# 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

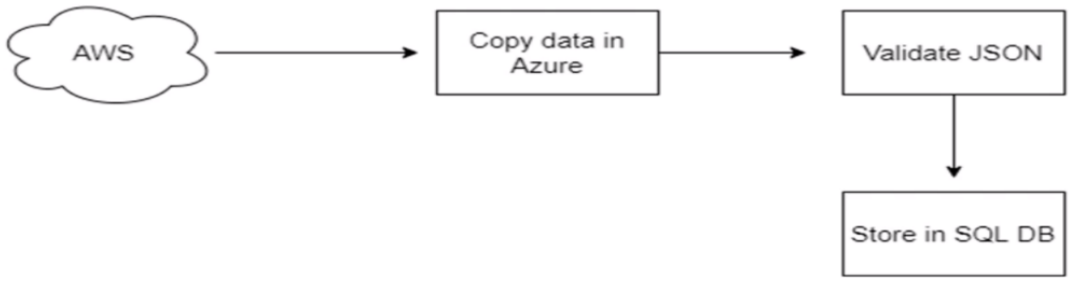
# Project 1 : Connected Vehicles

- 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.

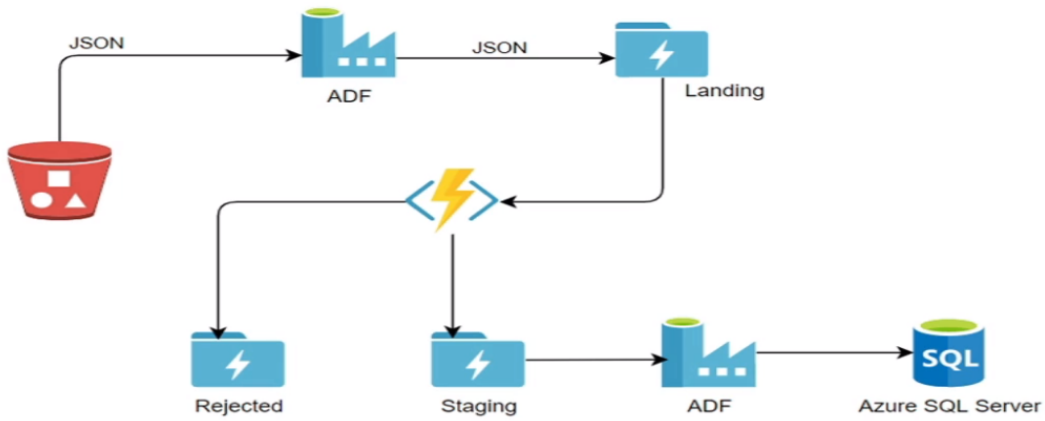

# Project 1 : Connected Vehicles

- 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



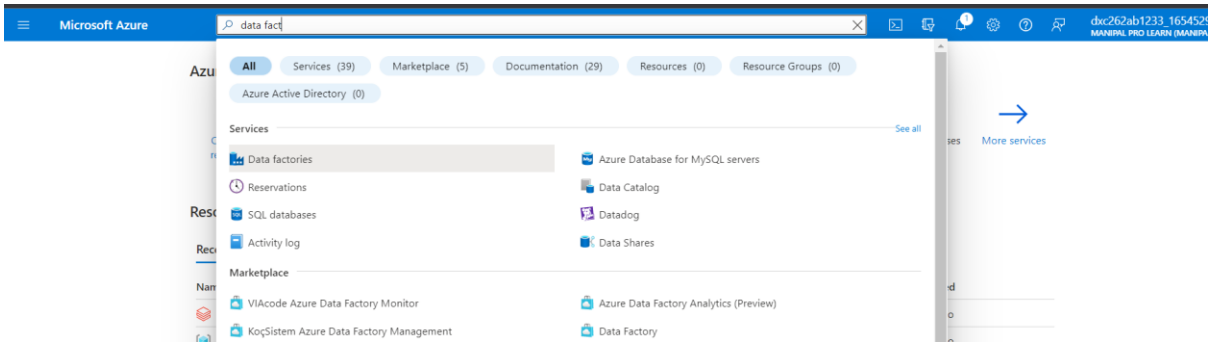

# Project 1 : Connected Vehicles



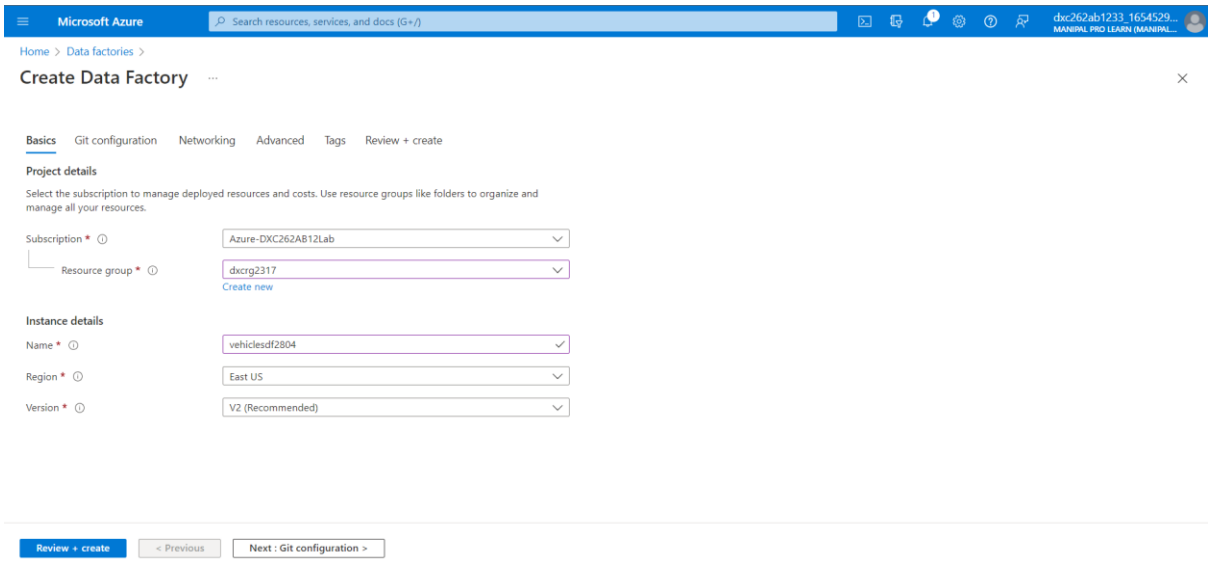

Architecture Diagram for Connected Vehicle Project

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

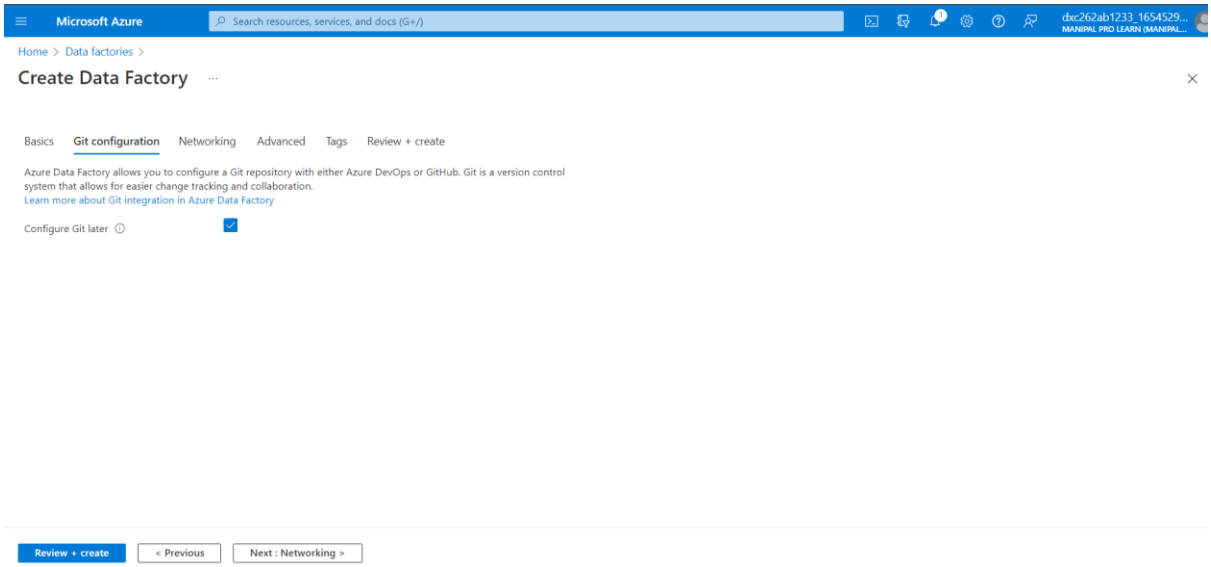
Step 1: First we need to open Azure and search for Data Factory and click on it



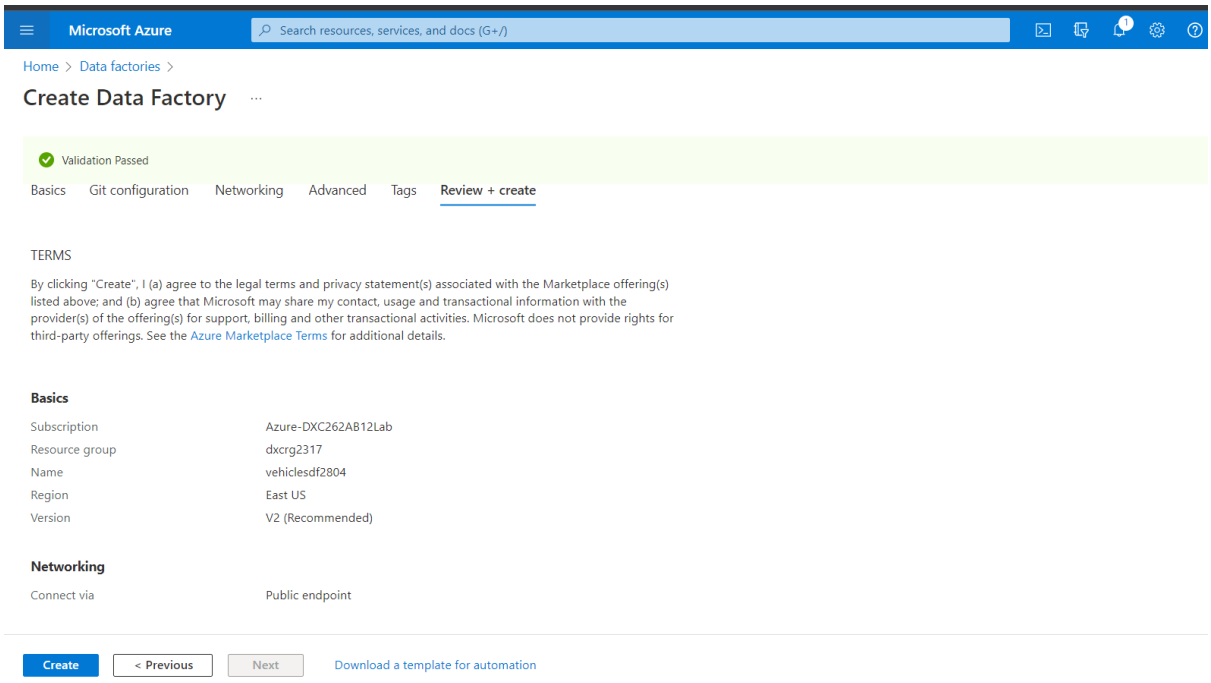

Step 2: Select data factory and click the + create and Give the Resource group name and Name vehiclesdf2804 and click next Git configuration



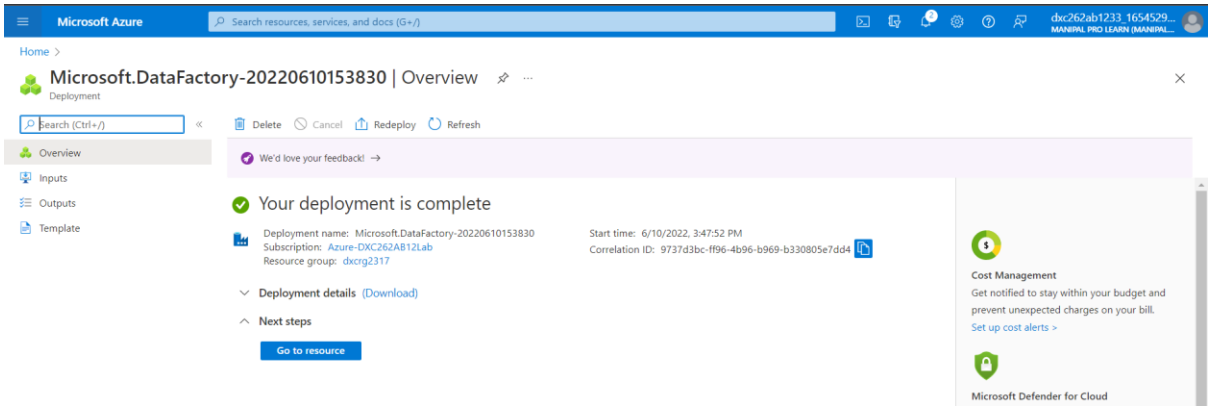

Step 3: Check box the git configuration later and all other are default and click Review + create



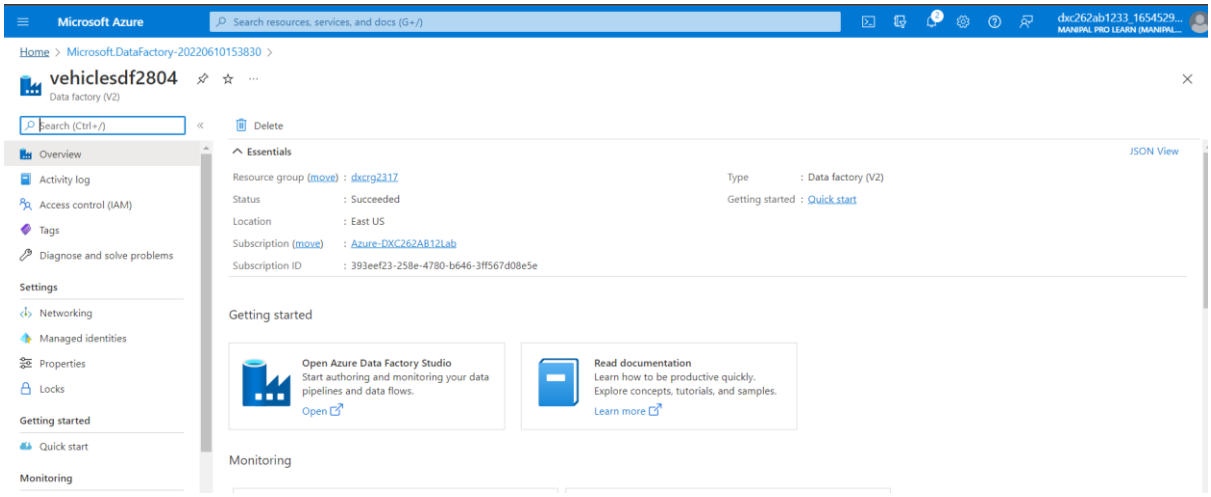

Step 4: After that we get the validation is passed and click on Create




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

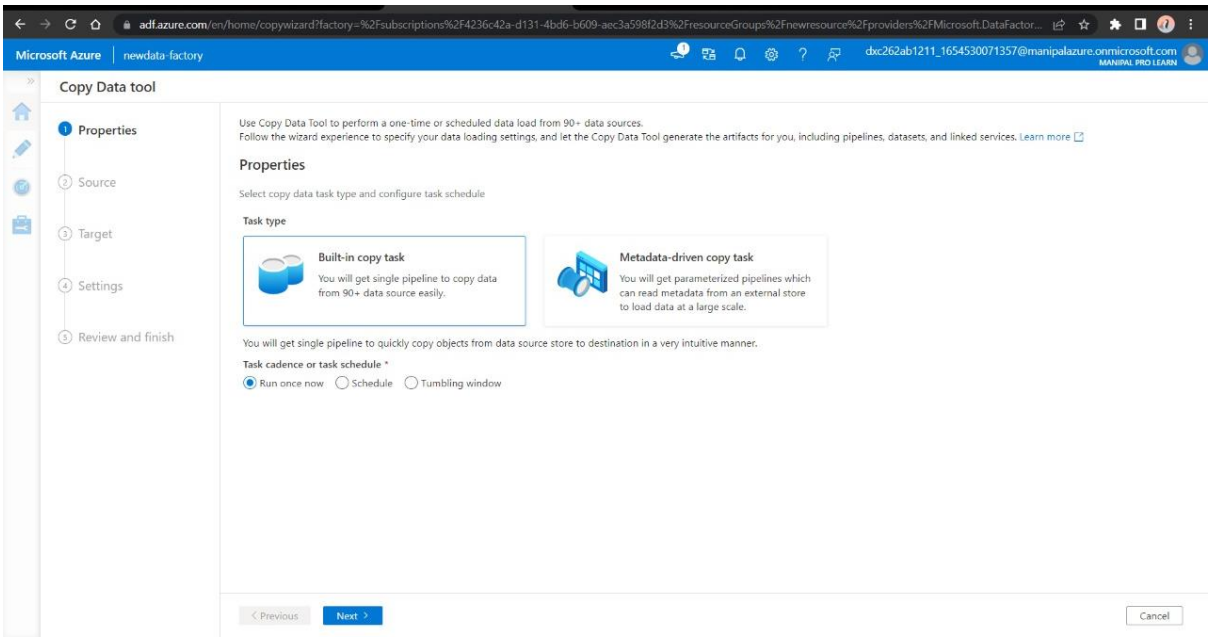



Step 6: click on Go to resource and in that click on open Azure data factory studio where data factory is created

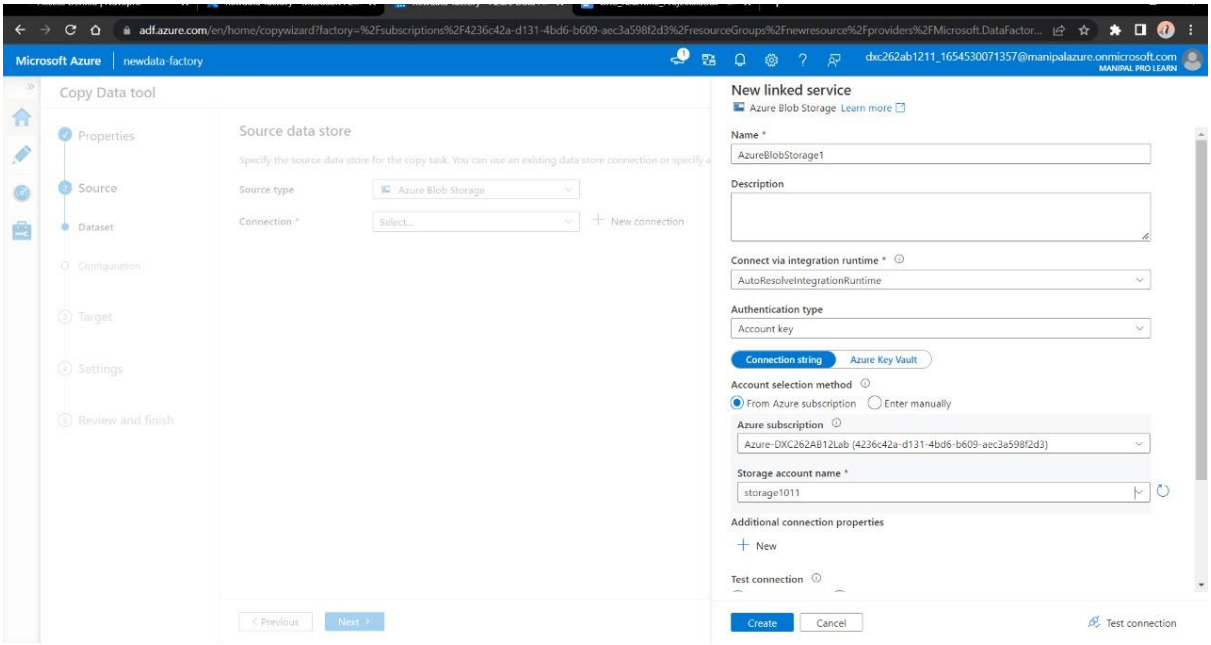


Practical Lab: Create **ADF Pipeline** End to end pipeline with triggers enabled

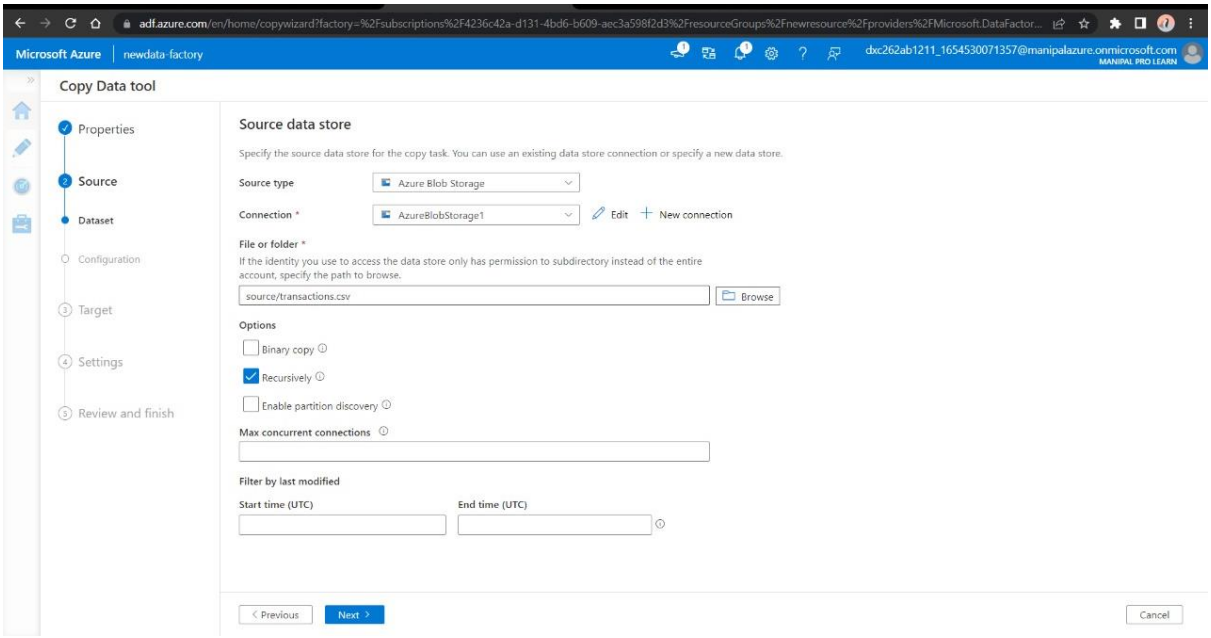
Step 1: Here we need to create a pipeline that will take data from blob storage and feed into the SQL database. Click on ingest



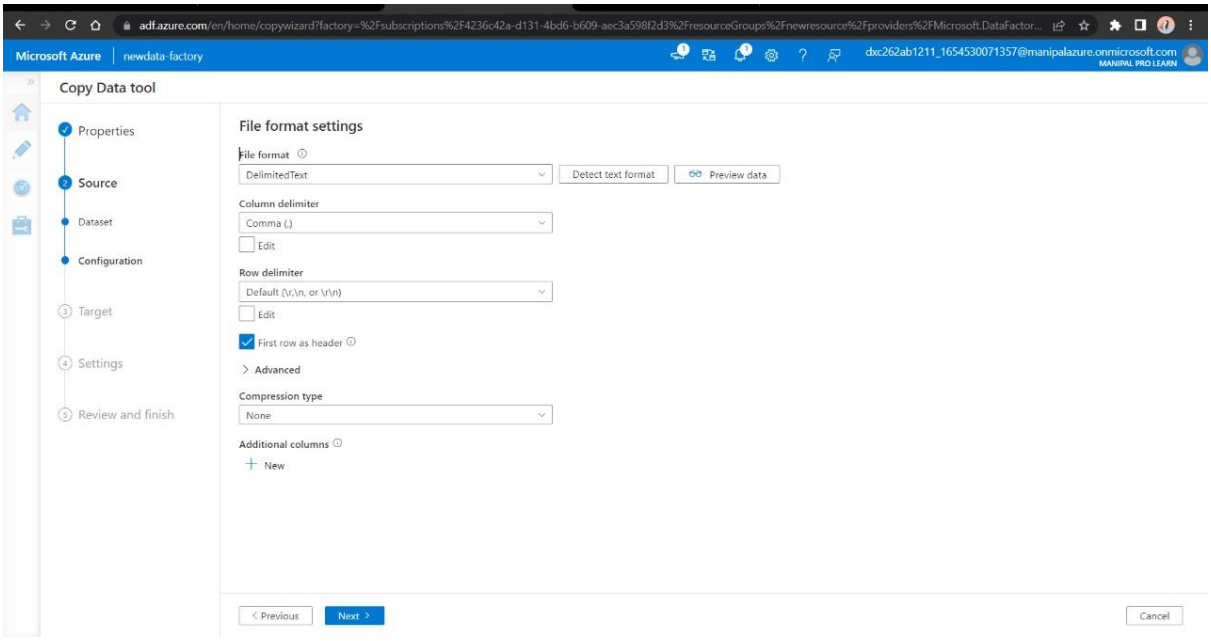

Step 2: We need to fill the name of Azure blob storage linked file. Select the name of storage account name we created in this and click on Create.



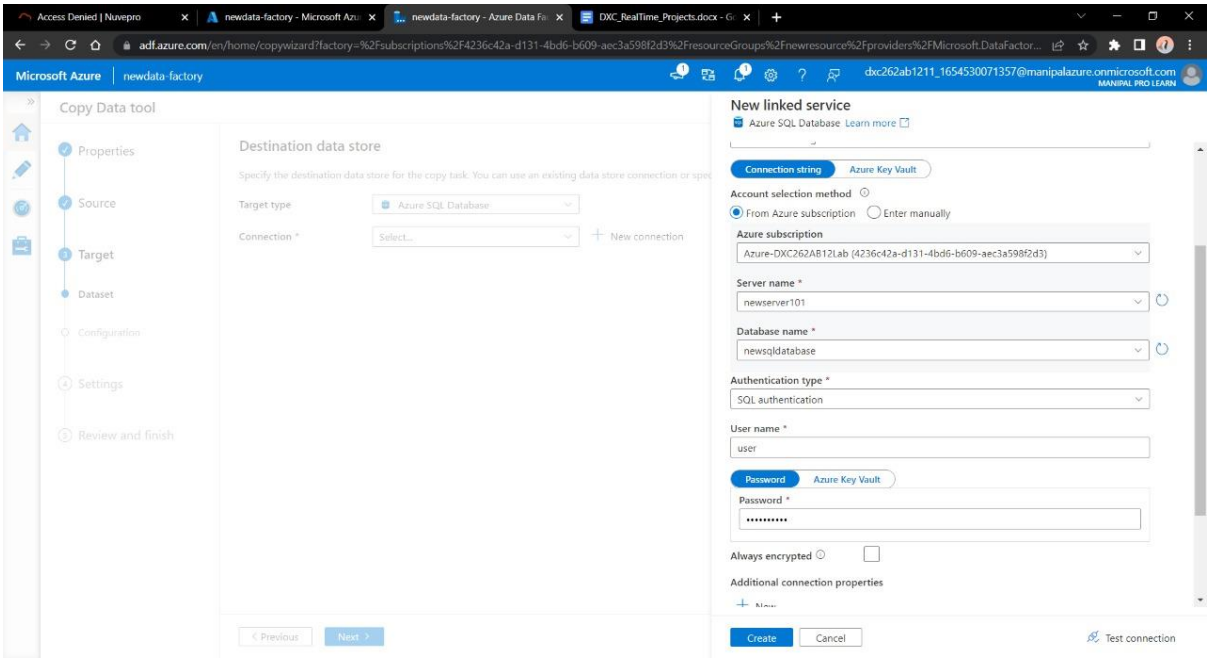

Step 3: Select the source type as blob storage and connection name that we created in previous step. Click  from Options. Click Next.



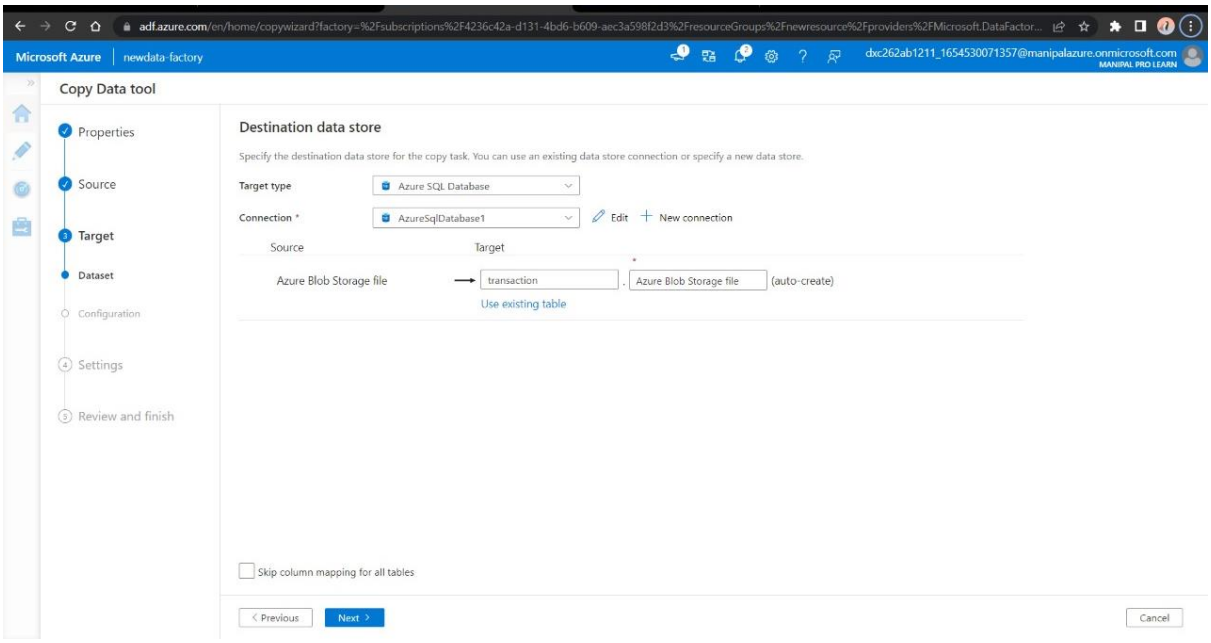

Choose the output format of the data



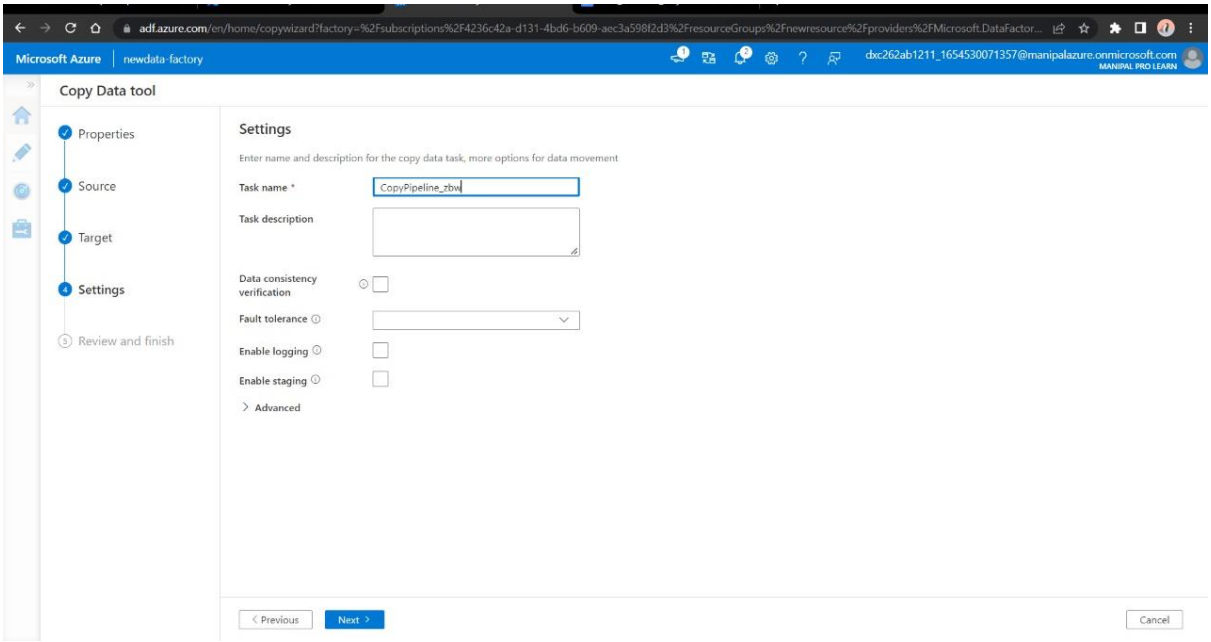

Step 4: Now, in the next step we will link the Data Factory with SQL database .Connect the database with new connection.



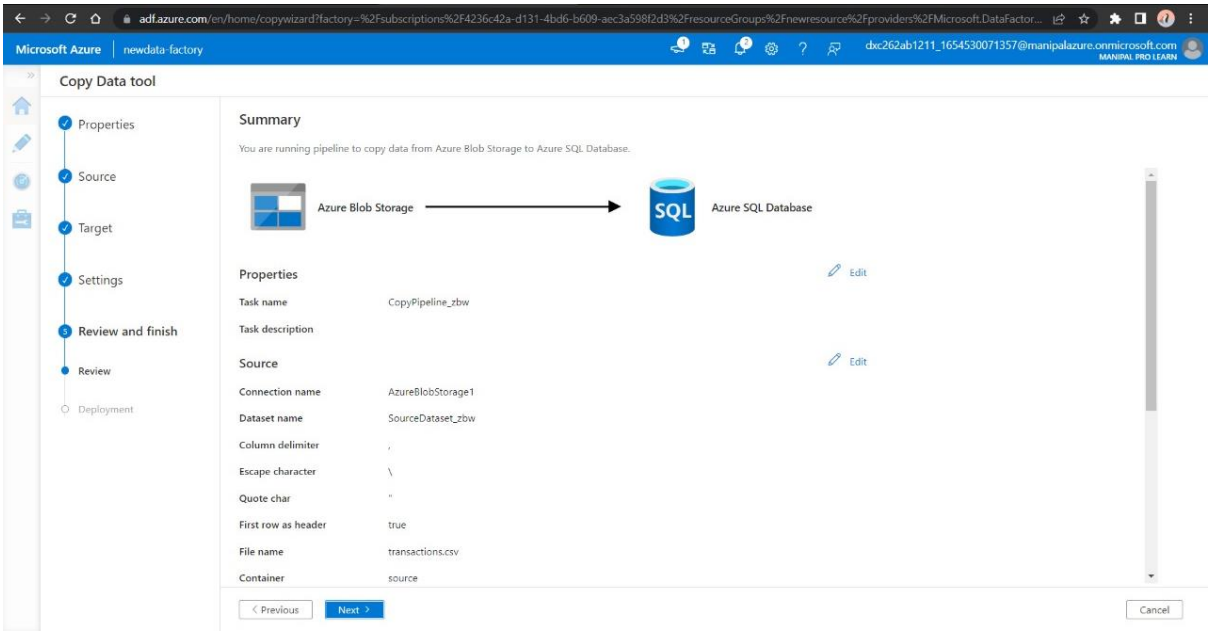
Step 6: Choose the file name you want to see in the SQL database. Click Next.



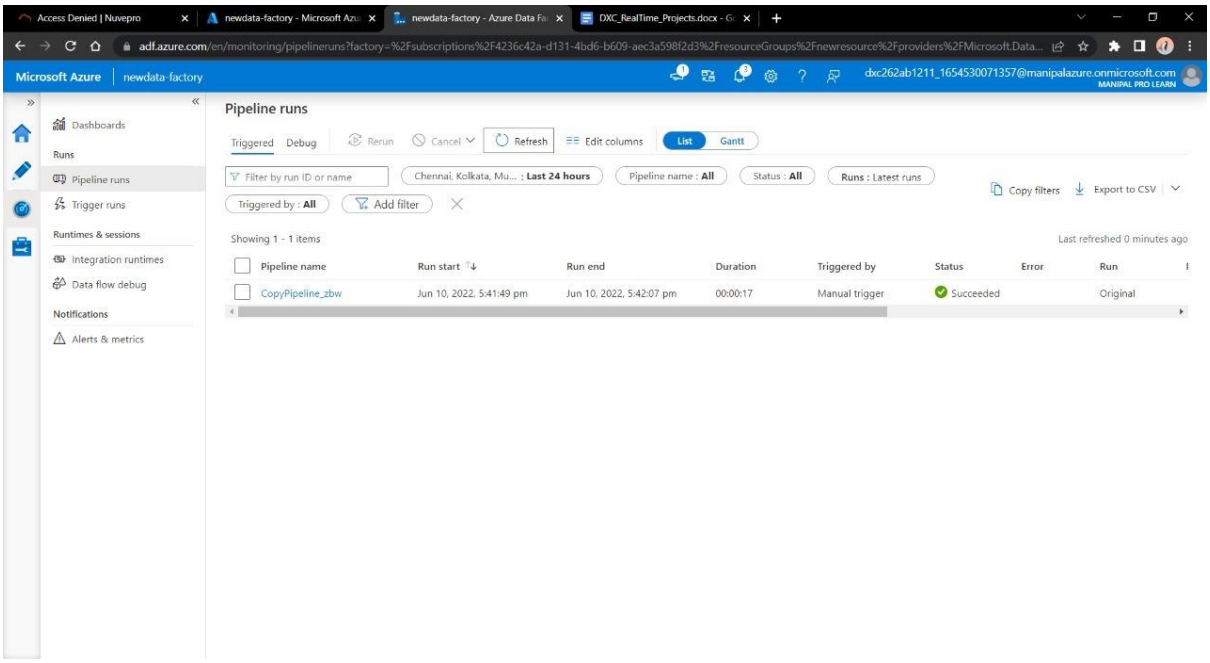

Step 7: Now we are the end stage of creating pipeline, click next.



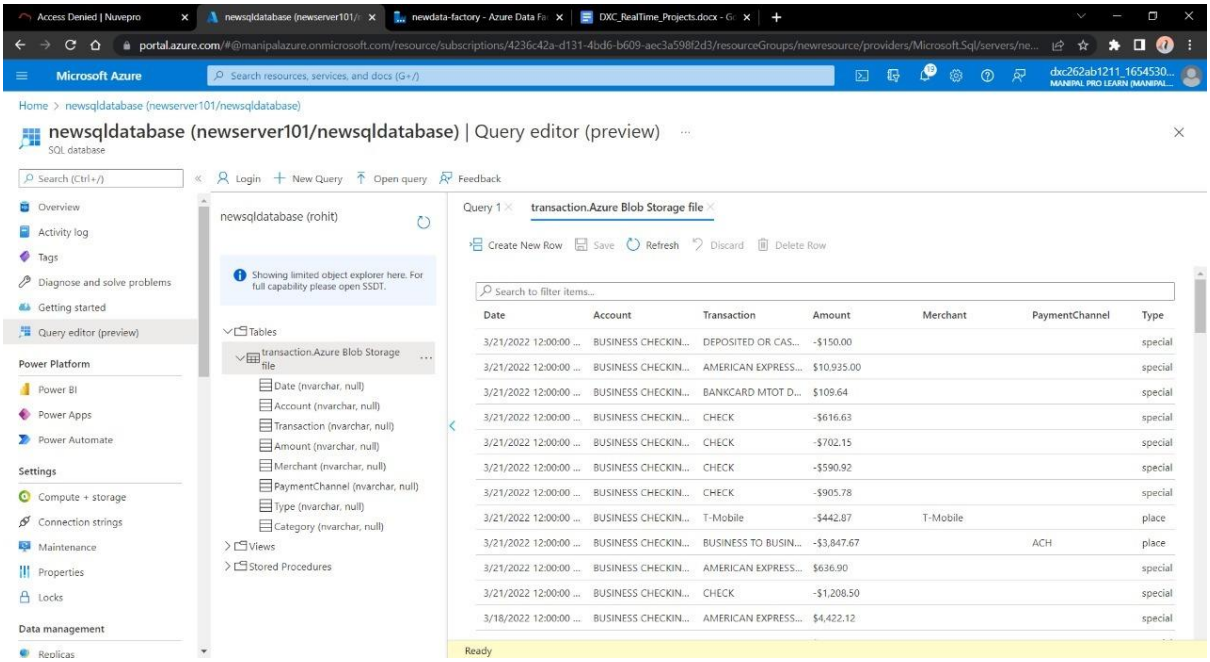
Step 8: Validate the  details and move forwards with Next




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

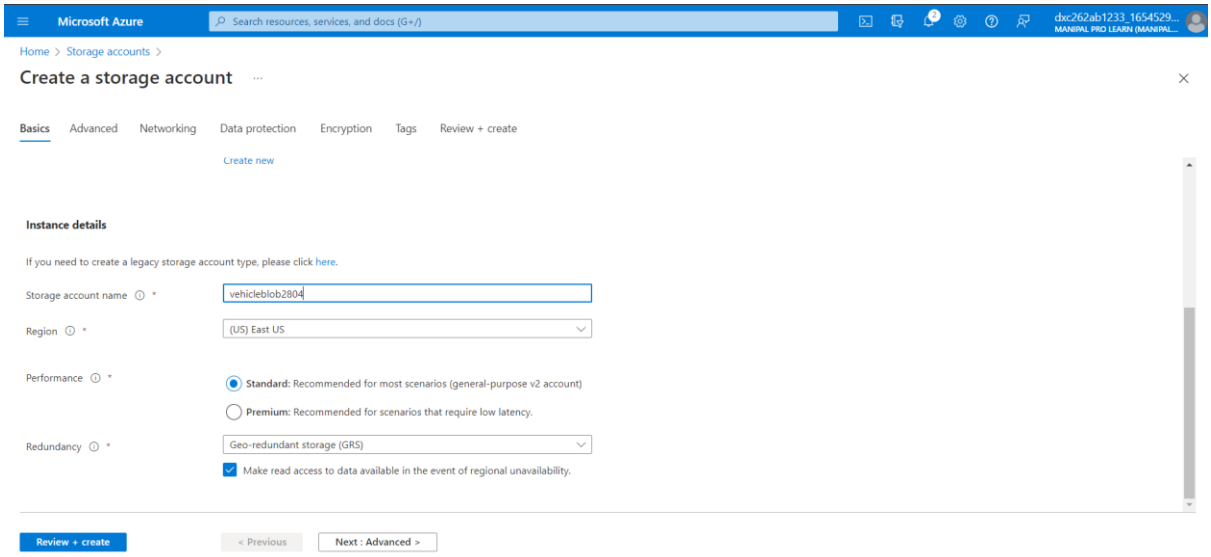



Step 10: Go to SQL database to know that the data is there

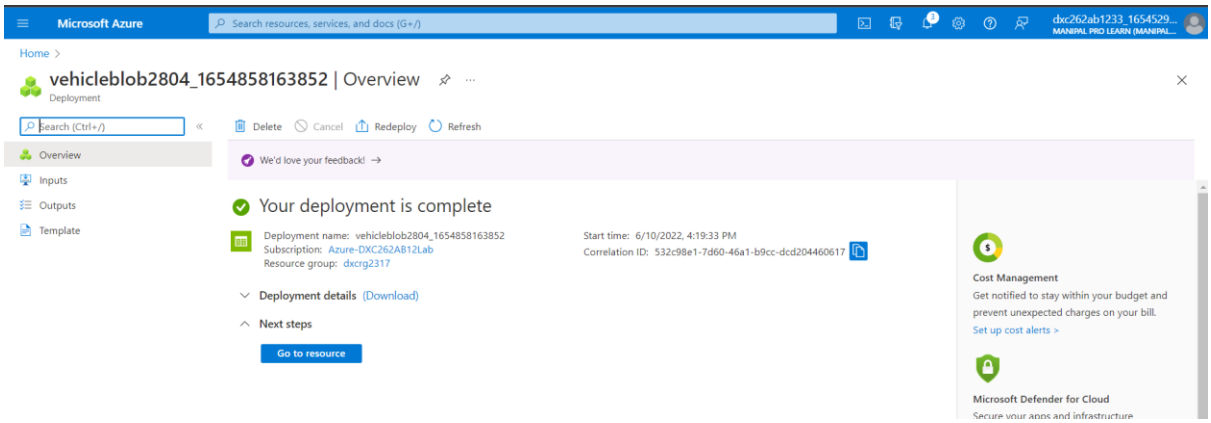

Access Denied | Nuvepro
newsqldatabase (newserver101/
newdata-factory - Azure Data Fa
DXC_RealTime_Projects.docx - G
portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/4236c42a-d131-4bd6-b609-aec3a598f2d3/resourceGroups/newresource/providers/Microsoft.Sql/servers/ne...
Microsoft Azure
Search resources, services, and docs (G+/)
dxc262ab1211_1654530...
MANIPAL PRO LEARN (MANIPAL...
Home > newsqldatabase (newserver101/newsqldatabase)
newsqldatabase (newserver101/newsqldatabase) | Query editor (preview)
SQL database
Search (Ctrl+/)
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
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

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

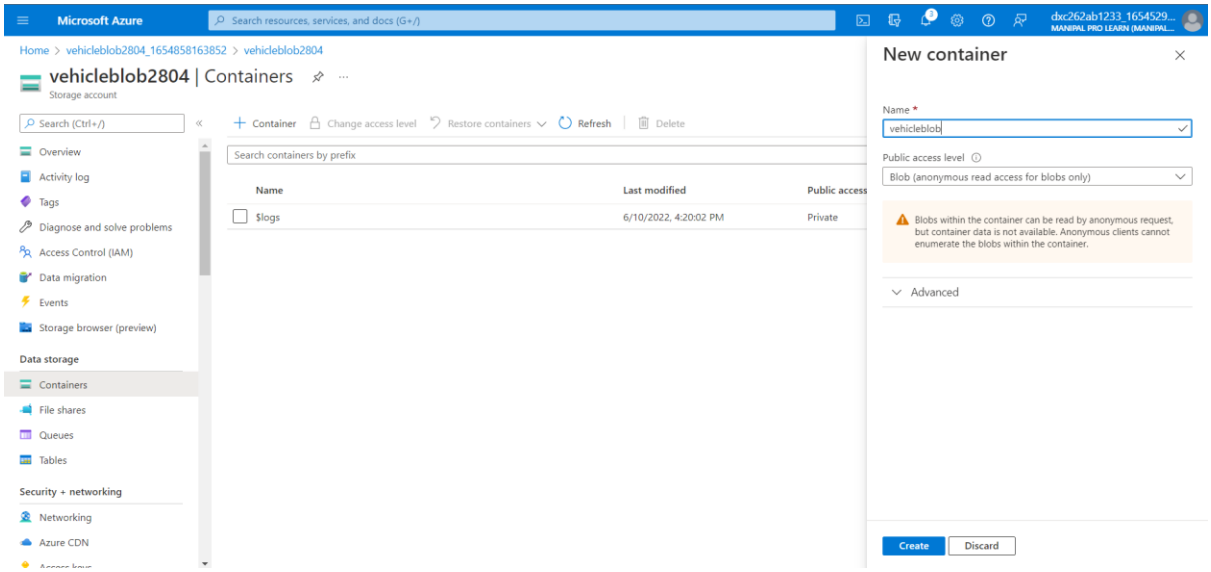
Step 1:  Search for Storage account and click + create



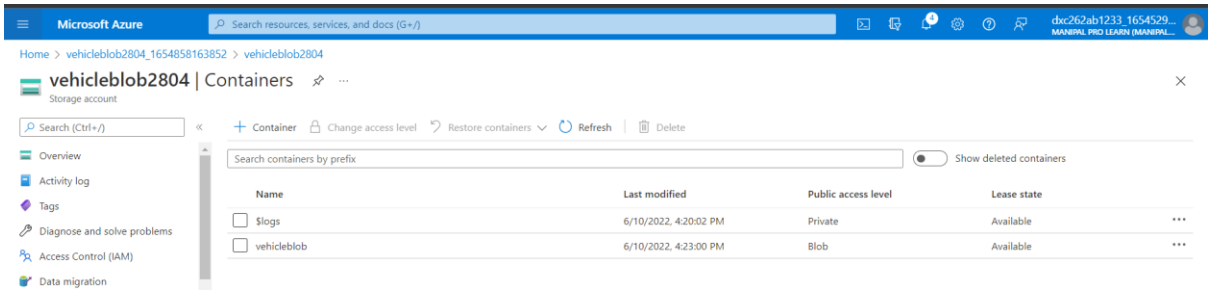

Step 2: Click on Review + create  and create




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

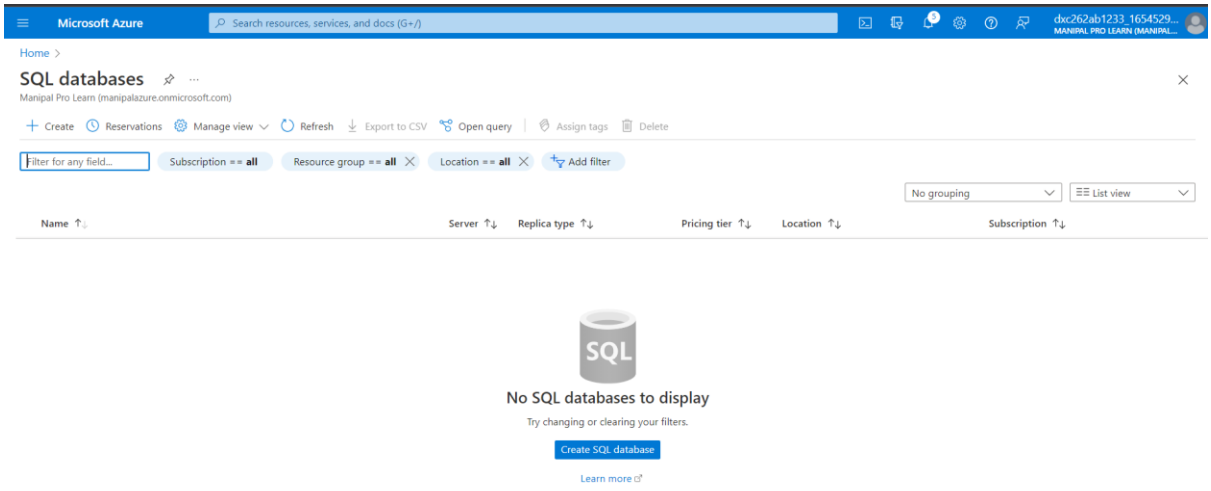



Step 4: Blob is created

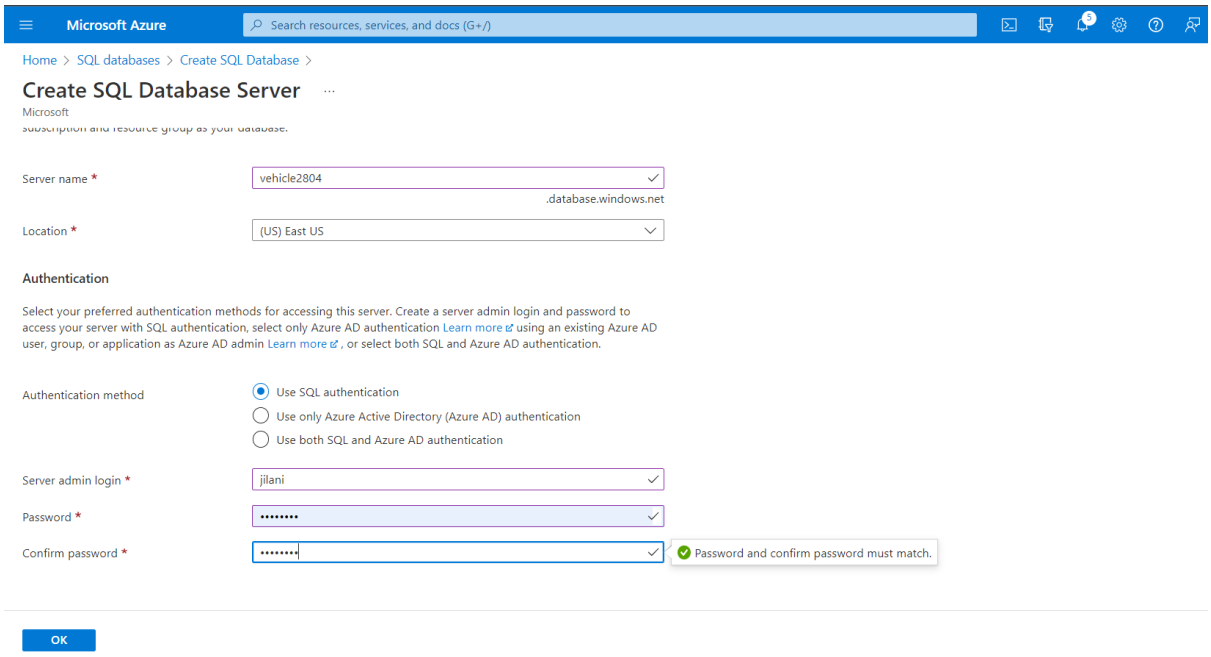



# Practical Lab: Create Azure **SQL Server and Database**

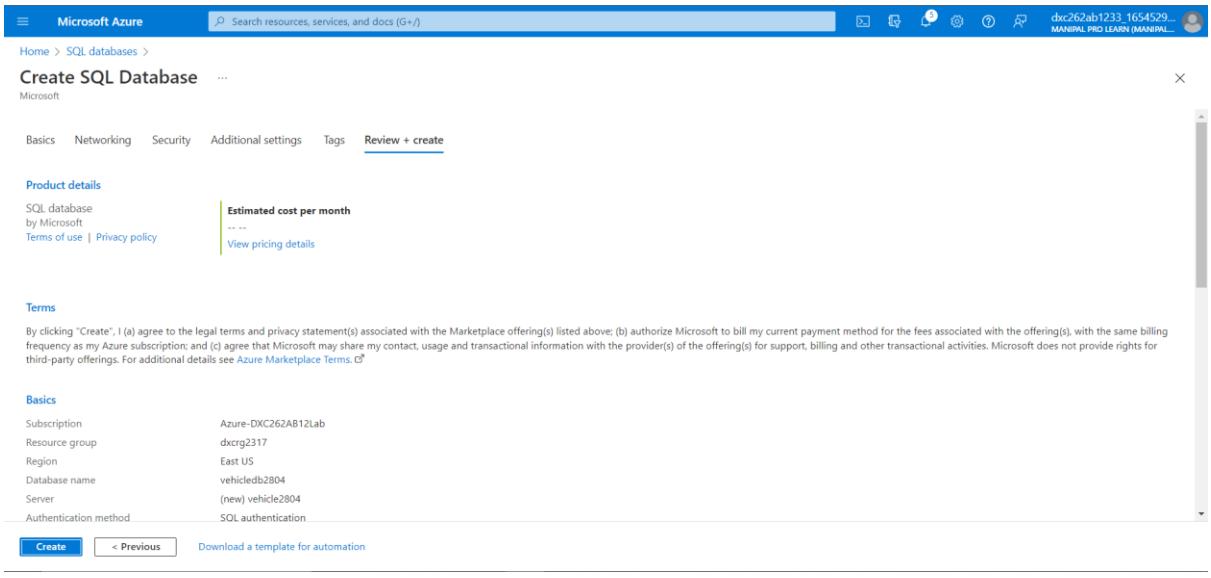
Step 1: search for sql data base and click + create



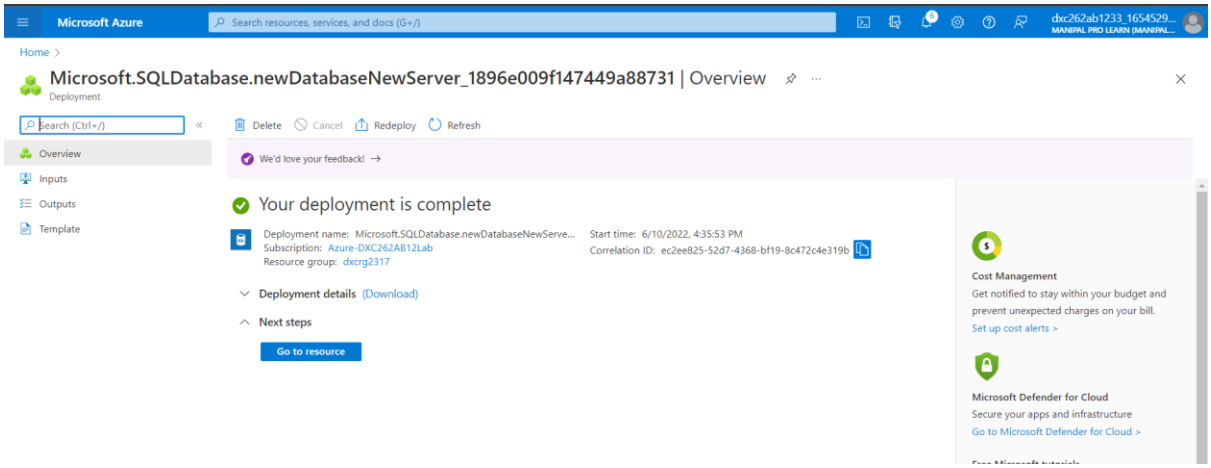
Step 2: Give the Resource name and Name of database and create new server give admin login and click ok



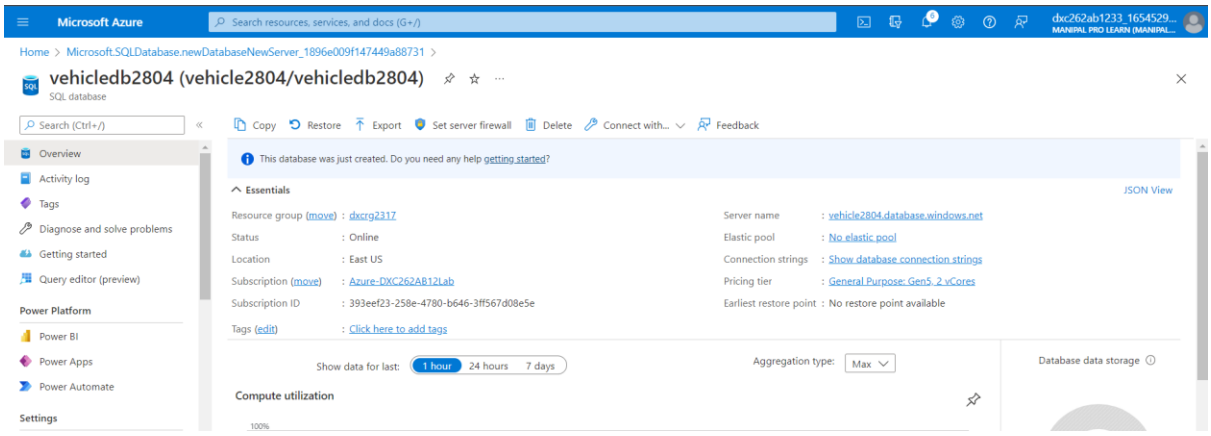

Step 3: All other options remain same and click review + create after that validation is passed click on create



Step 4: We will get Your deployment is complete




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



**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

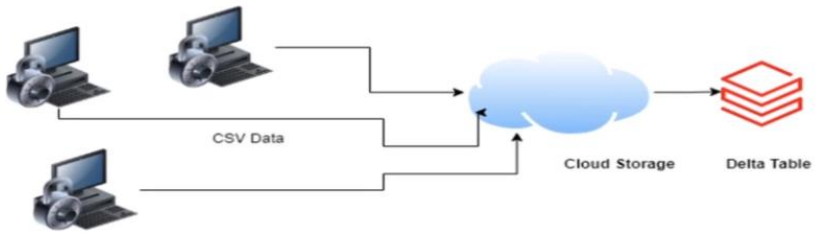
**Project2 Name: AP Morgan Data Platform**

Date: 10-6-2022

# Project 2: AP Morgan Data Platform

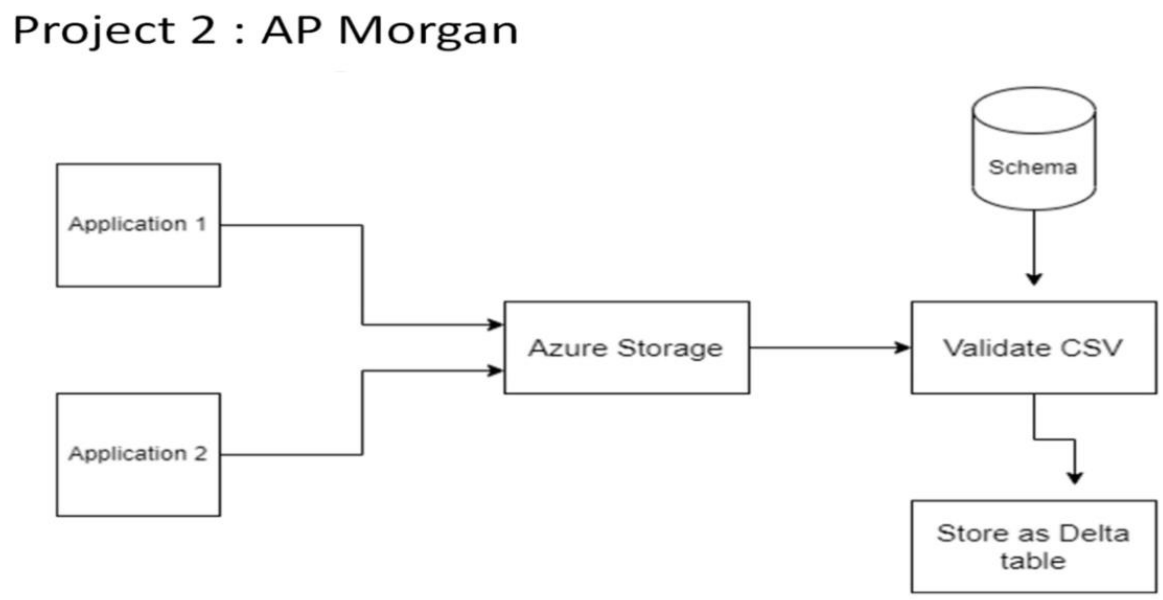
# Project 2 : AP Morgan

- 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.



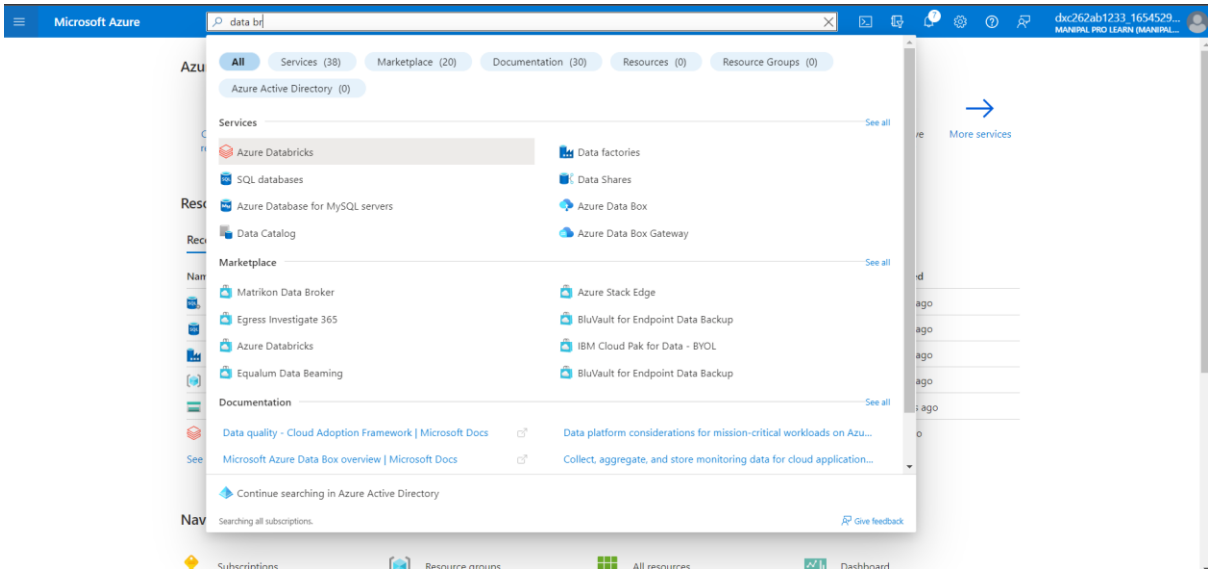

# Project 2 : AP Morgan- High Level Detail

- 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

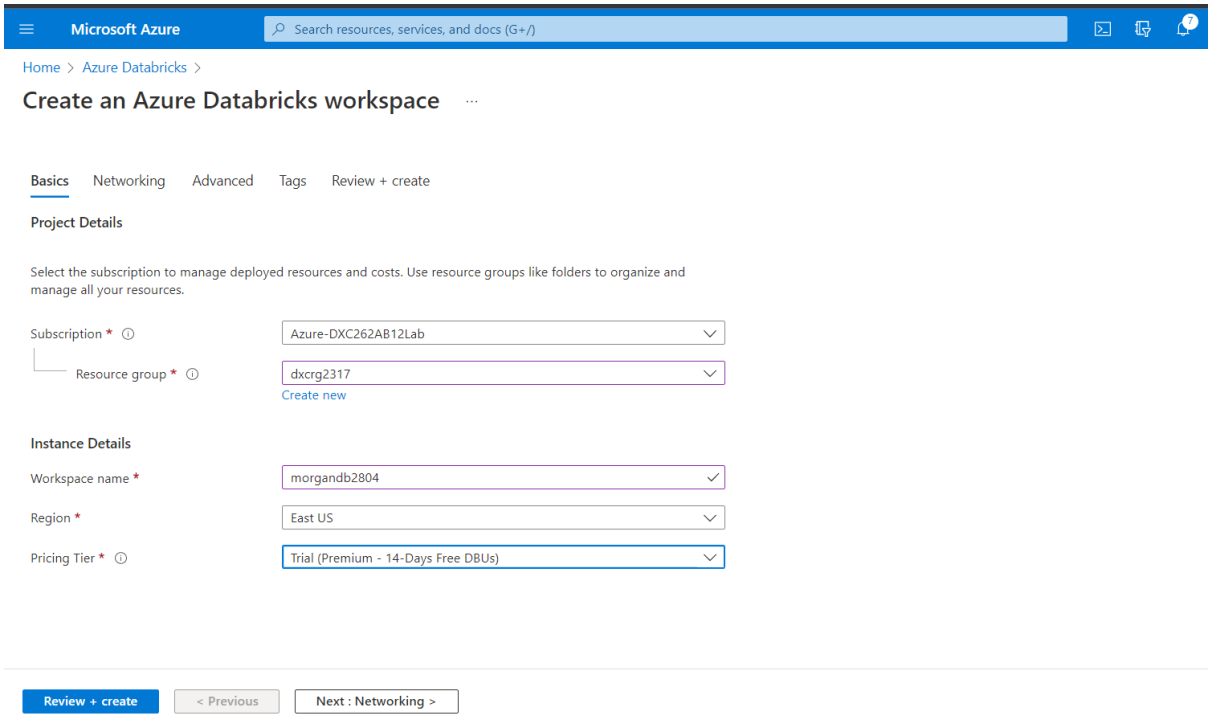



Practical Lab: Create a **Databricks**

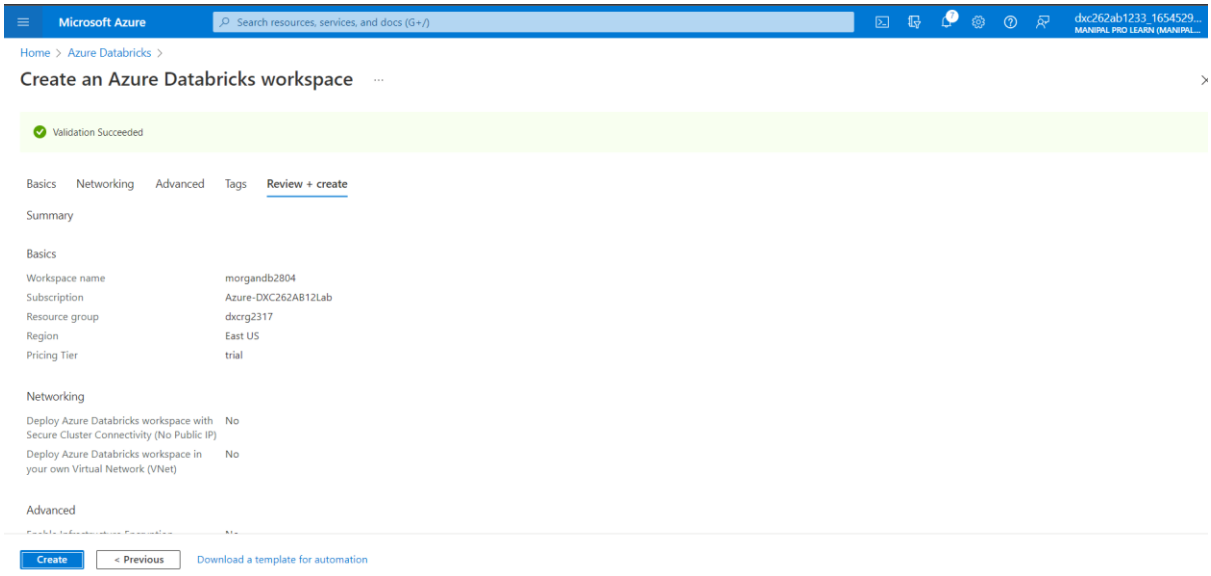
Step 1: Search for data bricks and select Azure data bricks



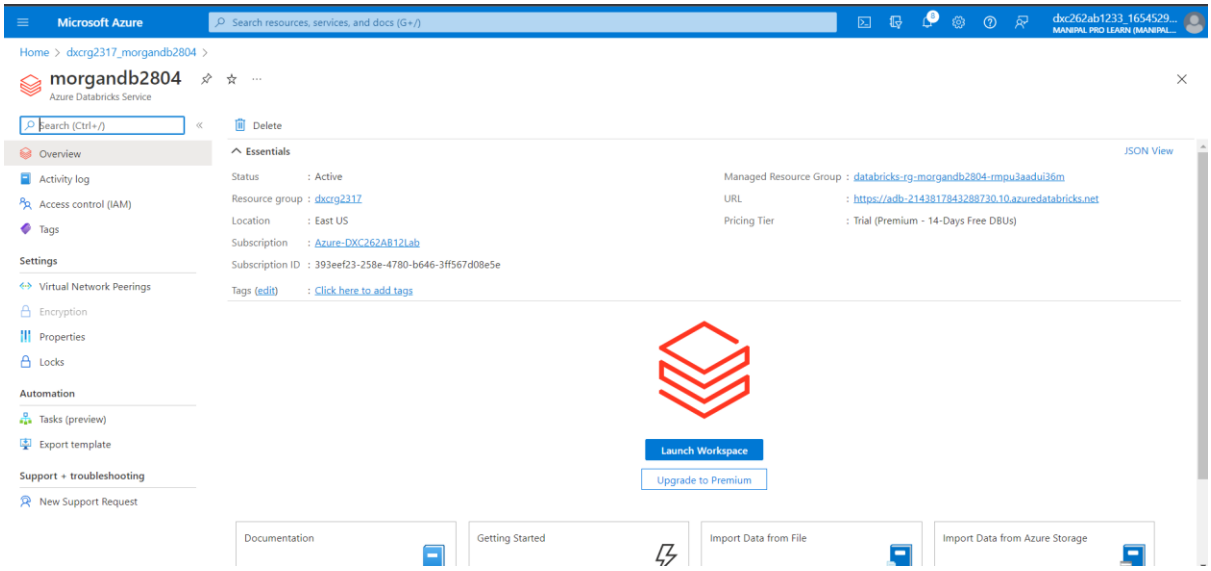

Step 2: Click on + create and fill the details in and pricing tier select Trial and all other are default and click review and create




# Step 3: After validation click on create

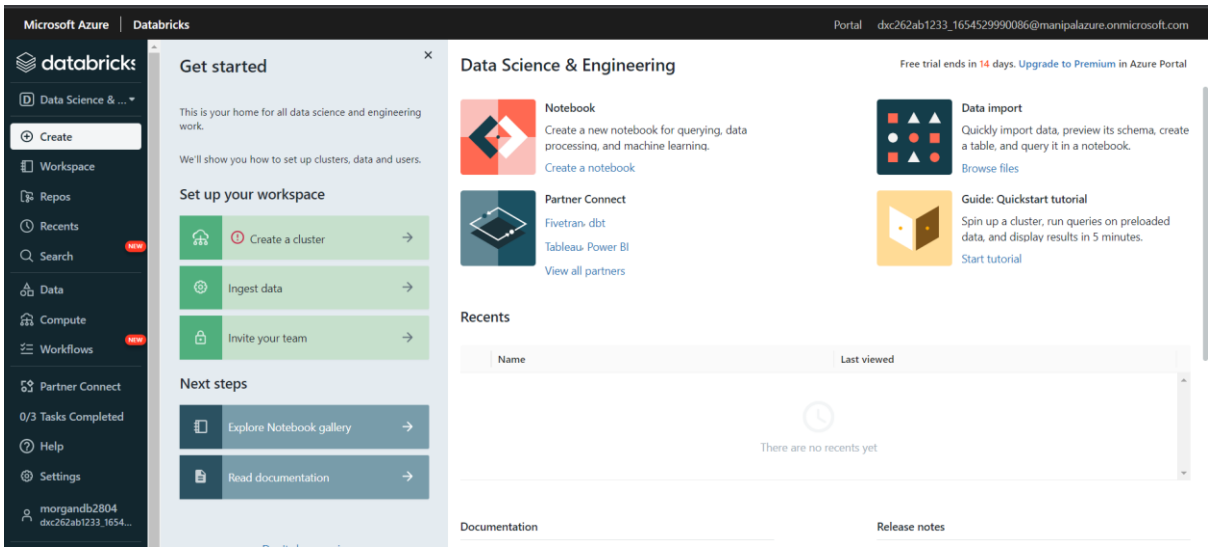



# Step 4: We get Your deployment is complete



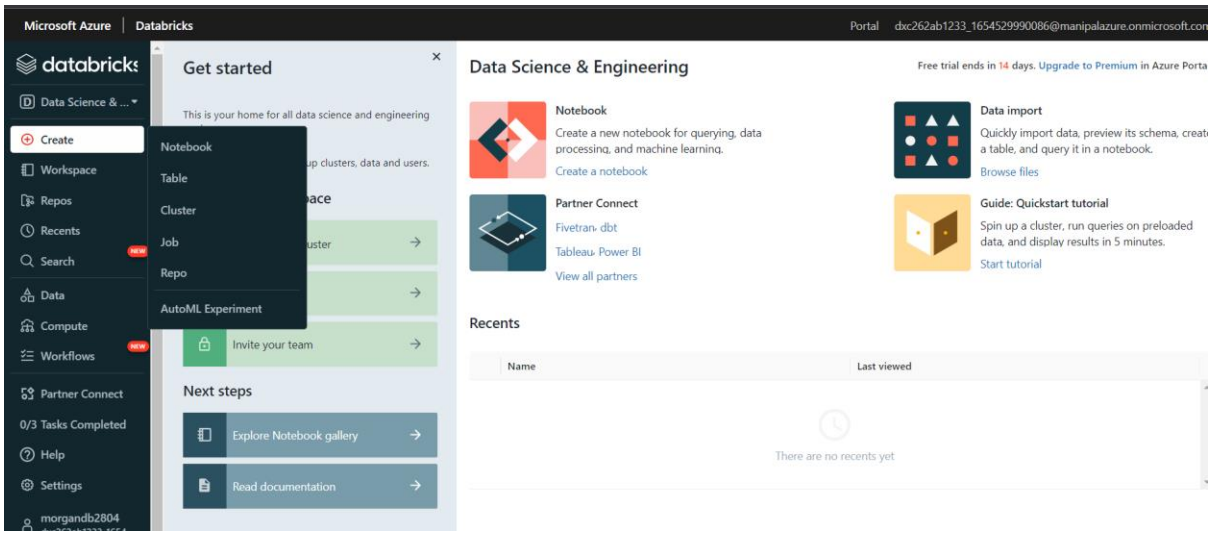

Step 5: Launch Workspace to go to data bricks that we created

This is the how page of data bricks we created

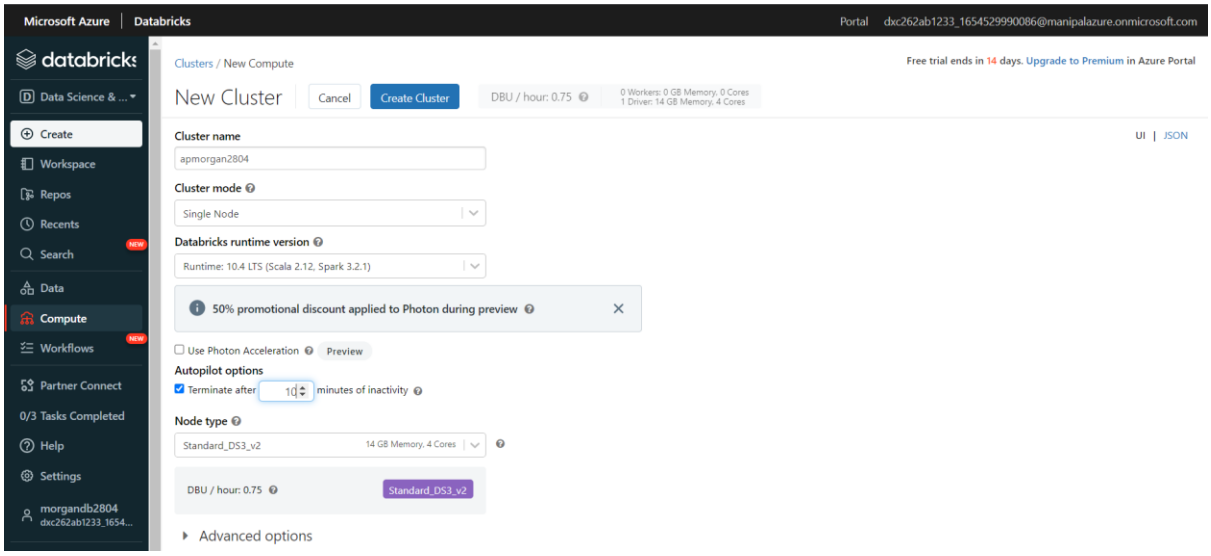


Practical Lab: Create **Cluster in Azure Databricks**

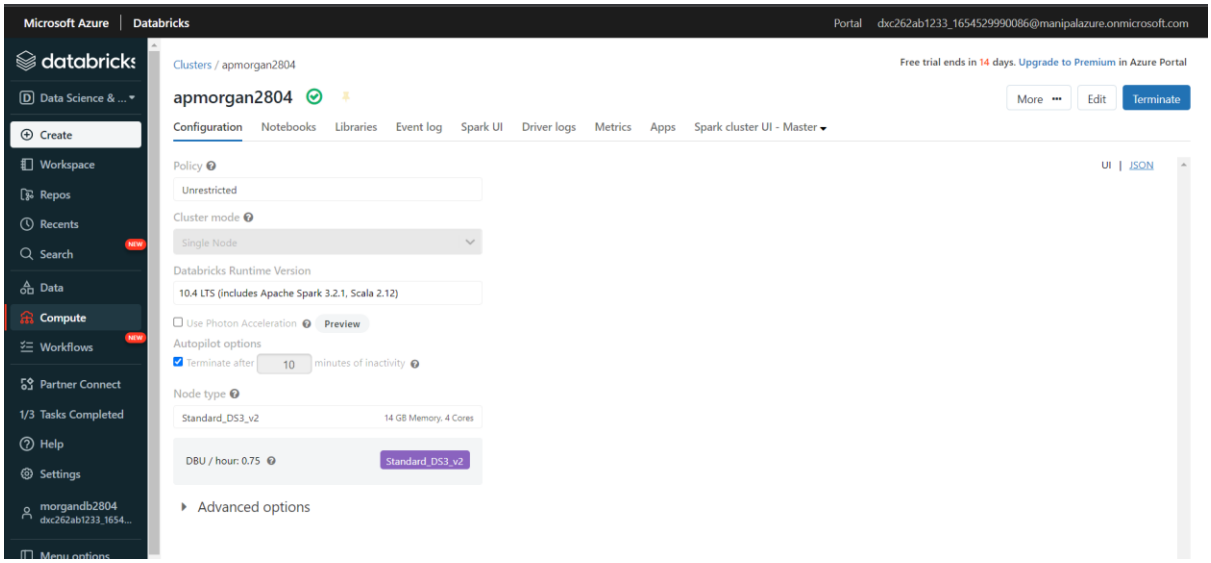
Step 1: After we create data bricks come to home page in that click on create cluster



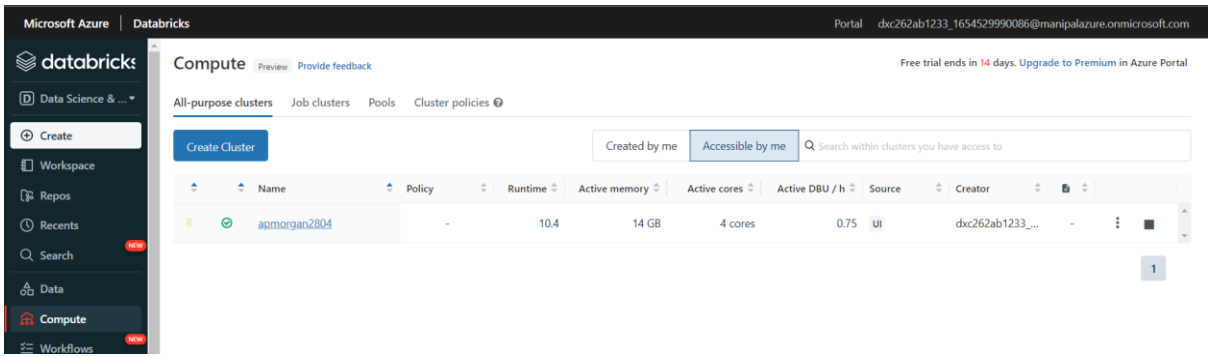

Step 2: Give the cluster name and cluster node single node and terminating time 10 minutes and create cluster




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

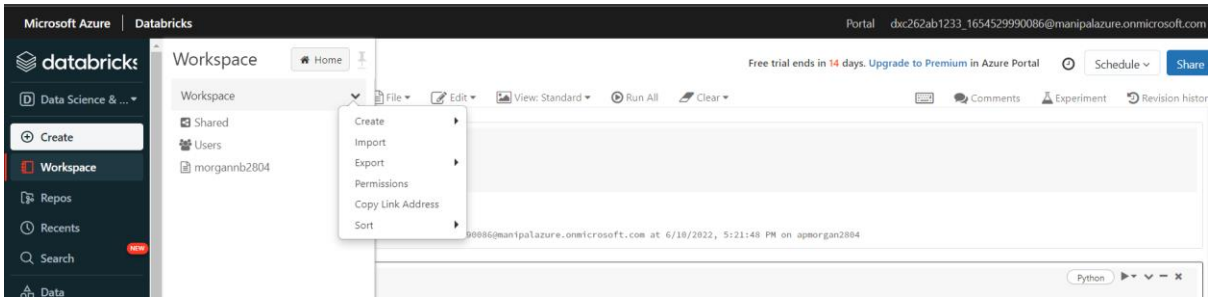



Step 4: To check the complete details of cluster we created go to compute and get the details of it

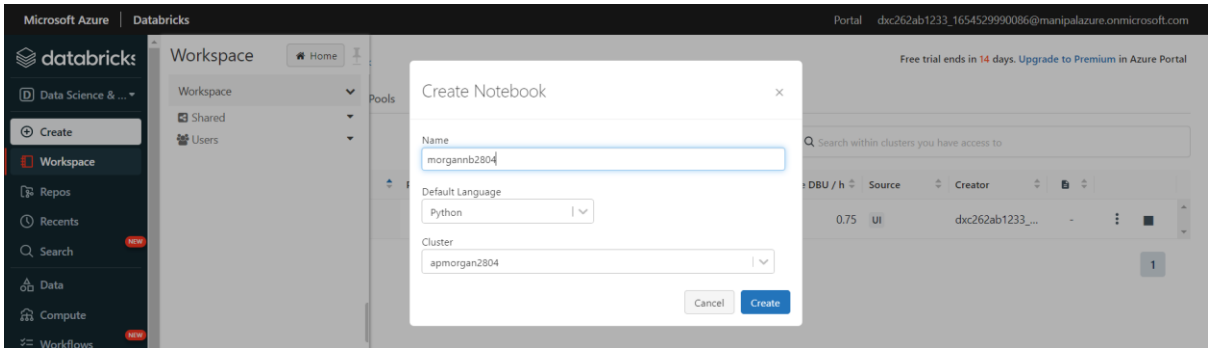


Practical Lab: **Add notebook in Databricks** and Implement the Business Logic

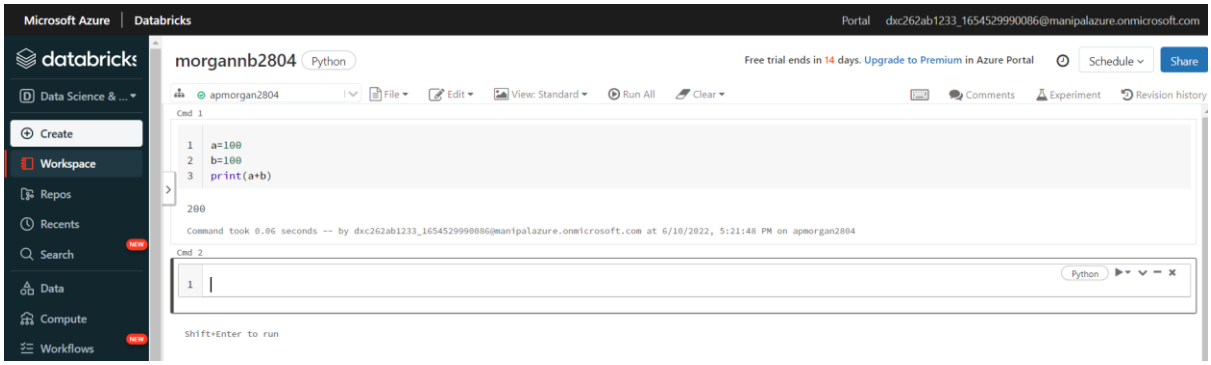
Step 1: To add notebook Go to Workspace-> create-> Notebook




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

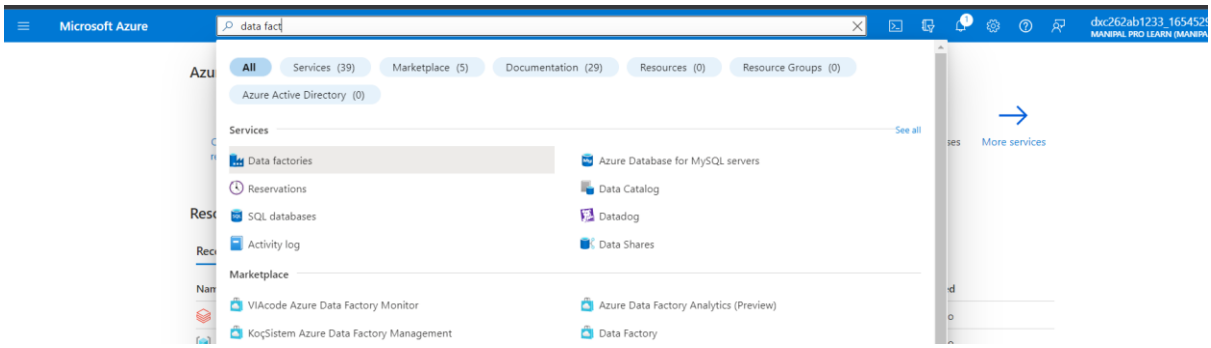



Step 3: to implement a notebook we created after create we will come to this page here we need to attach cluster and I have wrote a small program and run all it will show in seconds

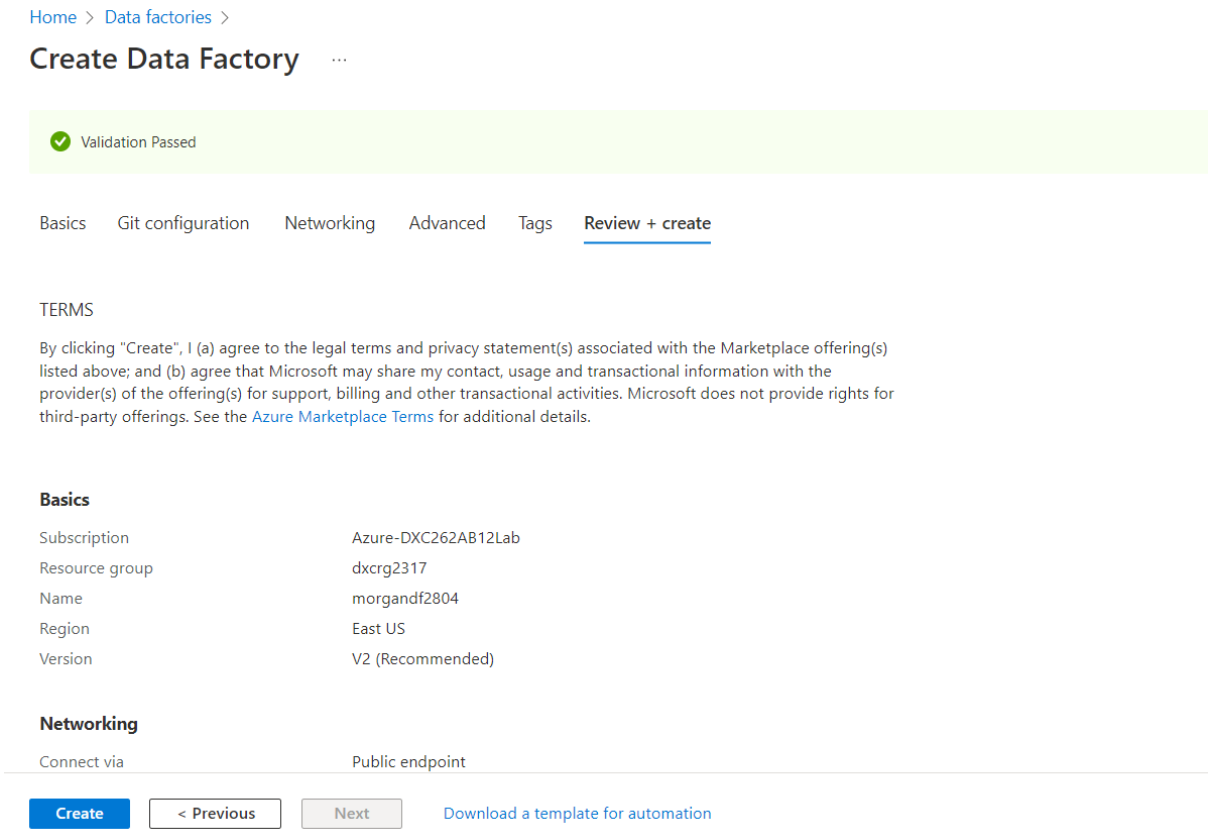


Practical Lab: **Azure Data Factory** For AP Morgan

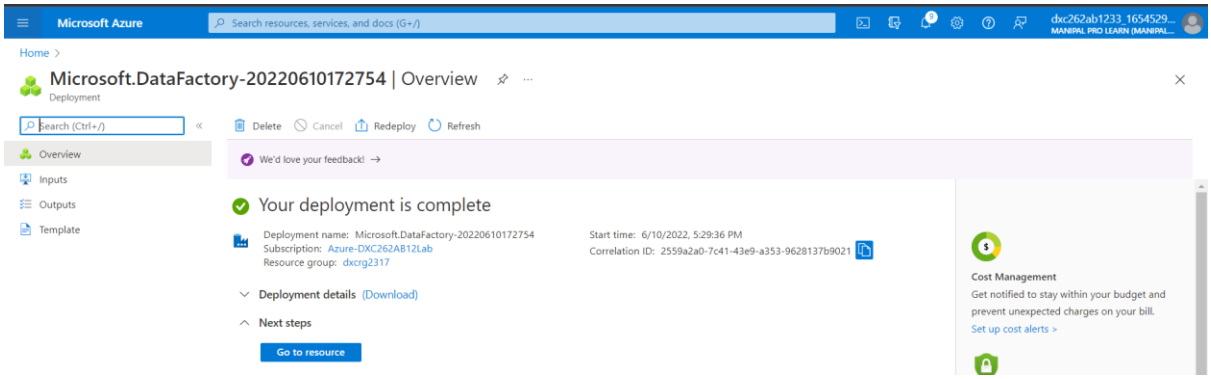
Step 1: First we need to open Azure and search for Data Factory and click on it



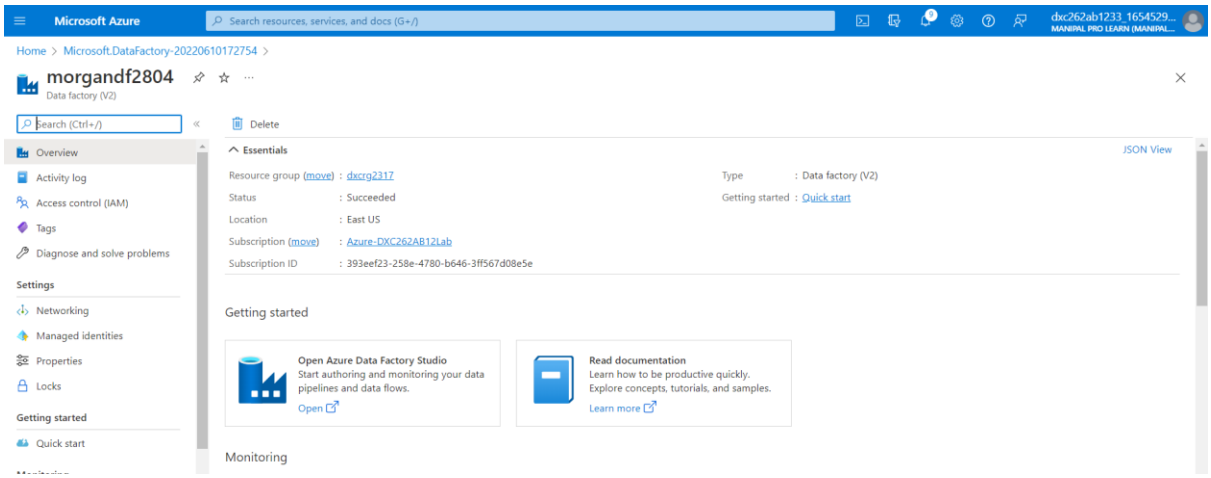

Step 2: Select data factory and click the + create and Give the Resource group name and Name morgandf2804 and click next Git configuration




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

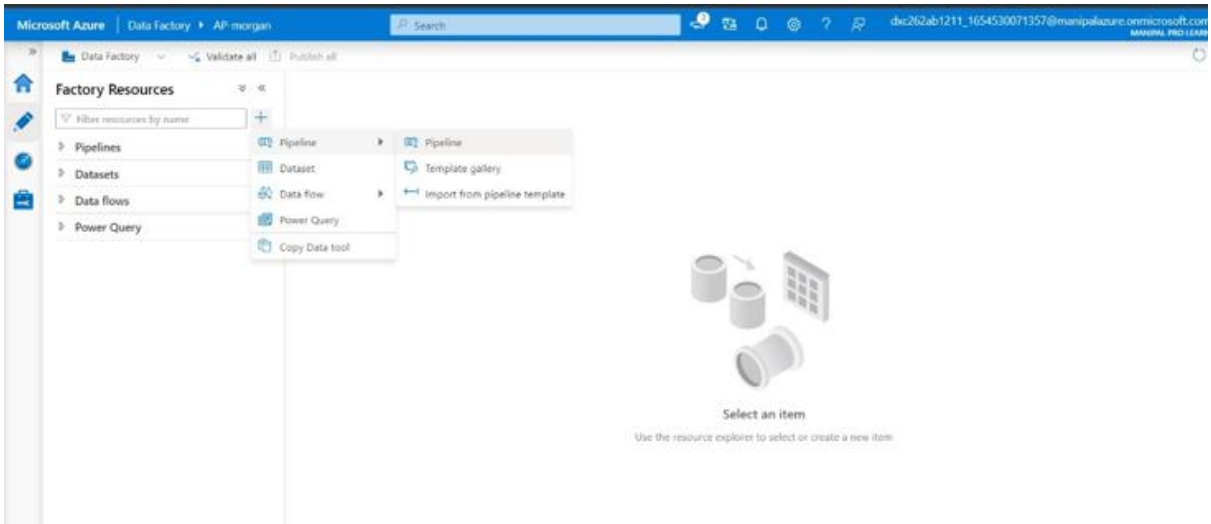



Step 4: Go to resource and open azure data factory to access

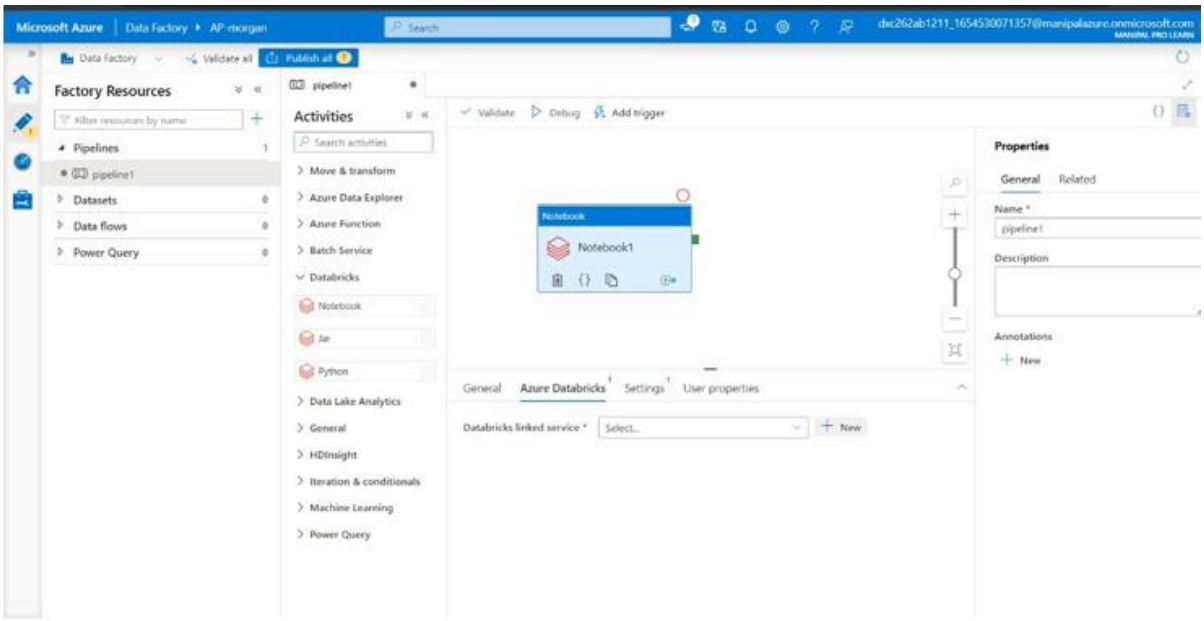



Practical Lab: **Create Azure Databricks** Linked Service in ADF

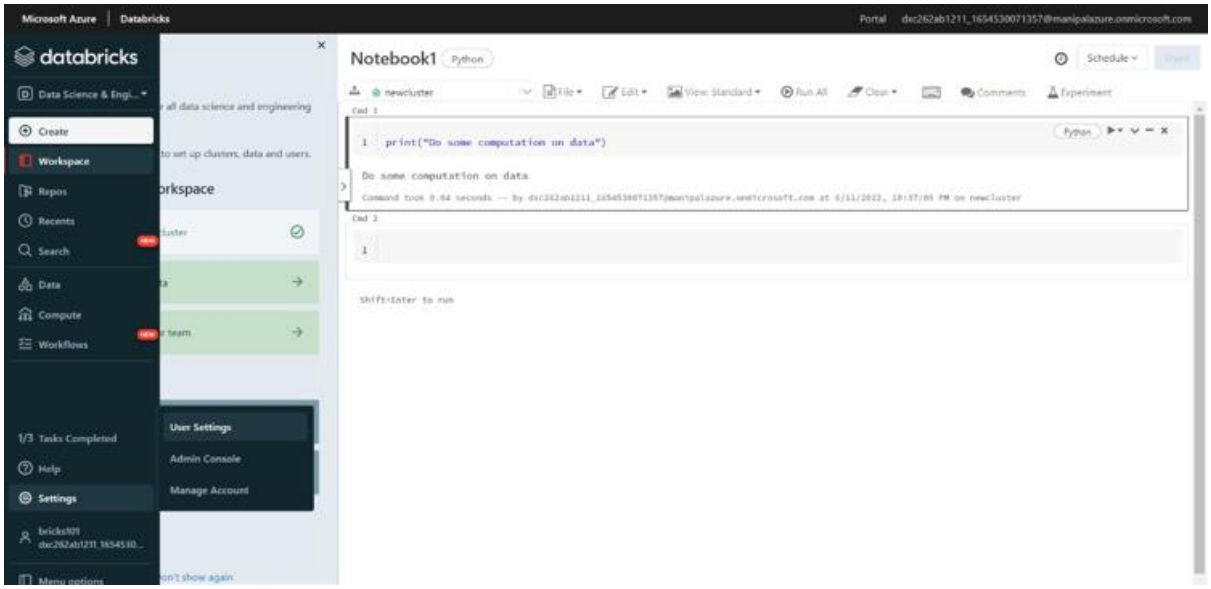
Step 1: In data factory we create a  new pipeline that will help the data bricks linked service



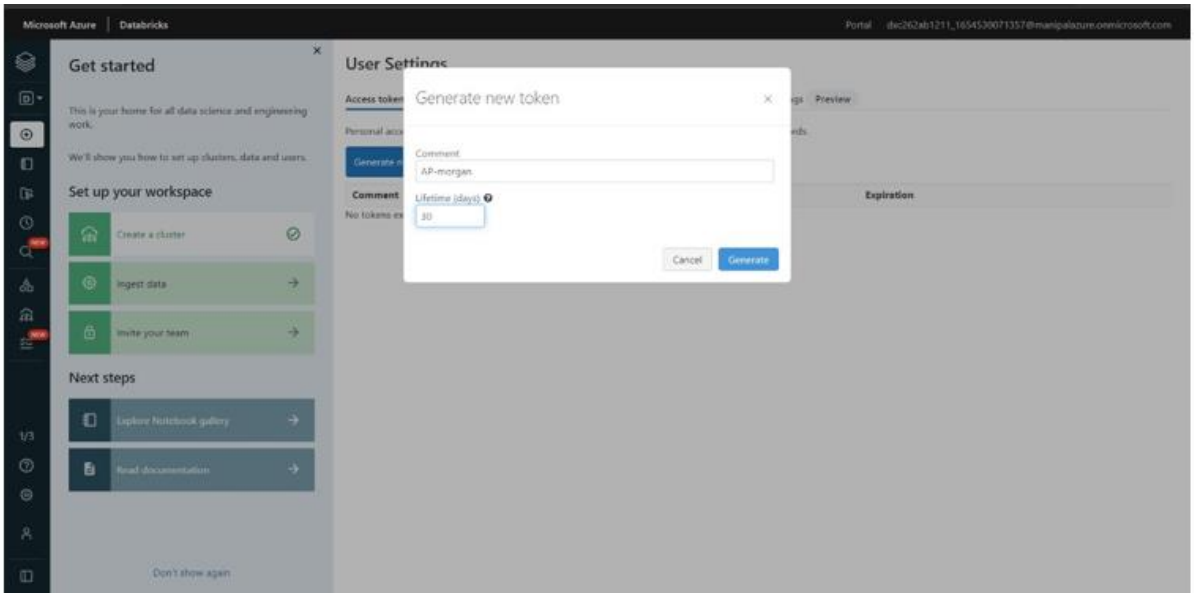
Step 2: In that pipeline drag and drop the notebook you have created from data bricks option



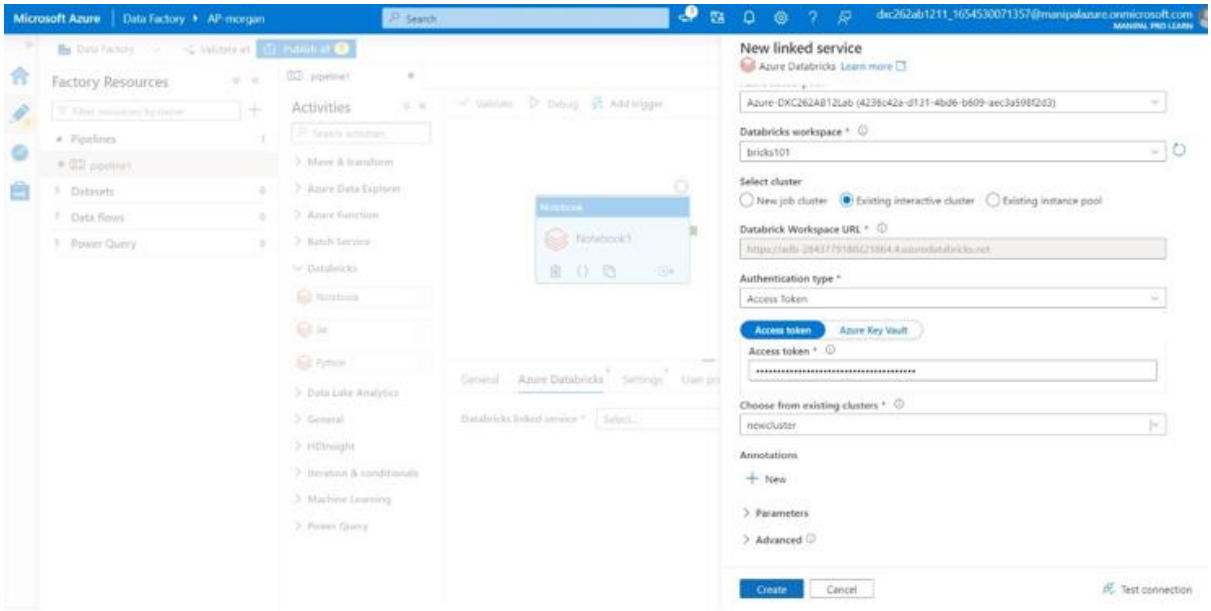

Step 3: First we need to go to data bricks and go to settings tab and user settings



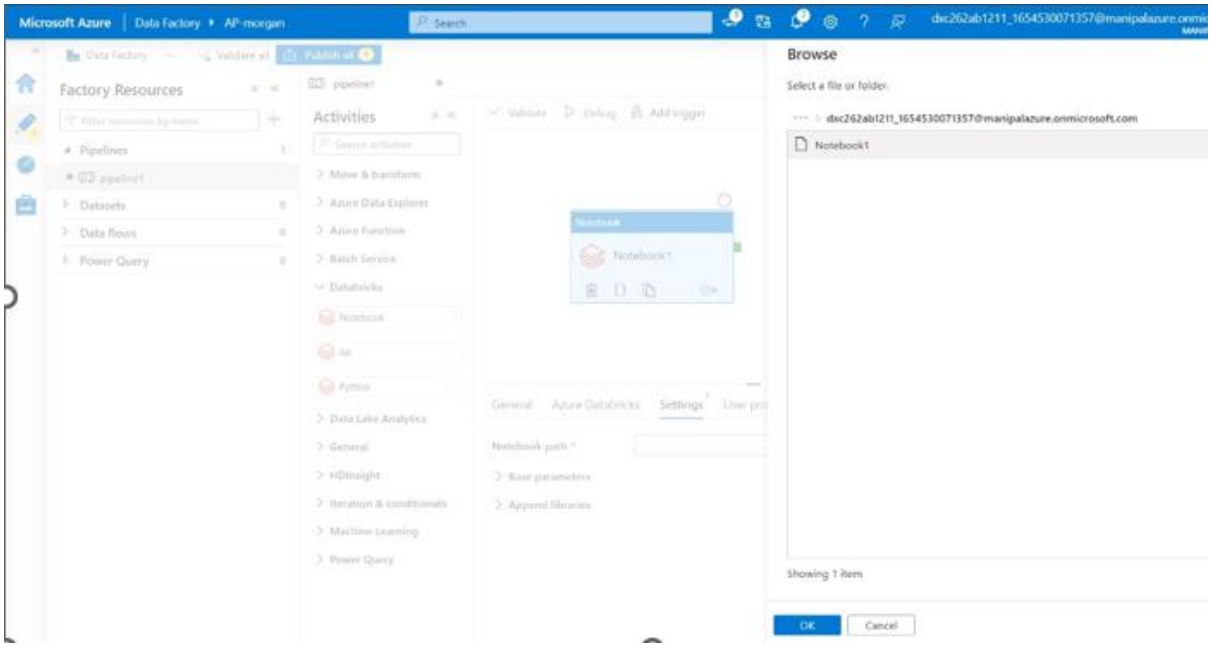

Step 4:  In user settings go to Generate new token by giving name and click on generate and we get the token copy that and paste it in the azure factory connection details



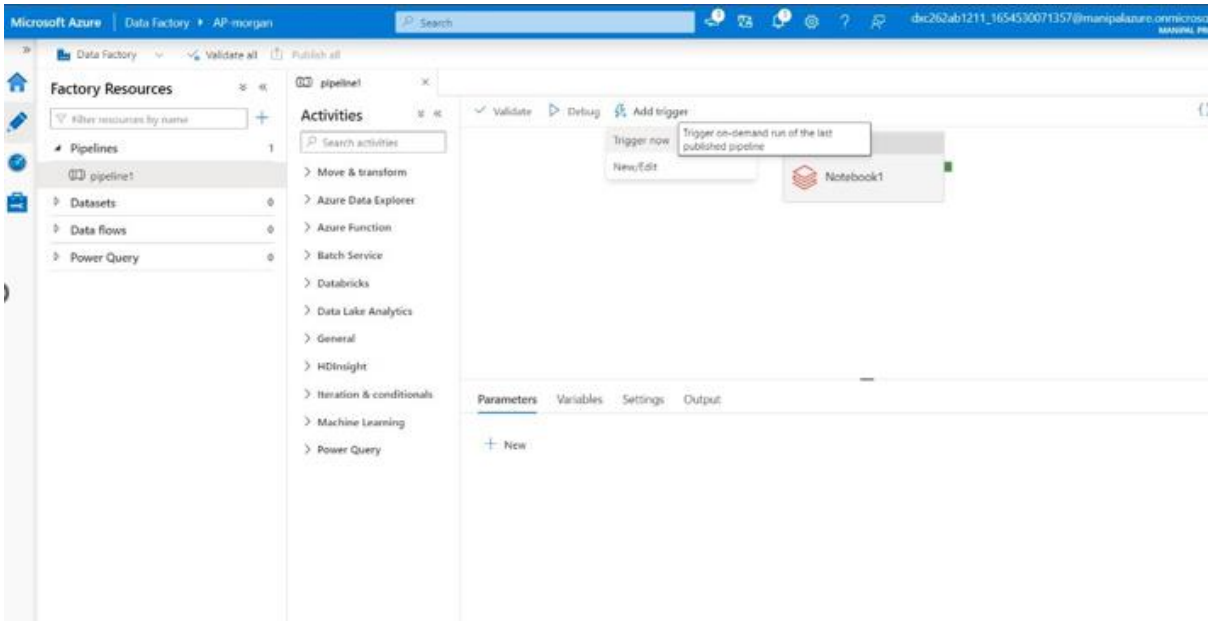

Step 5: Choose the cluster you have created on data bricks and click on create



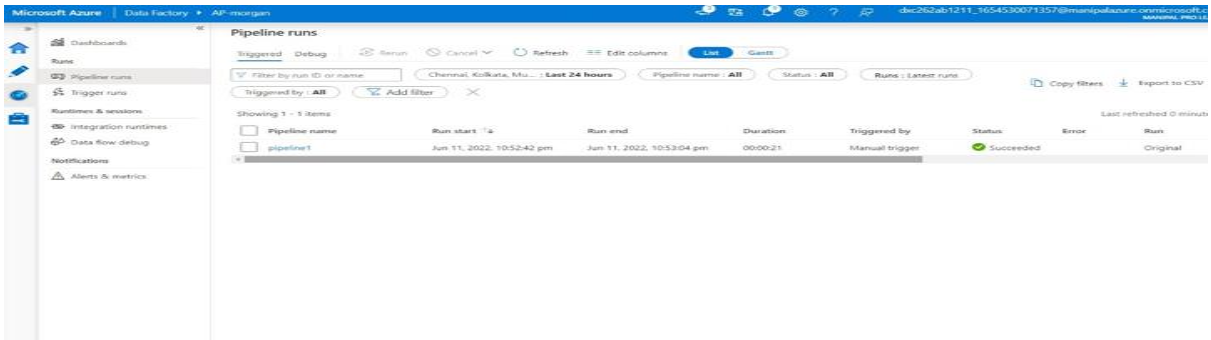

Step 6: After that choose the notebook in the pipeline tab to get trigger first publish the pipeline because it will trigger only when it get published




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 note book 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