

DXC REALTIME PROJECTS

AZ-900, DP - 203



JUNE 10, 2022
DXC TECHNOLOGY PVT.LTD.

Name: SUPRIYA BHARATHA

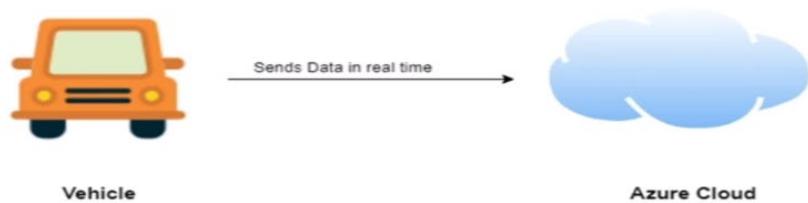
Reg No: DXCAB1211

Project 1 Name: Smart Vehicles

Date: 10th JUNE 2022

Project 1 : Connected Vehicles

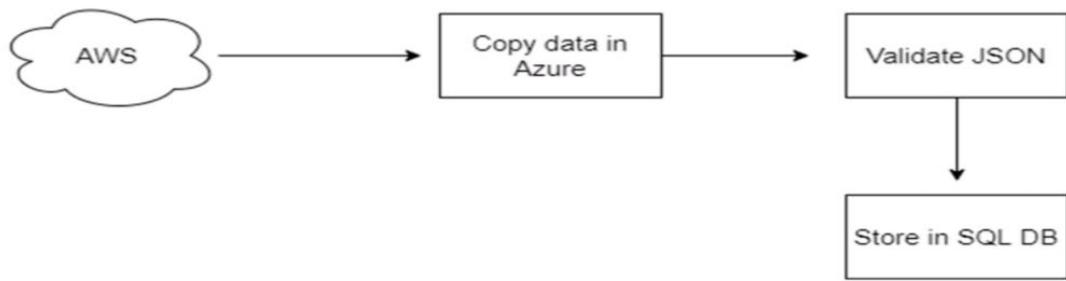
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



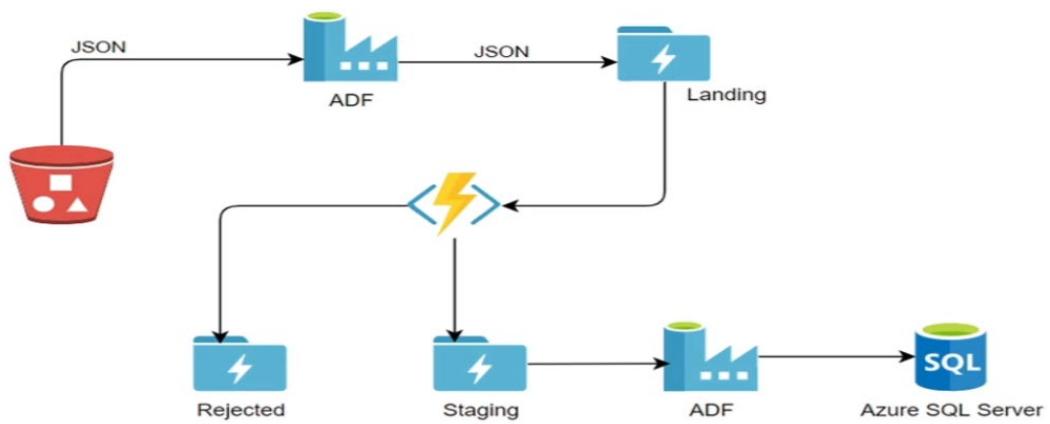
Project 1 : Connected Vehicles

- 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



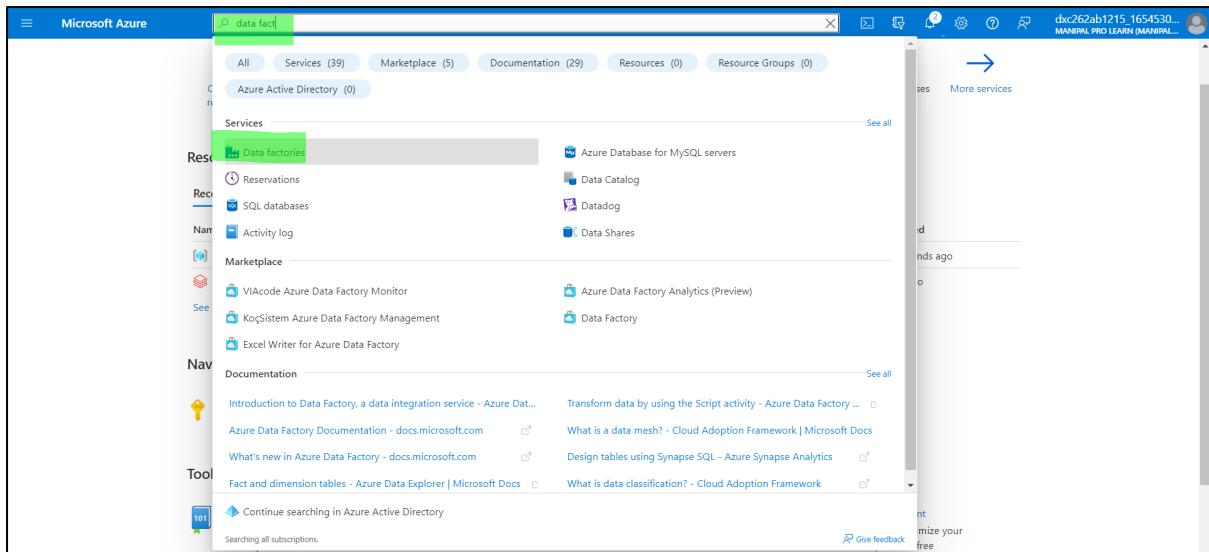
Project 1 : Connected Vehicles



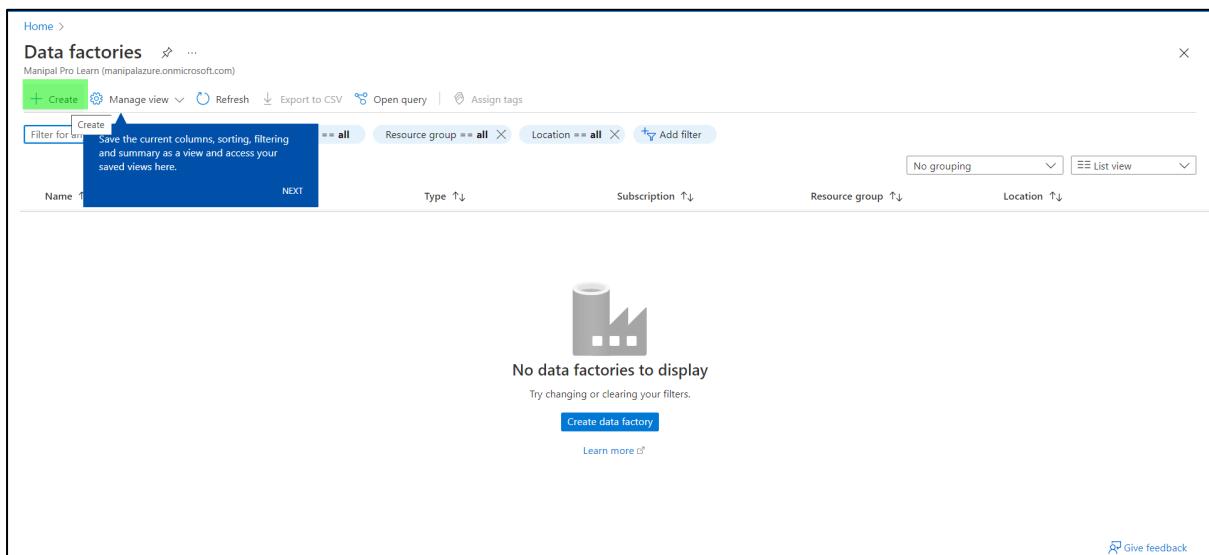
Architecture Diagram for Connected Vehicle Project

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

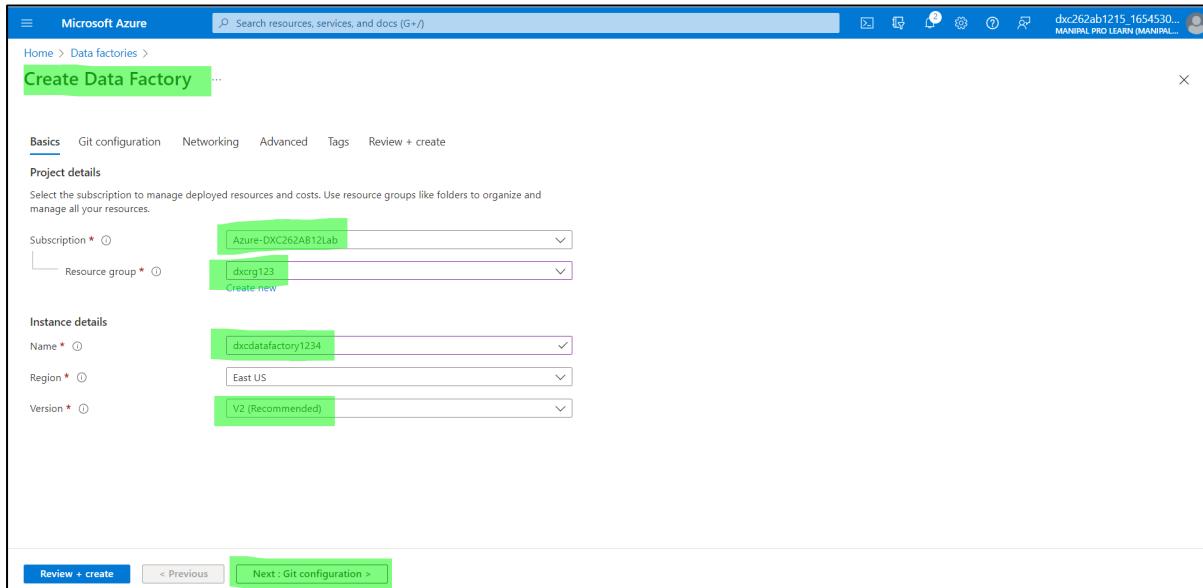
STEP 1: Go to azure portal and search for data factories and open select it.



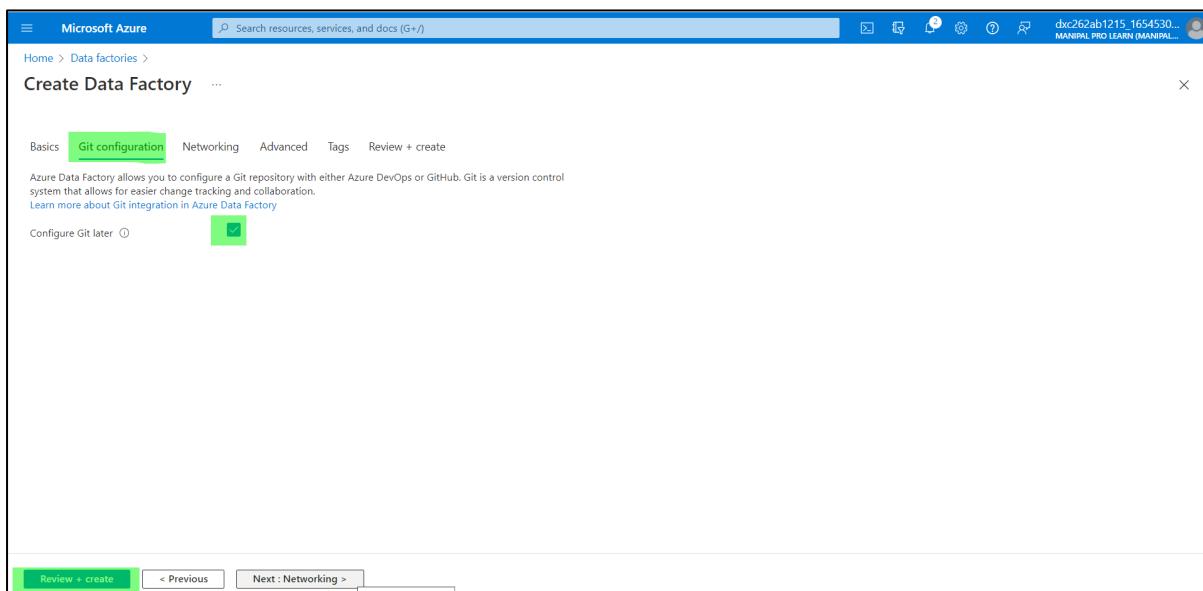
STEP 2: After selecting azure portal ,click on “+ create” for creating datafactory.



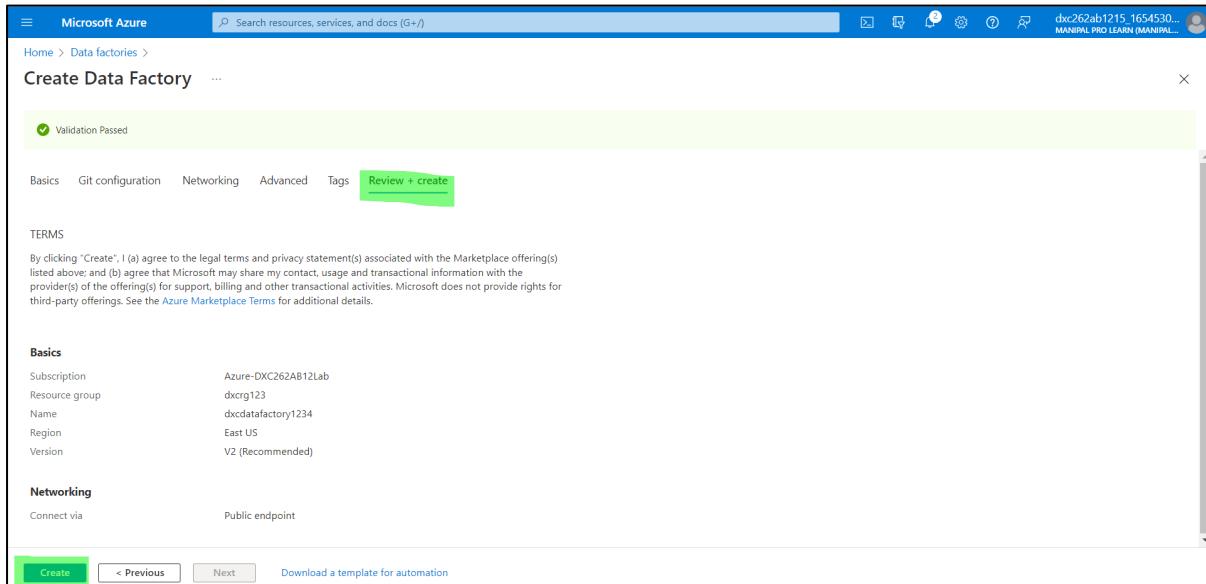
STEP 3: Here, we have give the details of resource group , Name ,version we used. Then next to git configuration.



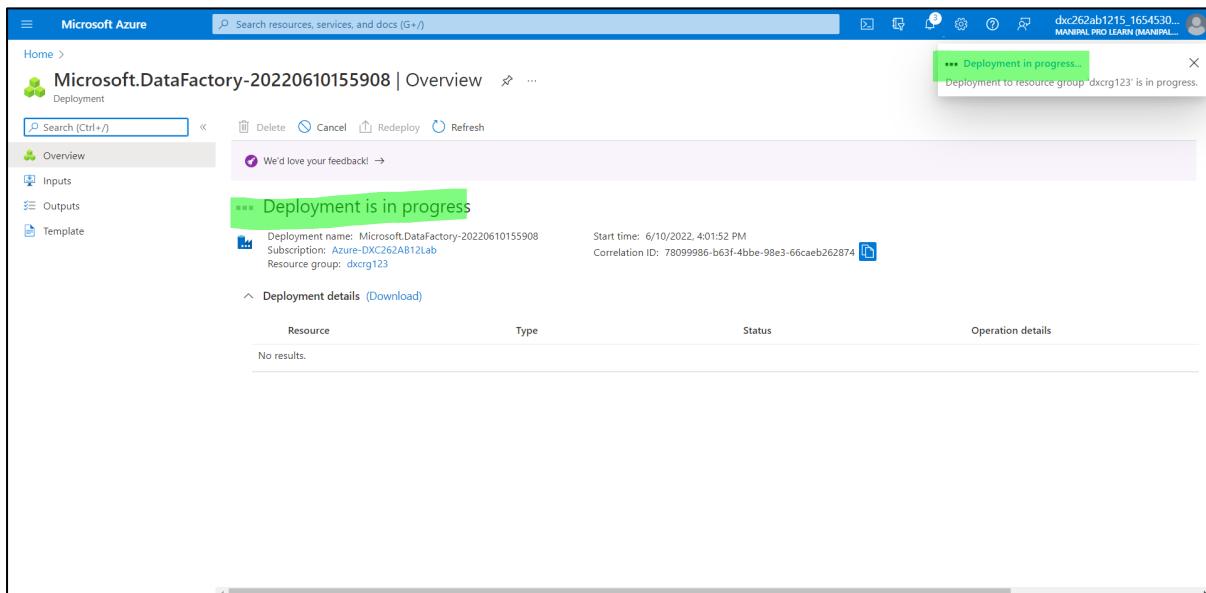
STEP 4: The details should be as shown in image for git configuration. Then, next to review+create.



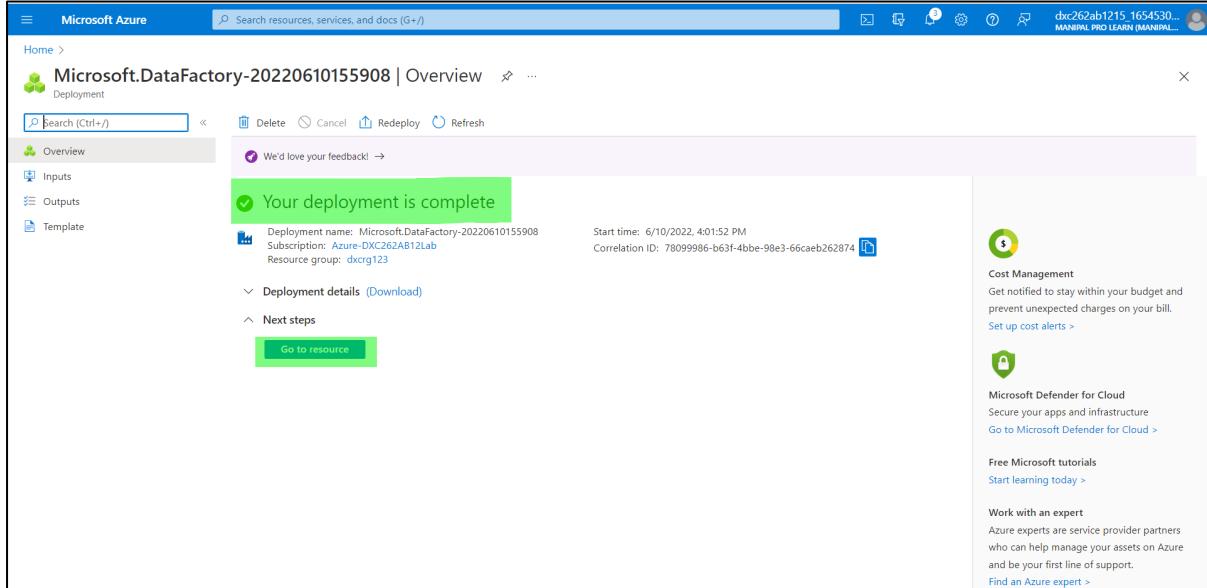
STEP 5: In review + create we need check and click on create below.



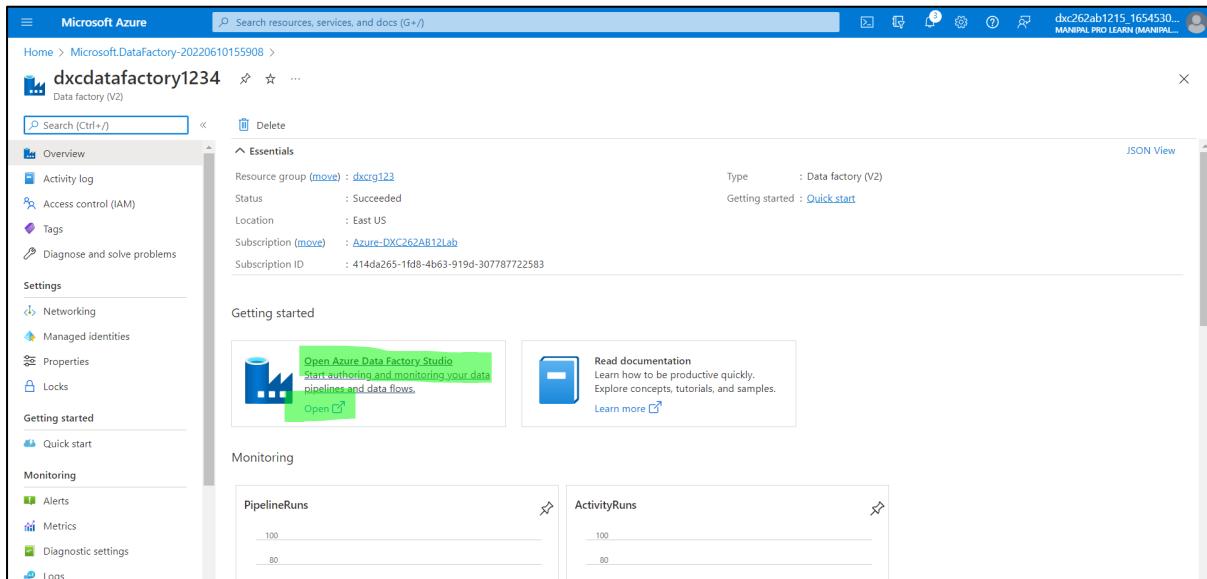
STEP 6: After clicking on create, deployment will be in progress.



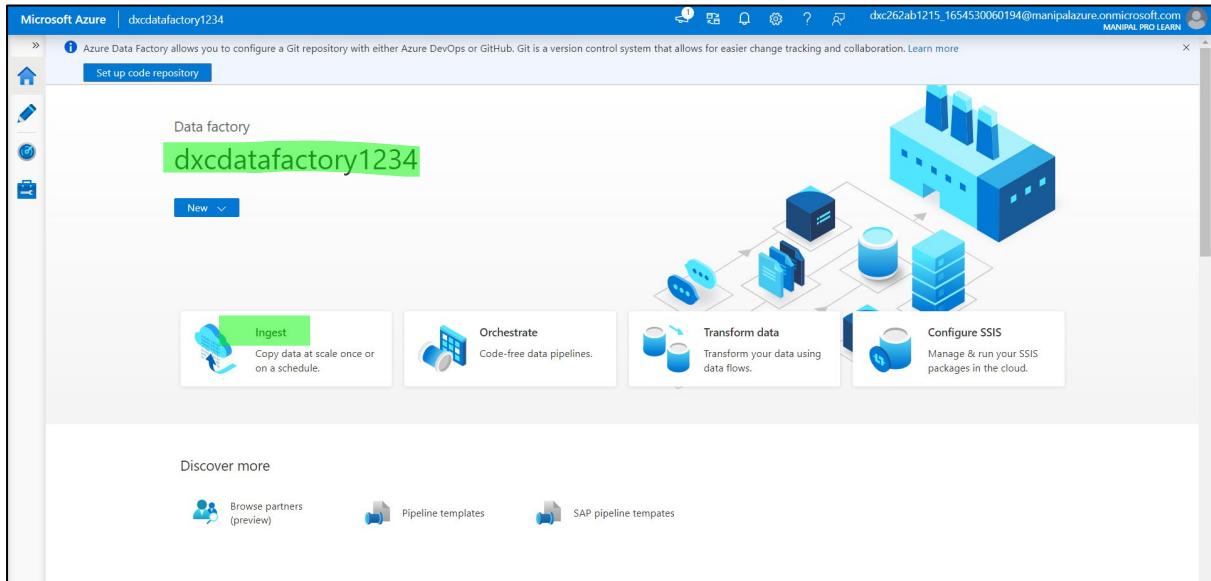
STEP 7: After few seconds or minutes the deployment is completed.



STEP 8: In resources , open azure data factory studio.

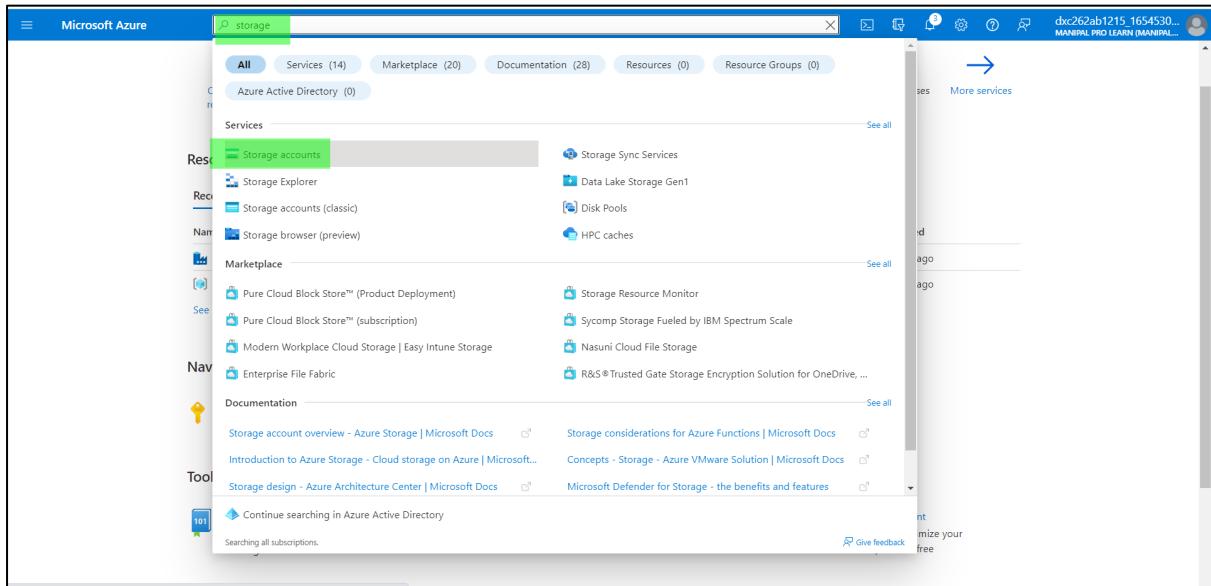


STEP 9: After opening data factory new tab, enter into ingest which is shown below.

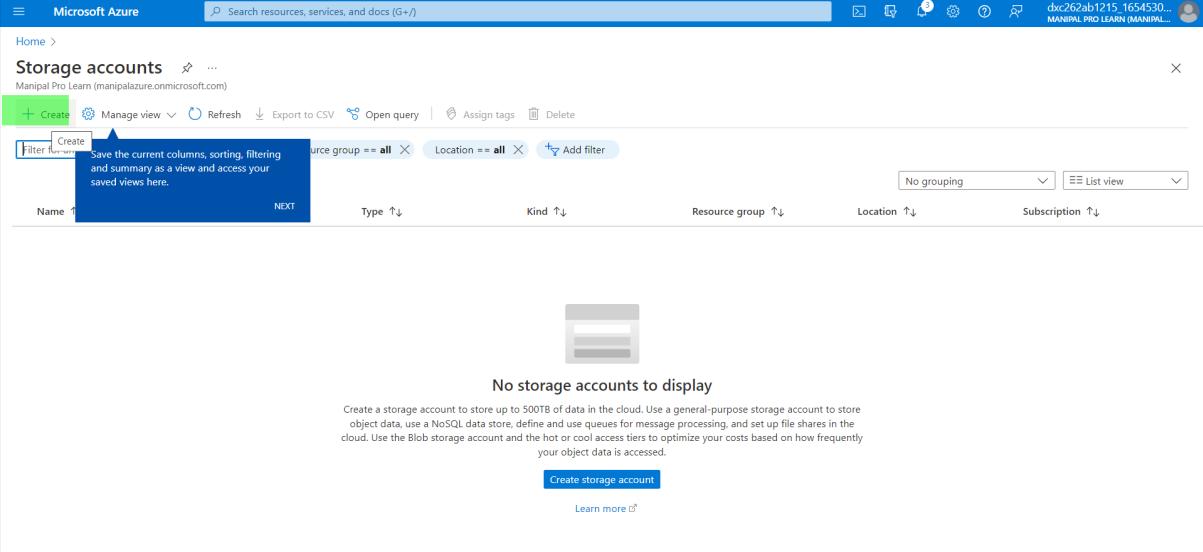


STEP 10: CREATING STORAGE

Search for Storage accounts and select it.

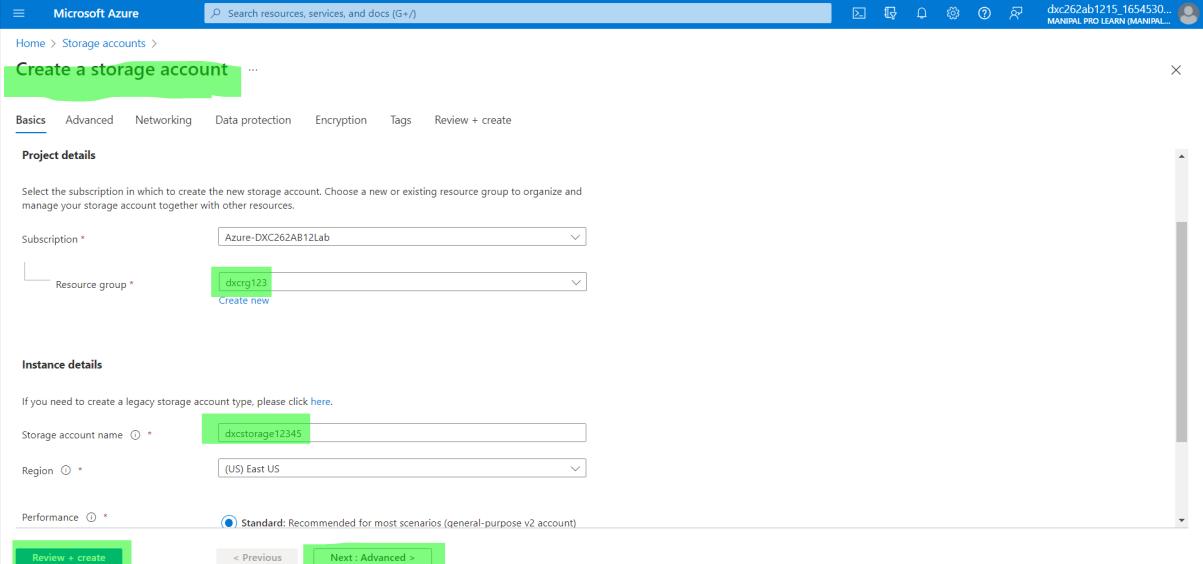


STEP 11: After entering into storage account click on “+ create” .



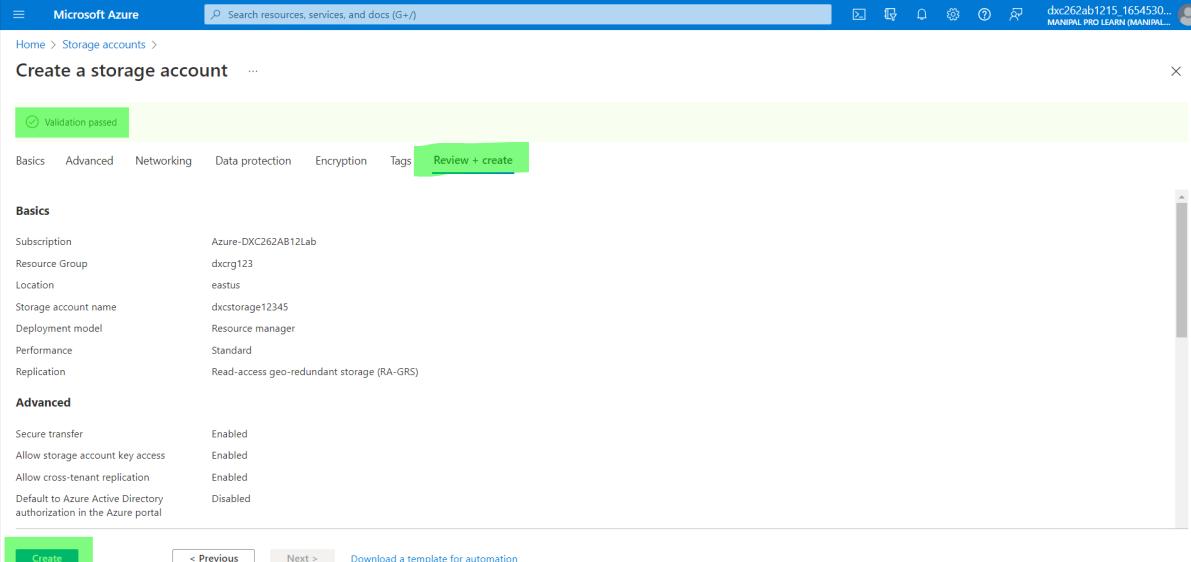
The screenshot shows the Microsoft Azure Storage accounts page. At the top, there is a search bar and various navigation links. Below the header, there is a toolbar with buttons for '+ Create', 'Manage view', 'Refresh', 'Export to CSV', 'Open query', 'Assign tags', and 'Delete'. A 'Filter' button is also present. The main area displays a message: 'No storage accounts to display' with a sub-message: 'Create a storage account to store up to 500TB of data in the cloud. Use a general-purpose storage account to store object data, use a NoSQL data store, define and use queues for message processing, and set up file shares in the cloud. Use the Blob storage account and the hot or cool access tiers to optimize your costs based on how frequently your object data is accessed.' Below this message are two buttons: 'Create storage account' and 'Learn more'. The '+ Create' button is highlighted with a green box.

STEP 12: For creating we need to give the basic details of resource group, storage account name . Then, next to review+create.



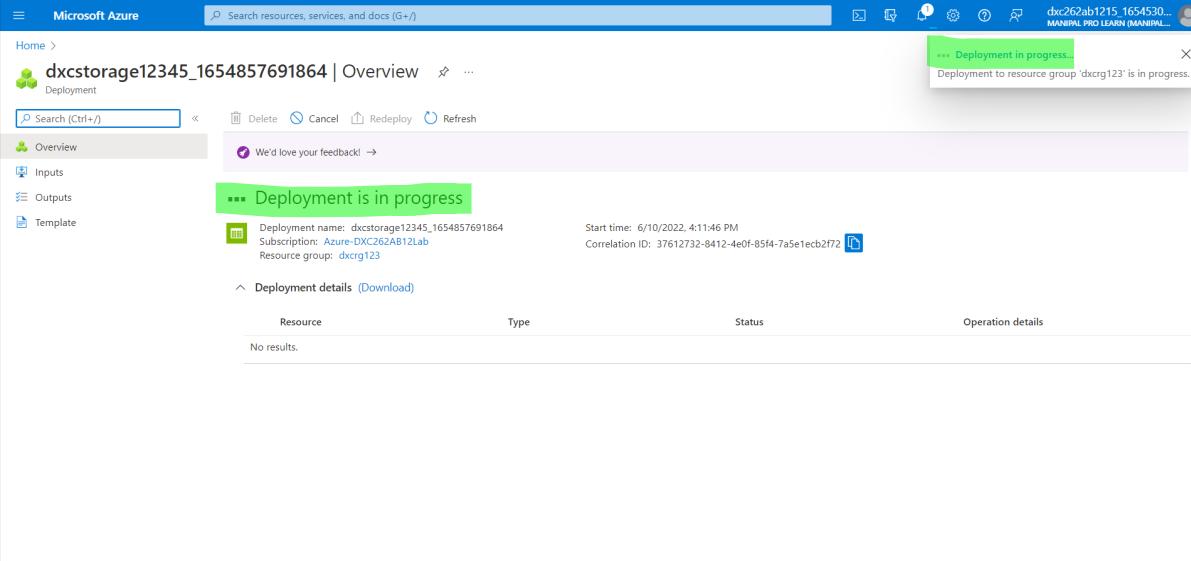
The screenshot shows the 'Create a storage account' wizard on the 'Basics' step. The title bar says 'Create a storage account'. Below it, there are tabs: Basics (which is selected), Advanced, Networking, Data protection, Encryption, Tags, and Review + create. The 'Project details' section asks to select a subscription and resource group. The 'Subscription' dropdown shows 'Azure-DXC262AB12Lab'. The 'Resource group' dropdown shows 'dxcrg123' with a 'Create new' link below it. The 'Instance details' section includes fields for 'Storage account name' (set to 'dxcstorage12345'), 'Region' (set to '(US) East US'), and 'Performance' (set to 'Standard: Recommended for most scenarios (general-purpose v2 account)'). At the bottom, there are buttons for 'Review + create' (highlighted with a green box) and 'Next : Advanced >'.

STEP 13: In review + create the validation is passed and then click on create.



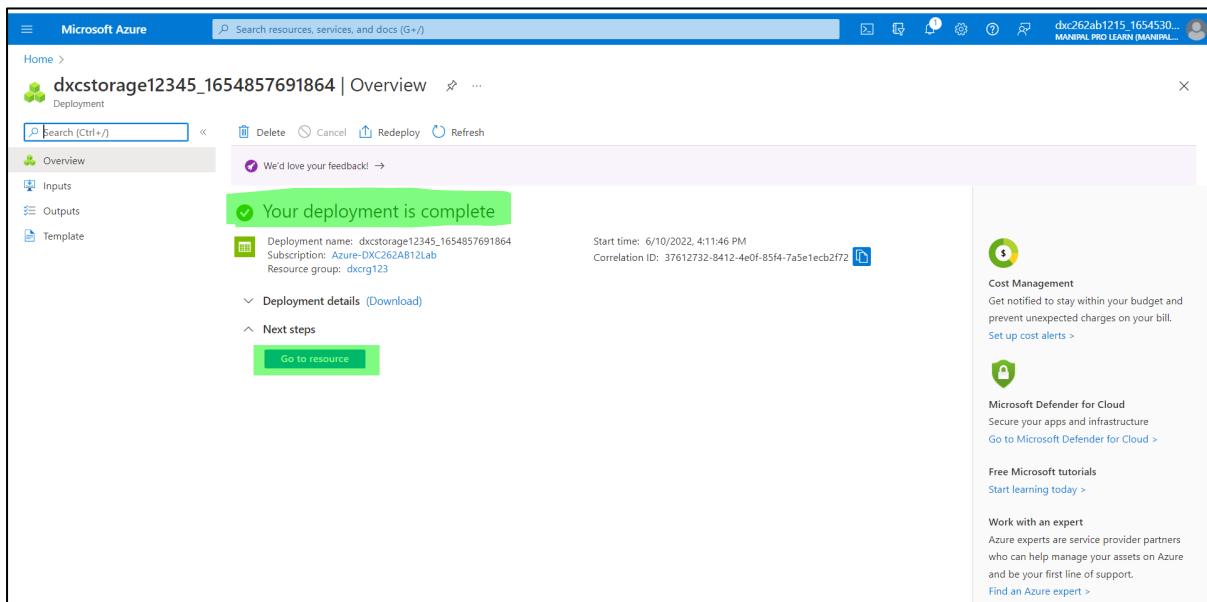
The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal. The validation status is 'Validation passed'. The 'Review + create' button is highlighted in green. The 'Basics' tab is selected, displaying configuration details like Subscription (Azure-DXC262AB12Lab), Resource Group (dxcrgr123), Location (eastus), Storage account name (dxcstorage12345), Deployment model (Resource manager), Performance (Standard), and Replication (Read-access geo-redundant storage (RA-GRS)). The 'Advanced' tab shows settings for Secure transfer, Allow storage account key access, Allow cross-tenant replication, and Default to Azure Active Directory authorization in the Azure portal. At the bottom, there are 'Create' and 'Next >' buttons, along with a link to 'Download a template for automation'.

STEP 14: After create the deployment will be in progress.

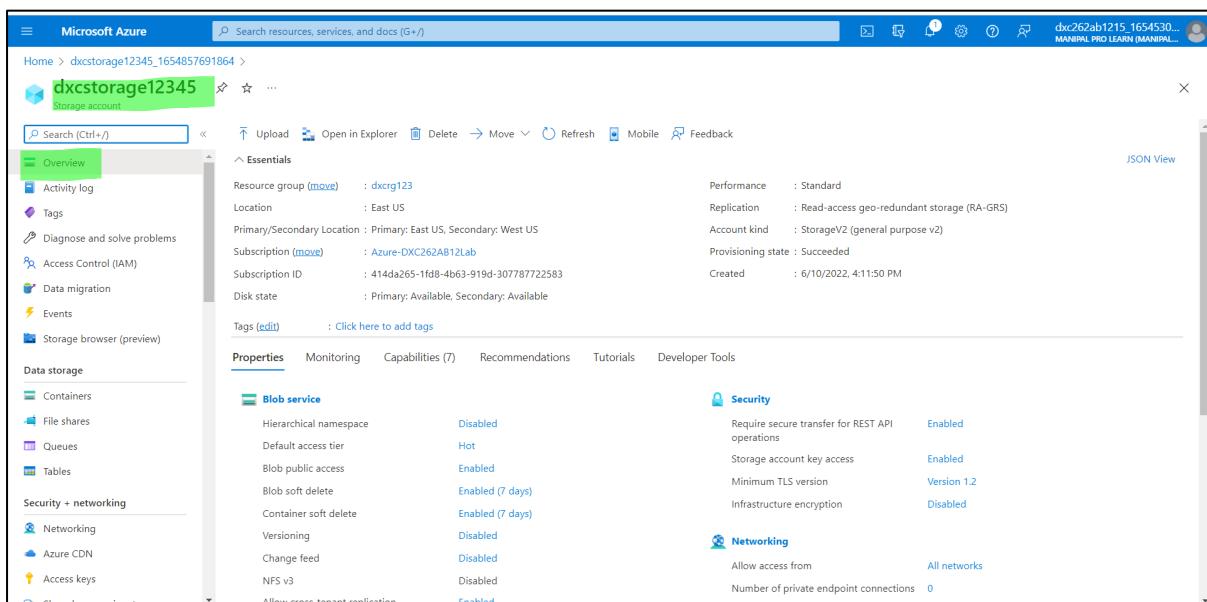


The screenshot shows the 'Overview' page for the storage account 'dxcstorage12345_1654857691864'. A green banner at the top indicates 'Deployment in progress...'. A message states 'Deployment to resource group 'dxcrgr123' is in progress.' On the left, there are tabs for Overview, Inputs, Outputs, and Template. The Overview tab is selected. It shows deployment details: Deployment name: dxcstorage12345_1654857691864, Subscription: Azure-DXC262AB12Lab, Resource group: dxcrgr123. The deployment started at 6/10/2022, 4:11:46 PM with Correlation ID: 37612732-8412-4e0f-85f4-7a5e1ecb2f72. Below this, a table titled 'Deployment details' shows 'No results.'

STEP 15: After few minutes the deployment is completed and proceeded. Then enter go to resource.



STEP 16: Here, the overview is appeared.



STEP 17: To create containers ,we select containers on left side. To add new containers we use “+ container”.

The screenshot shows the Microsoft Azure Storage account interface for 'dxcstorage12345'. The left sidebar has sections for Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, and Storage browser (preview). Under Data storage, 'Containers' is selected and highlighted with a green box. The main content area shows a table of existing containers:

Name	Last modified	Public access level	Lease state
Slogs	6/10/2022, 4:12:20 PM	Private	Available

At the top right of the main content area, there is a search bar labeled 'Search containers by prefix' and a button labeled '+ Container'.

STEP 18: Here, first we have to create the source container. I named source container as “Sourceblob1” and create it.

The screenshot shows the 'Containers' blade with a 'New container' dialog open. The 'Name' field is filled with 'sourceblob1'. The 'Public access level' dropdown is set to 'Blob (anonymous read access for blobs only)'. A warning message below the dropdown states: '⚠️ Blobs within the container can be read by anonymous request, but container data is not available. Anonymous clients cannot enumerate the blobs within the container.' At the bottom right of the dialog, there are 'Create' and 'Discard' buttons, with 'Create' highlighted with a green box.

STEP 19: As we can see the sourceblob 1 is created.

The screenshot shows the Microsoft Azure Storage account interface for 'dxcstorage12345'. In the left sidebar, 'Containers' is selected under 'Data storage'. The main pane displays a list of containers. One container, 'sourceblob1', is highlighted with a green box. The table columns are 'Name', 'Last modified', 'Public access level', and 'Lease state'. The 'sourceblob1' row shows 'sourceblob1' as the name, '6/10/2022, 4:14:16 PM' as the last modified date, 'Blob' as the public access level, and 'Available' as the lease state.

STEP 20: And second we have to create the destination container. I named source container as “destinationblob1” and create it.

The screenshot shows the Microsoft Azure Storage account interface for 'dxcstorage12345'. A new container dialog is open on the right. The 'Name' field contains 'destinationblob1'. The 'Public access level' dropdown is set to 'Blob (anonymous read access for blobs only)'. A warning message below states: 'Blobs within the container can be read by anonymous requests, but container data is not available. Anonymous clients cannot enumerate the blobs within the container.' At the bottom of the dialog are 'Create' and 'Discard' buttons.

STEP 21: And the destinationblob1 is also created.

The screenshot shows the Microsoft Azure Storage account interface for the container 'dxcstorage12345'. The left sidebar includes options like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, and Storage browser (preview). Under Data storage, 'Containers' is selected. The main pane displays a table of containers with columns: Name, Last modified, Public access level, and Lease state. A new container named 'destinationblob1' is listed, highlighted with a green box. A success message at the top right states 'Successfully created storage container' and 'Successfully created storage container 'destinationblob1''. The URL in the address bar is 'Home > dxcstorage12345_1654857691864 > dxcstorage12345'.

STEP 22: We can upload any details into the containers. Here, I uploaded the “1000_companies.csv” file in sourceblob1.

The screenshot shows the Microsoft Azure Storage account interface for the container 'sourceblob1'. The left sidebar includes Overview, Diagnose and solve problems, Access Control (IAM), Settings (Shared access tokens, Access policy, Properties, Metadata), and Authentication method (Access key, Switch to Azure AD User Account). The Location is set to 'sourceblob1'. The main pane shows a table with columns: Name, Modified, Access tier, Archive status, Blob type, and Size. A search bar at the top says 'Search blobs by prefix (case-sensitive)'. On the right, an 'Upload blob' dialog is open for 'sourceblob1/'. It shows a file input field containing '1000_Companies.csv', a checkbox for 'Overwrite if files already exist', and an 'Upload' button. The URL in the address bar is 'Home > dxcstorage12345_1654857691864 > dxcstorage12345 > sourceblob1'.

STEP 23: The file is uploaded as shown below.

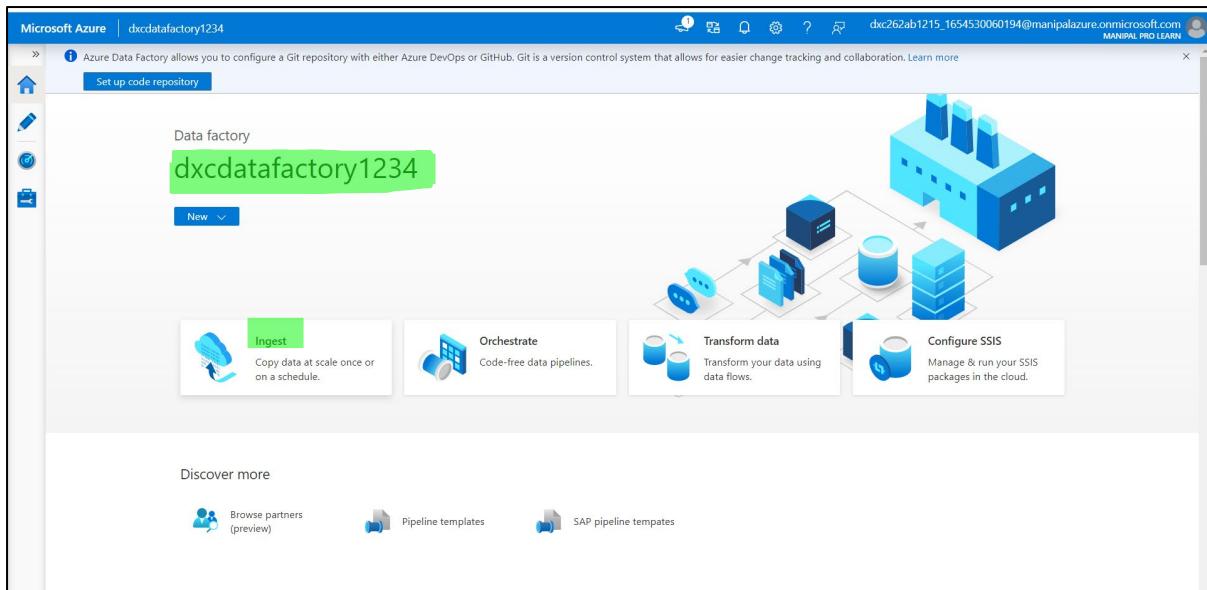
The screenshot shows the Microsoft Azure Storage Explorer interface. The left sidebar displays the navigation path: Home > dxcstorage12345_1654857691864 > dxstorage12345. The main area is titled "sourceblob1" and shows a single blob named "1000_Companies.csv". The blob details are as follows:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
1000_Companies.csv	6/10/2022, 4:22:05 PM	Hot (Inferred)		Block blob	50 KiB	Available

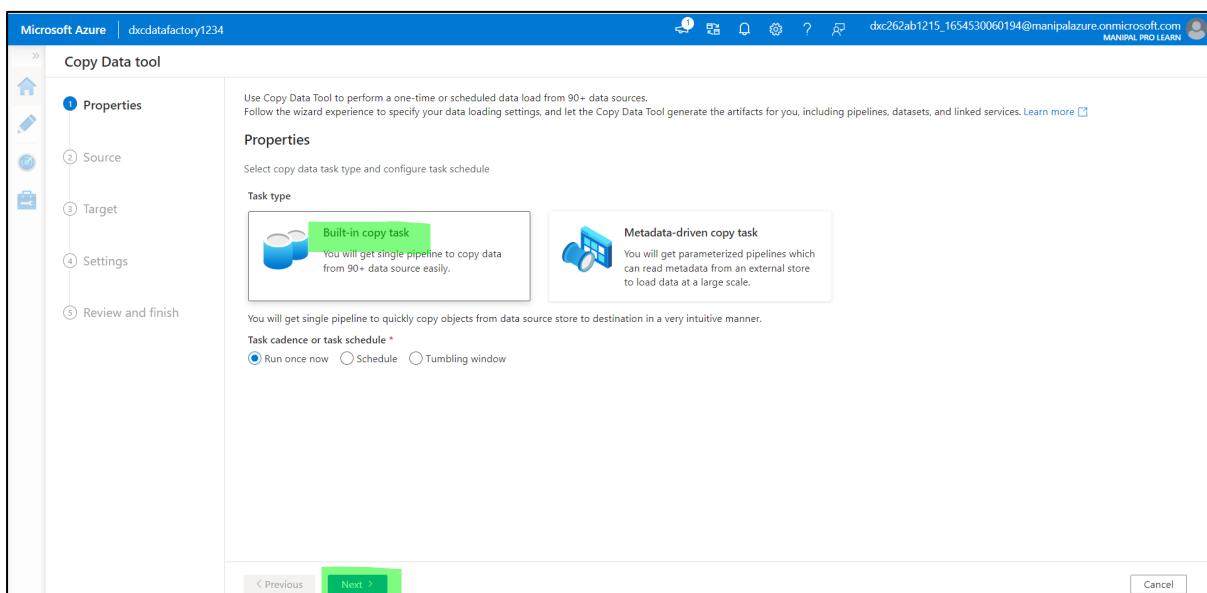
- Practical Lab: Create ADF Pipeline End to end pipeline with triggers enabled
- Practical Lab: Create Azure blob trigger logic

STEP 1: COMING BACK TO DATAFACTORY

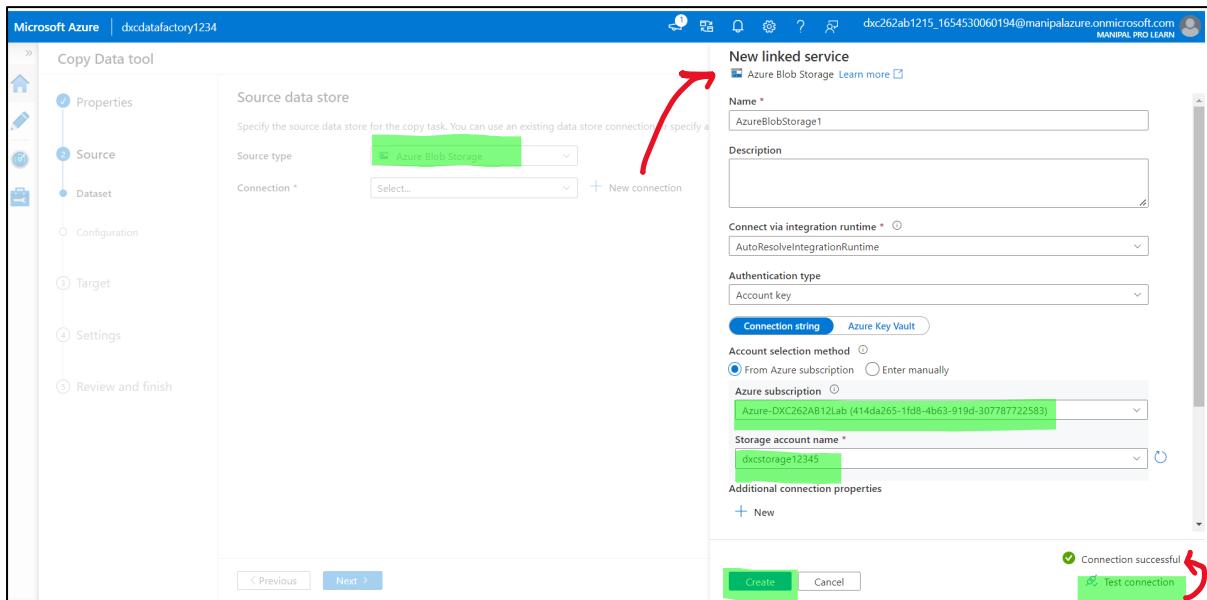
And coming to data factory that we created before.



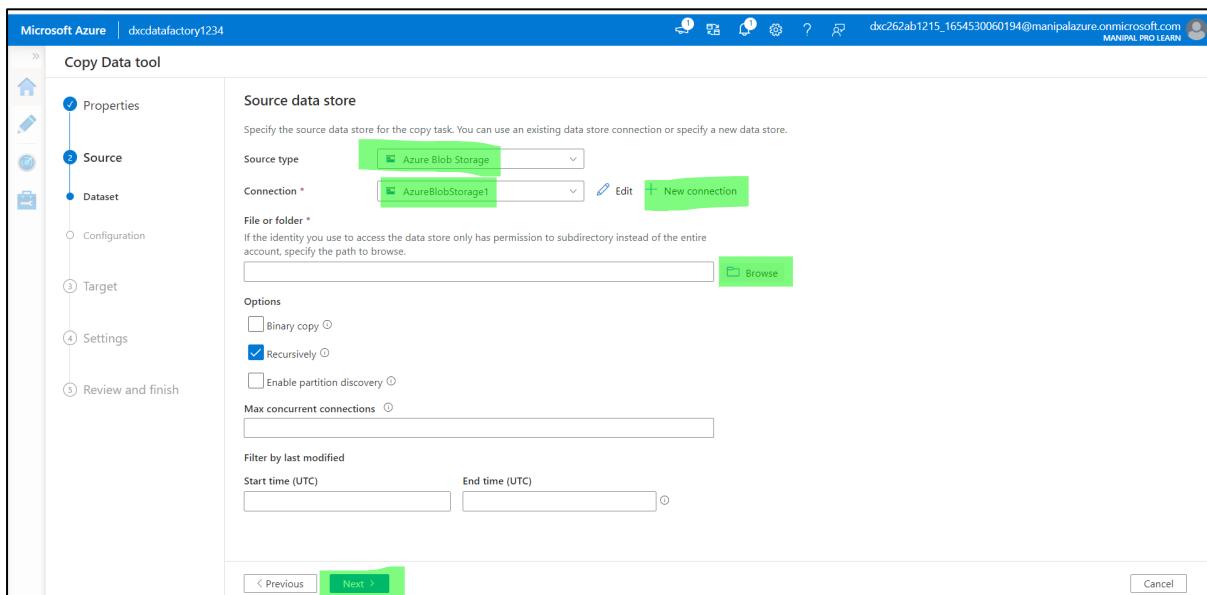
STEP 2: After entering into the ingest, we have to select built-in copy task. And then click on next.



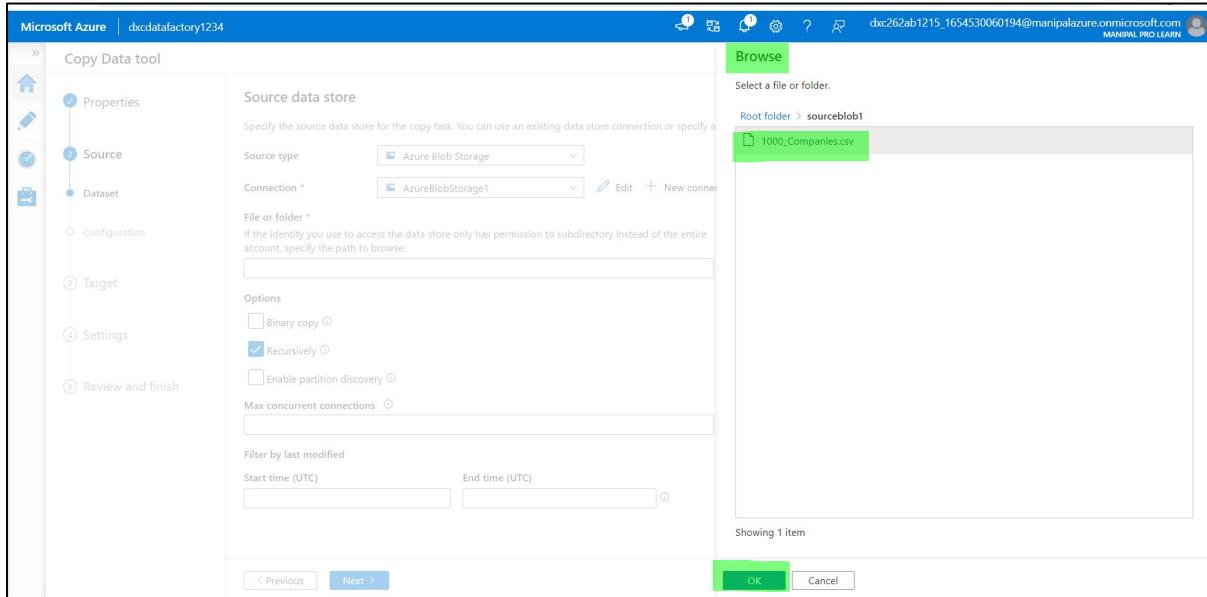
STEP 3: We can see the source data where source information is given.enter source type and add new connection as shown below.



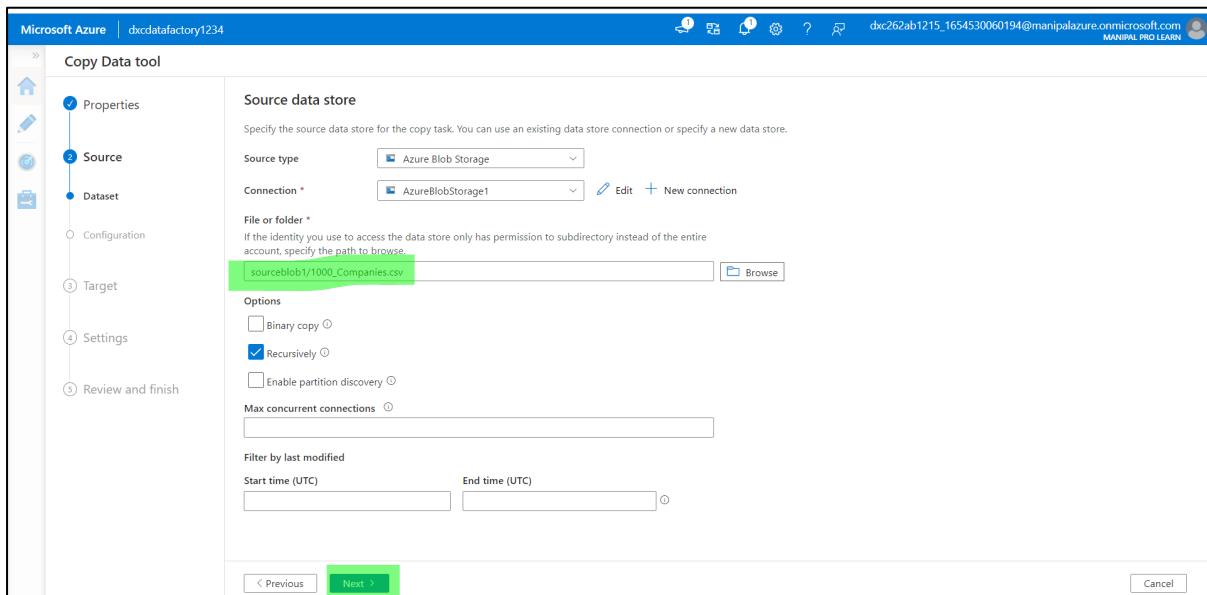
STEP 4: After connection, the file or folder need to uploaded.



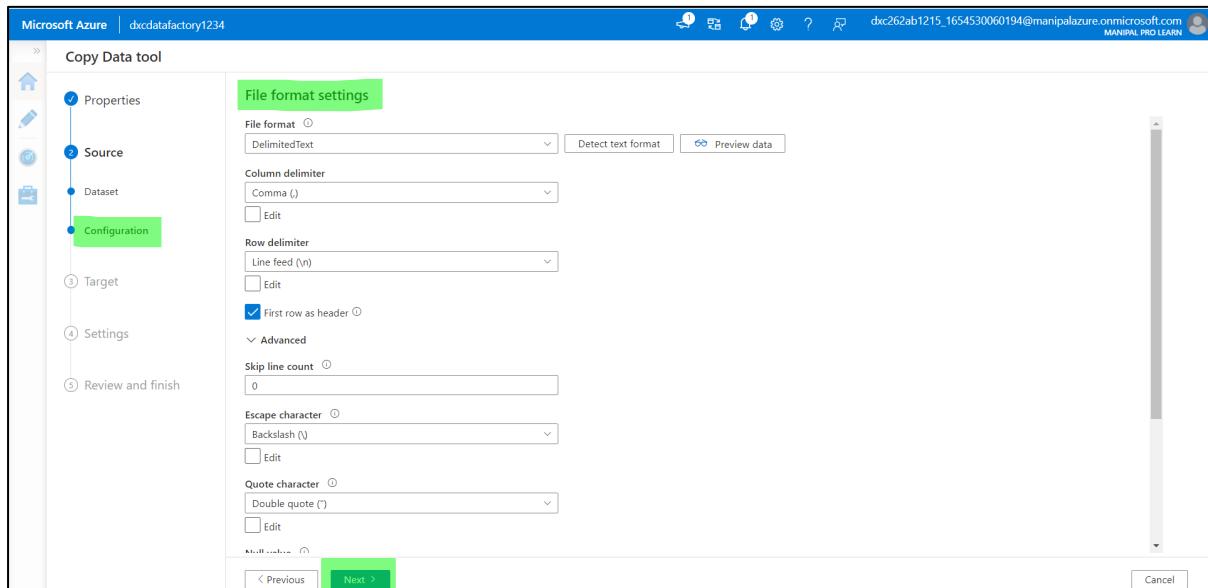
STEP 5: As previously we uploaded “1000_companies.csv” file in sourceblob1. Now, we can browse and select it.



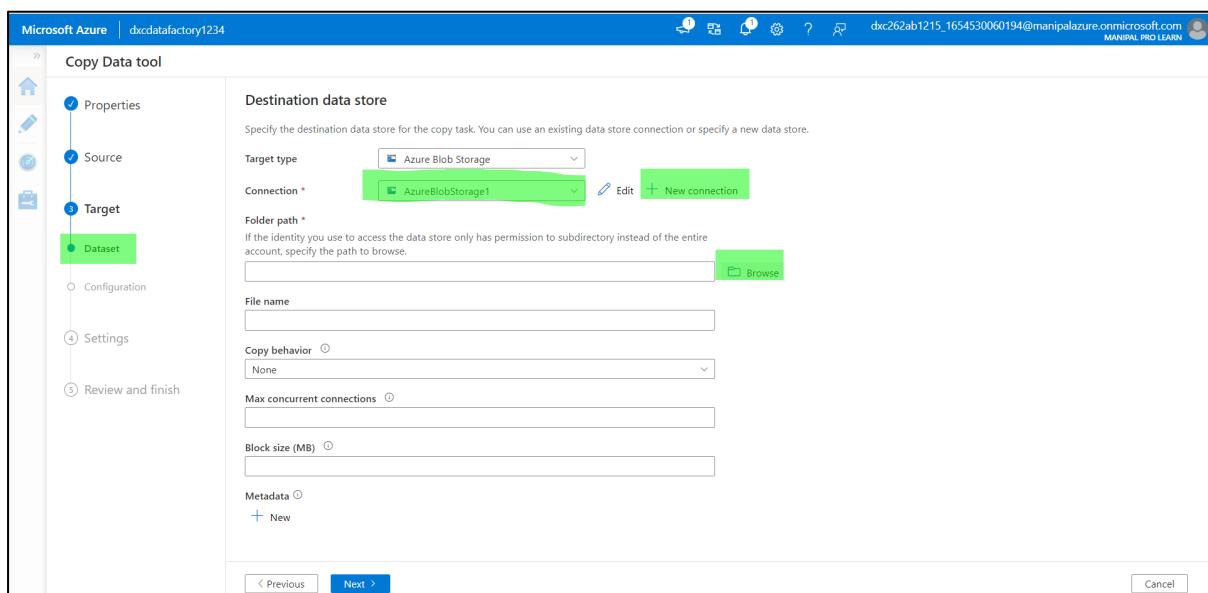
STEP 6: The file is uploaded and click on next.



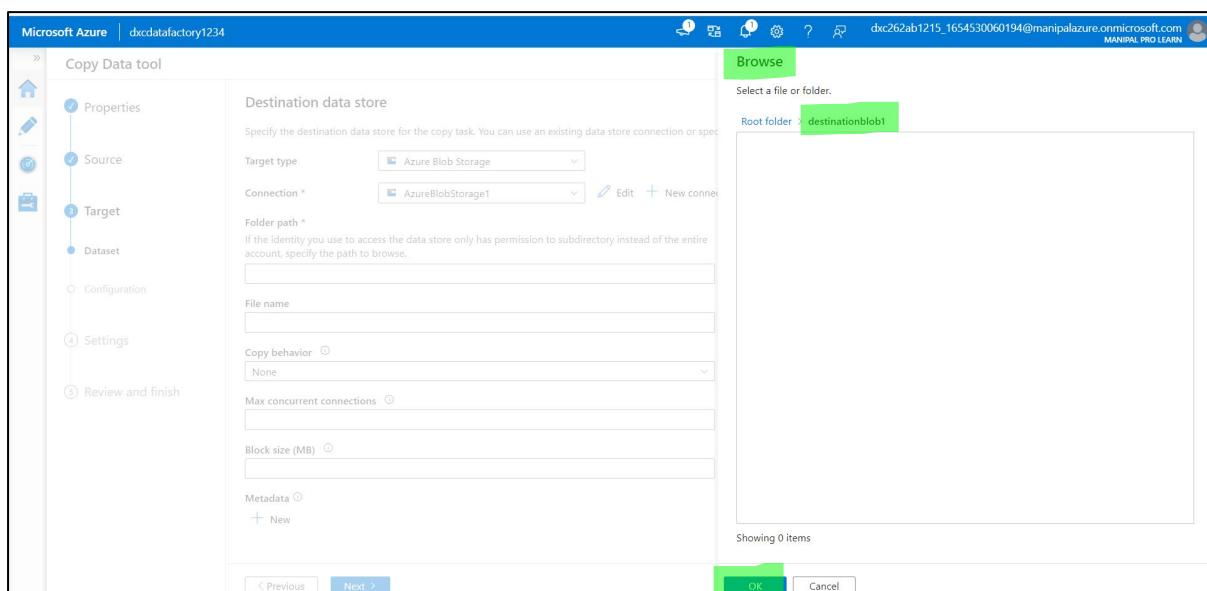
STEP 7: The file format settings are as shown.



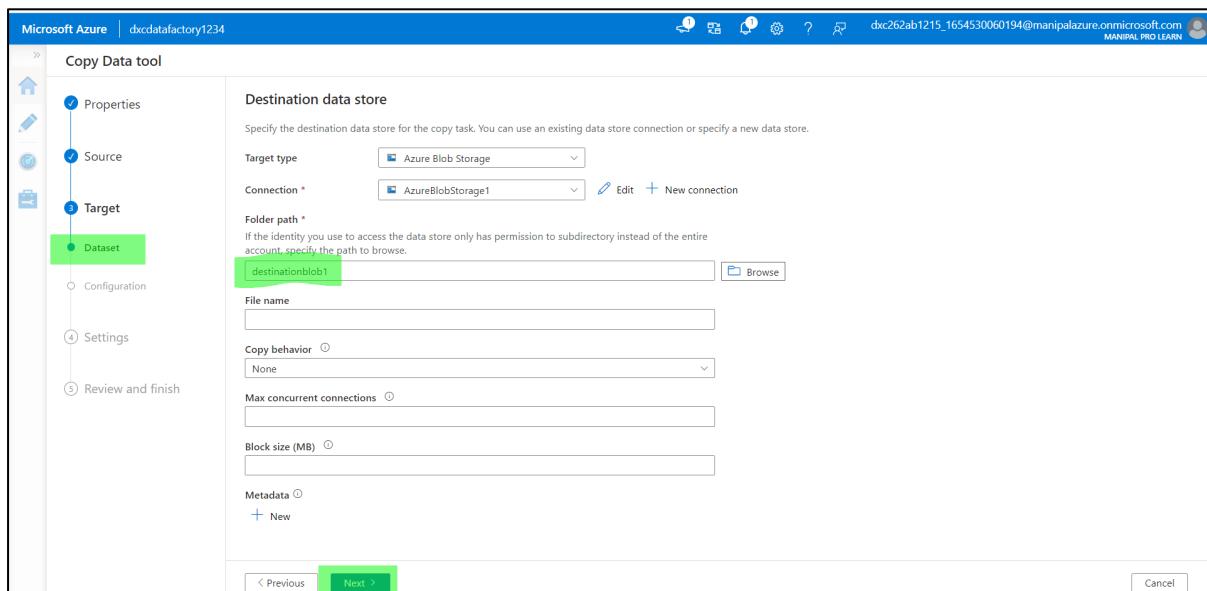
STEP 8: Now destination data store. Enter target type and connect.



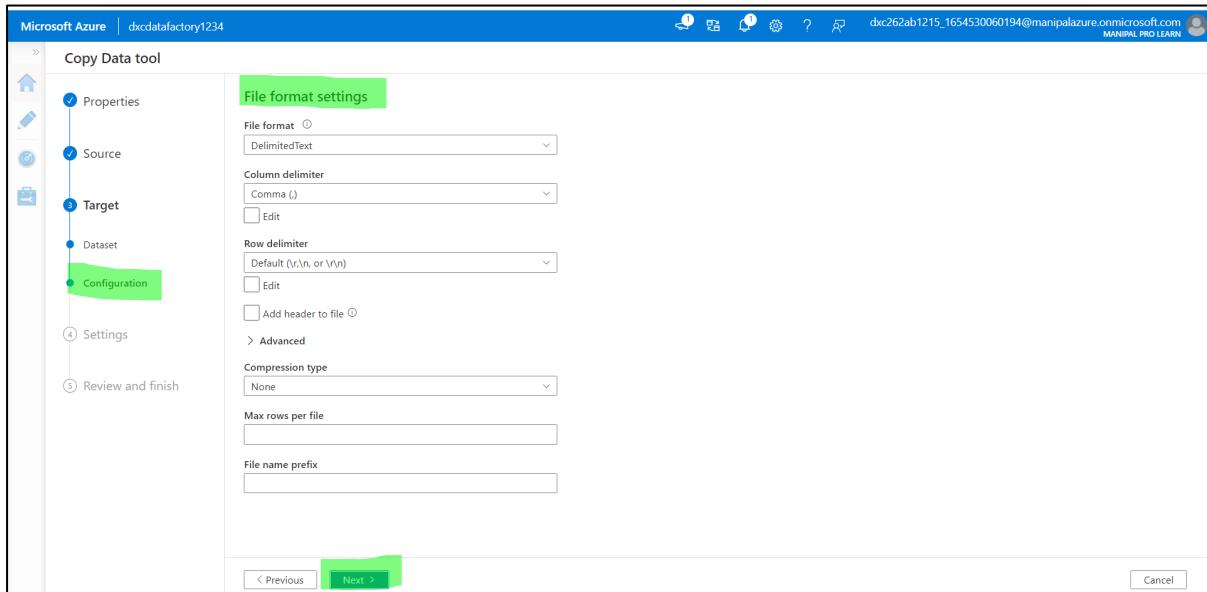
STEP 9: As the data need to be transferred from source to destination. No need to upload any file in it. Just need to select destinationblob1.



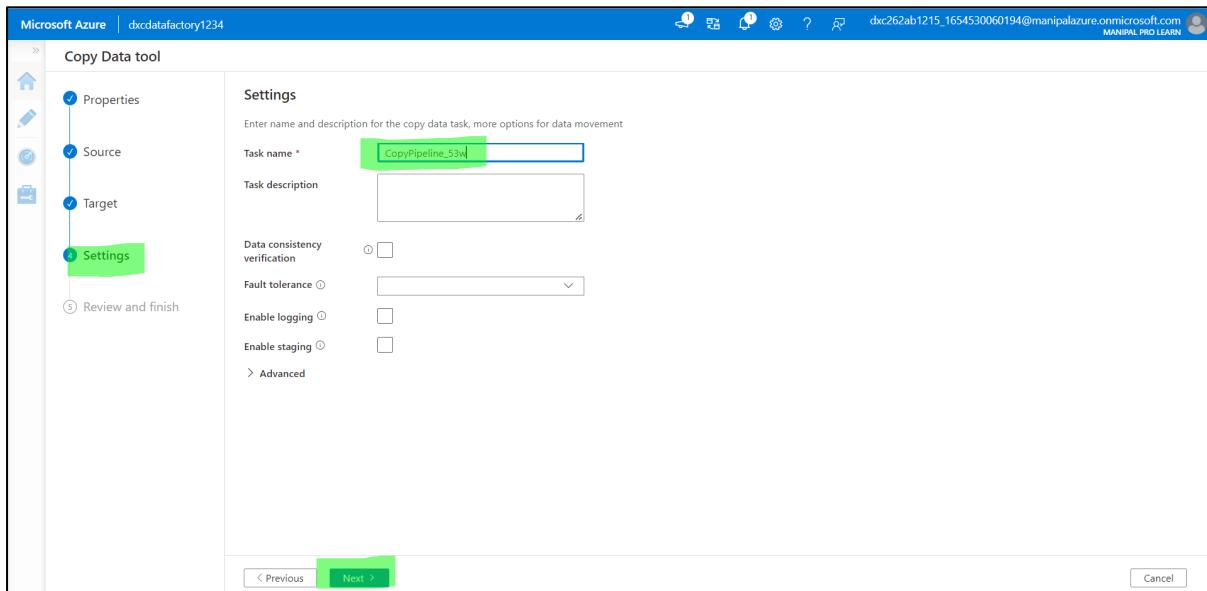
STEP 10: As the details are as shown and click on next.



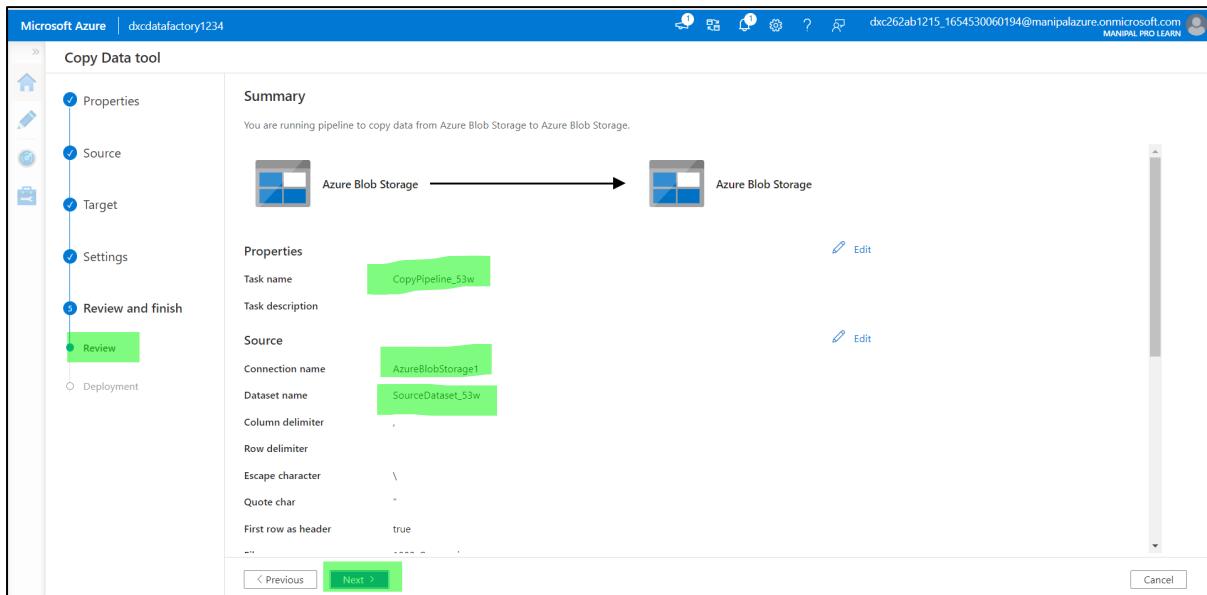
STEP 11: We enter into file format settings and details should as shown below.



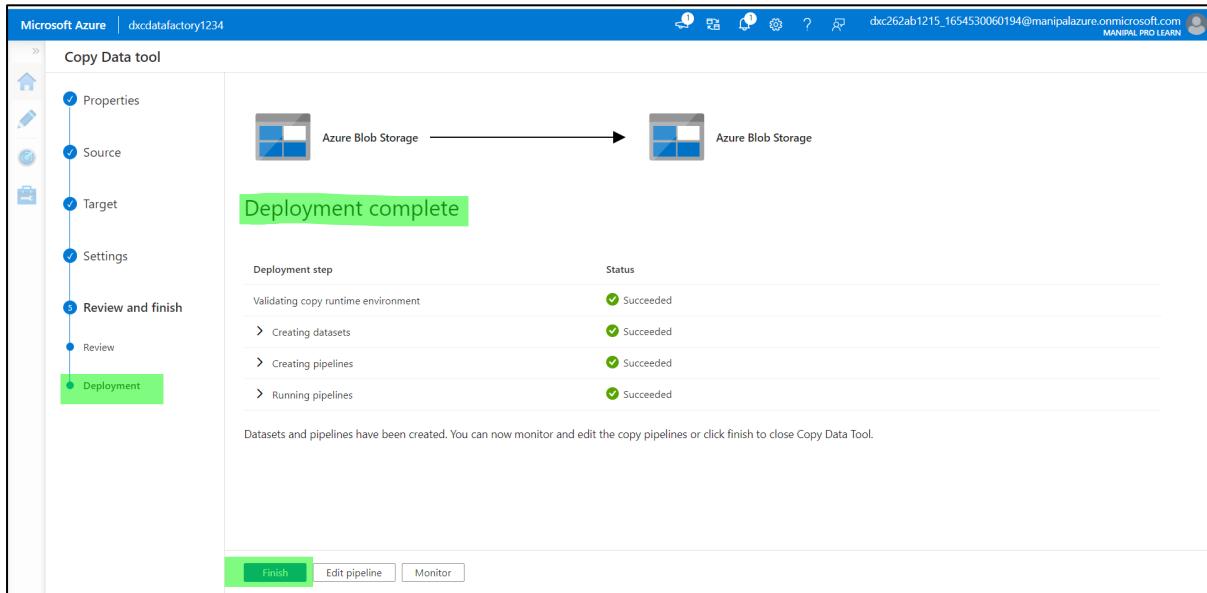
STEP 12: We enter into settings where we can change the name for pipelines or remain same.



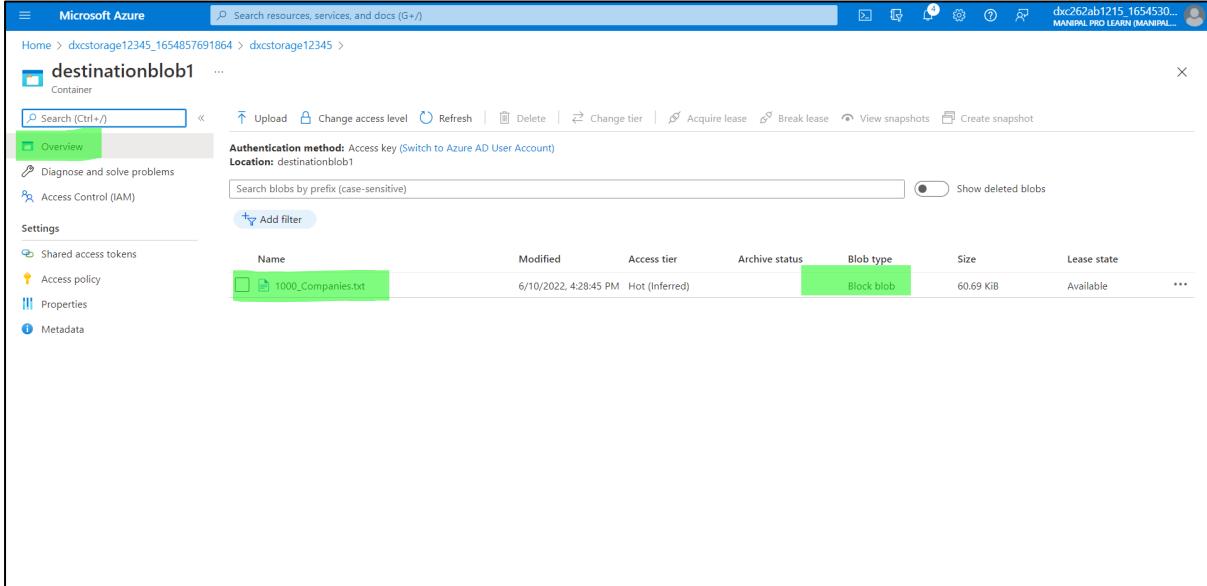
STEP 13: Here comes to summary we can see azure blob storage for source and azure blob storage for destination. And click on next.



STEP 14: Then deployment and click on finish.



STEP 15: After the pipeline created we can see the data (100_companies.csv) is in destinationblob1.

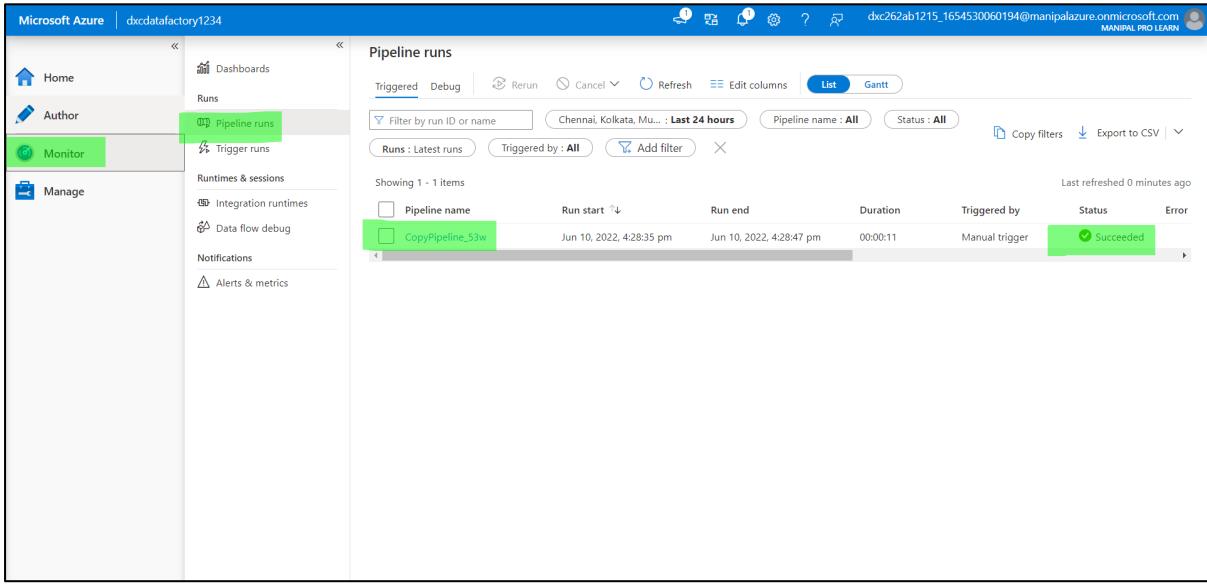


The screenshot shows the Microsoft Azure Storage Explorer interface. The left sidebar shows the navigation path: Home > dxstorage12345_1654857691864 > dxstorage12345 > destinationblob1. The main area is titled "destinationblob1 Container". The "Overview" tab is selected. It displays the following information:

- Authentication method: Access key (Switch to Azure AD User Account)
- Location: destinationblob1
- Search blobs by prefix (case-sensitive):
- Show deleted blobs:
- Add filter:
- Table of blobs:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
1000_Companies.txt	6/10/2022, 4:28:45 PM	Hot (Inferred)		Block blob	60.69 KiB	Available

STEP 16: For running a pipeline select monitor in datafactory and rerun it, it will get succeeded as shown below.

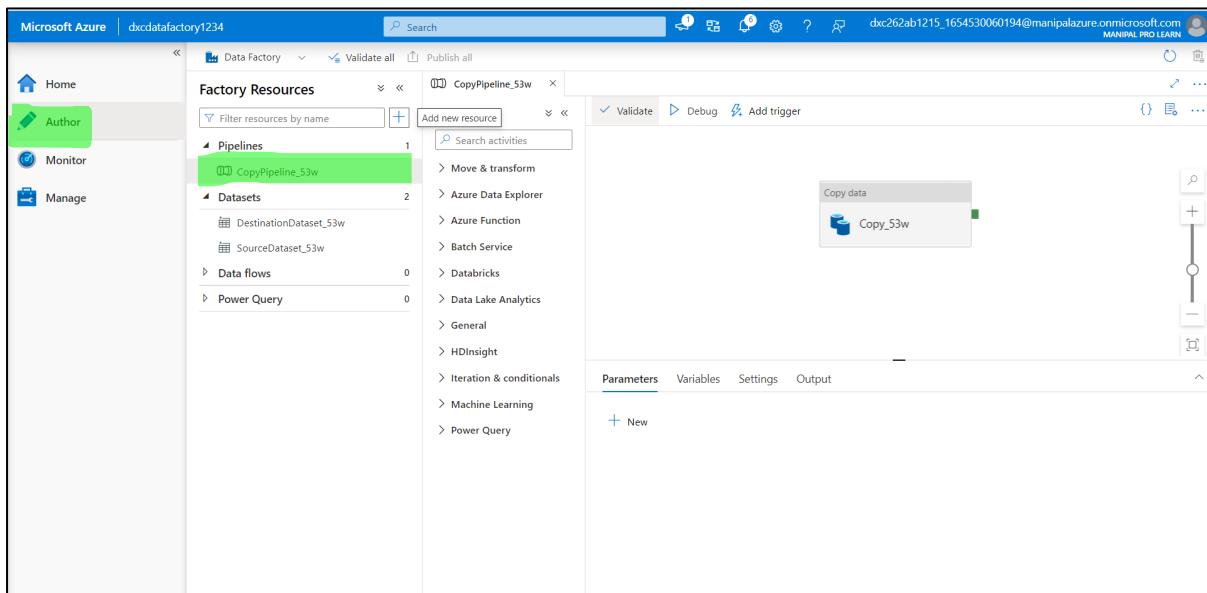


The screenshot shows the Microsoft Azure Data Factory Pipeline runs page. The left sidebar shows the navigation path: Home > dxdatafactory1234 > Monitor. The main area is titled "Pipeline runs". It displays the following information:

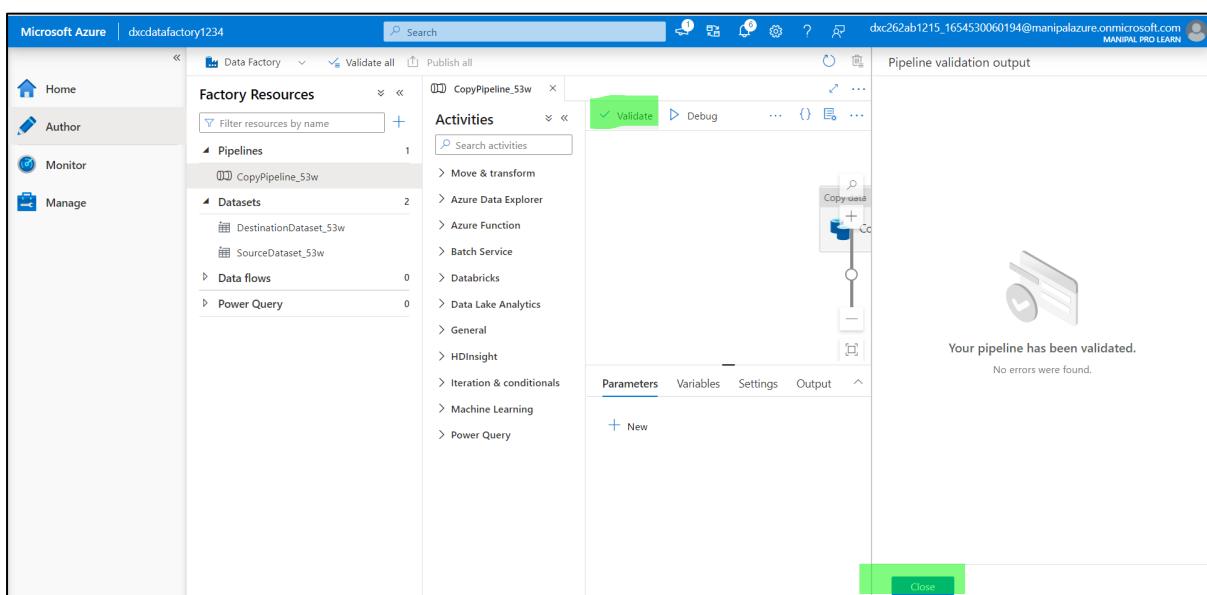
- Triggered, Debug, Rerun, Cancel, Refresh, Edit columns, List, Gantt buttons.
- Filter by run ID or name: Chennai, Kolkata, Mu... : Last 24 hours, Pipeline name: All, Status: All.
- Runs: Latest runs, Triggered by: All, Add filter.
- Last refreshed 0 minutes ago.
- Showing 1 - 1 items.
- Table of pipeline runs:

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error
CopyPipeline_53w	Jun 10, 2022, 4:28:35 pm	Jun 10, 2022, 4:28:47 pm	00:00:11	Manual trigger	Succeeded	

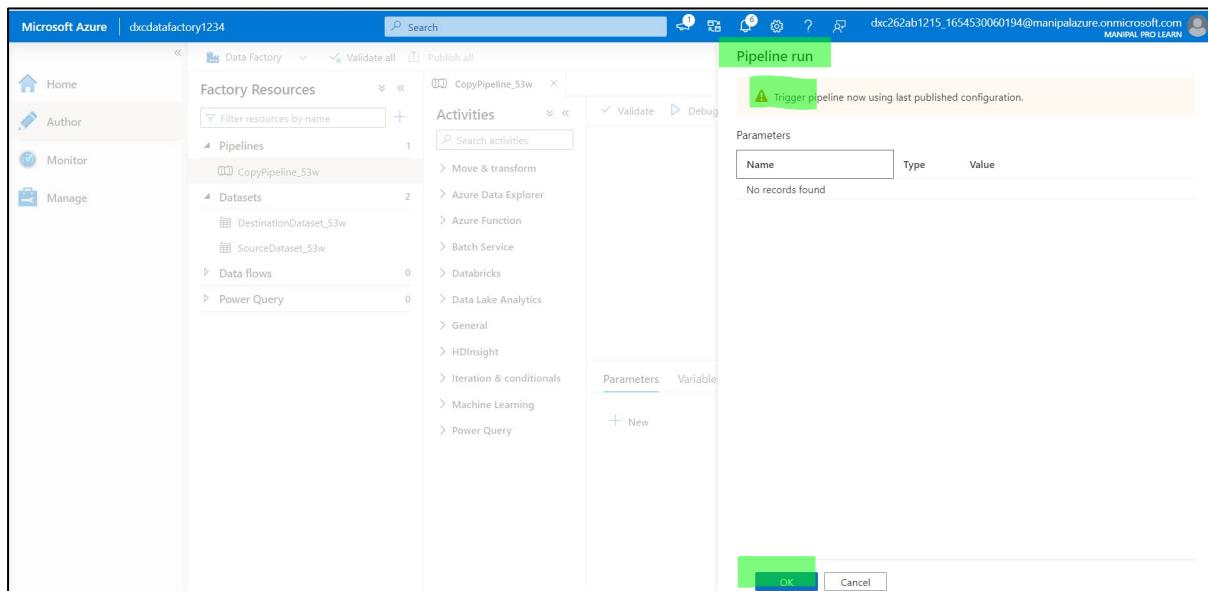
STEP 17: Select author and we can view the pipeline that created previously.



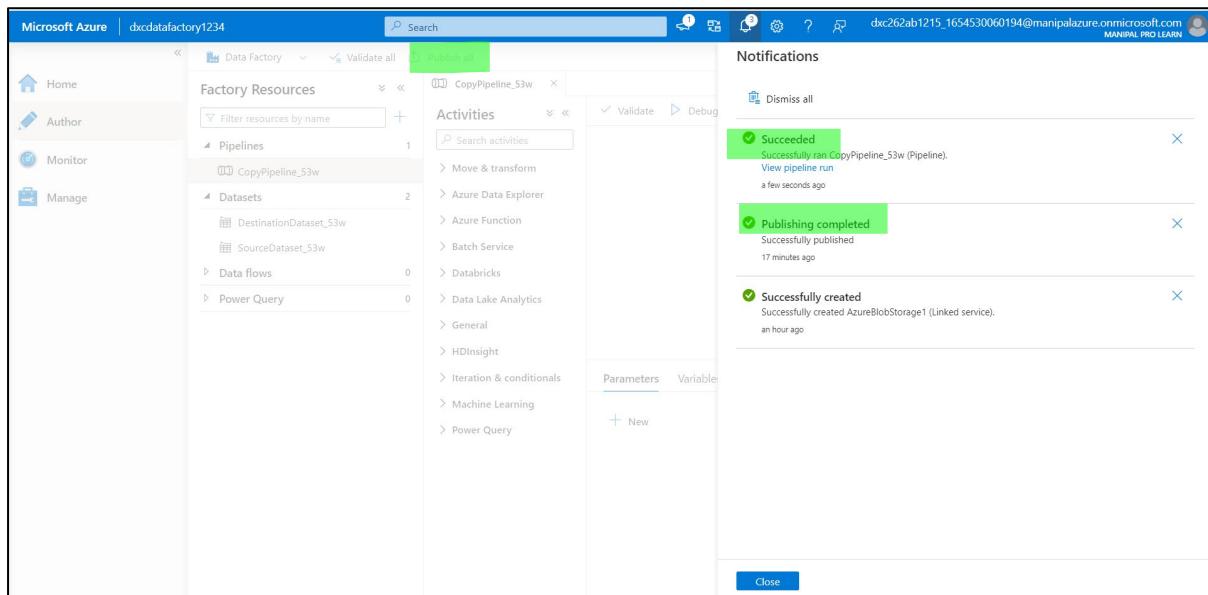
STEP 18: Validate the pipeline as shown and close it.



STEP 19: Then choose trigger now for pipeline or we can schedule to run by choosing new trigger option.

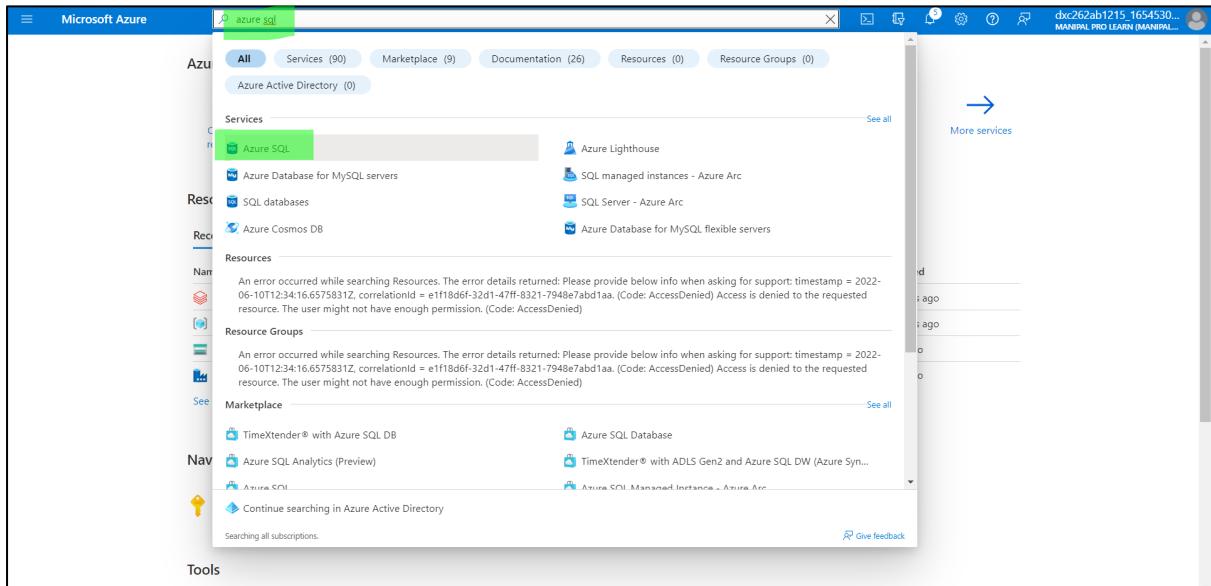


STEP 20: The trigger is ran successfully and publish the pipeline.

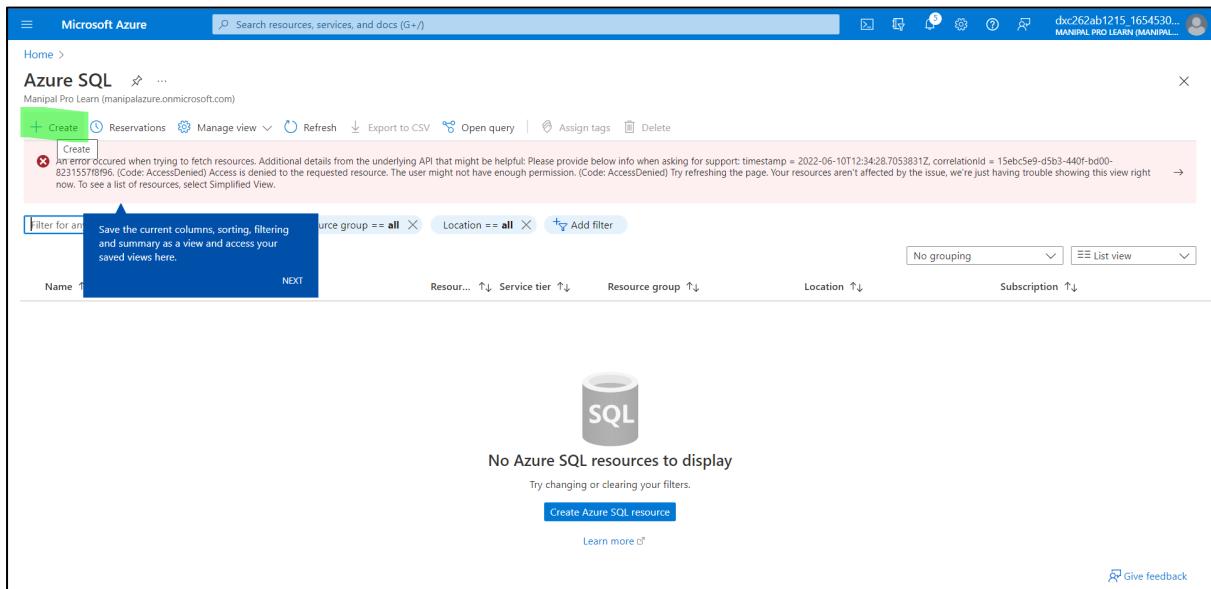


➤ Practical Lab: Create Azure SQL Server and Database

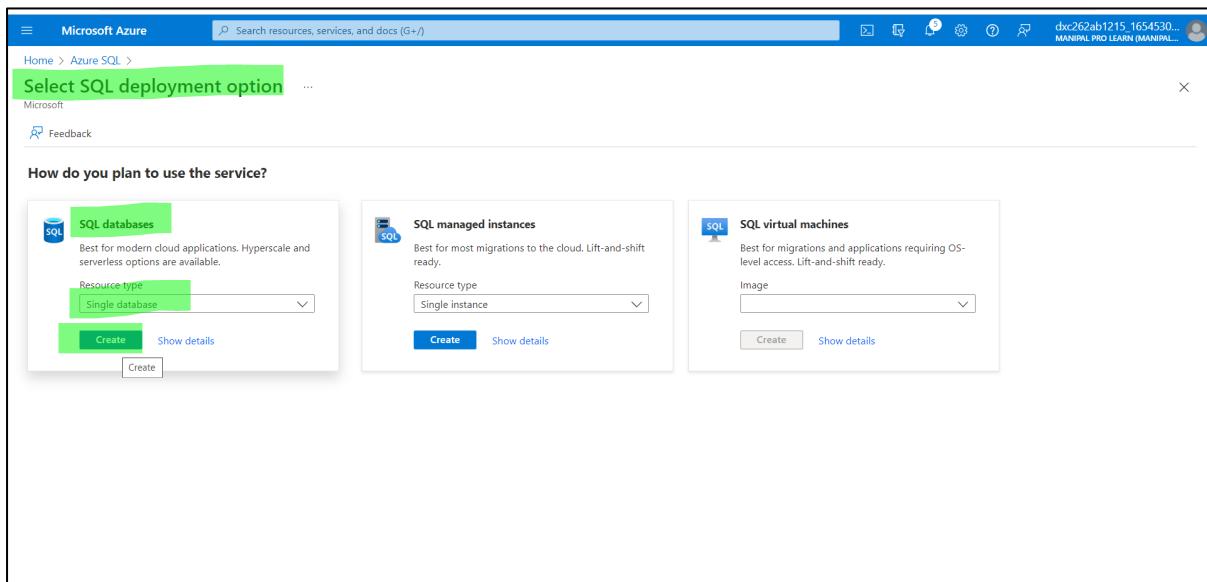
STEP 1: Search azure SQL and select it.



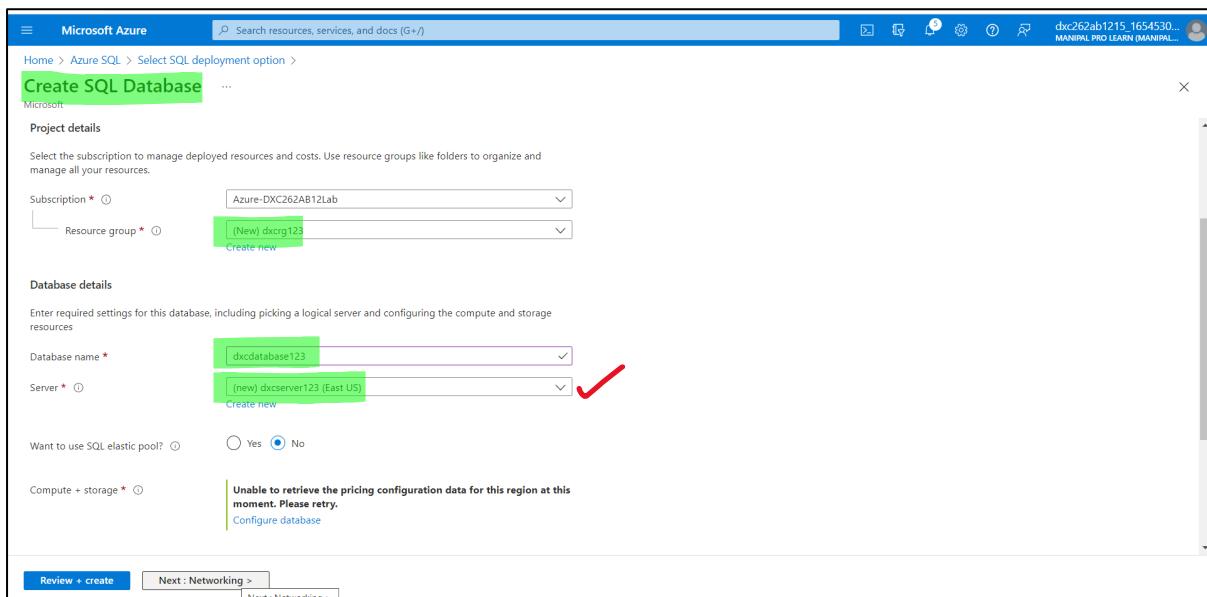
STEP 2: After selecting azure SQL . Click on “+ create”.



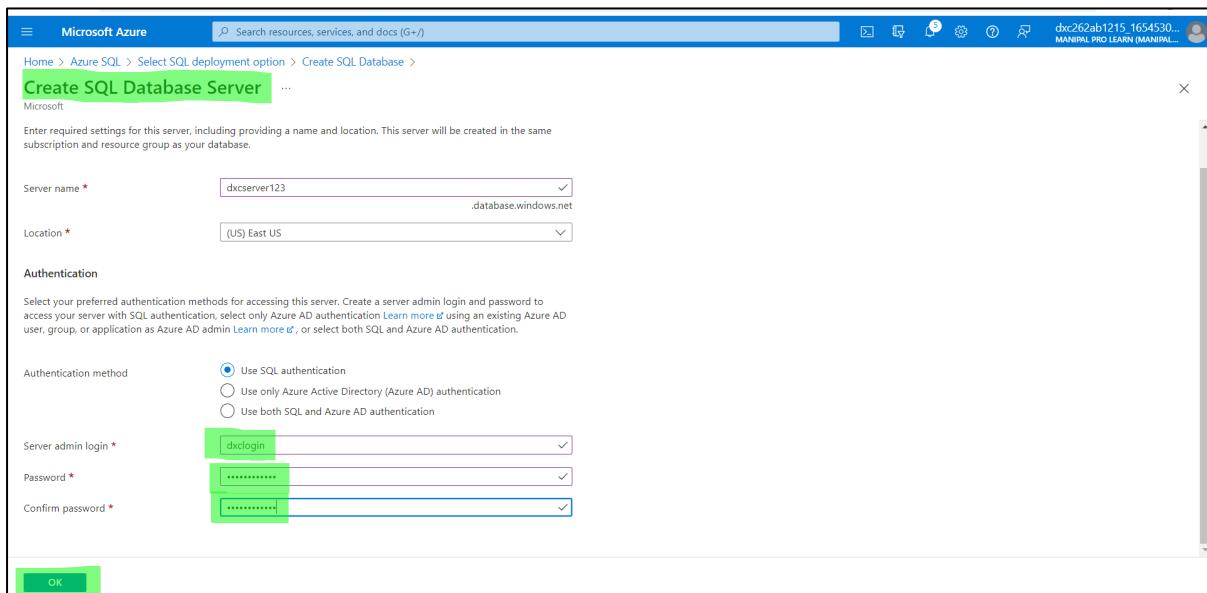
STEP 3: Select SQL databases and single databases and click on create.



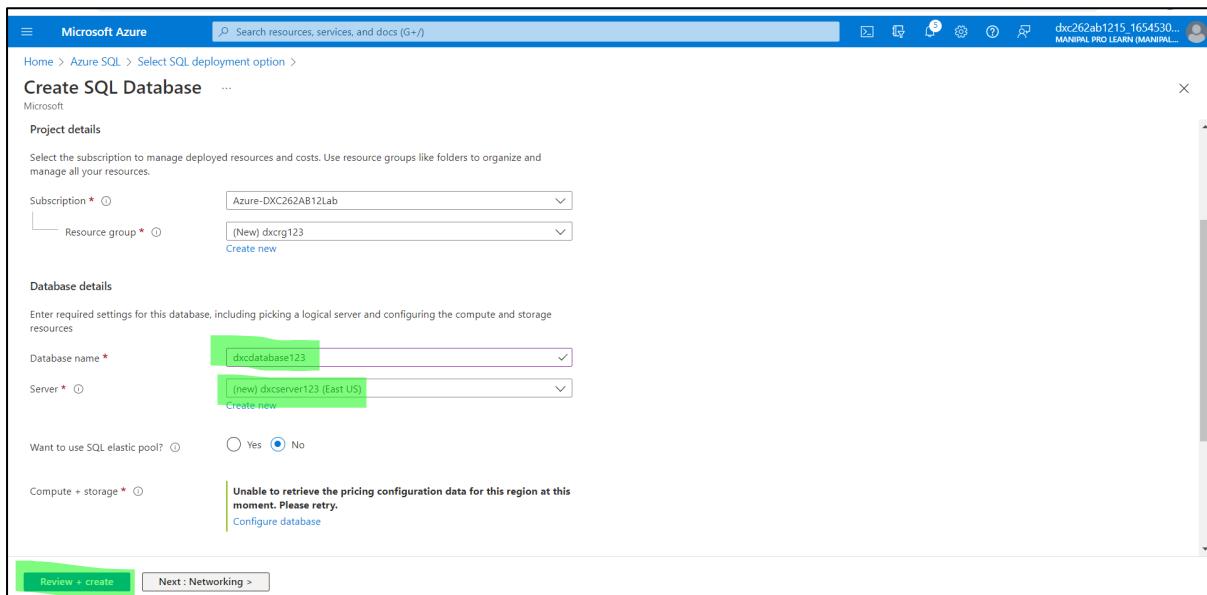
STEP 4: Enter the details of resource group, database name, Now select server.



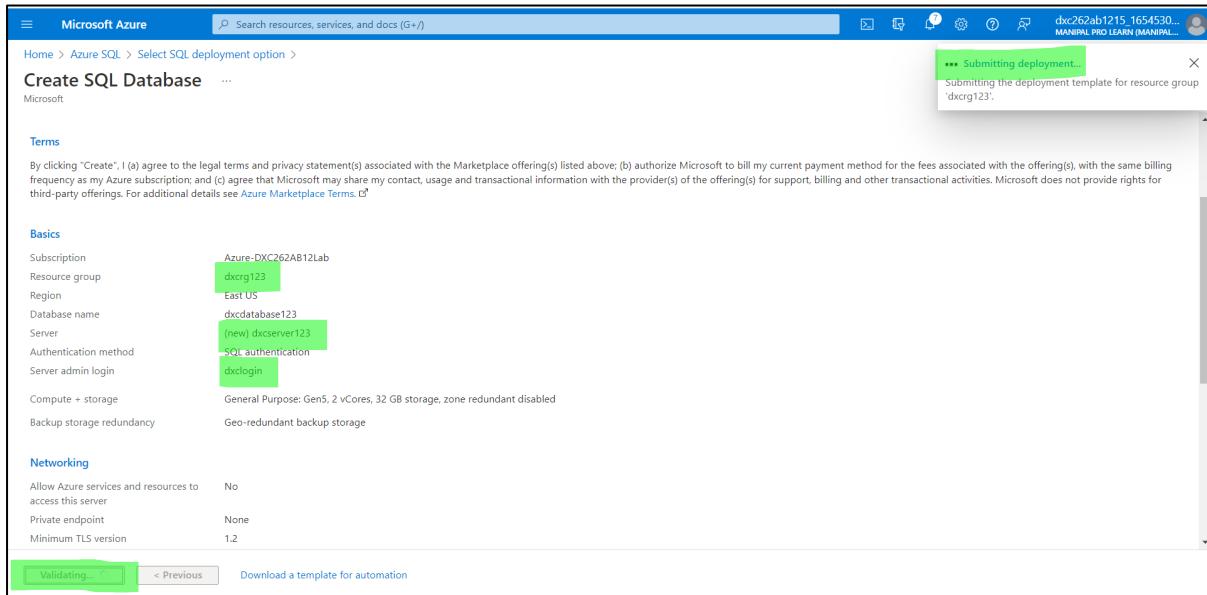
STEP 5: Now for creating the server enter the server name and server admin login, password, confirm password.



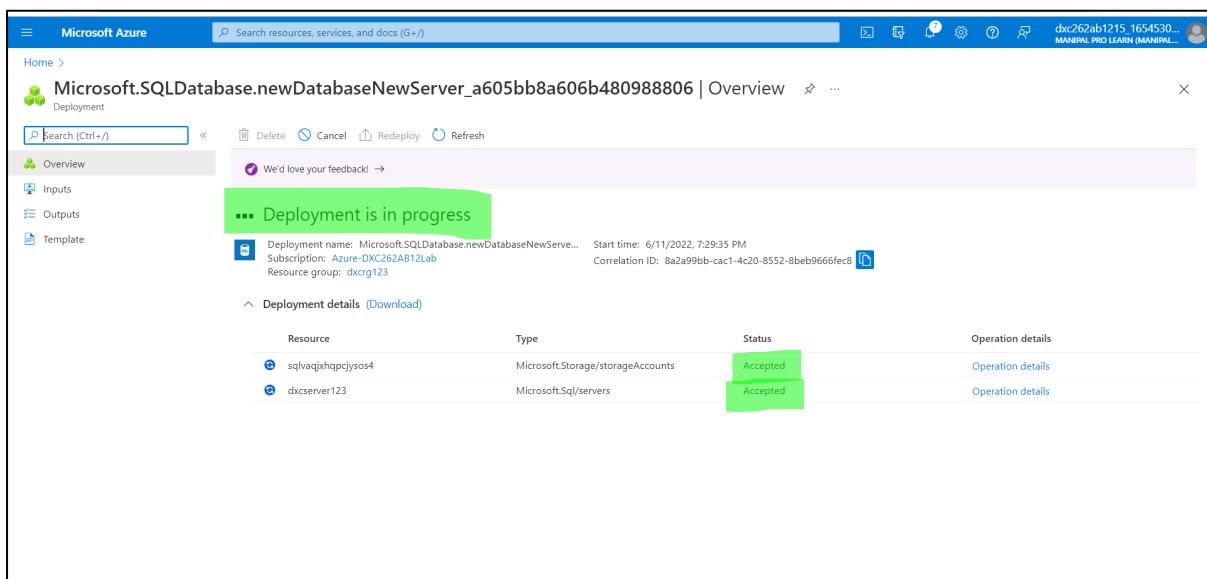
STEP 6: After entering server details , then click on review+create.



STEP 7: Click on create after reviewing the details and the process is validating.



STEP 8: Then deployment is in progress.



STEP 9: After few minutes the deployment is completed and then go to resource.

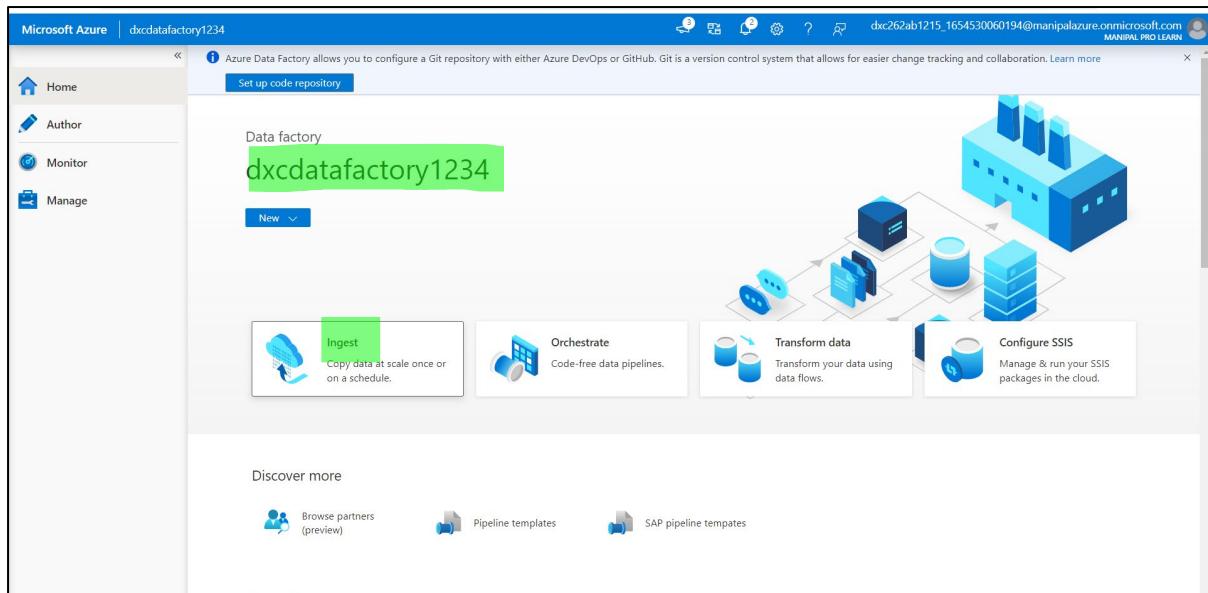
The screenshot shows the Microsoft Azure Deployment Overview page for a completed deployment named "Microsoft.SQLDatabase.newDatabaseNewServer_a605bb8a606b480988806". The deployment status is "Your deployment is complete". Deployment details include a start time of 6/11/2022, 7:29:35 PM, and a correlation ID of 8a2a99bb-cac1-4c20-8552-8beb9666fec8. A "Go to resource" button is visible. On the right side, there are promotional cards for "Cost Management", "Microsoft Defender for Cloud", "Free Microsoft tutorials", and "Work with an expert".

STEP 10: In resource we can observe overview the details of database.

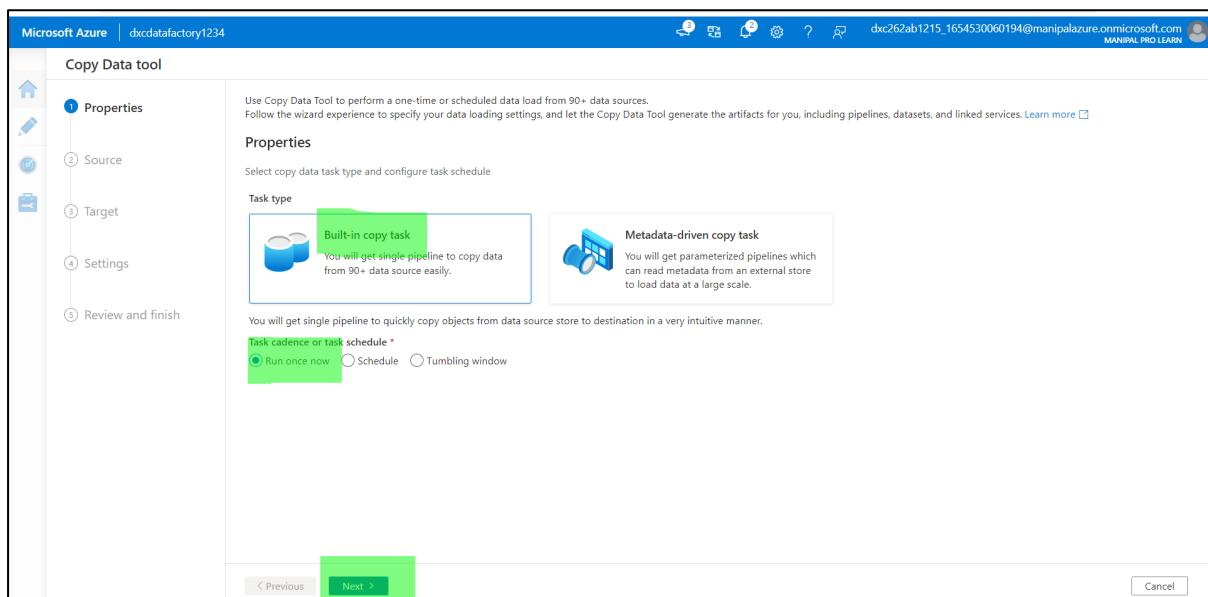
The screenshot shows the Microsoft Azure SQL Database Overview page for a database named "dxcdatabase123" located in "dxserver123/dxcdatabase123". The database was just created. The "Essentials" section displays details such as Resource group (move), Status (Online), Location (East US), Subscription (move), Subscription ID, Tags, and more. It also shows Compute utilization and Database data storage metrics, including a chart showing 0.01% used space (4 MB) over the last 1 hour.

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

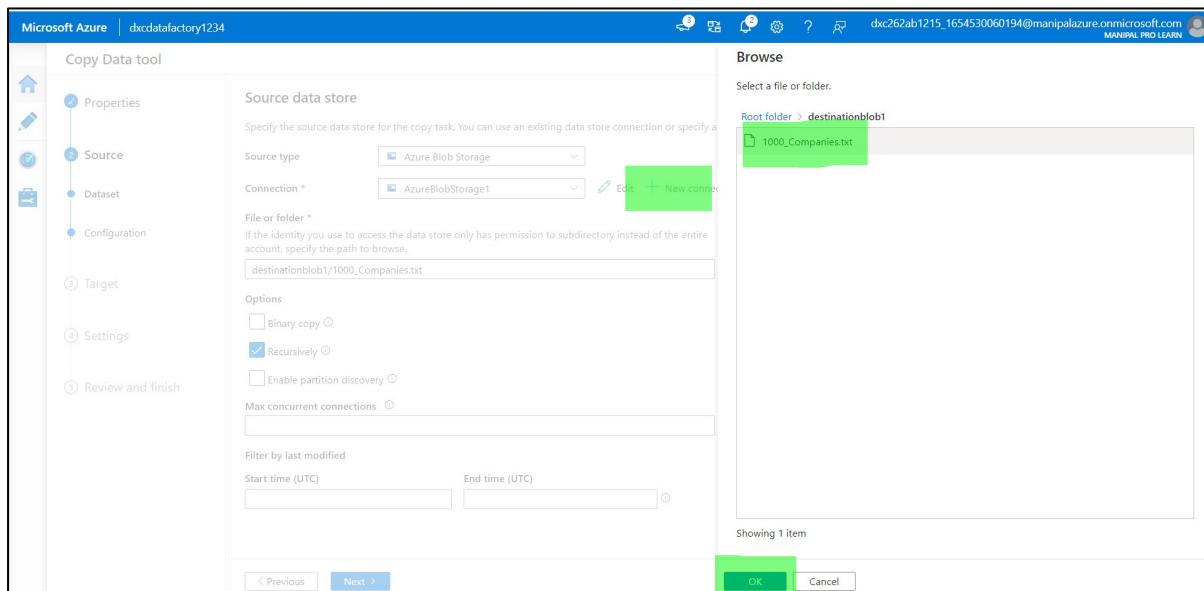
STEP 1: Now, in data factory we need to add another pipeline for staging the data to SQL DB. Click on ingest.



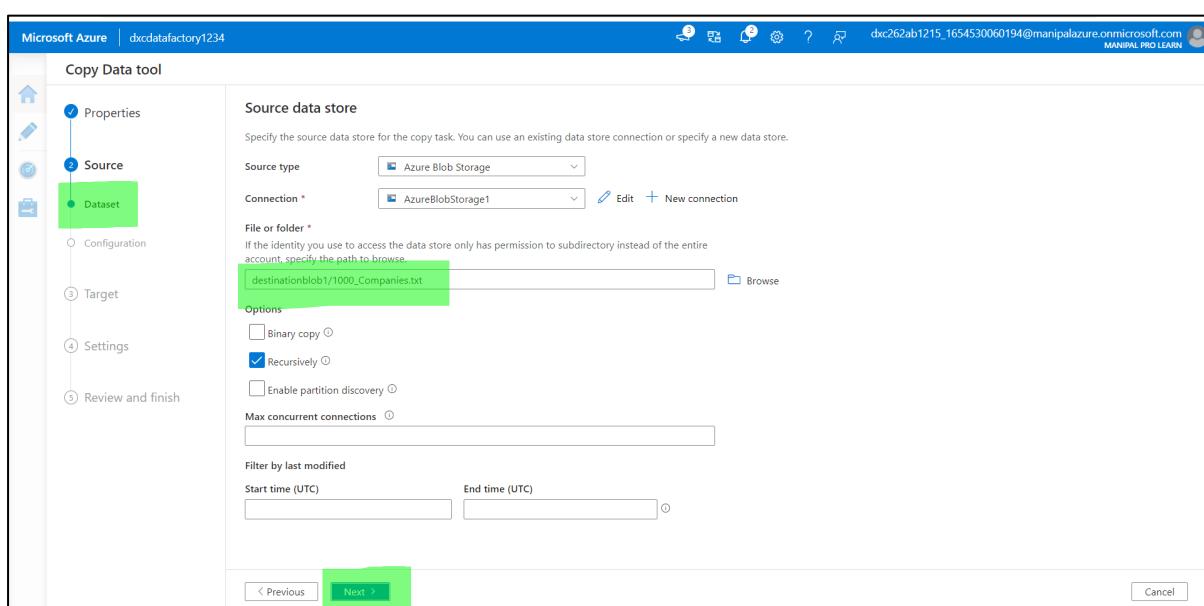
STEP 2: Click on next for built in copy task.



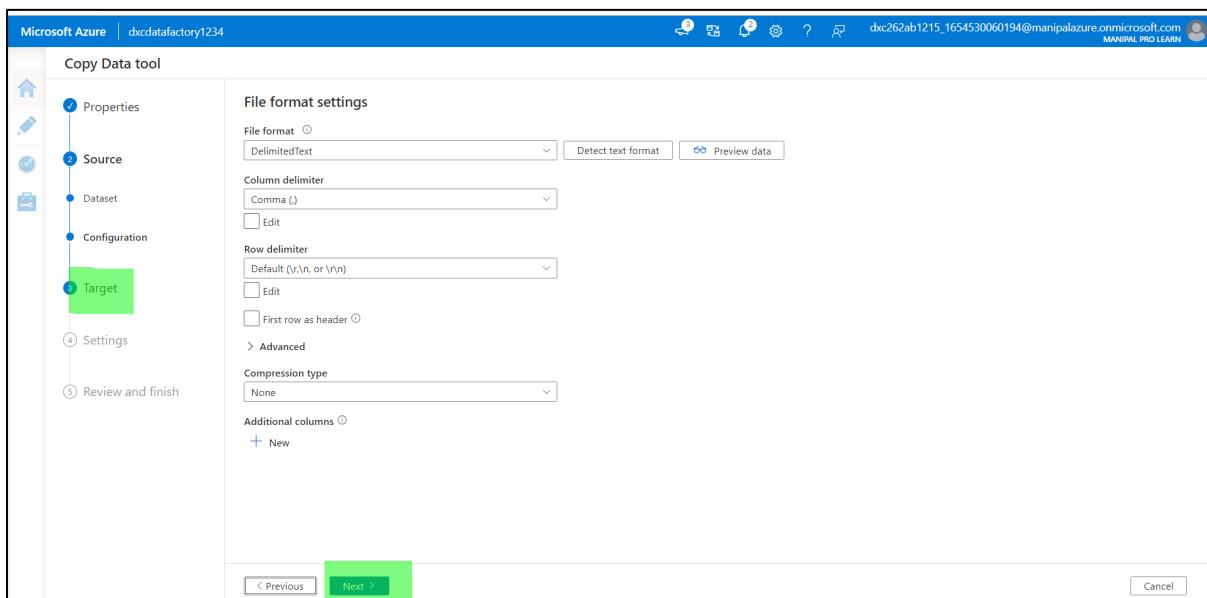
STEP 3: In source data store connect with blob storage and browse the data .



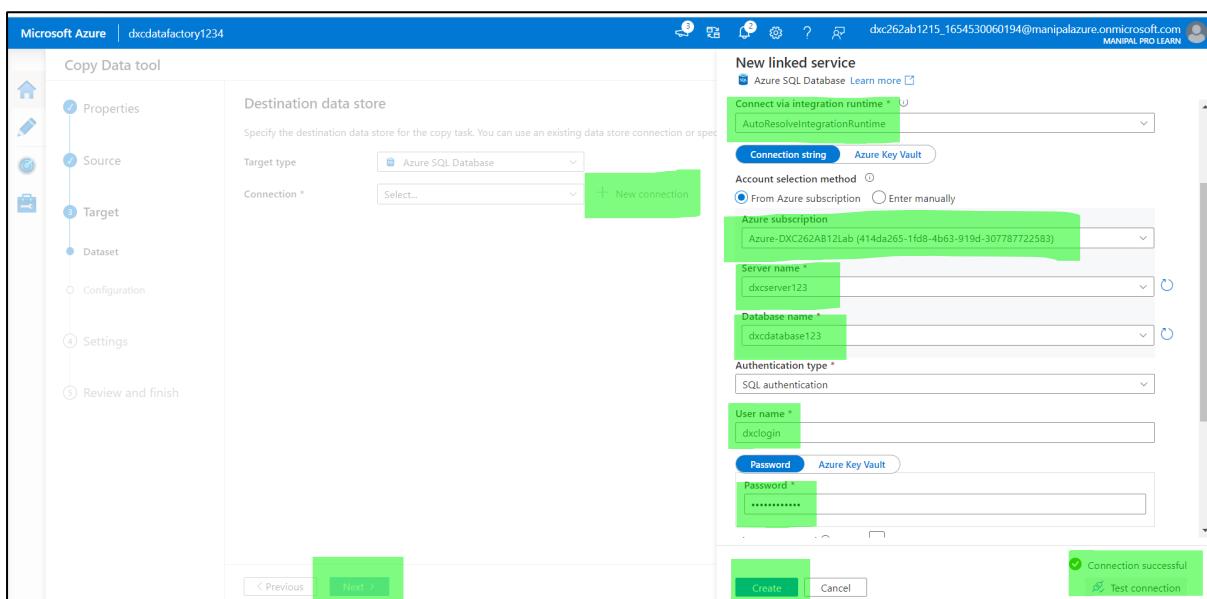
STEP 4: After browsing then click on next.



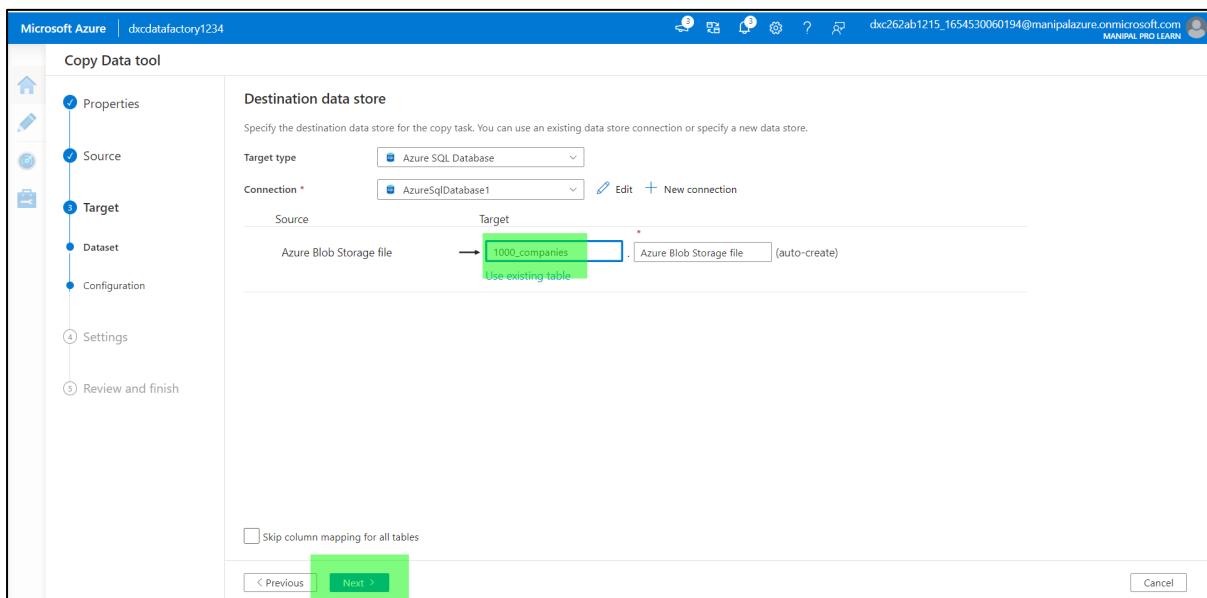
STEP 5: In target , click on next for file format settings.



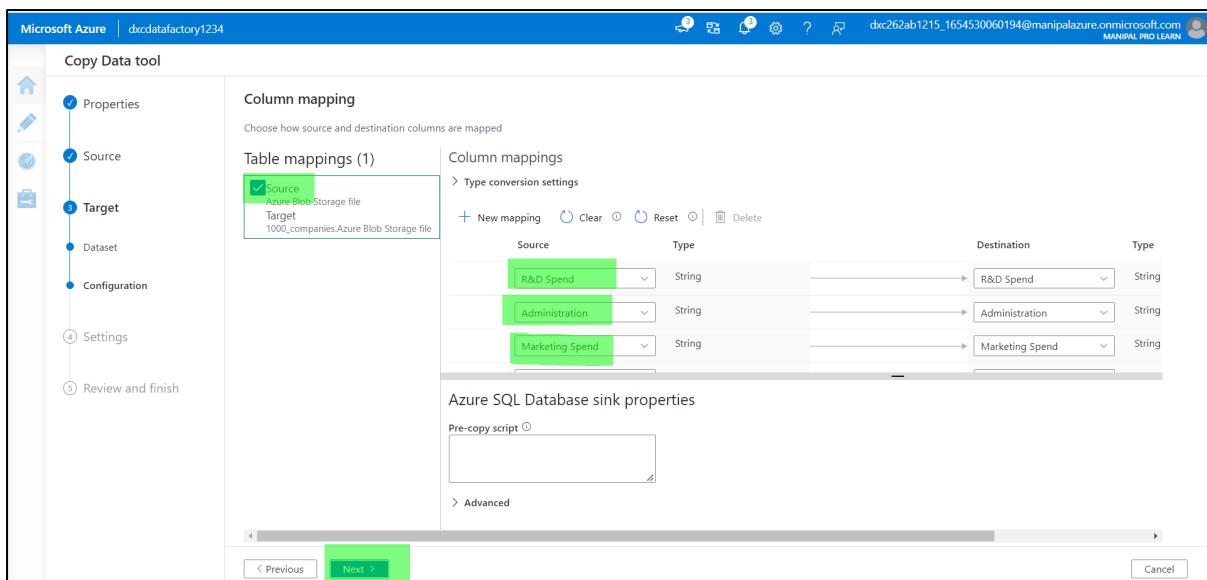
STEP 6: In destination data store create new connection , by new linked service by entering the details of server name, database name and test connection . Then, click on create.



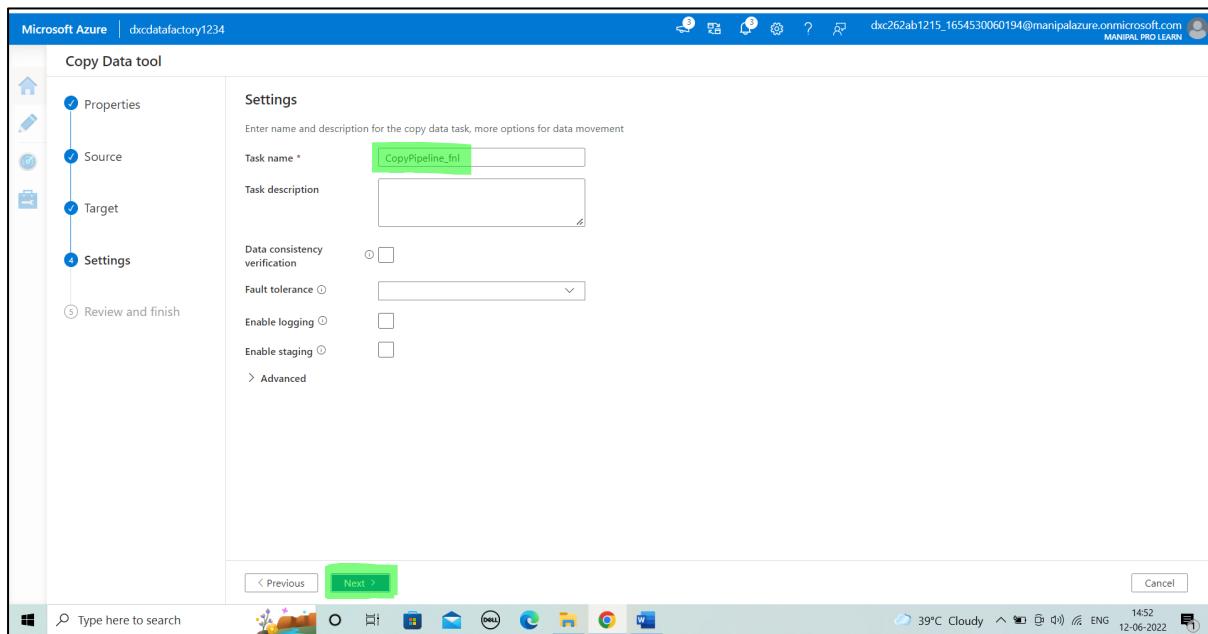
STEP 7: Enter the target details as the file that uploaded.



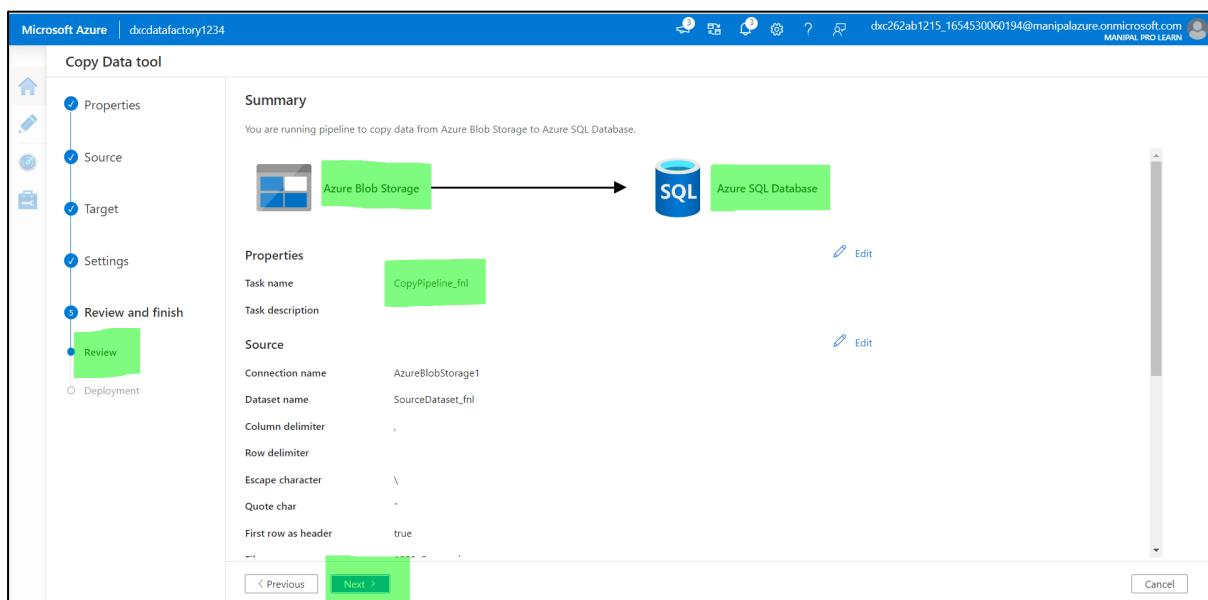
STEP 8: Now, check the column mapping and click on next.



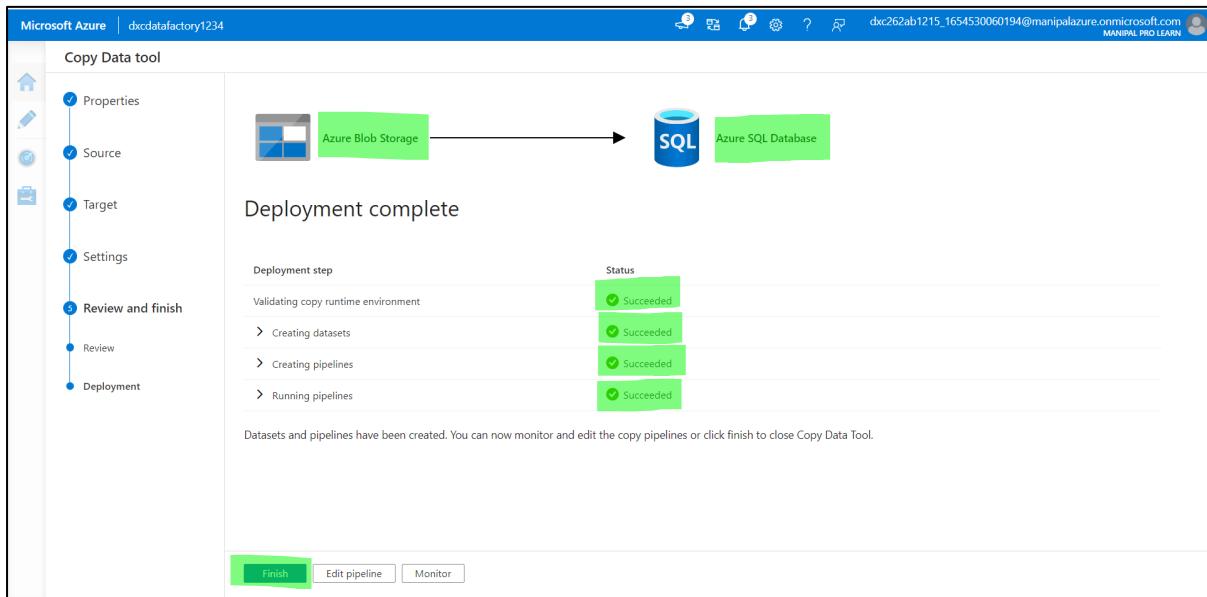
STEP 9: In setting task name is given as pipeline and then > next.



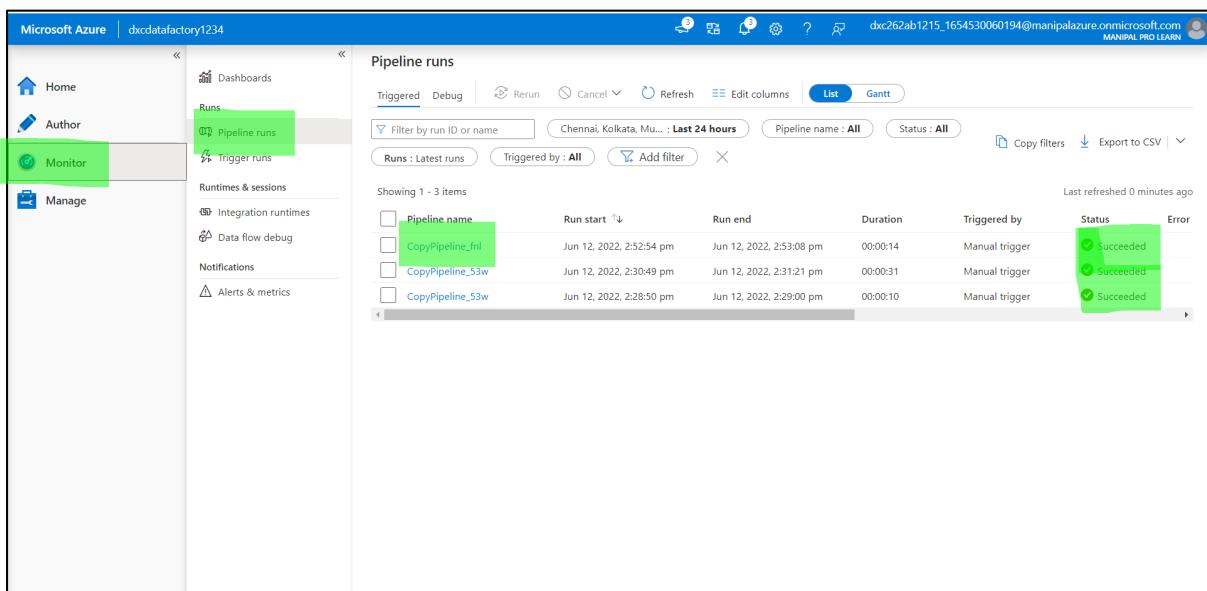
STEP 10: Check the summary details and then > next.



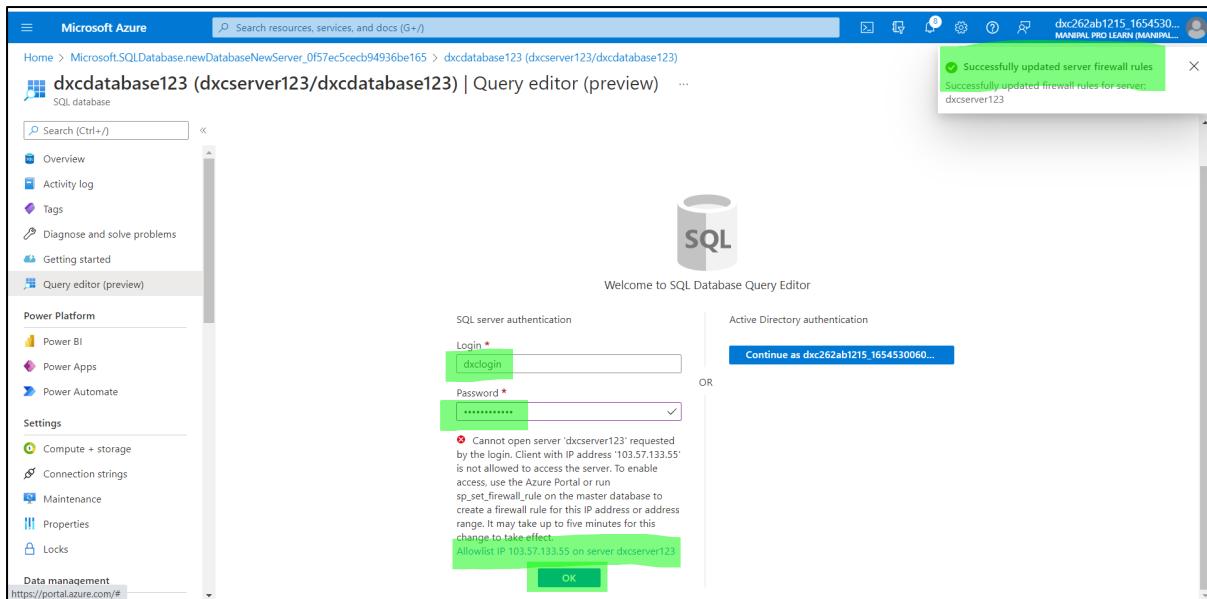
STEP 11: The deployment is completed. And click on finish.



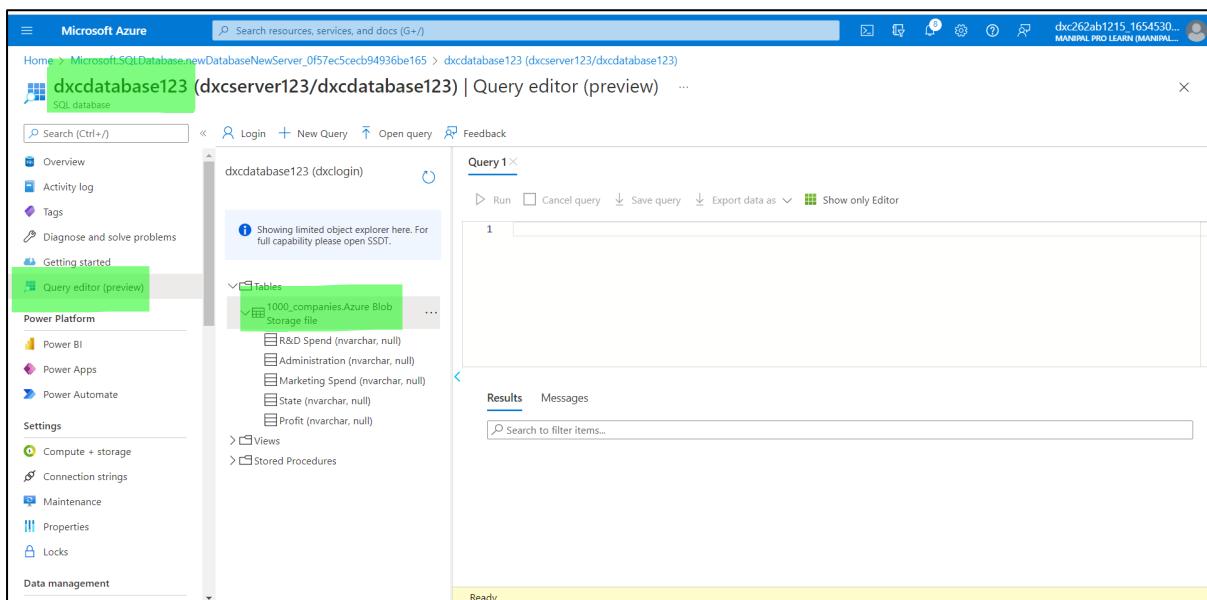
STEP 12: After creating the pipeline . In monitor the pipeline is viewed and it is succeeded.



STEP 13: Coming to database to check the pipeline successful by the data is available in it . Login with details shown and allow it by updating the firewall rules.



STEP 14: After login we can view the data in query editor .



STEP 15: Run the data to view table .

The screenshot shows the Microsoft Azure portal interface for a SQL database named 'dxdatabase123'. The left sidebar includes links for Overview, Activity log, Tags, Diagnose and solve problems, Getting started, and Query editor (preview). The main area displays a 'Query editor (preview)' window titled 'Query 4'. The query is:

```
1 SELECT TOP (1000) * FROM [1000_companies].[Azure Blob Storage file]
```

The results section shows a table with the following data:

R&D Spend	Administration	Marketing Spend	State	Profit
165349.2	136897.8	471784.1	New York	192261.83
162597.7	151377.59	443898.53	California	191792.06
153441.51	101145.55	407934.54	Florida	191050.39
144372.41	118671.85	383199.62	New York	182901.99
141107.24	61204.77	562160.13	Illinois	142107.04

A message at the bottom of the results table says 'Query succeeded | 11s'.

Result: In this project 1, We created a pipelines that will validate and copy the blob data from source to destination and then to SQL database by triggering each using Azure Data Factory.

Conclusion: The Blob data moved and stored into SQL database by using pipelines in datafactory.

Name: SUPRIYA BHARATHA

Reg No: DXCAB1211

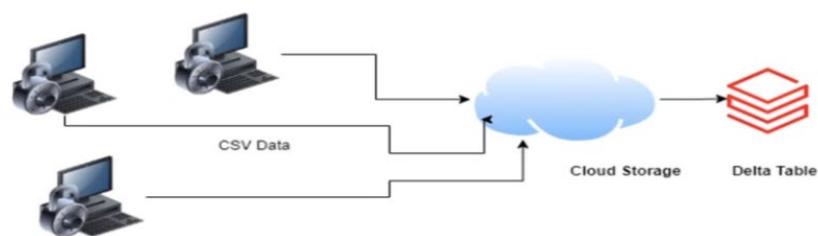
Project 2 Name: AP Morgan Data Platform

Date: 12th JUNE 2022

Project 2: AP Morgan Data Platform

Project 2 : AP Morgan

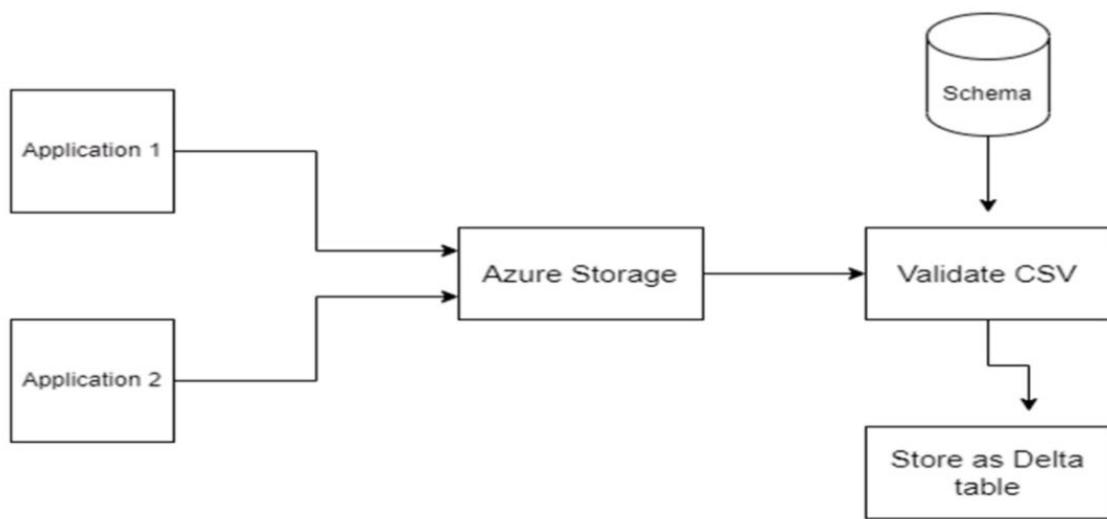
- Multiple Internal applications sends the data(huge size) in CSV format on daily basis in the cloud storage location. There are couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed data to be persisted as Delta table in Databricks for downstream system.



Project 2 : AP Morgan- High Level Detail

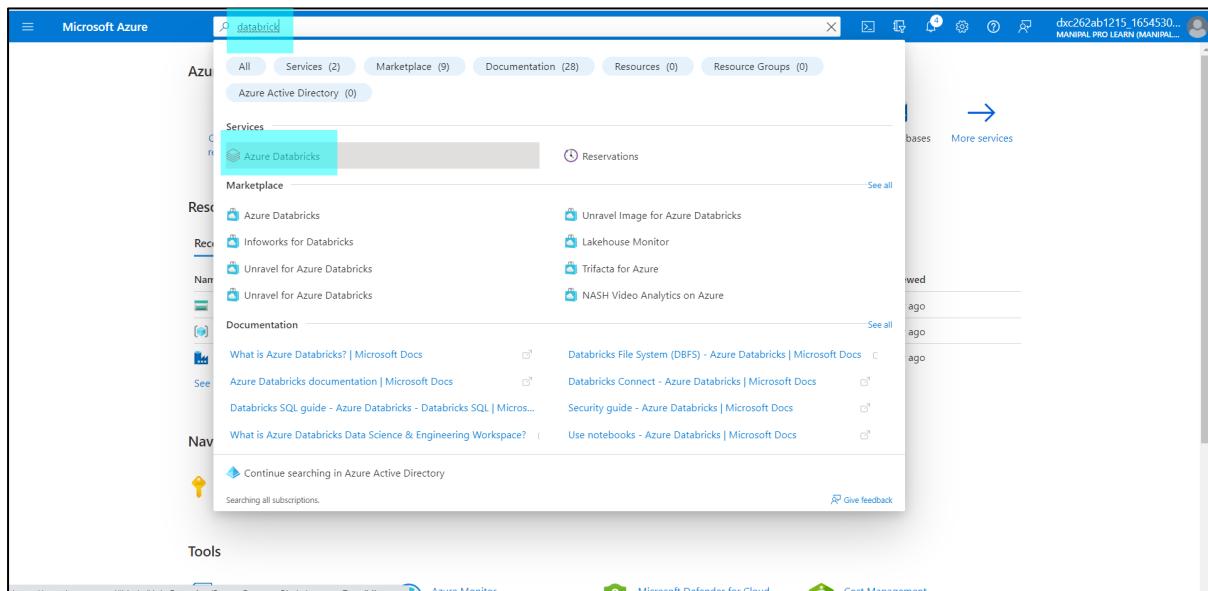
- Internal Application sends CSV file in Azure data lake storage.
- Validation needed to apply on this follows:
 - Check for duplicate rows. If it contains duplicate rows, file need to be rejected.
 - Need to validate the date format for all the date fields. Date column names and desired date format 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

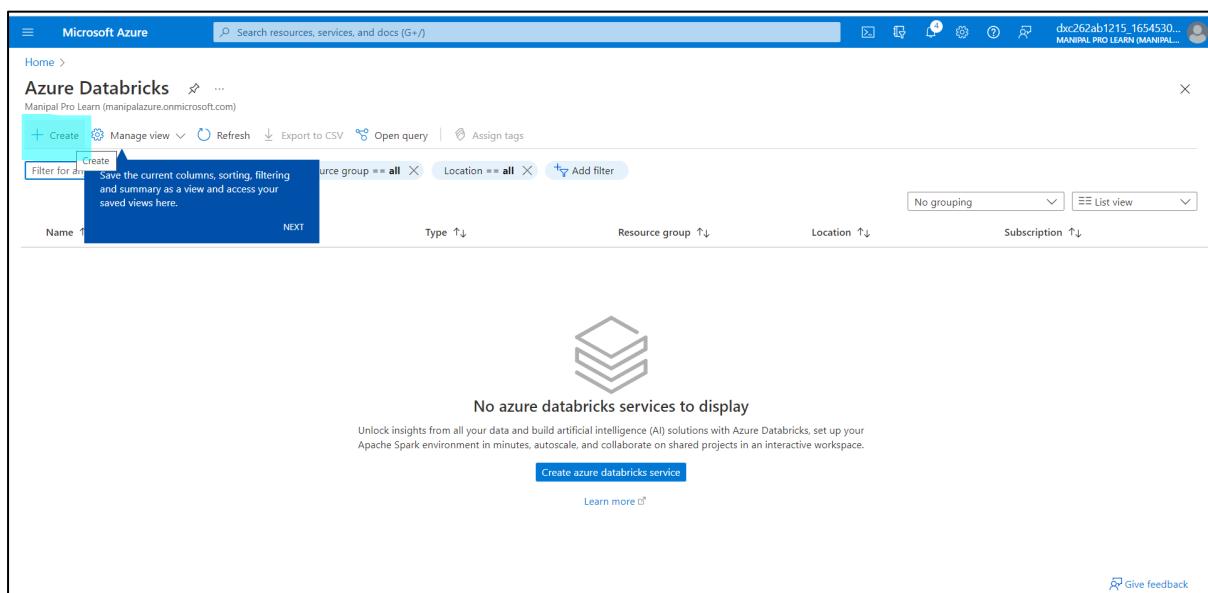


➤ Practical Lab: Create a Databricks

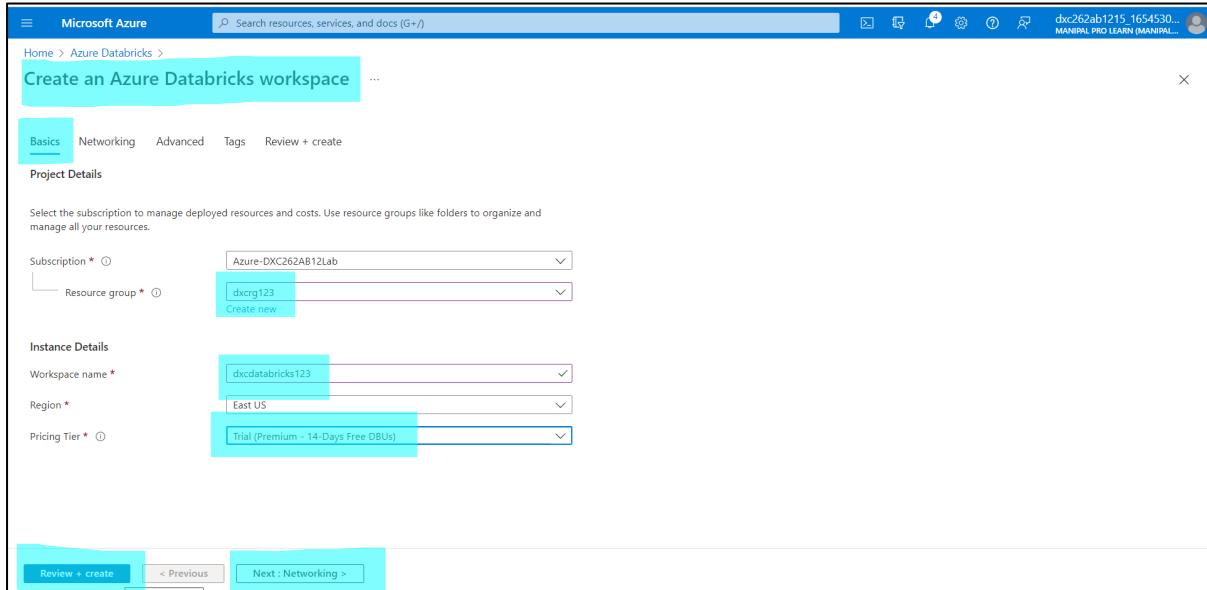
STEP 1: Search for Azure databricks and select it.



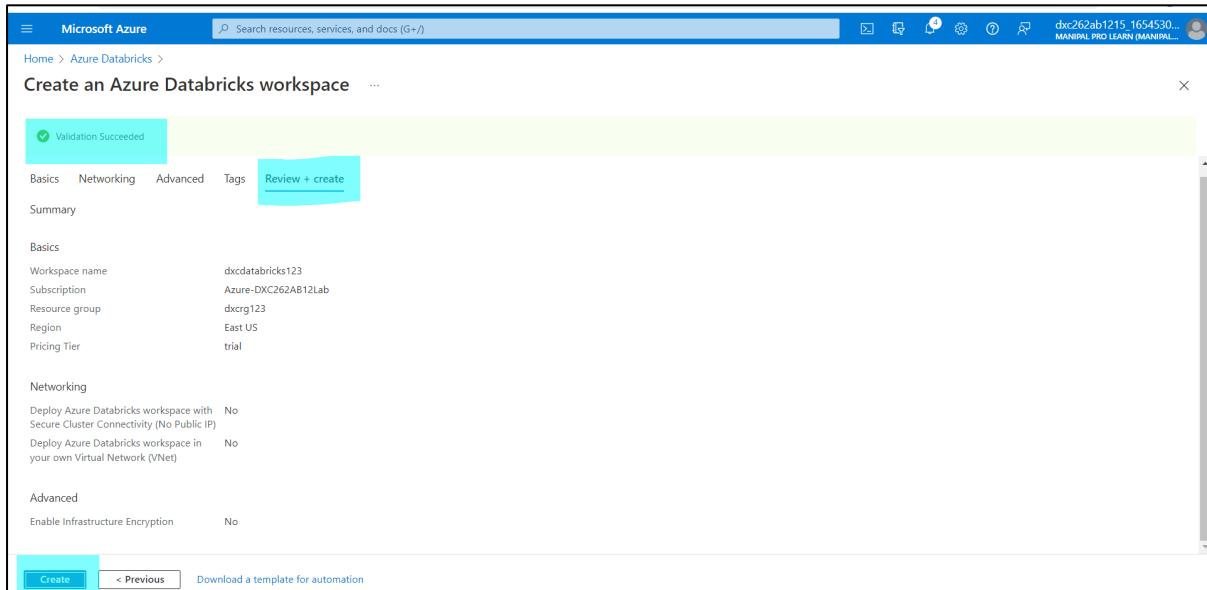
STEP 2: After selecting click on “+ create”.



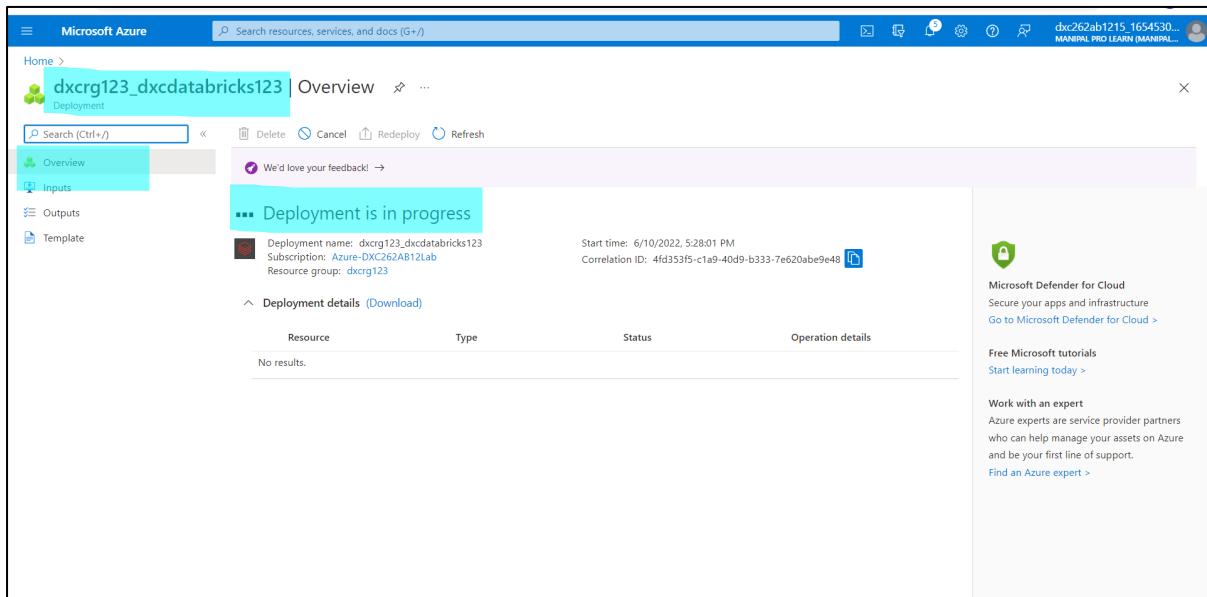
STEP 3: For creating enter the details of resource group, workspace name, pricing tier. And click on review + create.



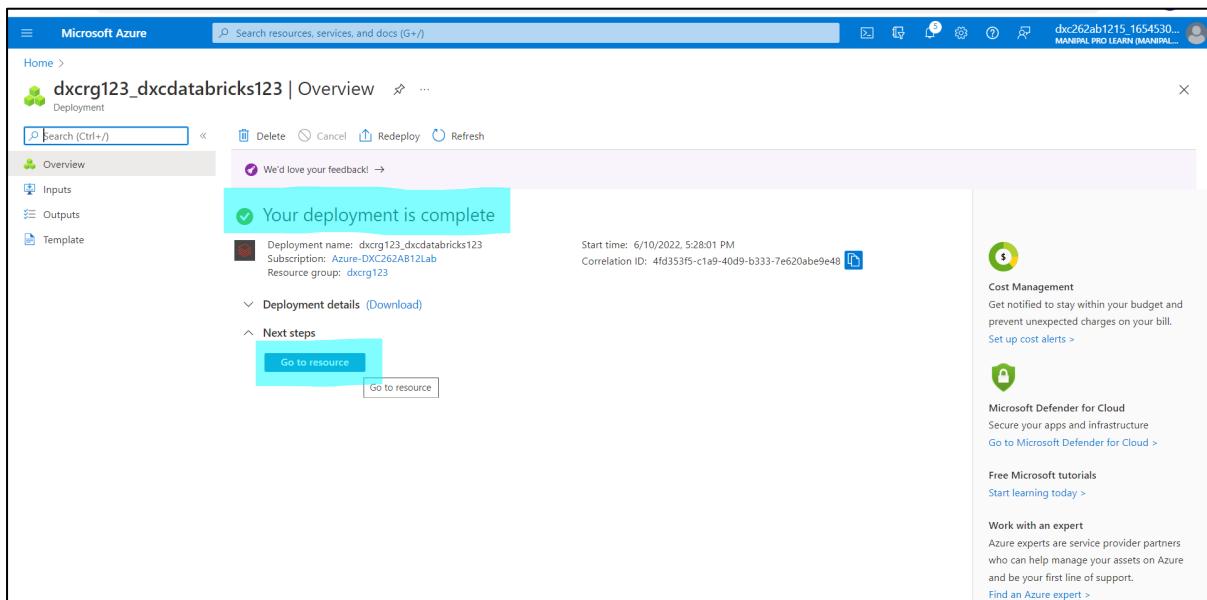
STEP 4: In review + create check the details, validation is succeeded. And click on create.



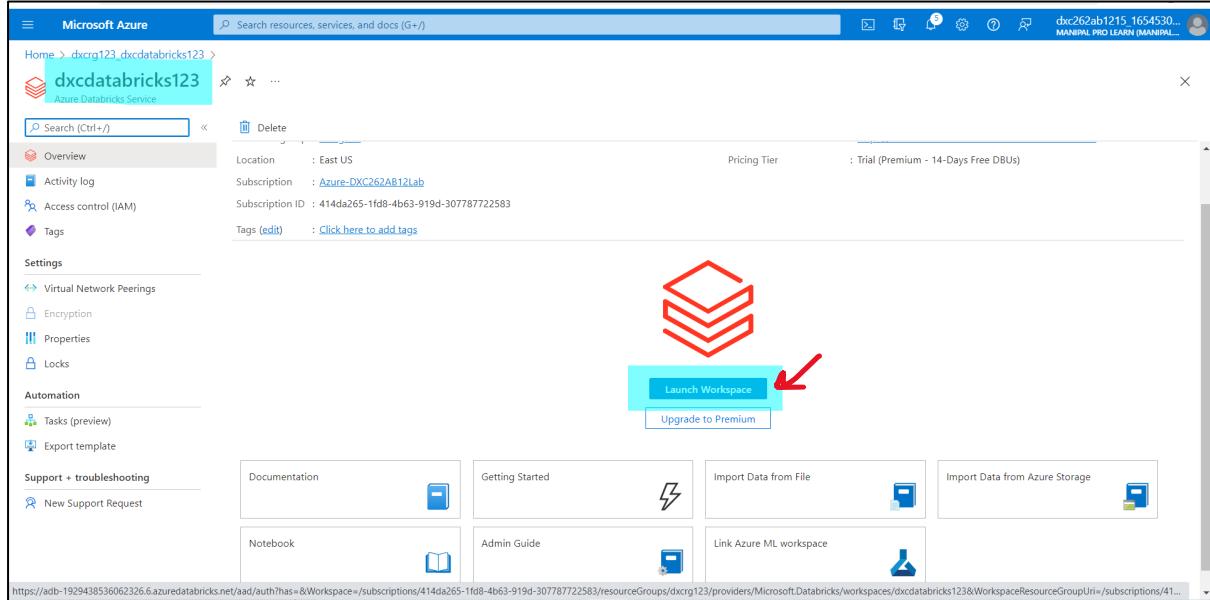
STEP 5: After selecting create, deployment will be in progress.



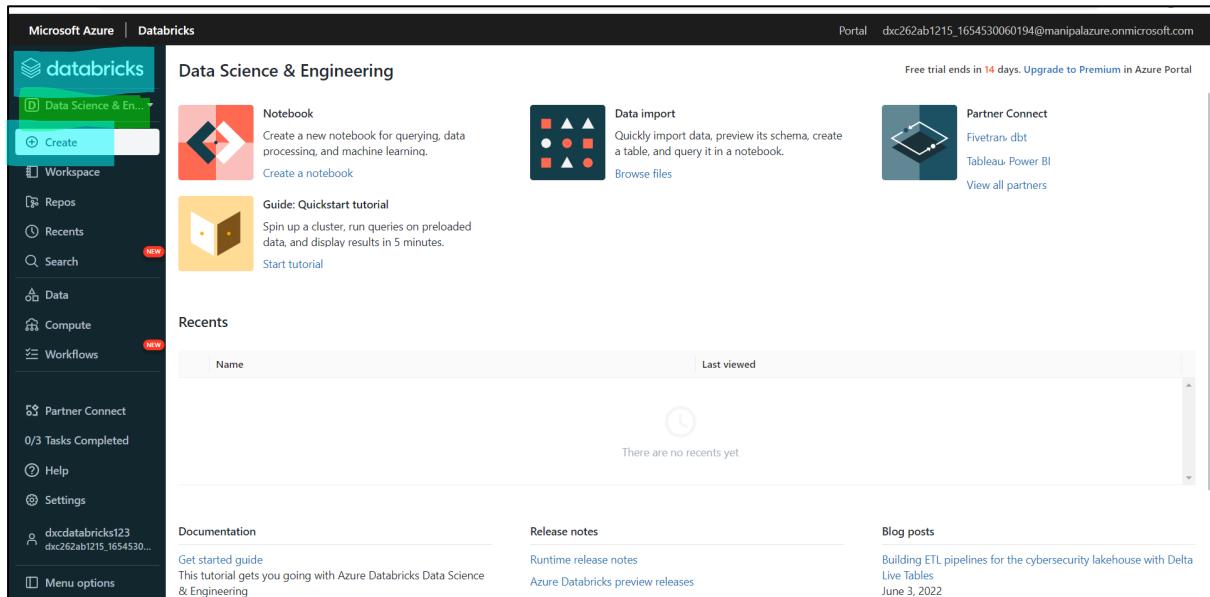
STEP 6: After few minutes the deployment is completed and then go to resource.



STEP 7: In resource we can observe the overview of databricks and we need to launch the workspace.

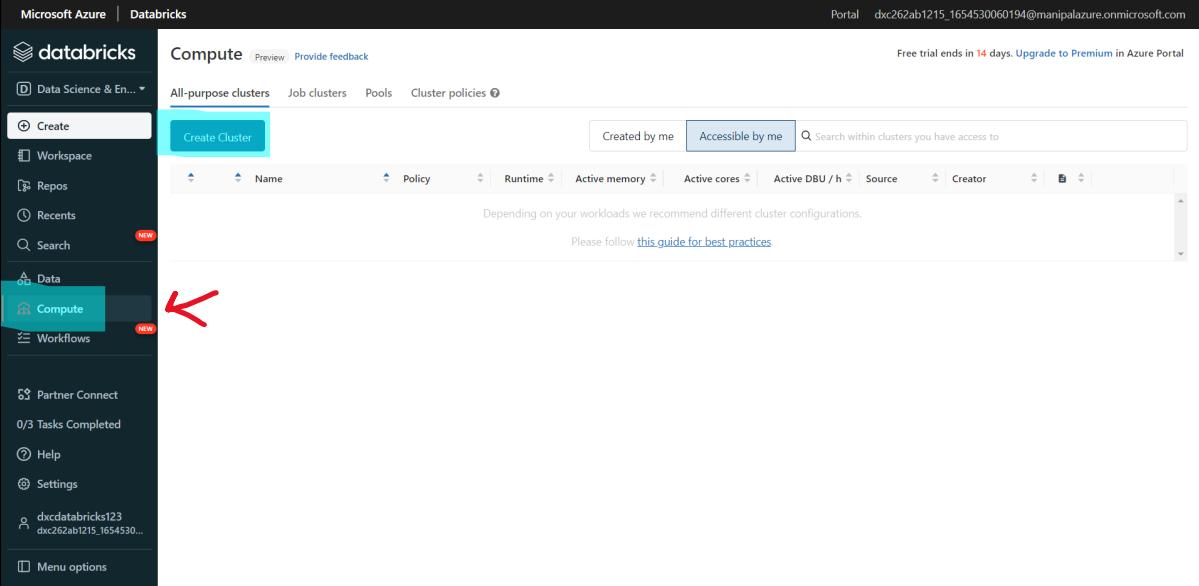


STEP 8: After launching the workspace we can see the page as shown below.



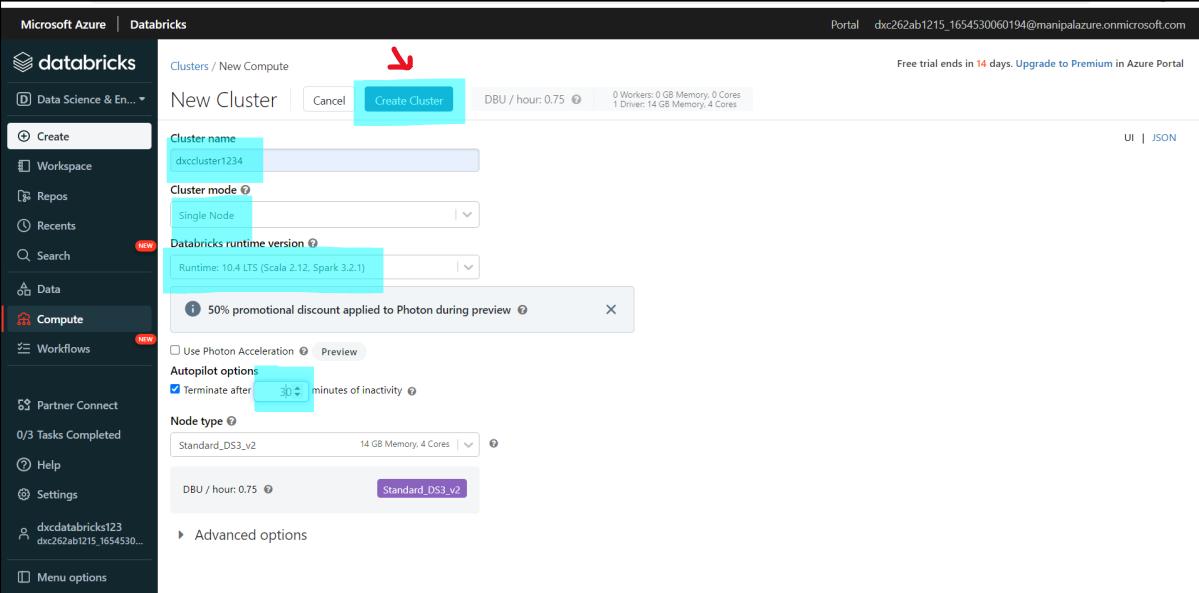
➤ Practical Lab: Create Cluster in Azure Databricks

STEP 1: For creating cluster, select compute on left side of tab shown and click on create cluster.



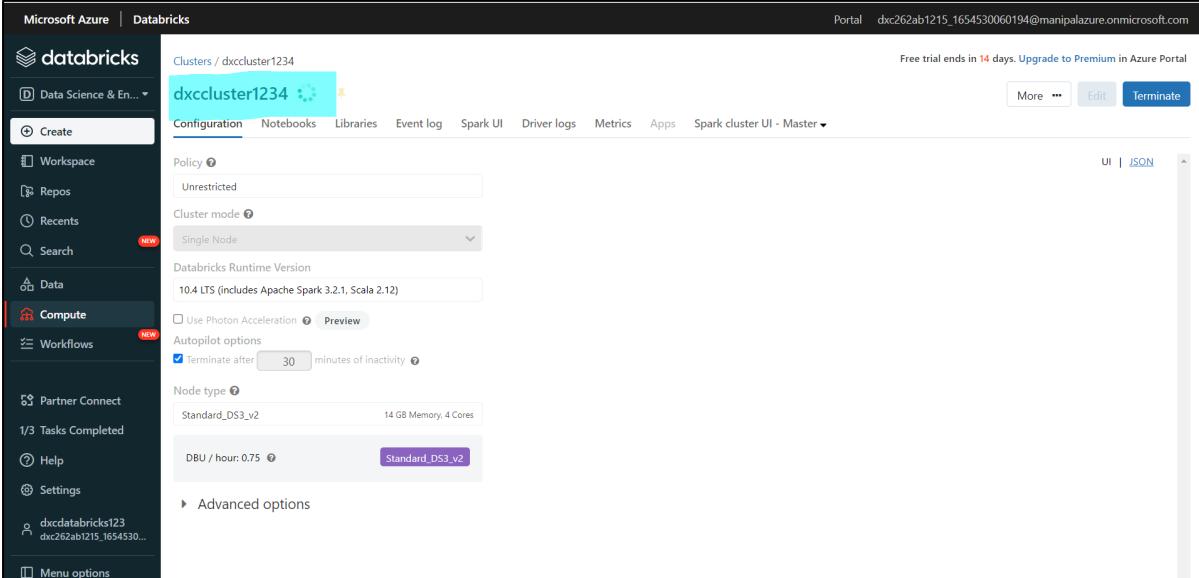
The screenshot shows the Microsoft Azure Databricks Compute interface. On the left sidebar, under the 'Compute' section, there is a 'Create' button. A red arrow points from the text in Step 1 to this 'Create' button. The main area is titled 'Compute' and shows tabs for 'All-purpose clusters', 'Job clusters', 'Pools', and 'Cluster policies'. Below these tabs is a large 'Create Cluster' button. To the right of the button is a search bar and a message about recommended cluster configurations. The top right corner of the page displays a free trial notice and upgrade options.

STEP 2: We need to enter the details of cluster name, cluster mode, runtime version, terminate time for new cluster . And click on create.



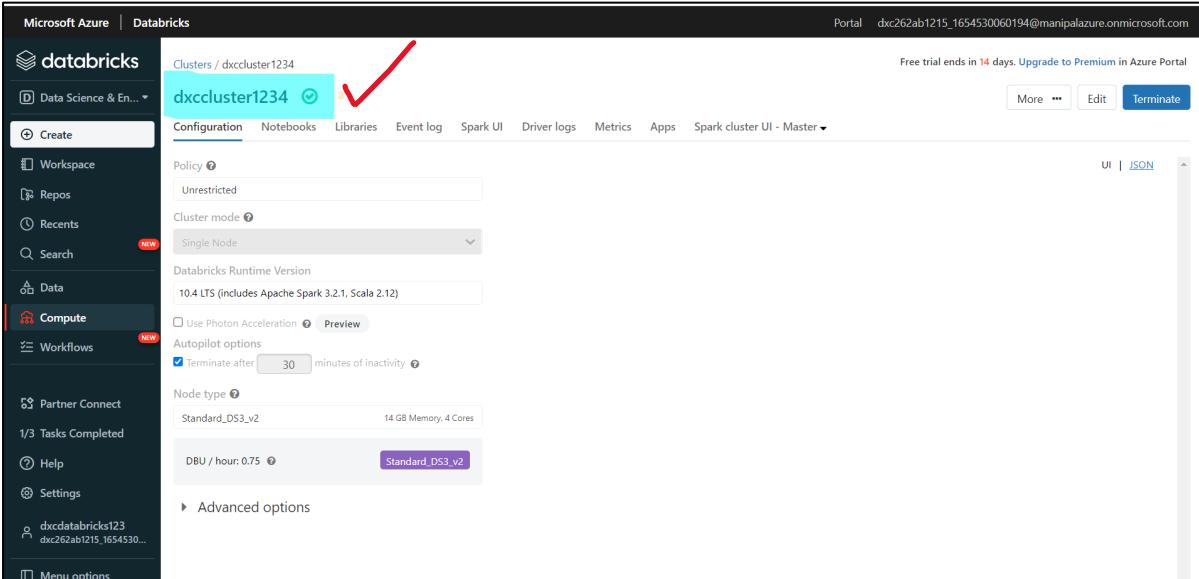
The screenshot shows the 'New Compute' dialog box for creating a cluster. The 'Cluster name' field is filled with 'dxccluster1234'. The 'Cluster mode' dropdown is set to 'Single Node'. The 'Databricks runtime version' dropdown is set to 'Runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1)'. There is a note indicating a 50% promotional discount applied to Photon during preview. Under 'Autopilot options', the 'Terminate after' field is set to '30 minutes of inactivity'. The 'Node type' dropdown is set to 'Standard_DS3_v2'. The 'DBU / hour' value is listed as '0.75'. At the bottom of the dialog, there is a 'Create Cluster' button, which is highlighted with a red arrow. The top right corner of the dialog box shows a UI/JSON link and a note about a free trial.

STEP 3: The cluster is getting created..



The screenshot shows the Microsoft Azure Databricks interface. On the left, there's a sidebar with various options like Workspace, Repos, Recents, Search, Data, Compute (which is currently selected), Workflows, Partner Connect, Tasks Completed, Help, Settings, and Menu options. In the main area, it says 'Clusters / dxcluster1234'. Below that, there are tabs for Configuration, Notebooks, Libraries, Event log, Spark UI, Driver logs, Metrics, Apps, and Spark cluster UI - Master. The Configuration tab is active. It shows settings for Policy (Unrestricted), Cluster mode (Single Node), Databricks Runtime Version (10.4 LTS), and Autopilot options (Terminate after 30 minutes of inactivity checked). It also shows Node type (Standard_DS3_v2) with 14 GB Memory and 4 Cores, and DBU / hour: 0.75. A 'More' button, an 'Edit' button, and a 'Terminate' button are at the top right. A note at the top right says 'Free trial ends in 14 days. Upgrade to Premium in Azure Portal'. There's also a 'UI | JSON' link.

STEP 4: The cluster is created which is shown by tick mark after creating.



This screenshot is identical to the one above, showing the Microsoft Azure Databricks Cluster configuration page. The 'Create' button in the sidebar is highlighted with a red box, and a large red checkmark is placed over the cluster name 'dxcluster1234' in the main list area. The rest of the interface, including the tabs, settings, and status bar, remains the same.

STEP 5: The created cluster can be viewed as shown below.

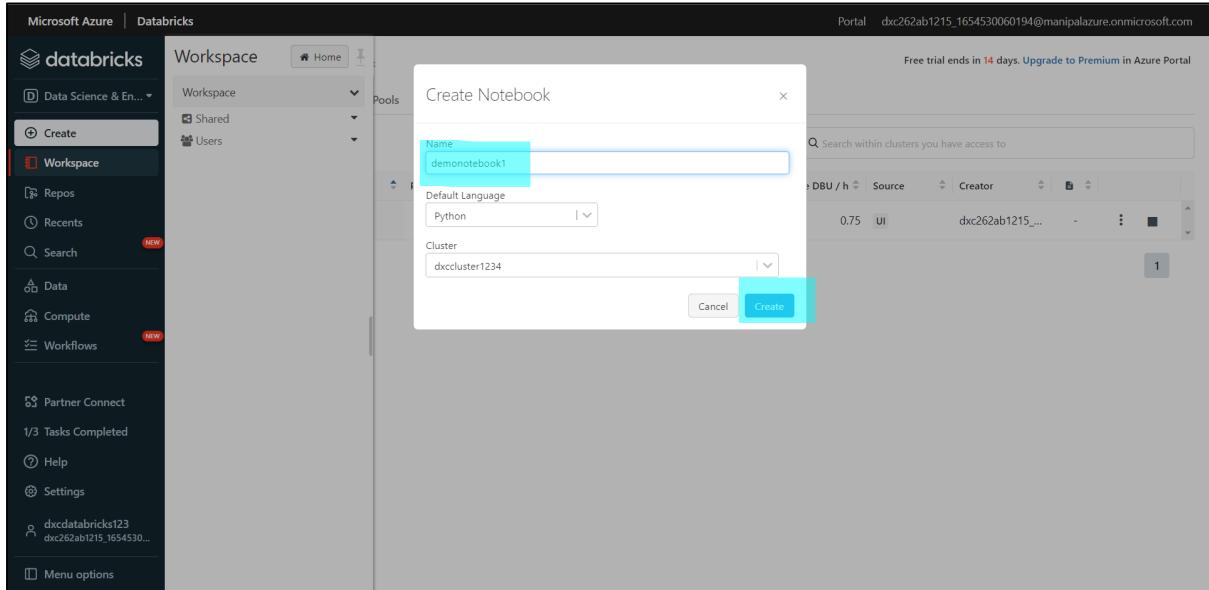
The screenshot shows the Microsoft Azure Databricks Compute blade. On the left, there's a sidebar with options like 'Data Science & ML', 'Create', 'Workspace', 'Repos', 'Recents', 'Search', 'Data', 'Compute', 'Workflows', 'Partner Connect', 'Tasks Completed', 'Help', 'Settings', and a 'Menu options' button. The main area is titled 'Compute' and shows the 'All-purpose clusters' tab. It lists a single cluster named 'decluster1234'. The cluster details are as follows: Name: decluster1234, Policy: -, Runtime: 10.4, Active memory: 14 GB, Active cores: 4 cores, Active DBU / h: 0.75, Source: UI, and Creator: dxc262ab1215_1654530060194@manipalazure.onmicrosoft.com. There are also buttons for 'Created by me', 'Accessible by me', and a search bar. The 'Creator' field is highlighted with a red box.

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

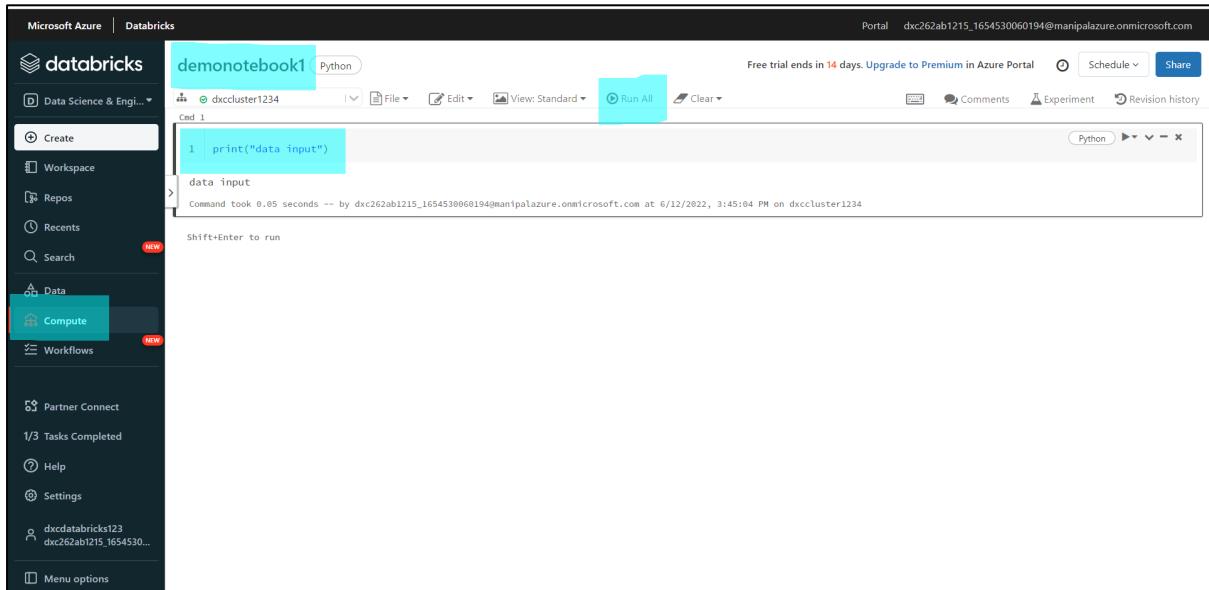
STEP 1: For adding notebook into the databricks , we need to select workspace > Create > Notebook.

The screenshot shows the Microsoft Azure Databricks workspace blade. On the left, there's a sidebar with options like 'Data Science & ML', 'Create', 'Workspace', 'Repos', 'Recents', 'Search', 'Data', 'Compute', 'Workflows', 'Partner Connect', 'Tasks Completed', 'Help', 'Settings', and a 'Menu options' button. The main area is titled 'Workspace' and shows a 'Create' button. A red arrow points from the 'Create' button in the sidebar to the 'Notebook' option in the 'Create' dropdown menu. The 'Create' dropdown menu is open, showing 'Notebook' as the selected item, along with other options like 'Import', 'Export', 'Permissions', and 'MLflow Experiment'. The rest of the page shows the workspace interface with tabs for 'Workspace', 'Shared', and 'Users'.

STEP 2: As, shown we can create a notebook by entering the name and click on create.

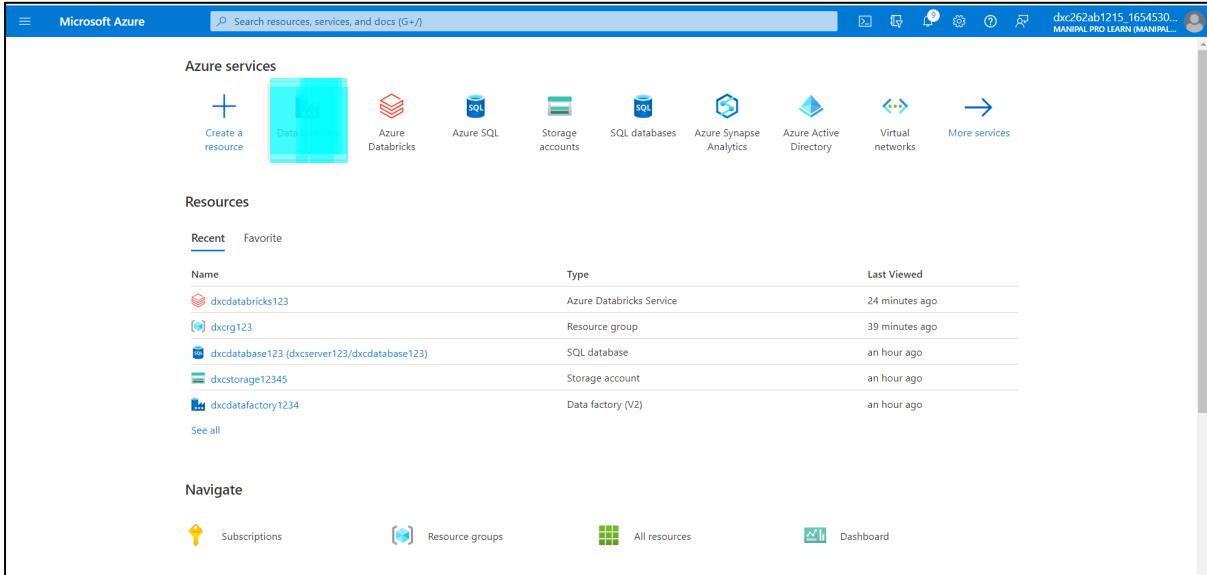


STEP 3: After notebook is created we can enter code and run it as shown below and view the result and run time.



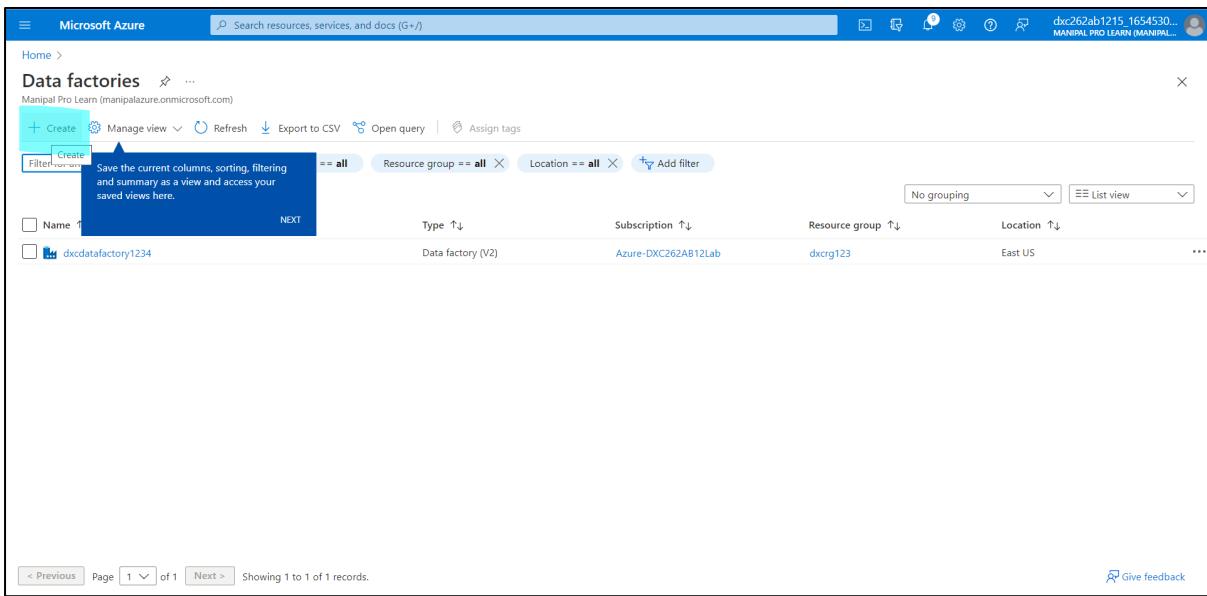
➤ Practical Lab: Azure Data Factory For AP Morgan

STEP 1: Search for datafactories and select it.



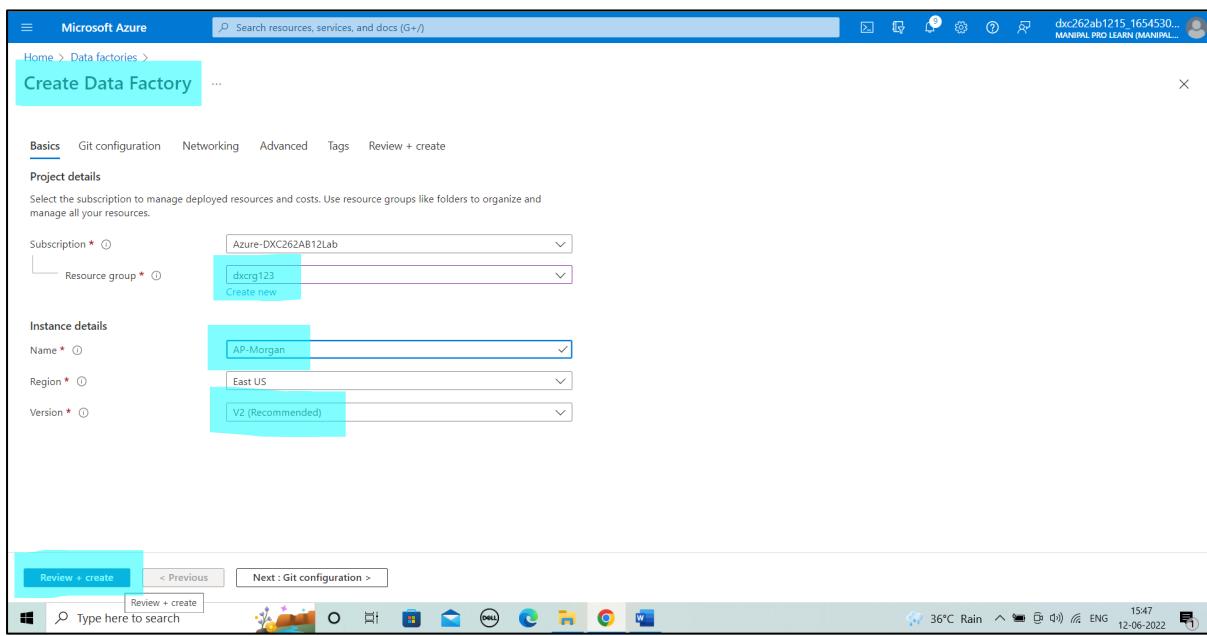
The screenshot shows the Microsoft Azure portal homepage. At the top, there's a search bar and a navigation bar with icons for account management and help. Below the search bar, there's a section titled "Azure services" with a "Create a resource" button and a "Data factory" icon highlighted in blue. Other service icons include Azure Databricks, Azure SQL, Storage accounts, SQL databases, Azure Synapse Analytics, Azure Active Directory, Virtual networks, and More services. The main content area is titled "Resources" and has tabs for "Recent" and "Favorite". It lists recent resources: "dxdatabricks123" (Azure Databricks Service), "dxcrg123" (Resource group), "dxdatabase123 (dxserver123/dxdatabase123)" (SQL database), "dxdstorage12345" (Storage account), and "dxdfactory1234" (Data factory (V2)). A "See all" link is also present. At the bottom, there's a "Navigate" section with links for Subscriptions, Resource groups, All resources, and Dashboard.

STEP 2: After selecting click on “+ create”.

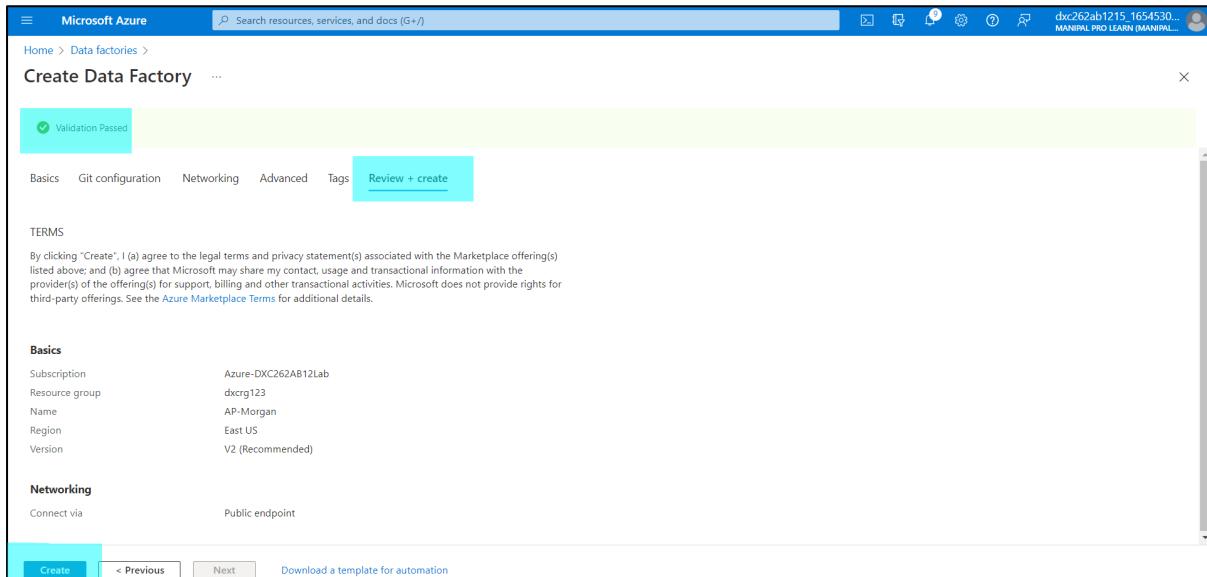


The screenshot shows the "Data factories" blade in the Microsoft Azure portal. At the top, there's a header with "Data factories" and a "Create" button, which is highlighted with a blue box and an arrow pointing to it. Below the header, there are filter options: "Filter" (dropdown), "Save the current columns, sorting, filtering and summary as a view and access your saved views here.", "Manage view", "Refresh", "Export to CSV", "Open query", and "Assign tags". There are also dropdowns for "No grouping" and "List view". The main table lists one data factory: "dxdfactory1234" (Name), "Data factory (V2)" (Type), "Azure-DXC262AB12Lab" (Subscription), "dxcrg123" (Resource group), and "East US" (Location). At the bottom, there are navigation links for "Previous", "Page 1 of 1", "Next >", and "Showing 1 to 1 of 1 records". A "Give feedback" link is also present in the bottom right corner.

STEP 3: Now for creating Ap morgan enter name and resource group.



STEP 4: Then, click on review + create and we can view the details. Validation is passed and then create.



STEP 5: The deployment is completed and go to resource.

The screenshot shows the Microsoft Azure Data Factory Overview page for a deployment named 'Microsoft.DataFactory-20220612154648'. A prominent message says 'Your deployment is complete' with a green checkmark icon. Deployment details include the name, subscription ('Azure-DXC262AB12Lab'), and resource group ('dxcrg123'). The start time was 6/12/2022, 3:48:16 PM, and the correlation ID is 1f744ac8-9748-48db-afda-7bfe4d2aab2. On the right side, there are promotional cards for Cost Management, Microsoft Defender for Cloud, and Free Microsoft tutorials.

STEP 6: In resource we can view overview and then open azure data factory studio.

The screenshot shows the Microsoft Azure Data Factory resource overview for a resource named 'AP-Morgan'. The 'Overview' tab is selected. Key details shown include the resource group ('dxcrg123'), status ('Succeeded'), location ('East US'), subscription ('Azure-DXC262AB12Lab'), and subscription ID ('414da265-1fd8-4b63-919d-307787722583'). The 'Getting started' section features links to 'Open Azure Data Factory Studio' and 'Read documentation'. The 'Monitoring' section displays two charts: 'PipelineRuns' and 'ActivityRuns', both showing values around 100. The bottom of the screen shows the Windows taskbar with various pinned icons.

STEP 7: The data factory page of AP-Morgan is shown in new page.

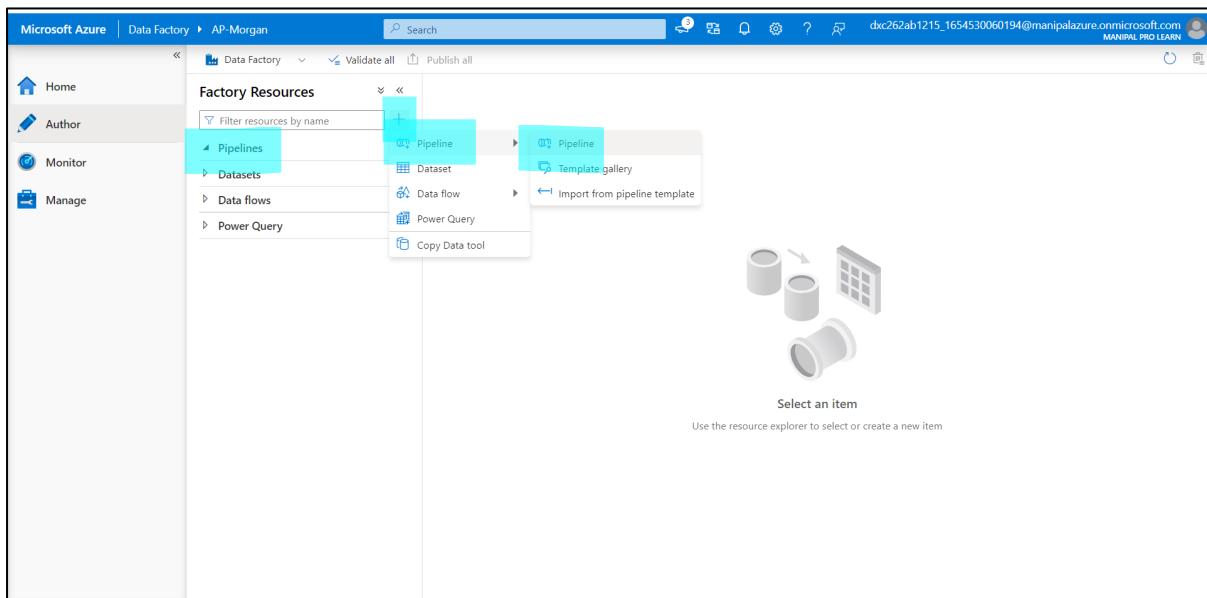
The screenshot shows the Microsoft Azure Data Factory interface for the 'AP-Morgan' data factory. The left sidebar has 'Home' selected. The main area displays the data factory name 'AP-Morgan' and a large 3D icon representing data flow. Below the icon are four cards: 'Ingest' (Copy data at scale once or on a schedule), 'Orchestrate' (Code-free data pipelines), 'Transform data' (Transform your data using data flows), and 'Configure SSIS' (Manage & run your SSIS packages in the cloud). At the bottom, there are links for 'Discover more' (Browse partners, Pipeline templates, SAP pipeline templates) and 'Recent resources'.

➤ Practical Lab: Create Azure Databricks Linked Service in ADF

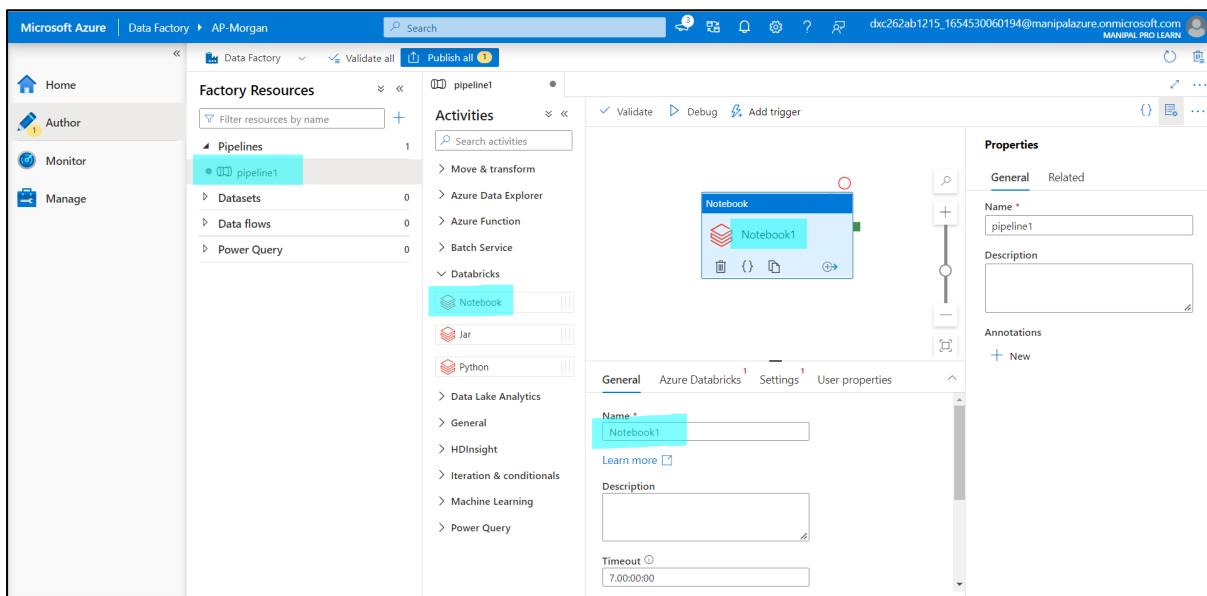
STEP 1: Click on author .

This screenshot is identical to the one above, but the 'Author' option is highlighted in the left sidebar, indicating it is the active section. The rest of the interface, including the data factory name, 3D icon, and functional cards, remains the same.

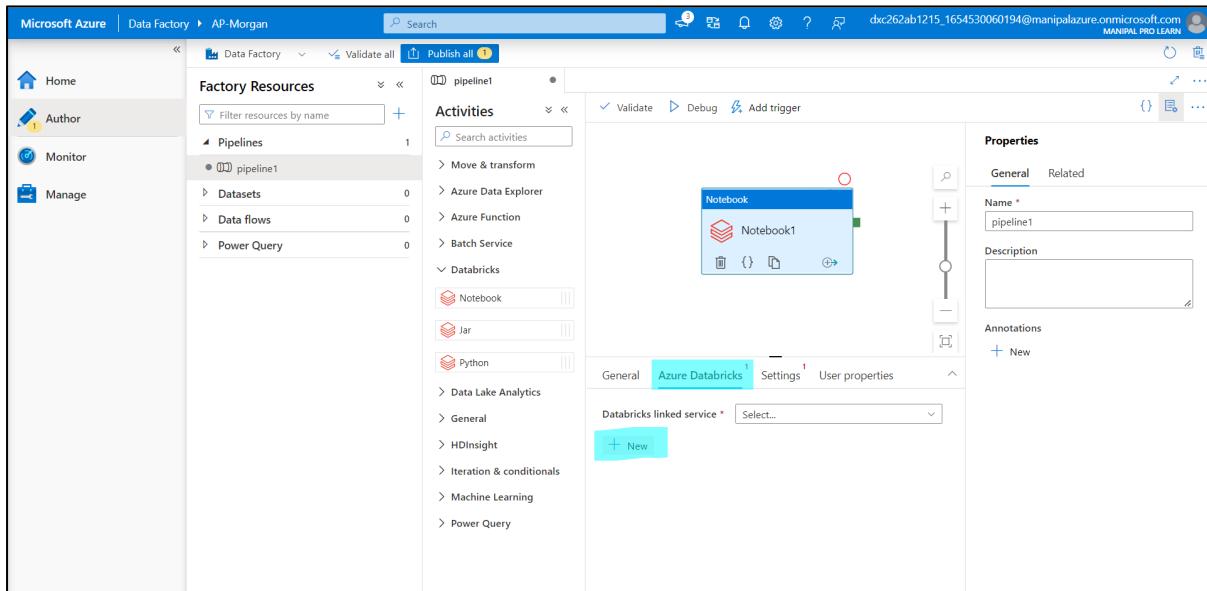
STEP 2: In author, click on “+” > pipeline to create a pipeline that will connect Data Bricks notebook with Data Factory.



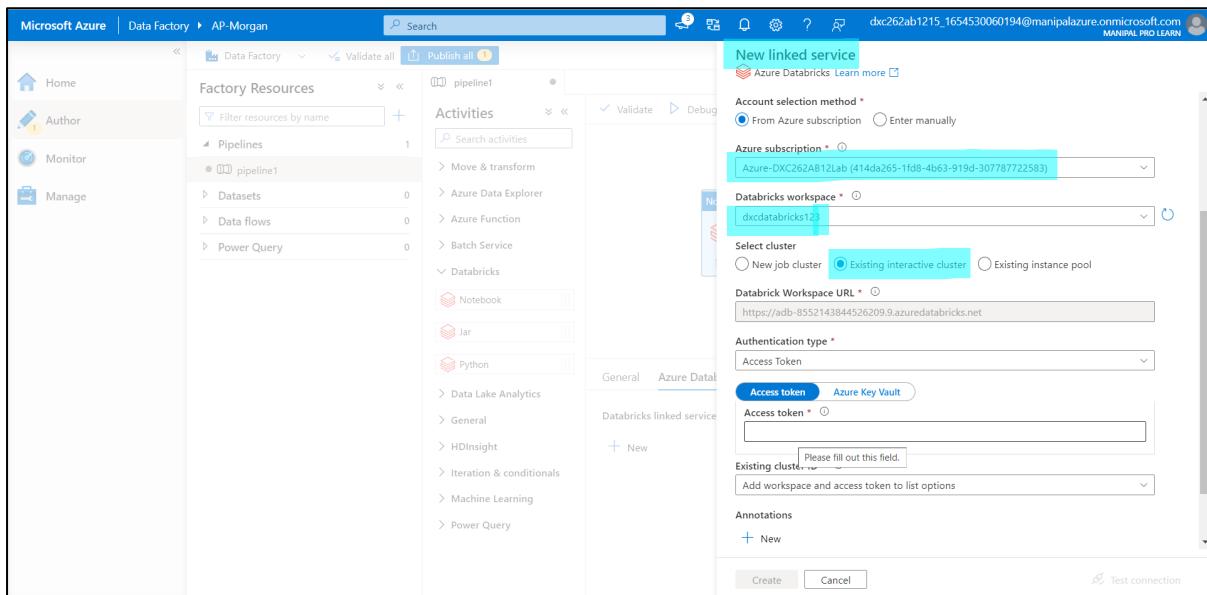
STEP 3: Create a pipeline by dragging the notebook from databricks.



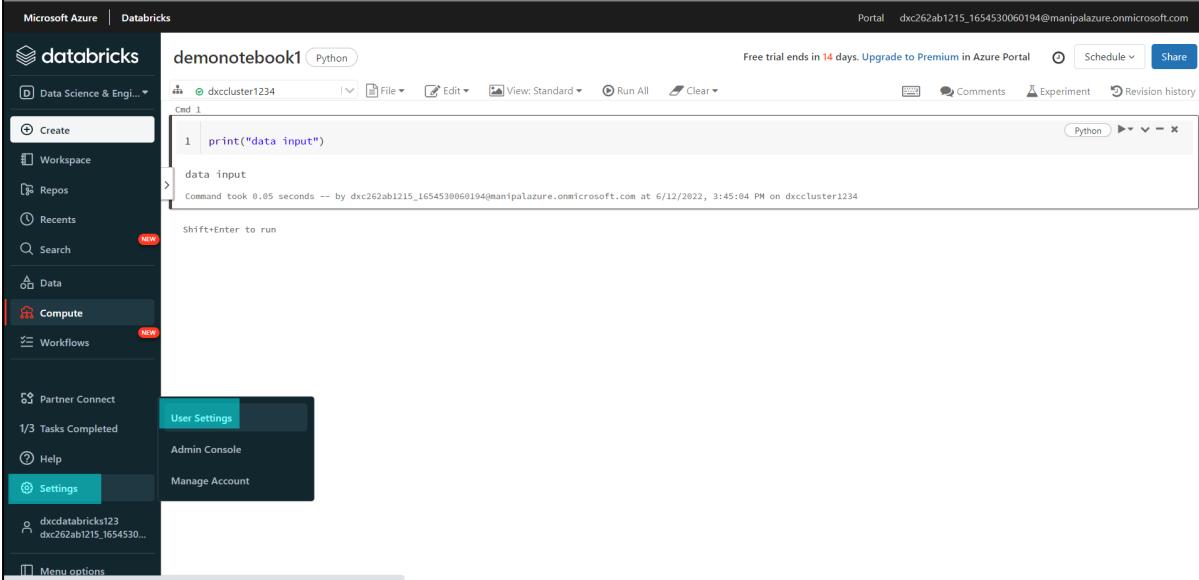
STEP 4: In azure databricks we can add new databricks linked service .



STEP 5: In databricks linked service we can enter the details shown and we need the access token.



STEP 6: To get the token details comeback to databricks and click on settings > Usersettings > generate new token.

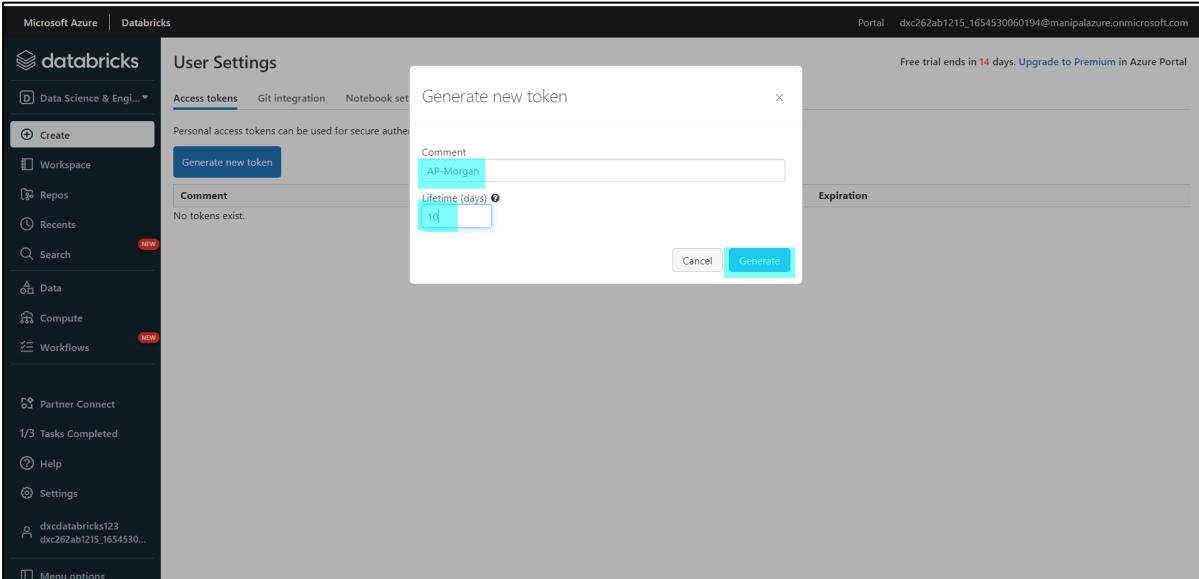


The screenshot shows a Microsoft Azure Databricks interface. On the left, the navigation sidebar includes options like Data Science & Engineering, Create, Workspace, Repos, Recents, Search, Data, Compute (which is selected), Workflows, Partner Connect, Tasks Completed, Help, Settings (which is selected), and Menu options. In the center, a notebook titled "demonotebook1" in Python is open. The code cell contains the following Python code:

```
1 print("data input")
data input
```

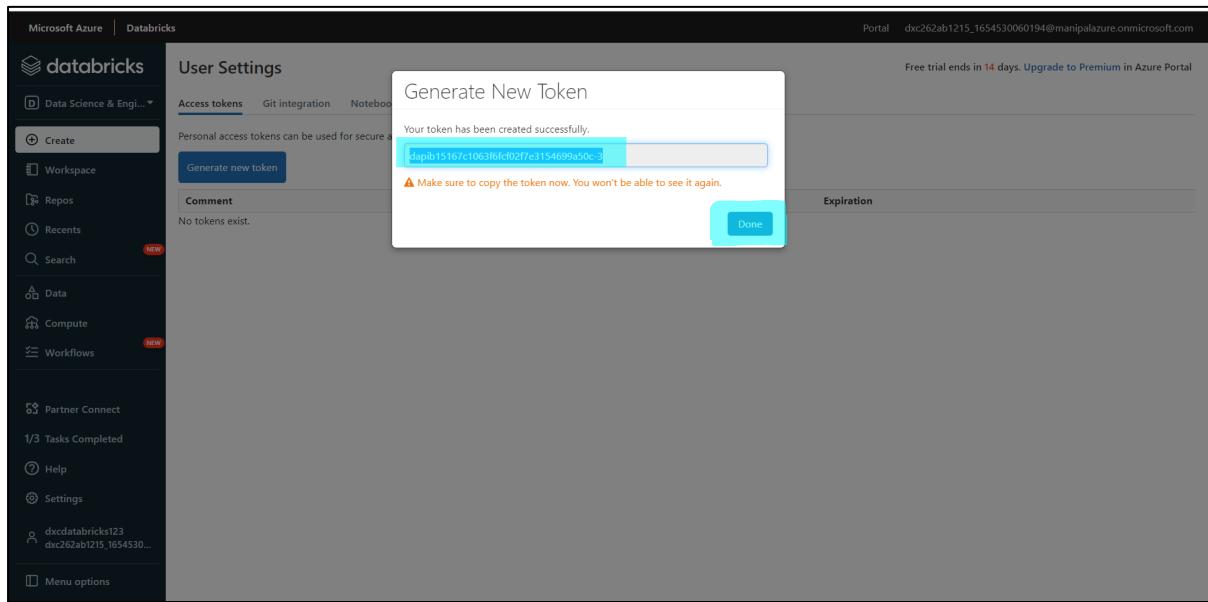
Below the code cell, it says "Command took 0.85 seconds -- by dxc26ab1215_1654530060194@manipalazure.onmicrosoft.com at 6/12/2022, 3:45:04 PM on dxccluster1234". A tooltip "Shift+Enter to run" is shown above the cell. At the top right, there are buttons for Schedule and Share. The URL in the browser bar is <https://anu-25521130445062098.dustedbrick.net/Dx-05521140445262098#setting/account>.

STEP 7: Comment the Ap-morgan and lifetime the generate.

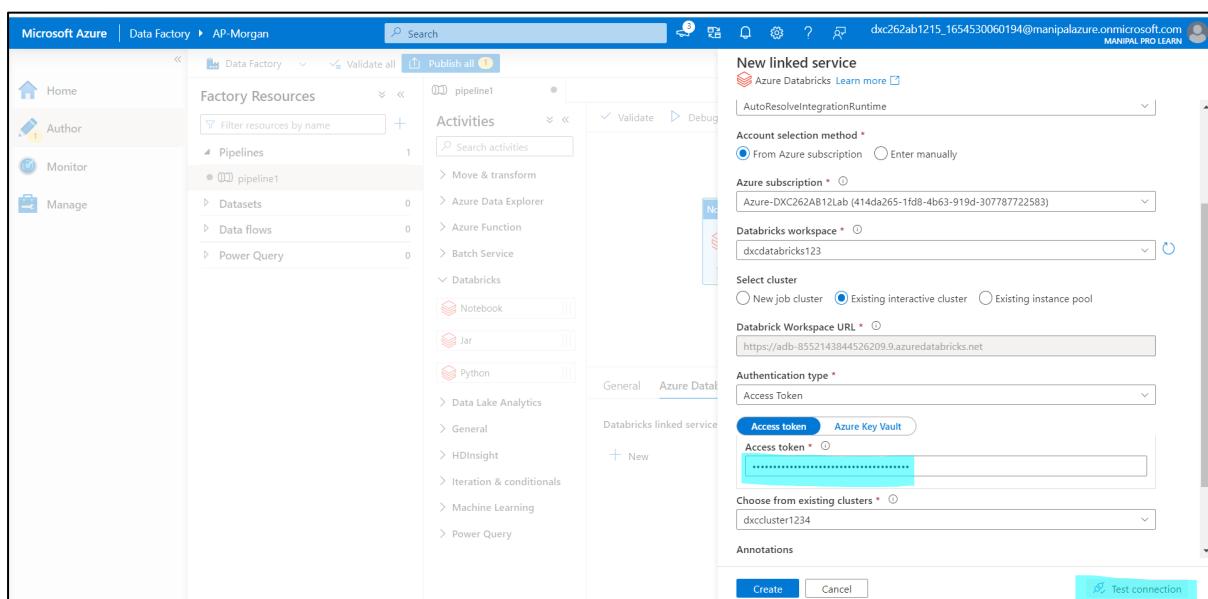


The screenshot shows the "User Settings" page in the Databricks interface. The left sidebar is identical to the previous screenshot. The main area displays the "Access tokens" section with a sub-section for "Generate new token". A modal window titled "Generate new token" is open. It contains fields for "Comment" (with "AP-Morgan" typed in) and "Lifetime (days)" (with "10" typed in). Below these fields are "Cancel" and "Generate" buttons. The URL in the browser bar is <https://anu-25521130445062098.dustedbrick.net/Dx-05521140445262098#setting/account>.

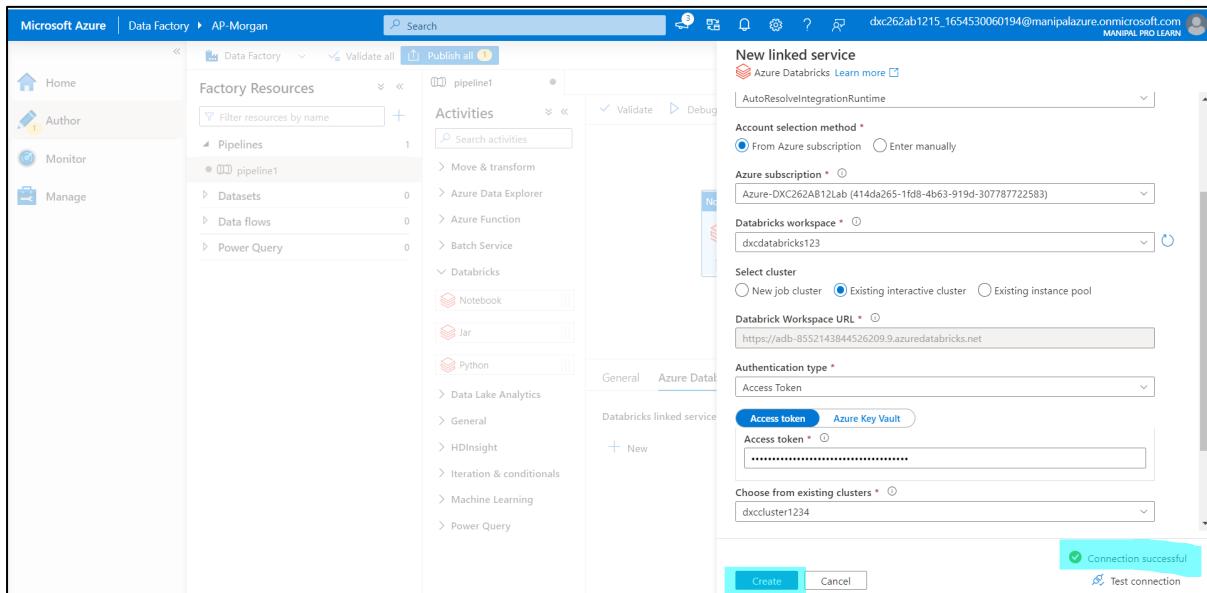
STEP 8: We get the token.



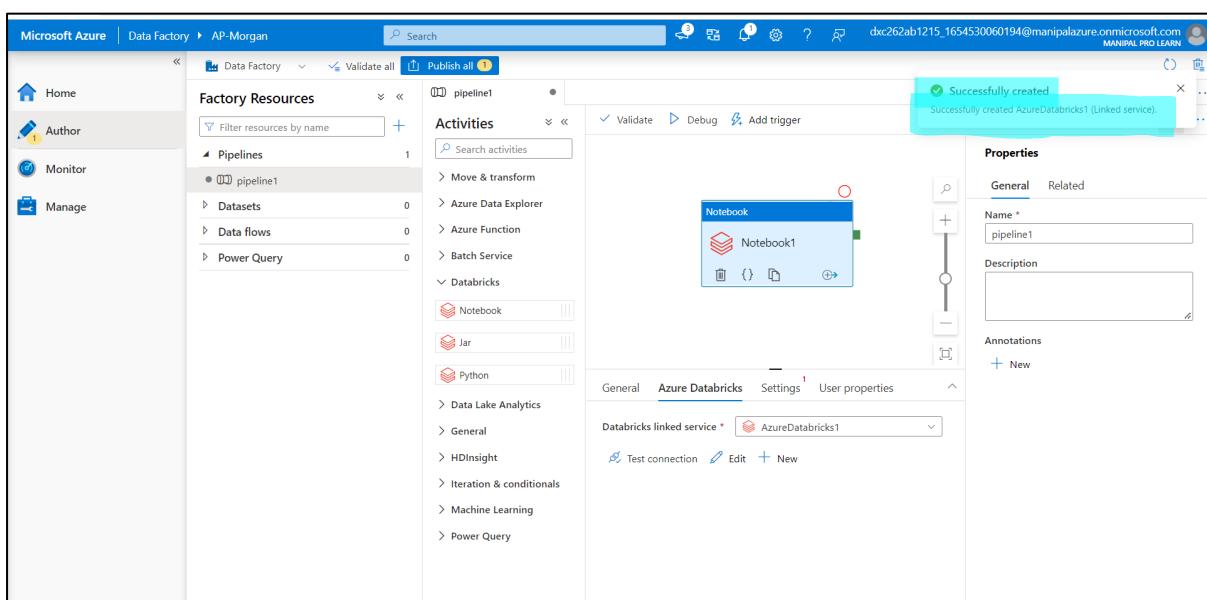
STEP 9: Enter the token details in previous new linked service page then test connection.



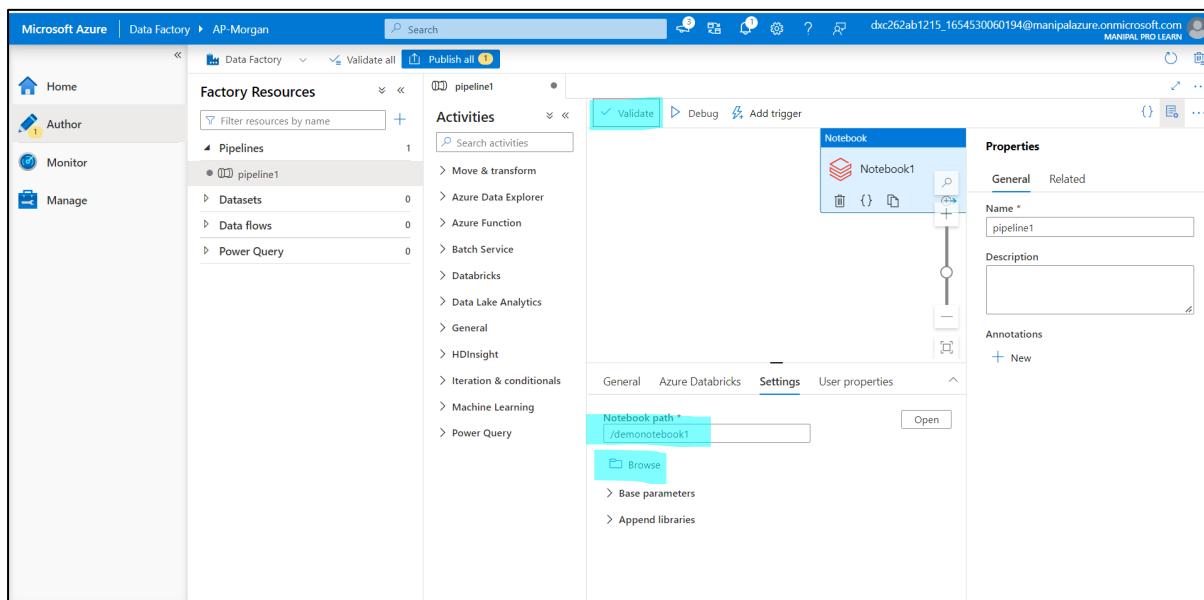
STEP 10: After the connection is successful , click on create.



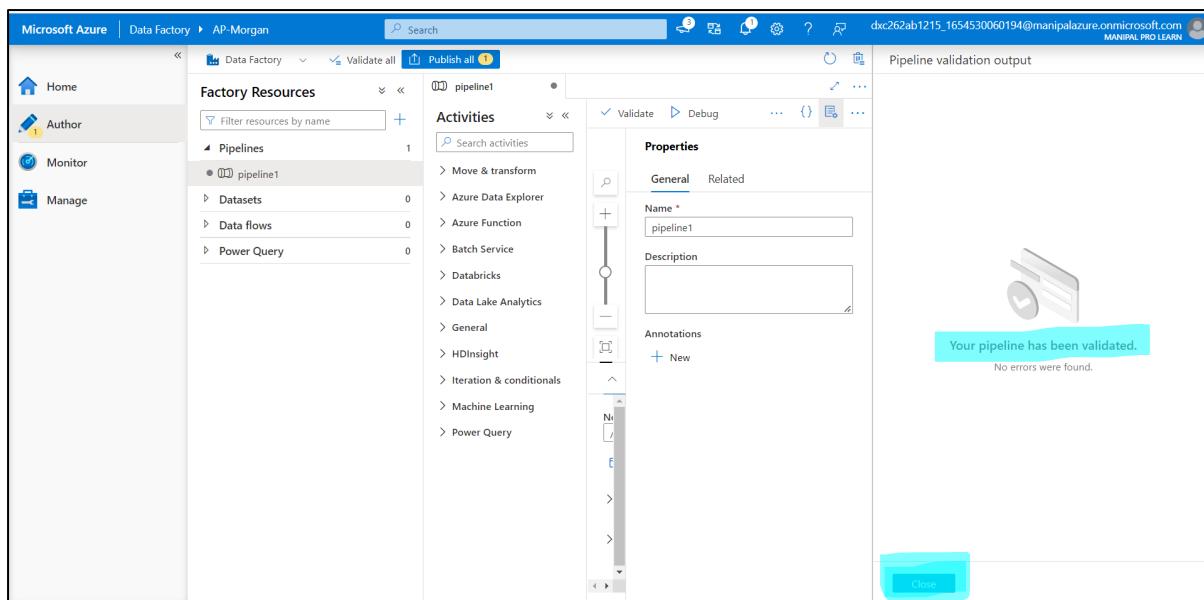
STEP 11: It is successfully created then enter the notebook path by browsing.



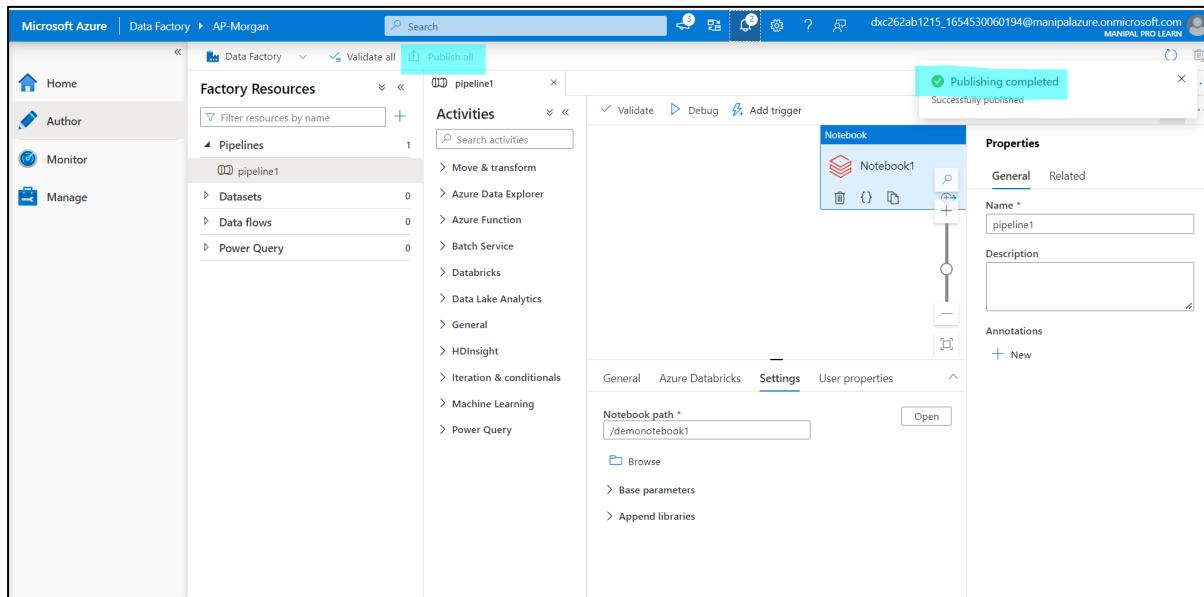
STEP 12: Validate the notebook.



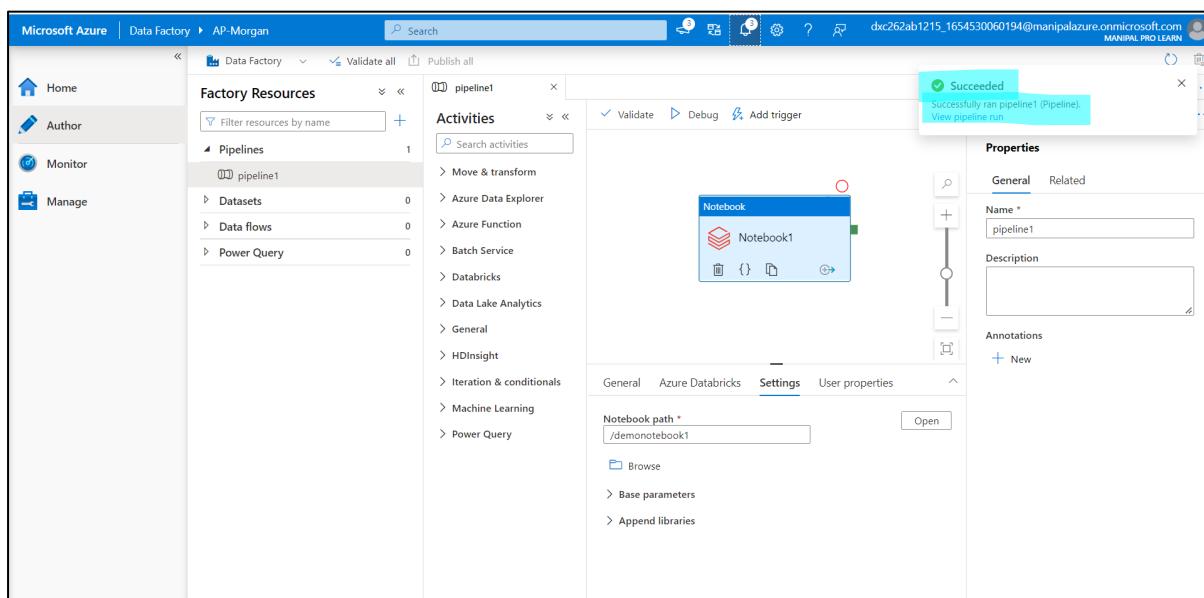
STEP 13: After the pipeline is validated click on publish.



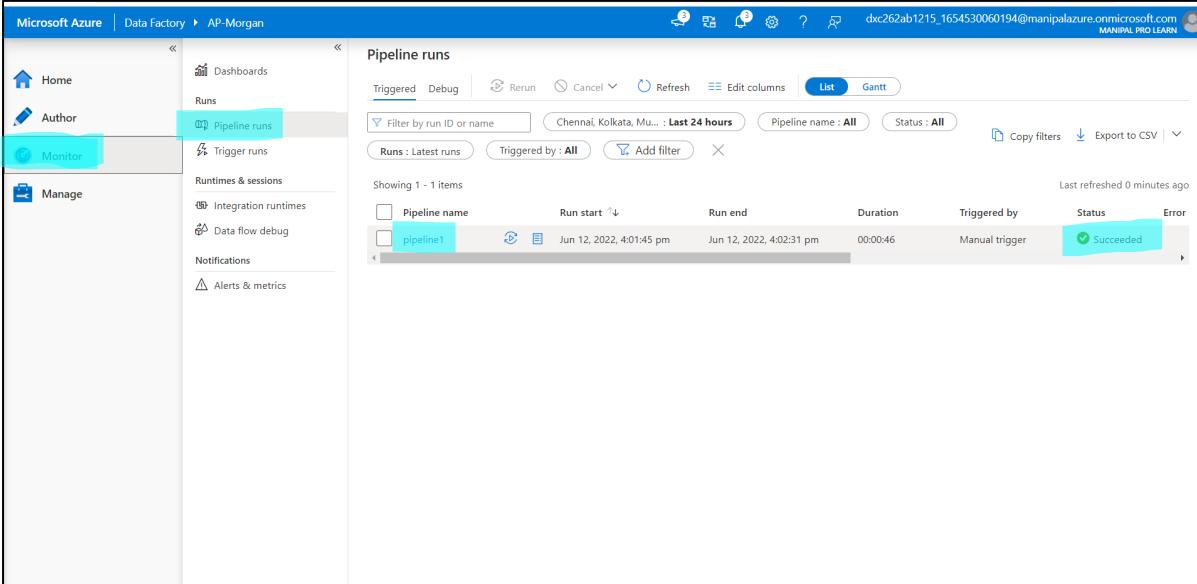
STEP 14: The publishing is done .



STEP 15: Click on trigger > trigger now. The pipeline is successfully ran.



STEP 16: We can view the pipeline runs in monitor .



The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has tabs for Home, Author, Monitor (which is selected and highlighted in blue), and Manage. Under the Monitor tab, there are sub-options: Dashboards, Runs (selected and highlighted in blue), Trigger runs, Runtimes & sessions, Integration runtimes, Data flow debug, Notifications, and Alerts & metrics. The main area is titled "Pipeline runs" and shows a table of runs. The table has columns: Pipeline name, Run start, Run end, Duration, Triggered by, Status, and Error. One run is listed: pipeline1, triggered by Manual trigger, started at Jun 12, 2022, 4:01:45 pm, ended at Jun 12, 2022, 4:02:31 pm, duration 00:00:46, status Succeeded, and no error. There are buttons for Triggered, Debug, Rerun, Cancel, Refresh, Edit columns, List (selected), Gantt, Filter by run ID or name (Chennai, Kolkata, Mu... : Last 24 hours), Pipeline name: All, Status: All, Copy filters, and Export to CSV.

Result: In this project 2, We created the azure databricks notebook and linked it to azure datafactory which is named as AP-Morgan.

Conclusion: Azure datafactory is linked with the databricks.