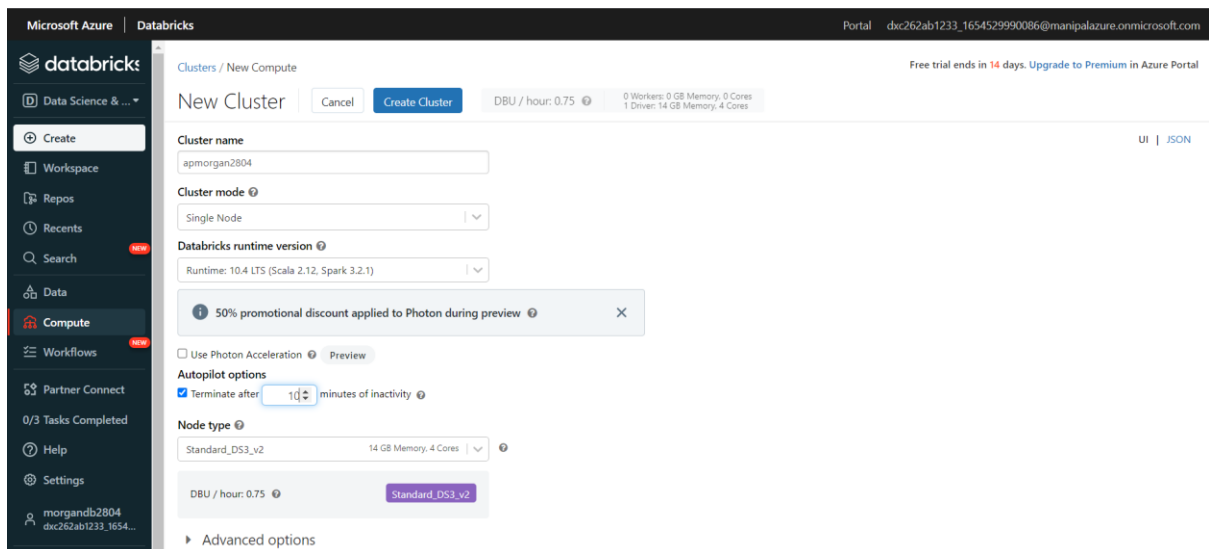


Practical Lab: Create **Cluster** in Azure Databricks

Step 1: After we create data bricks come to home page in that click on create cluster



Step 2: Give the cluster name and cluster node single node and terminating time 10 minutes and create cluster



Step 3: We get the tick mark with cluster name and cluster is created

Microsoft Azure | Databricks

Portal dxc262ab1233_1654529990086@manipalazure.onmicrosoft.com

Free trial ends in 14 days. Upgrade to Premium in Azure Portal

Clusters / apmorgan2804

apmorgan2804

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

Autopilot options

Terminate after 10 minutes of inactivity

Node type Standard_DS3_v2 14 GB Memory, 4 Cores

DBU / hour: 0.75

Advanced options

Step 4: To check the complete details of cluster we created go to compute and get the details of it

Microsoft Azure | Databricks

Portal dxc262ab1233_1654529990086@manipalazure.onmicrosoft.com

Free trial ends in 14 days. Upgrade to Premium in Azure Portal

Compute

All-purpose clusters Job clusters Pools Cluster policies

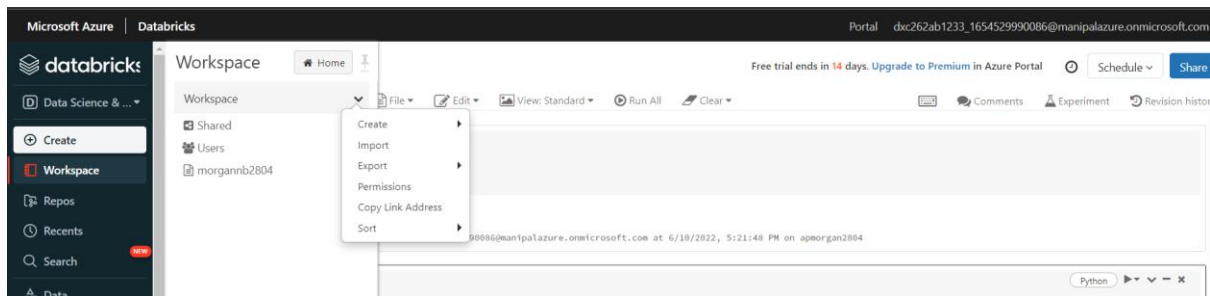
Create Cluster

Created by me Accessible by me

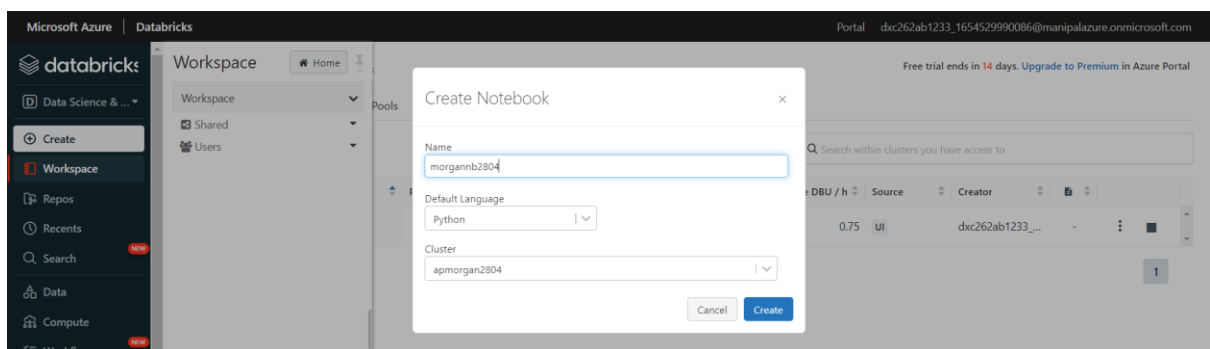
Name	Policy	Runtime	Active memory	Active cores	Active DBU / h	Source	Creator
apmorgan2804	-	10.4	14 GB	4 cores	0.75	UI	dxc262ab1233_...

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

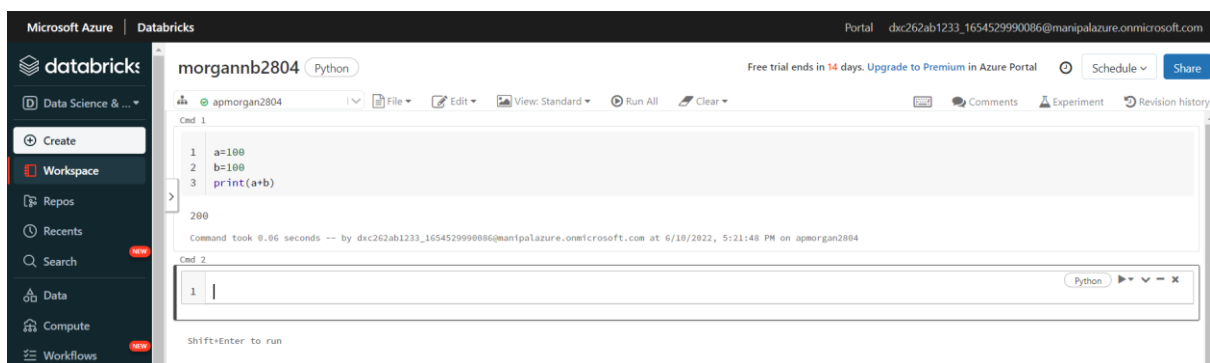
Step 1: To add notebook Go to Workspace-> create-> Notebook



Step 2: give name of notebook and cluster you created and click on create

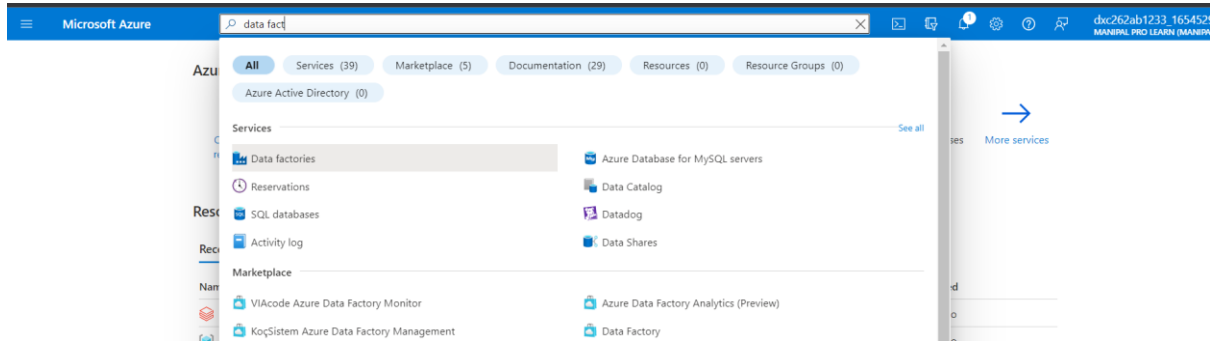


Step 3: to implement a notebook we created after create we will come to this page here we need to attach cluster and I have wrote a small program and run all it will show in seconds

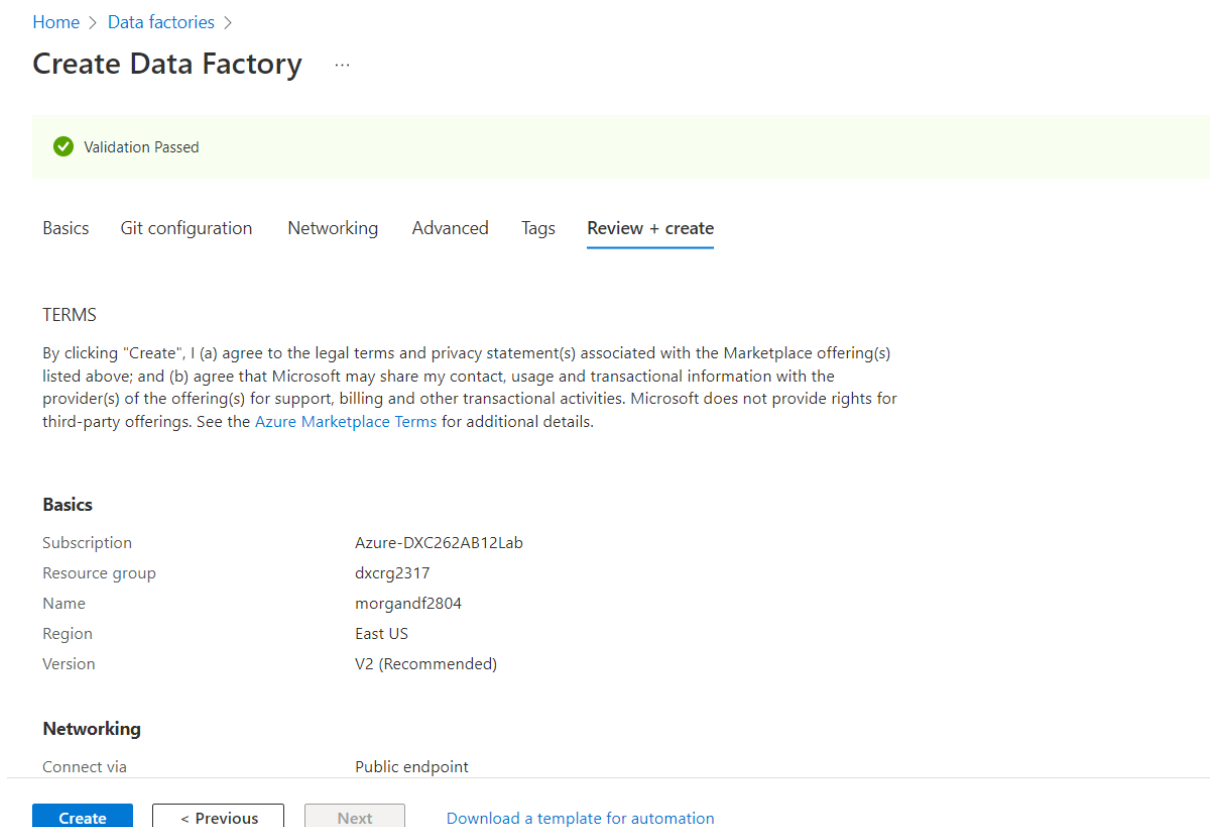


Practical Lab: Azure Data Factory For AP Morgan

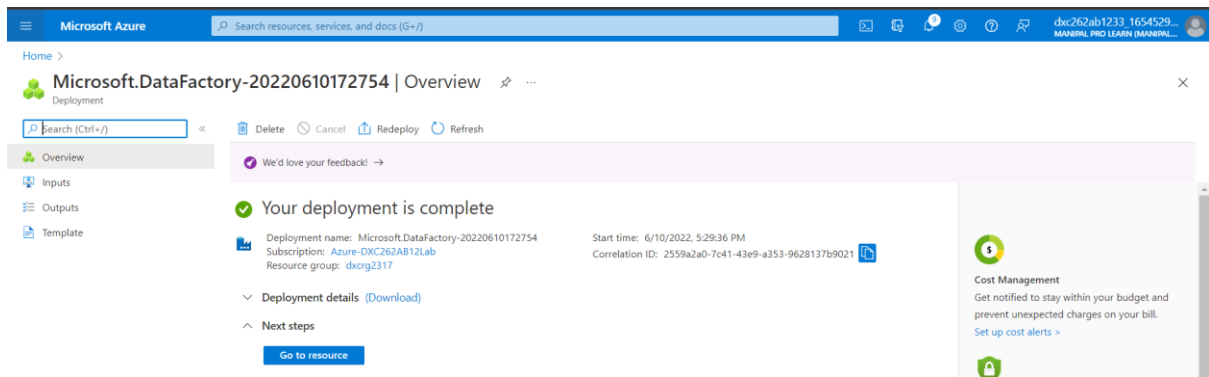
Step 1: First we need to open Azure and search for Data Factory and click on it



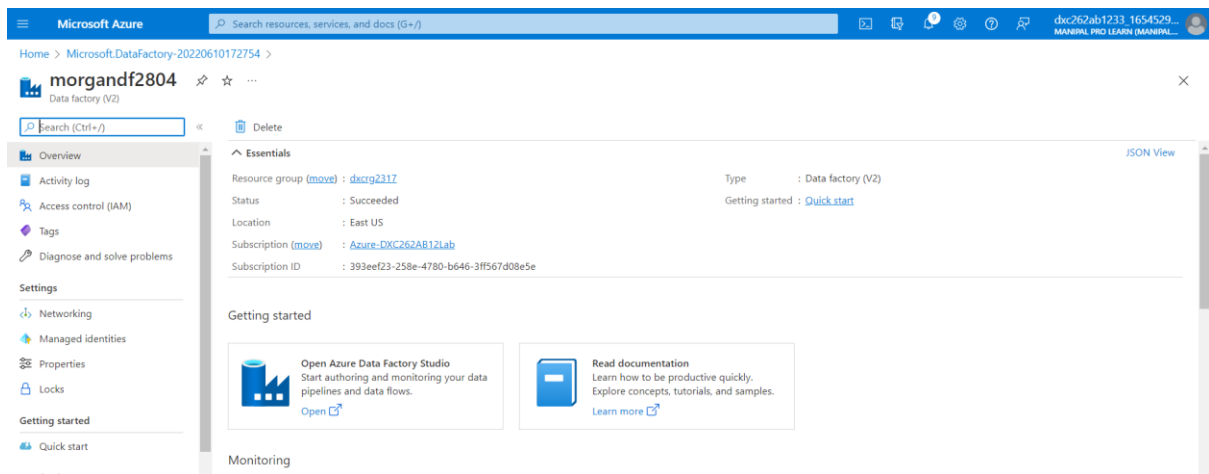
Step 2: Select data factory and click the + create and Give the Resource group name and Name morgandf2804 and click next Git configuration



Step 3: After validation click on create we get Your deployment us complete

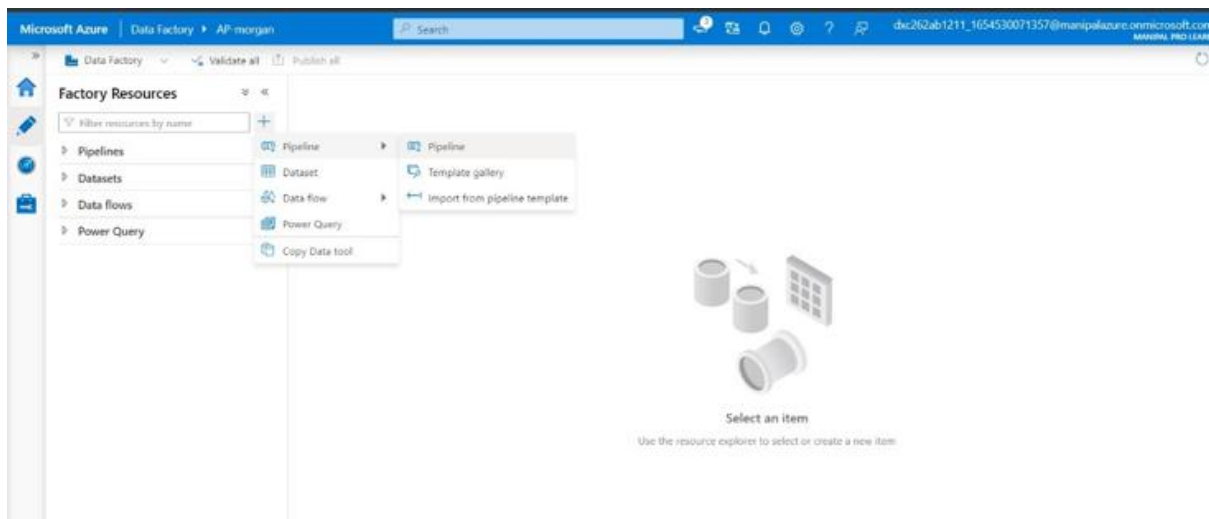


Step 4: Go to resource and open azure data factory to access

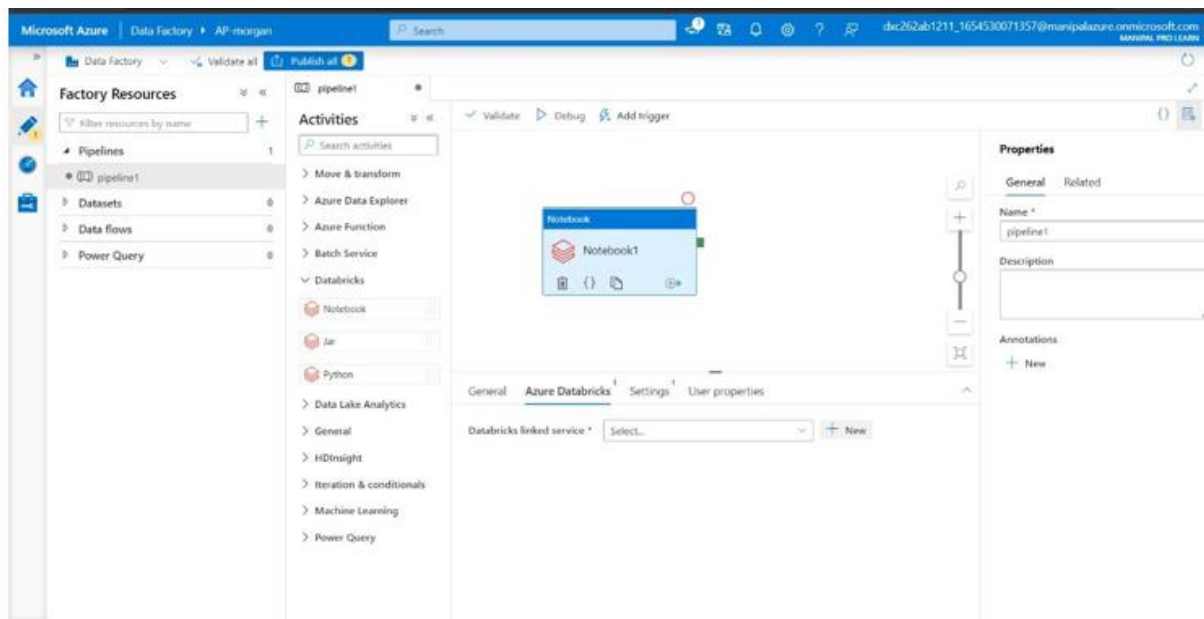


Practical Lab: Create Azure Databricks Linked Service in ADF

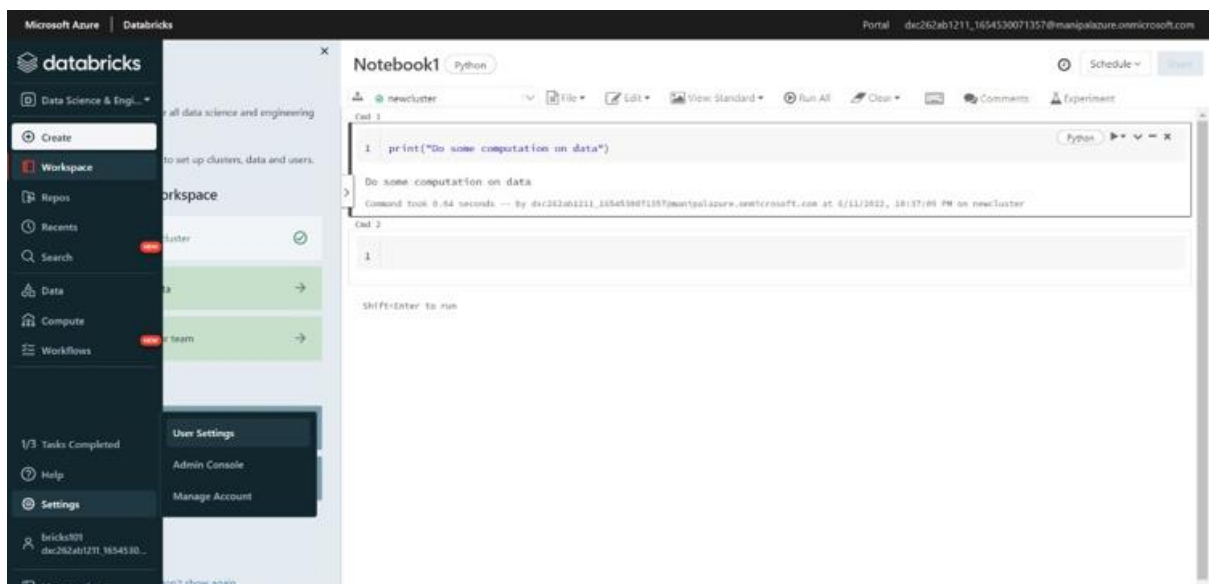
Step 1: In data factory we create a new pipeline that will help the data bricks linked service



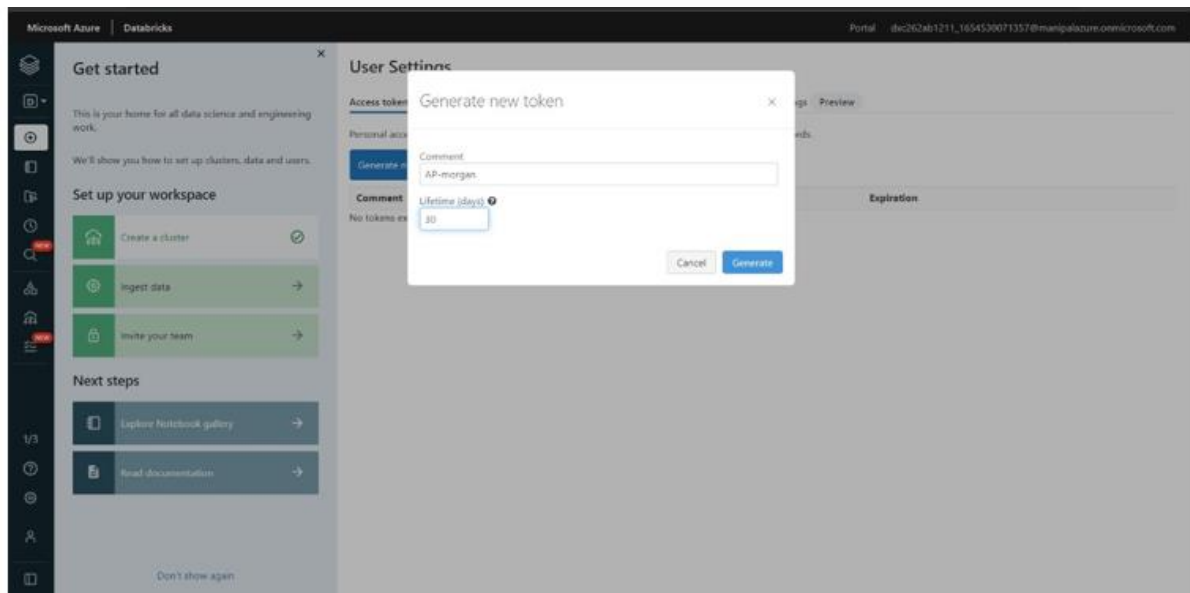
Step 2: In that pipeline drag and drop the notebook you have created from data bricks option



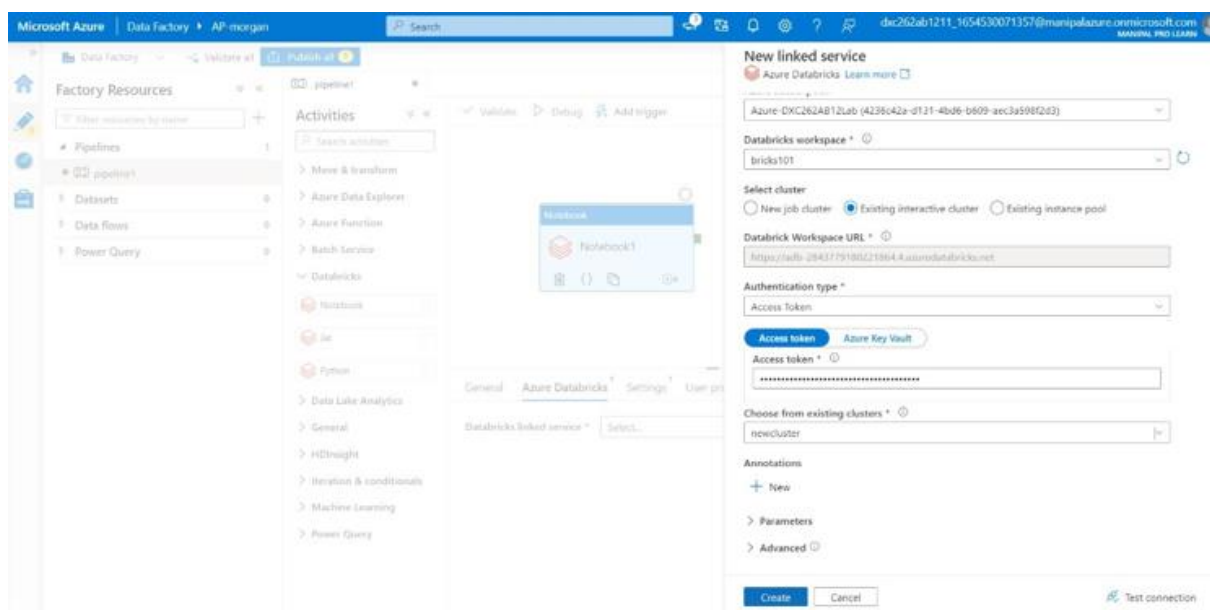
Step 3: First we need to go to data bricks and go to settings tab and user settings



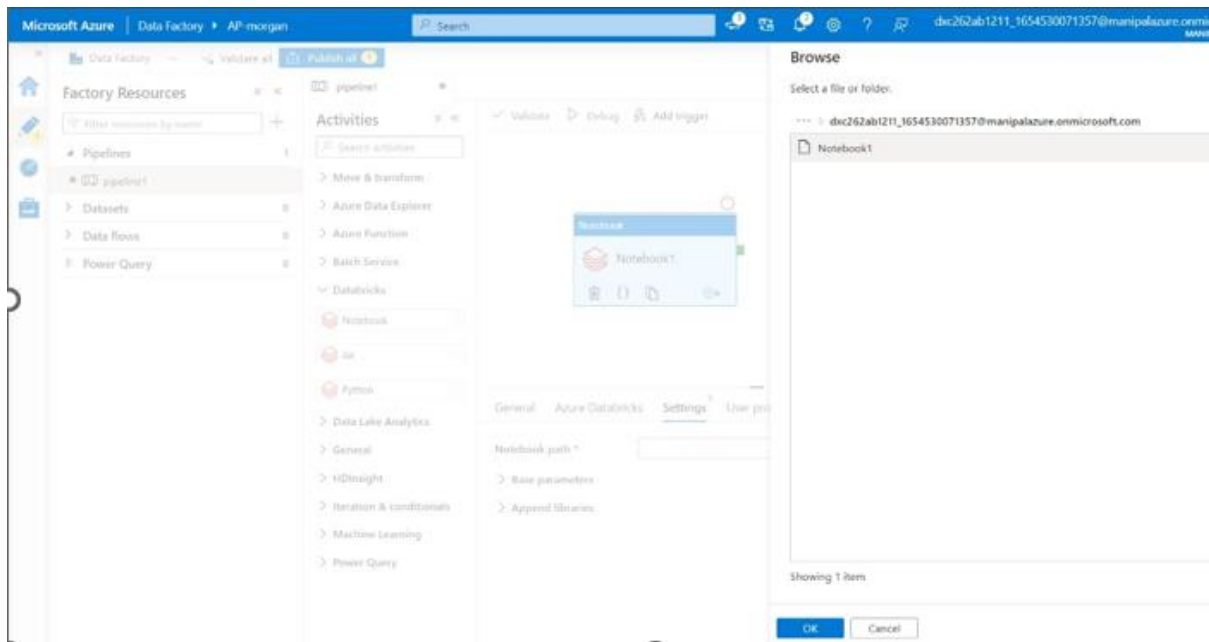
Step 4: In user settings go to Generate new token by giving name and click on generate and we get the token copy that and paste it in the azure factory connection details



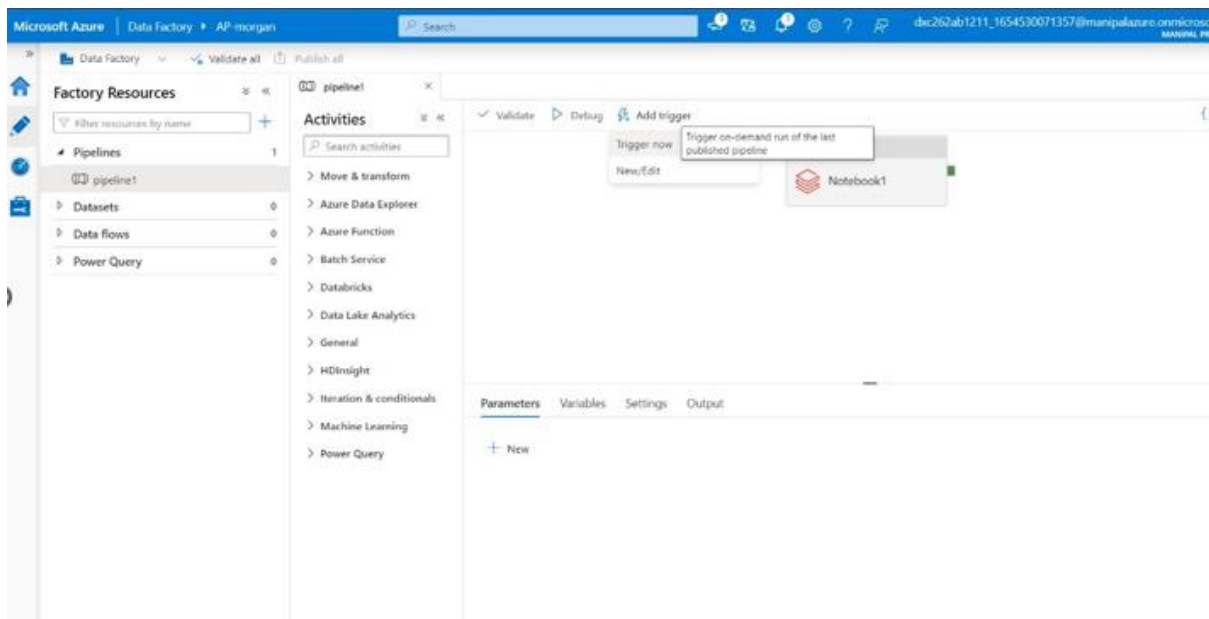
Step 5: Choose the cluster you have created on data bricks and click on create



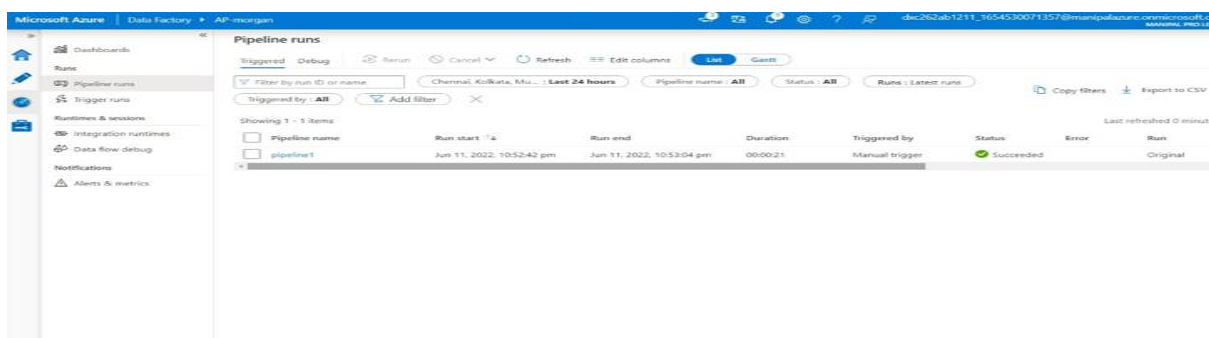
Step 6: After that choose the notebook in the pipeline tab to get trigger first publish the pipeline because it will trigger only when it get published



Step 7: Click on add trigger and go to trigger now



Step 8: In data Factory studio go to monitor tab here we can check the execution



Here from this we have successfully linked notebook from data bricks to data factory

Result: created the data bricks , cluster in data bricks and notebook in data bricks and data factory

Conclusion: From all this we got to know how to create a data brick, cluster in data bricks and notebook and how to run the code in data bricks and also got how to create a data factory what are the details to give and how to create

References: The class said by sir