



DXC REALTIME PROJECTS

AZ-900, DP - 203



JUNE 10, 2022
DXC TECHNOLOGY

Name: Shan

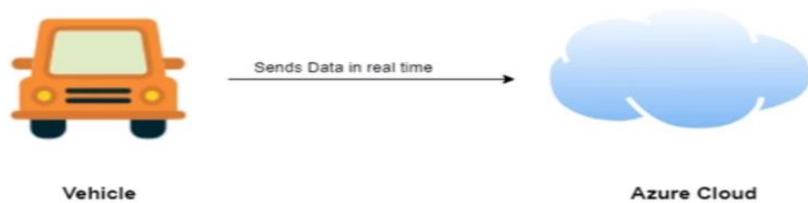
Reg No: DXC262AB1228

Project1 Name: Smart Vehicles

Date: 10TH June 2022

Project 1 : Connected Vehicles

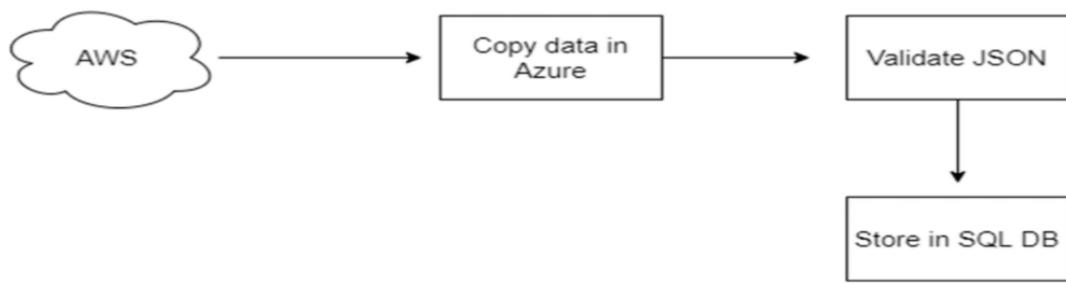
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



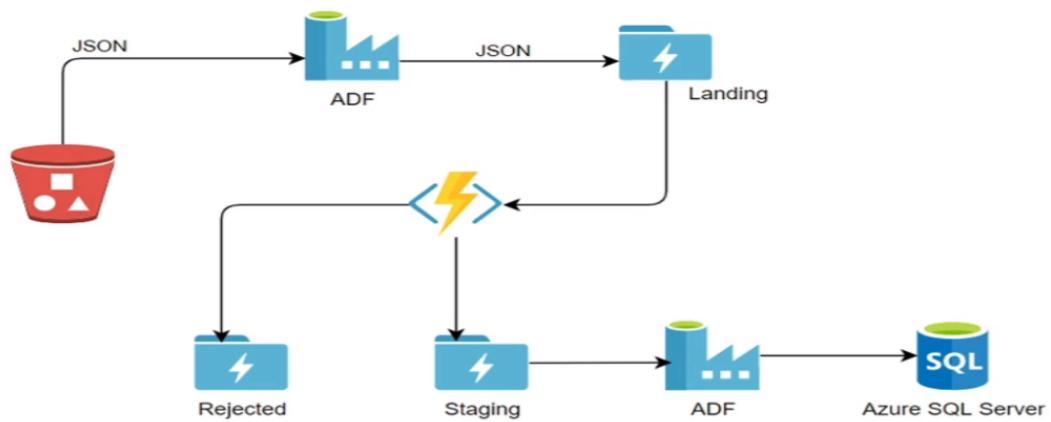
Project 1 : Connected Vehicles

- 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

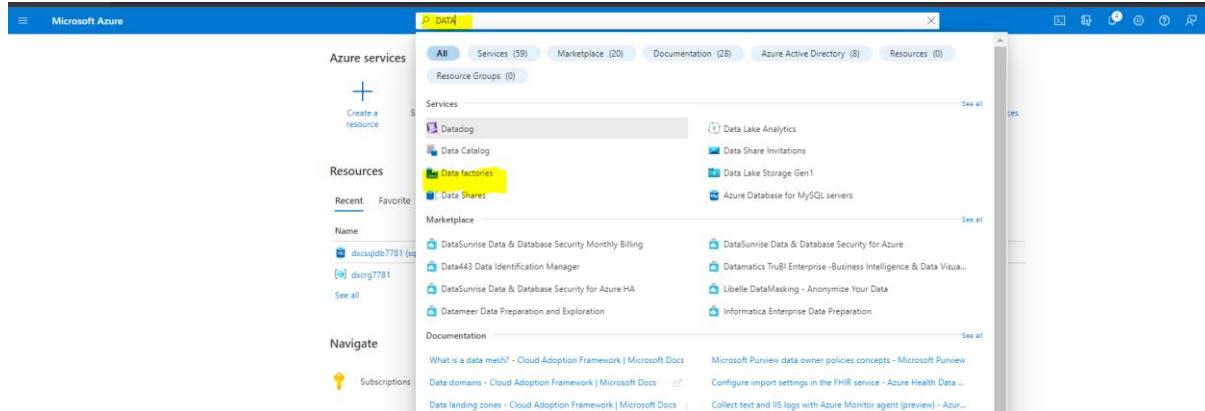


Architecture Diagram for Connected Vehicle Project

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

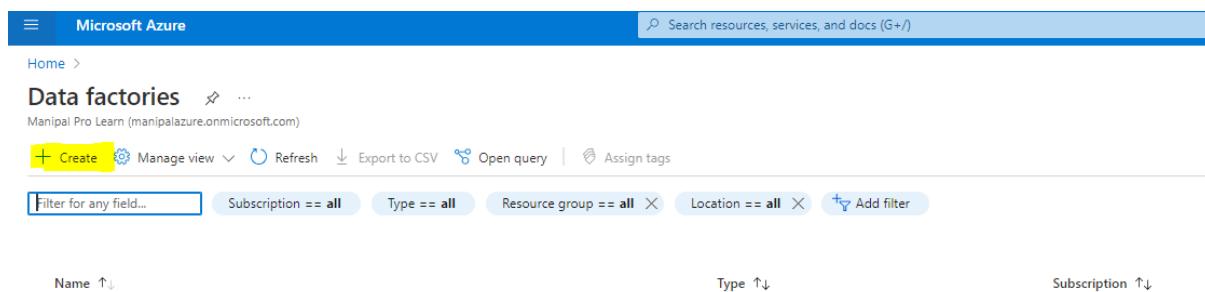
STEPS TO CREATE AZURE DATA FACTORY ACCOUNT FOR DATA PIPELINES.

STEP 1:-CLICK ON SEARCH MENU THEN CLICK ON DATA FACTORY.



The screenshot shows the Microsoft Azure search interface. The search bar at the top contains the text 'Data'. Below the search bar, there are several tabs: All, Services (59), Marketplace (20), Documentation (28), Azure Active Directory (8), and Resources (0). The 'Services' tab is selected. In the main pane, under the 'Services' heading, the 'Data factories' option is highlighted with a yellow box. Other listed services include Datadog, Data Catalog, Data Lake Analytics, Data Share Invitations, Data Lake Storage Gen1, Data Shares, and Azure Database for MySQL servers. There are also sections for 'Marketplace' and 'Documentation'.

STEP 2:-CLICK ON CREATE.



The screenshot shows the 'Data factories' blade in the Azure portal. At the top left, there is a 'Create' button highlighted with a yellow box. Below it, there are other buttons for 'Manage view', 'Refresh', 'Export to CSV', 'Open query', and 'Assign tags'. A search bar is located at the top right. The main area displays filtering options: 'Filter for any field...', 'Subscription == all', 'Type == all', 'Resource group == all', 'Location == all', and 'Add filter'. The columns for the data table are 'Name', 'Type', and 'Subscription'.

STEP 3:-FILL ALL THE PROJECT DETAILS AND CLICK ON NEXT.

Home > Data factories >

Create Data Factory

Basics Git configuration Networking Advanced Tags Review + create

Project details

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

Subscription *

Resource group * [Create new](#)

Instance details

Name *

Region *

Version *

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

STEP 4:-TICK ON THE BOX TO CONFIGURE GIT LATER.

Microsoft Azure

Home > Data factories >

Create Data Factory

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

Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration.
[Learn more about Git integration in Azure Data Factory](#)

Configure Git later

STEP 5:- AFTER THE VALIDATION IS PASSED CLICK ON CREATE.

The screenshot shows the Microsoft Azure 'Create Data Factory' wizard. At the top, there's a blue header bar with the Microsoft Azure logo and a search bar. Below the header, the breadcrumb navigation shows 'Home > Data factories > Create Data Factory'. The main title 'Create Data Factory' is centered above a green banner that says 'Validation Passed' with a checkmark icon. Below the banner, there are tabs: Basics, Git configuration, Networking, Advanced, Tags, and Review + create. The 'Review + create' tab is underlined, indicating it's the active step. Under the 'TERMS' section, there's a paragraph of legal text about agreeing to terms and privacy statements. In the 'Basics' section, the following details are listed:

- Subscription: Azure-DXC262AB12Lab
- Resource group: dxcrgr7781
- Name: datafactory7781
- Region: East US
- Version: V2 (Recommended)

In the 'Networking' section, it shows 'Connect via: Public endpoint'. At the bottom, there are buttons for 'Create' (highlighted in yellow), '< Previous', 'Next', and 'Download a template for automation'.

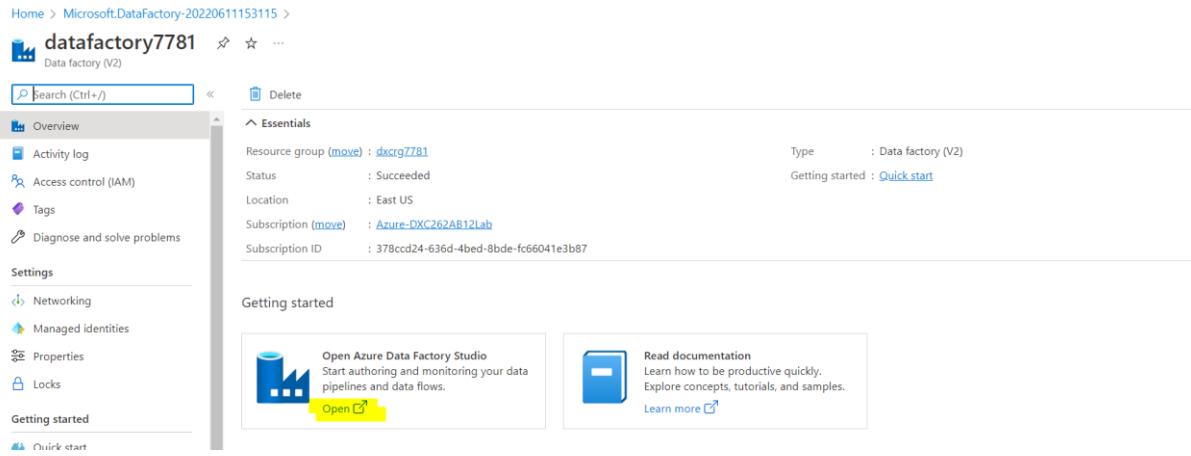
STEP 6:-AFTER THE DEPLOYMENT IS COMPLETE CLICK ON GO TO RESOURCE.

The screenshot shows the Microsoft Azure 'Overview' page for a deployed Data Factory. The top navigation bar includes the Microsoft Azure logo, a search bar, and icons for refresh, copy, and delete. The main title is 'Microsoft.DataFactory-20220611153115 | Overview'. On the left, there's a sidebar with navigation links: Home, Overview (which is selected and highlighted in grey), Inputs, Outputs, and Template. A feedback message 'We'd love your feedback!' is displayed. The main content area features a green banner with a checkmark icon and the text 'Your deployment is complete'. Below this, deployment details are listed:

- Deployment name: Microsoft.DataFactory-20220611153115
- Subscription: Azure-DXC262AB12Lab
- Resource group: dxcrgr7781

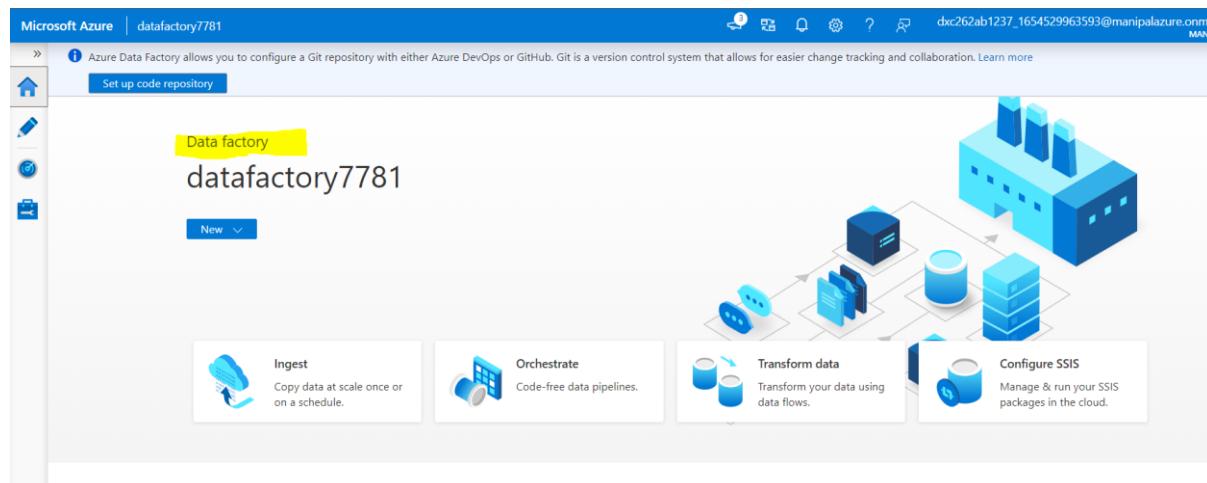
On the right, deployment logs show 'Start time: 6/11/2022, 3:37:18 PM' and 'Correlation ID: eee5aac6-50ef-4953-94af-ab4e2d6199e1'. There are also 'Deployment details (Download)' and 'Next steps' sections, with a 'Go to resource' button at the bottom.

STEP 7:- CLICK ON OPEN AZURE DATA FACTORY STUDIO.



The screenshot shows the Azure Data Factory V2 overview page for a resource named 'datafactory7781'. The 'Getting started' section contains two cards: 'Open Azure Data Factory Studio' and 'Read documentation'. The 'Open Azure Data Factory Studio' card has a yellow box around its 'Open' button.

STEP 8:- AZURE DATA FACTORY IS OPEN IN NEW TAB.

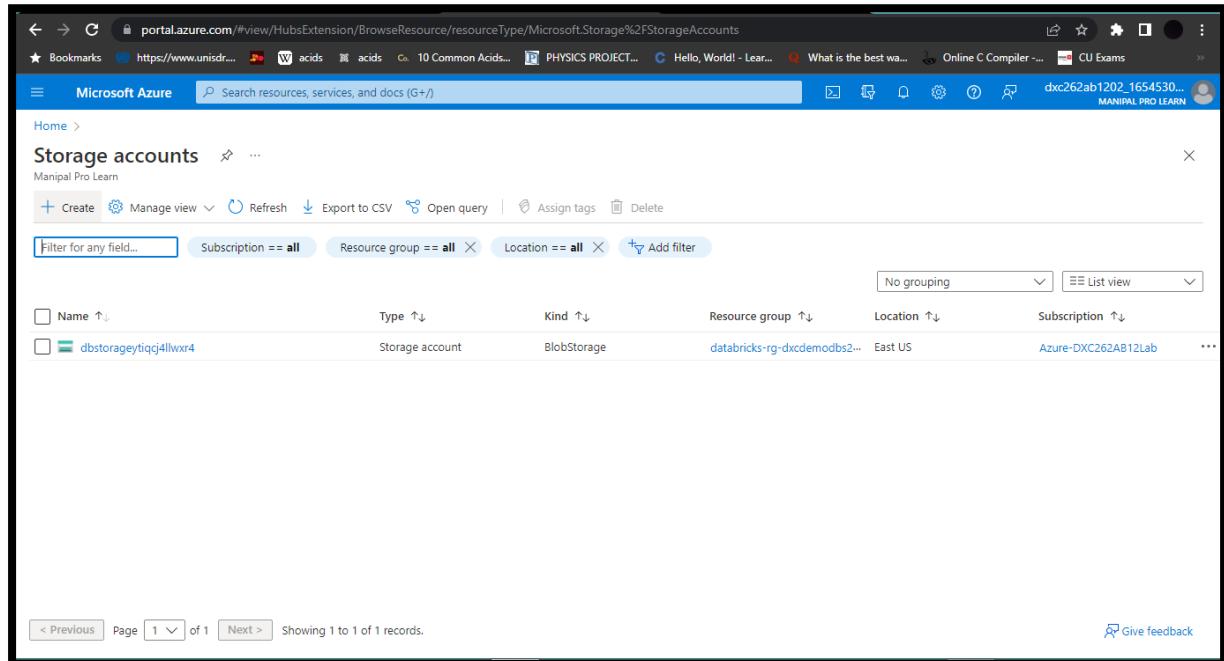


The screenshot shows the Azure Data Factory studio interface. The 'Data factory' tab is selected, displaying a dashboard with four main components: 'Ingest', 'Orchestrate', 'Transform data', and 'Configure SSIS'. A large 3D diagram of industrial pipes and tanks is visible in the background.

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

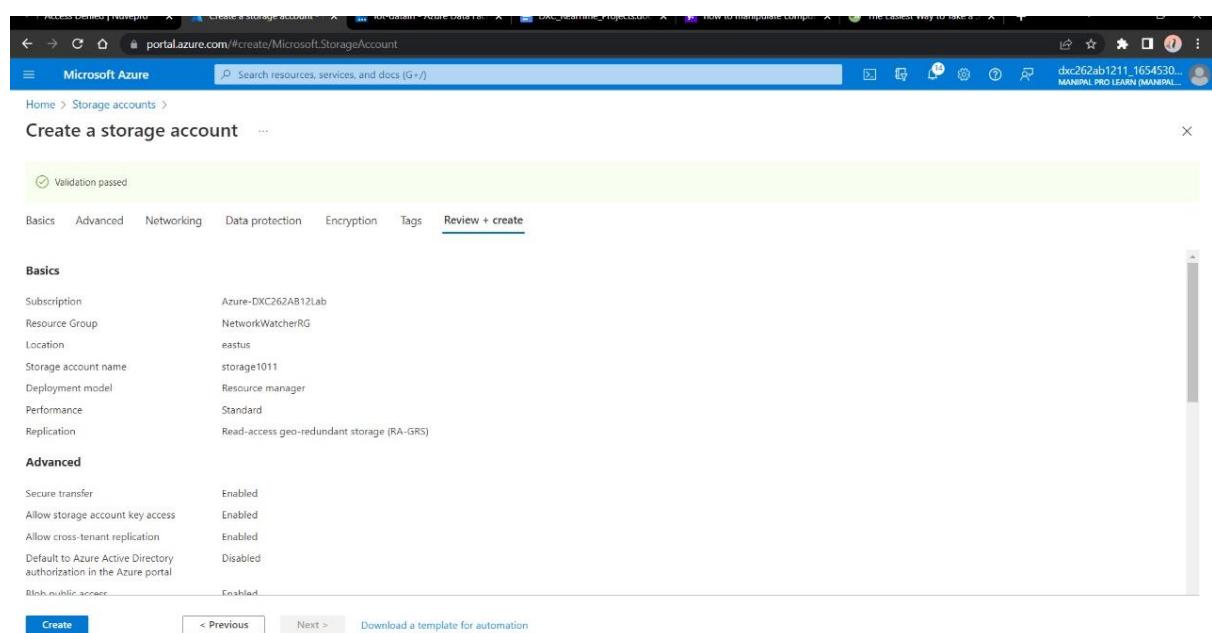
Practical Lab: Create Azure blob trigger logic.

Step 1:-Go to Azure, search for storage. Click on 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 table listing storage accounts. The table has columns for Name, Type, Kind, Resource group, Location, and Subscription. One account is listed: dbstorageytjqcj4llwxr4, which is a BlobStorage type located in East US, associated with databricks-rg-dxcdemodbs2... and Azure-DXC262AB12Lab. There are also filter and sorting options at the top of the table. At the bottom, there are pagination controls and a feedback link.

Step 2:-Fill up all the details and click on Review+Create. Deploy it .



The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal. The current step is 'Review + create'. The page displays validation status and configuration details across several tabs: Basics, Advanced, Networking, Data protection, Encryption, Tags, and Review + create. Under the Basics tab, the account is named 'storage1011' and is set to Standard performance level. Under the Advanced tab, secure transfer and allow storage account key access are enabled. At the bottom, there are 'Create' and 'Next >' buttons, along with a link to download a template for automation.

Step 3:-Once deployment is done click on Resource Group.

The screenshot shows the Microsoft Azure portal with the URL https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F4236c42a-d131-4bd6-b609-aec3a598f2d3%2FresourceGroups%2FNetworkWatcherRG/providers/Microsoft.Storage/deployments/storage1011_1654861967486. The page displays a deployment status message: "Your deployment is complete". It includes deployment details such as name, subscription, and resource group. A "Go to resource" button is present. On the right side, there are promotional cards for Cost Management, Microsoft Defender for Cloud, and Free Microsoft tutorials.

Step 4:-Navigate towards containers. Click on + Container.

The screenshot shows the Microsoft Azure portal with the URL <https://portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/4236c42a-d131-4bd6-b609-aec3a598f2d3/resourcegroups/NetworkWatcherRG/providers/Microsoft.Storage/storageAccounts/storage1011>. The page displays the "Containers" blade for the storage account "storage1011". It shows a table with one container named "\$logs". The left sidebar contains navigation links for Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser (preview), Data storage (Containers, File shares, Queues, Tables), Security + networking (Networking, Azure CDN, Access keys, Shared access signature), and a search bar.

Step 5:-Container is used in order to store the data. Click create.

The screenshot shows the Microsoft Azure portal interface for a storage account named 'storage1011'. On the left, there's a navigation sidebar with options like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, and Storage browser (preview). The main area displays a list of existing containers: 'Logs' (Last modified: 6/10/2022, 5:23:53 PM, Public access level: Private). A 'New container' dialog box is open on the right, asking for a name ('source') and setting the 'Public access level' to 'Private (no anonymous access)'. At the bottom right of the dialog are 'Create' and 'Discard' buttons.

Step 6:-Give a name to the container Click on upload.

This screenshot shows the same Microsoft Azure Storage account 'storage1011' after a new container named 'source' has been created. The 'Containers' blade now lists three containers: 'Logs', 'source', and 'source' (the newly created one). The 'source' container was created on 6/10/2022 at 5:26:01 PM with a 'Private' public access level and is currently 'Available'. The rest of the interface remains consistent with the previous screenshot, including the sidebar and the 'New container' dialog.

Step 7:-Considering we don't have the source of the data we will make this container as the source of our blob data and will fill in a sample file to paste into the sql database.

The screenshot shows the Microsoft Azure Storage Container blade for the 'source' container. The left sidebar includes links for Overview, Diagnose and solve problems, Access Control (IAM), Settings (Shared access tokens, Access policy, Properties, Metadata), and a Search bar. The main area displays a table of blobs:

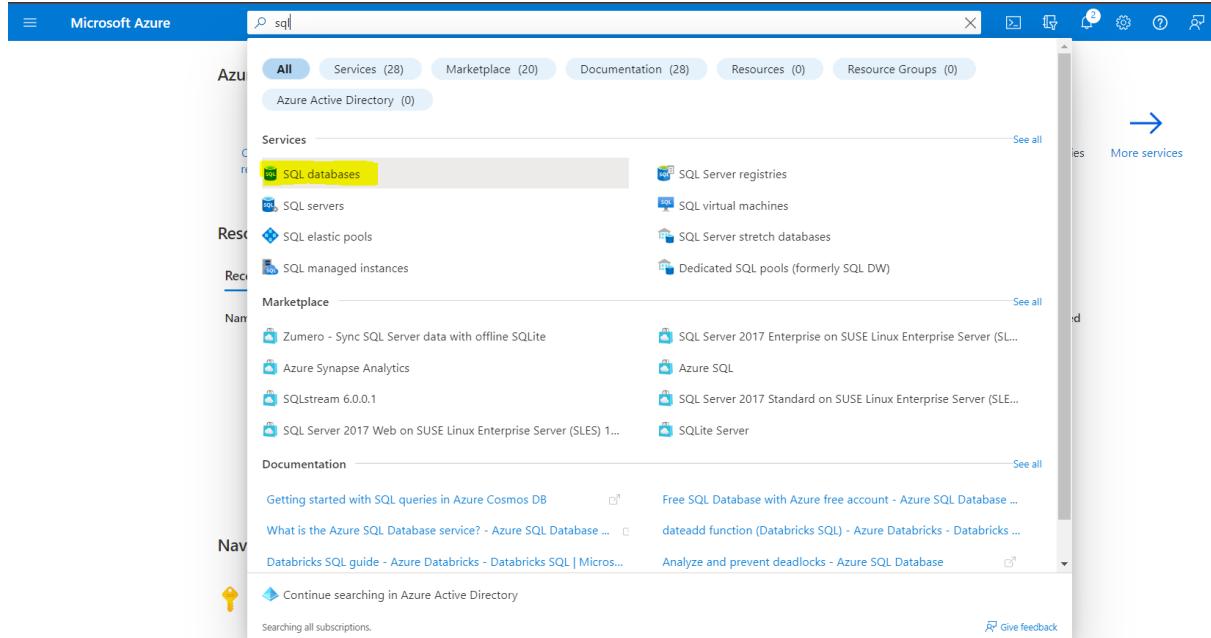
Name	Modified	Access tier	Archive status	Blob type
transactions.csv	6/10/2022, 5:27:16 PM	Hot (inferred)		Block blob

A modal dialog on the right shows a successful upload of 'transactions.csv' from 'storage1011'. The file is 880.14 KB. Below the modal, a 'Current uploads' section shows the same file with a progress bar indicating it's completed at 880 KiB / 880 KiB.

Practical Lab: Create Azure SQL Server and Database

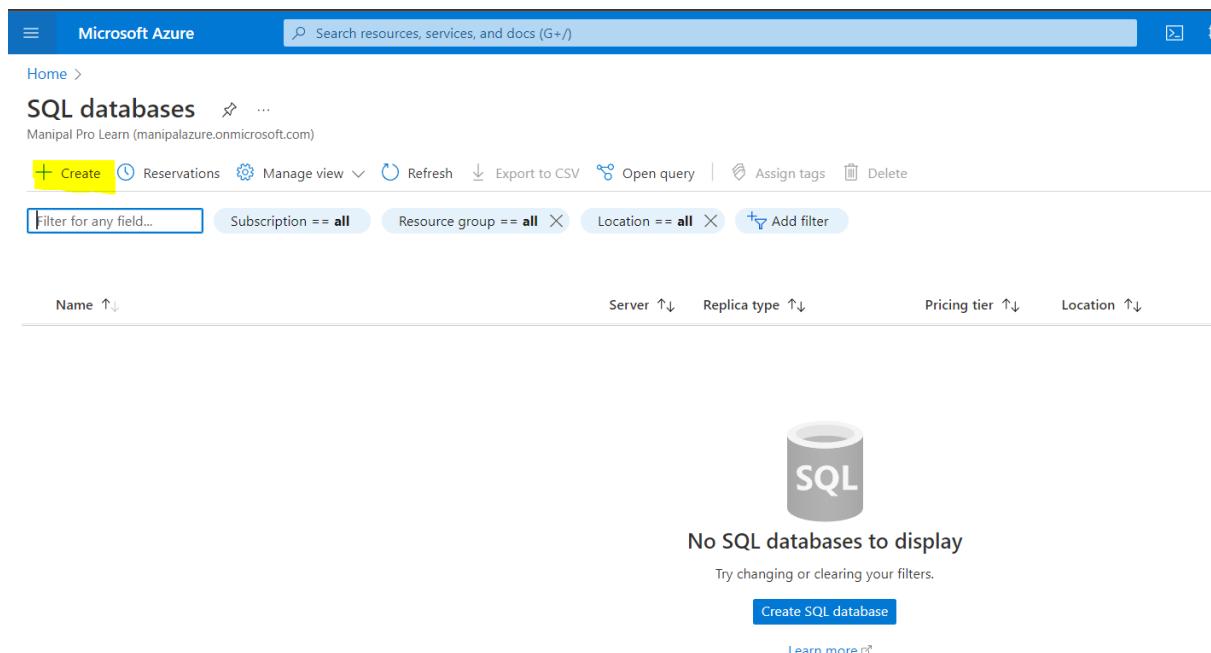
STEPS TO CREATE SQL SERVER. AND DATABASES.

STEP 1: - CLICK ON SEARCH THEN SEARCH FOR SQL DATABASES.



The screenshot shows the Microsoft Azure search interface. The search bar at the top contains the query 'sql'. Below the search bar, there are several navigation tabs: 'All', 'Services (28)', 'Marketplace (20)', 'Documentation (28)', 'Resources (0)', and 'Resource Groups (0)'. The 'Services' tab is selected. Under the 'Services' section, the 'SQL databases' item is highlighted with a yellow box. Other service options listed include 'SQL servers', 'SQL elastic pools', and 'SQL managed instances'. To the right of the main search results, there are additional service categories like 'SQL Server registries', 'SQL virtual machines', 'SQL Server stretch databases', and 'Dedicated SQL pools (formerly SQL DW)'. The overall interface is clean and modern, typical of the Azure portal.

STEP 2:-CLICK ON CREATE.



The screenshot shows the 'SQL databases' management page in the Azure portal. At the top, there is a search bar and a breadcrumb trail indicating the user is in the 'Home > SQL databases' section. Below the header, there are several filter and action buttons: '+ Create', 'Reservations', 'Manage view', 'Refresh', 'Export to CSV', 'Open query', 'Assign tags', and 'Delete'. There are also filters for 'Subscription', 'Resource group', and 'Location'. The main table area has columns for 'Name', 'Server', 'Replica type', 'Pricing tier', and 'Location'. A large 'SQL' icon is centered on the page, and the text 'No SQL databases to display' is shown, along with a note to 'Try changing or clearing your filters.' A prominent blue 'Create SQL database' button is located at the bottom of the page.

Step 3:-FILL ALL THE PROJECT DETAILS.

The screenshot shows the 'Create SQL Database' wizard in the Microsoft Azure portal. The top navigation bar includes 'Microsoft Azure' and a search bar. The breadcrumb path is 'Home > SQL databases > Create SQL Database'. The main title is 'Create SQL Database' with a '...' link. Below the title, it says 'Project details'.

Project details

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

Subscription * (dropdown)

Resource group * (dropdown)
Create new

Database details

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

Database name * (dropdown)

Server * (dropdown)
Create new

Want to use SQL elastic pool? Yes No

Compute + storage * (dropdown)
Gen5, 2 vCores, 32 GB storage, zone redundant disabled
[Configure database](#)

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

STEP 4:-CLICK ON CREATE NEW SQL DATABASE SERVER.

Microsoft Azure

Home > SQL databases > Create SQL Database >

Create SQL Database Server

Enter required settings for this server, including providing a name and location. This server will be created in the same subscription and resource group as your database.

Server name *

Location *

Authentication

Select your preferred authentication methods for accessing this server. Create a server admin login and password to access your server with SQL authentication, select only Azure AD authentication [Learn more](#) using an existing Azure AD user, group, or application as Azure AD admin [Learn more](#), or select both SQL and Azure AD authentication.

Authentication method

Use SQL authentication

Use only Azure Active Directory (Azure AD) authentication

Use both SQL and Azure AD authentication

Server admin login *

shari

>Password *

Confirm password *

OK

STEP 5:-CLICK ON CONFIGURE DATABASE IN COMPUTE+STORAGE AND SELECT THE SERVICE TIER COMPUTE TIER TO SERVERLESS.

Microsoft Azure

Home > SQL databases > Create SQL Database >

Configure

Feedback

Service and compute tier

Select from the available tiers based on the needs of your workload. The vCore model provides a wide range of configuration controls and offers Hyperscale and Serverless to automatically scale your database based on your workload needs. Alternately, the DTU model provides set price/performance packages to choose from for easy configuration. [Learn more](#)

Service tier

General Purpose (Scalable compute and storage options)

Compare service tiers

Compute tier

Provisioned - Compute resources are pre-allocated. Billed per hour based on vCores configured.

Serverless - Compute resources are auto-scaled. Billed per second based on vCores used.

Compute Hardware

Select the hardware configuration based on your workload requirements. Availability of compute optimized, memory optimized, and confidential computing hardware depends on the region, service tier, and compute tier.

Hardware Configuration

Gen5
up to 40 vCores, up to 120 GB memory
[Change configuration](#)

Max vCores

Min vCores

Apply

STEP 6:-CHOOSE MAX AND MIN CORES AND THEN CLICK ON APPLY.

The screenshot shows the 'Configure' step of creating a SQL database in Microsoft Azure. The 'Max vCores' slider is set to 1, and the 'Min vCores' slider is set to 0.5. Below the sliders, it displays '2.02 GB MIN MEMORY' and '3 GB MAX MEMORY'. Under 'Auto-pause delay', the 'Enable auto-pause' checkbox is checked, and the 'Hours' dropdown is set to 1. In the 'Data max size (GB)' section, the slider is at 1, and it shows '307.2 MB LOG SPACE ALLOCATED'. At the bottom, there is a question about making the database zone redundant with 'Yes' and 'No' options, and a large green 'Apply' button.

Microsoft Azure

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

Home > SQL databases > Create SQL Database >

Configure

Feedback

Max vCores

Min vCores

2.02 GB MIN MEMORY 3 GB MAX MEMORY

Auto-pause delay

The database automatically pauses if it is inactive for the time period specified here, and automatically resumes when database activity recurs. Alternatively, auto-pausing can be disabled.

Enable auto-pause

Days Hours Minutes

0 1 0

Data max size (GB) ①

Would you like to make this database zone redundant? ①

Yes No

Apply

STEP 7:- NOW CLICK ON REVIEW AND CREATE.

The screenshot shows the 'Create SQL Database' wizard in the Microsoft Azure portal. The top navigation bar includes 'Microsoft Azure' and a search bar. The breadcrumb path is 'Home > SQL databases > Create SQL Database'. The main title is 'Create SQL Database' with a '...' link. Below the title, it says 'Project details'.

Project details

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

Subscription * (dropdown)

Resource group * (dropdown)
Create new

Database details

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

Database name * (dropdown)

Server * (dropdown)
Create new

Want to use SQL elastic pool? Yes No

Compute + storage * (dropdown)
Gen5, 2 vCores, 32 GB storage, zone redundant disabled
[Configure database](#)

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

STEP 8: - NOW CLICK ON CREATE AND THEN SQL DATABASE WILL BE CREATED.

The screenshot shows the Microsoft Azure portal interface for creating a new SQL database. The top navigation bar includes the Microsoft Azure logo, a search bar, and a 'Search resources, services, and docs (G+ /)' field. Below the navigation, the breadcrumb trail shows 'Home > SQL databases > Create SQL Database'. The main title is 'Create SQL Database' with a 'Microsoft' logo underneath. The 'Review + create' tab is highlighted in yellow at the top of the page. The page content is divided into sections: 'Product details', 'Terms', 'Basics', 'Networking', and 'Actions' at the bottom.

Product details

SQL database by Microsoft
Terms of use | Privacy policy

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

Terms
By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rig

Basics

Subscription	Azure-DXC262AB12Lab
Resource group	dxcrg7781
Region	East US
Database name	dxcsql7781
Server	(new) sqlserver7781
Authentication method	SQL authentication
Server admin login	shan
Compute + storage	General Purpose - Serverless: Gen5, 1 vCore, 1 GB storage, zone redundant disabled
Backup storage redundancy	Geo-redundant backup storage

Networking

Allow Azure services and resources to access this server	No
Private endpoint	None
Minimum TLS version	1.2
Connection Policy	Default

Actions

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

STEP 9:-WHEN THE DEPLOYEMENT IS COMPLETED.THEN CLICK ON GO TO RESOURCE.

The screenshot shows the Microsoft Azure Overview page for a deployment named "Microsoft.SQLDatabase.newDatabaseNewServer_f5eb82df81f547408ecc7". The main message is "Your deployment is complete". It provides deployment details: Deployment name: Microsoft.SQLDatabase.newDatabaseNewServer_f5eb82df81f547408ecc7, Start time: 6/11/2022, 2:35:09 PM, Subscription: Azure-DKC262AB12Lab, Correlation ID: a508ccfb-a3f1-4596-b8e6-5b4abce309c4, Resource group: dxcrg7781. There is a "Go to resource" button at the bottom.

STEP 10:-CLICK ON QUERY EDITOR THEN GIVE YOUR LOGIN CREDENTIALS AND THEN CLICK ON OK.

The screenshot shows the "Query editor (preview)" section of the Microsoft SQL Database. The left sidebar lists various options like Overview, Activity log, Tags, etc., with "Query editor (preview)" highlighted. The main area shows a "Welcome to SQL Database Query Editor" message and a login form. The "Login" field contains "shan" and the "Password" field contains a masked password. Below the fields are "OK" and "Cancel" buttons. To the right, there are sections for "SQL server authentication" and "Active Directory authentication", with a "Continue as dxc262ab1237_1654529963..." link.

Microsoft Azure

Home > Microsoft.SQLDatabase.newDatabaseNewServer_f5eb82df81f54740becc7 > dxcsqldb7781 (sqlserver7781/dxcsqldb7781)

dxcsqldb7781 (sqlserver7781/dxcsqldb7781) | Query editor (preview) ...

Search (Ctrl+F)

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

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

Tables Views Stored Procedures

Query 1

```

1 CREATE TABLE address( address_id INTEGER NOT NULL, address_building_number VARCHAR(55) NOT NULL, address_street VARCHAR(55) NOT NULL, address_locality VARCHAR(55), address_city VARCHAR(55), address_zip_postal VARCHAR(55), address_state_province_county VARCHAR(55), address_country VARCHAR(55) NOT NULL, CONSTRAINT PK_address PRIMARY KEY (address_id) );
2 CREATE TABLE email_address( email_address_id INTEGER NOT NULL, email_address_person_id INTEGER, email_address VARCHAR(55) NOT NULL, CONSTRAINT PK_email_address PRIMARY KEY (email_address_id) );
3 CREATE TABLE person( person_id INTEGER NOT NULL, person_first_name VARCHAR(55) NOT NULL, person_last_name VARCHAR(55) NULL, person_contacted_number INTEGER NOT NULL, person_date_last_contacted DATETIME, person_address_id INTEGER NOT NULL, person_address_person_id INTEGER NOT NULL, person_address_address_id INTEGER NOT NULL, CONSTRAINT PK_person PRIMARY KEY (person_id) );
4 CREATE TABLE person_address( person_address_id INTEGER NOT NULL, person_address_person_id INTEGER NOT NULL, person_address_address_id INTEGER NOT NULL, CONSTRAINT PK_person_address PRIMARY KEY (person_address_id) );
5 CREATE TABLE phone_number( phone_number_id INTEGER NOT NULL, phone_number_person_id INTEGER NOT NULL, phone_number VARCHAR(55) NOT NULL, phone_number_address_id INTEGER NOT NULL, CONSTRAINT PK_phone_number PRIMARY KEY (phone_number_id) );
6

```

Results Messages

Search to filter items...

Home > Microsoft.SQLDatabase.newDatabaseNewServer_f5eb82df81f54740becc7 > dxcsqldb7781 (sqlserver7781/dxcsqldb7781)

dxcsqldb7781 (sqlserver7781/dxcsqldb7781) | Query editor (preview) ...

Search (Ctrl+F)

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

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

Tables Views Stored Procedures

Query 1

```

1 INSERT INTO address (address_id, address_building_number, address_street, address_locality, address_city, address_zip_postal, address_state_province_county, address_country) VALUES (1, 555, 'azure203Demo', 'Boones', 'Somewhere', '11111', 'Maine', 'US');
2 INSERT INTO address (address_id, address_building_number, address_street, address_locality, address_city, address_zip_postal, address_state_province_county, address_country) VALUES (2, 555, 'azure203Demo', 'Toronto', 'Ontario', '7777', 'Ontario', 'CA');
3 INSERT INTO address (address_id, address_building_number, address_street, address_locality, address_city, address_zip_postal, address_state_province_county, address_country) VALUES (3, 555, 'azure203Demo', 'Boones', 'Somewhere', '11111', 'Maine', 'US');
4 INSERT INTO address (address_id, address_building_number, address_street, address_locality, address_city, address_zip_postal, address_state_province_county, address_country) VALUES (4, 555, 'azure203Demo', 'San Francisco', 'California', '91001', 'California', 'US');
5
6 INSERT INTO email_address (email_address_id, email_address_person_id, email_address) VALUES (1, 1, 'jon.flanders@mail.com');
7 INSERT INTO email_address (email_address_id, email_address_person_id, email_address) VALUES (2, 1, 'jon@andtherest.com');
8 INSERT INTO email_address (email_address_id, email_address_person_id, email_address) VALUES (3, 2, 'frizz@mail.com');
9 INSERT INTO email_address (email_address_id, email_address_person_id, email_address) VALUES (4, 3, 'wron@mail.com');
10
11 INSERT INTO person (person_id, person_first_name, person_last_name, person_contacted_number, person_date_last_contacted, person_data_added) VALUES (1, 'Jon', 'Flanders', 5, '2013-09-14');
12 INSERT INTO person (person_id, person_first_name, person_last_name, person_contacted_number, person_date_last_contacted, person_data_added) VALUES (2, 'Shannon', 'Ahern', 6, '2013-09-14');
13 INSERT INTO person (person_id, person_first_name, person_last_name, person_contacted_number, person_date_last_contacted, person_data_added) VALUES (3, 'Frizz', 'Onion', 1, '2013-07-14');
14
15 INSERT INTO person_address (person_address_id, person_address_person_id, person_address_address_id) VALUES (1, 1, 1);

```

Results Messages

Query succeeded: Affected rows: 0

Home > Microsoft.SQLDatabase.newDatabaseNewServer_f5eb82df81f54740becc7 > dxcsqldb7781 (sqlserver7781/dxcsqldb7781)

dxcsqldb7781 (sqlserver7781/dxcsqldb7781) | Query editor (preview) ...

Search (Ctrl+F)

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

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

Tables Views Stored Procedures

Query 1

```

1 SELECT * FROM address

```

Results Messages

Search to filter items...

address_id	address_building_number	address_street	address_locality	address_city	address_zip_postal	address_state_province_county	address_country
1	555	azure203Demo		Los Angeles	91001	California	US
2	555	azure203Demo		Toronto	7777	Ontario	Canada
3	555	azure203Demo	Boones	Somewhere	11111	Maine	US
4	555	azure203Demo		San Francisco	91001	California	US

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

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

Step 1:- Click on ingest.

The screenshot shows the 'Copy Data tool' wizard in the Microsoft Azure portal. The current step is 'Properties'. The left sidebar lists steps 1 through 5: Properties, Source, Target, Settings, and Review and finish. The main area displays two options under 'Task type': 'Built-in copy task' (selected) and 'Metadata-driven copy task'. Both options are described as generating single pipelines for copying data from 90+ data sources. Below these, 'Task cadence or task schedule' is set to 'Run once now'. At the bottom are 'Previous' and 'Next >' buttons, and a 'Cancel' button on the right.

STEP 2:- Enter the name of Azure blob storage linked file. Select the name of storage account name we created in this and click on Create.

The screenshot shows the 'Copy Data tool' wizard in the Microsoft Azure portal, currently at the 'Source' step. The left sidebar shows steps 1 through 5: Properties, Source, Dataset, Configuration, Target, Settings, and Review and finish. The main area is titled 'Source data store' and specifies 'Azure Blob Storage' as the source type. It includes fields for 'Connection' (with 'Select...' and 'New connection' buttons), 'Name' (set to 'AzureBlobStorage1'), 'Description', 'Connect via integration runtime' (set to 'AutoResolveIntegrationRuntime'), 'Authentication type' (set to 'Account key'), and 'Storage account name' (set to 'storage1011'). A 'Create' button is at the bottom, along with 'Test connection' and 'Cancel' buttons.

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

Copy Data tool

Properties

Source

Dataset

Configuration

Target

Settings

Review and finish

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: source/transactions.csv

Options

Binary copy

Recursively

Enable partition discovery

Max concurrent connections: (empty)

Filter by last modified

Start time (UTC): (empty)

End time (UTC): (empty)

< Previous Next > Cancel

Step 4:- Choose the output format of the data

Copy Data tool

Properties

Source

Dataset

Configuration

Target

Settings

Review and finish

File format settings

File format: DelimitedText

Column delimiter: Comma (,)

Edit

Row delimiter: Default (\r\n, or \n\r)

First row as header

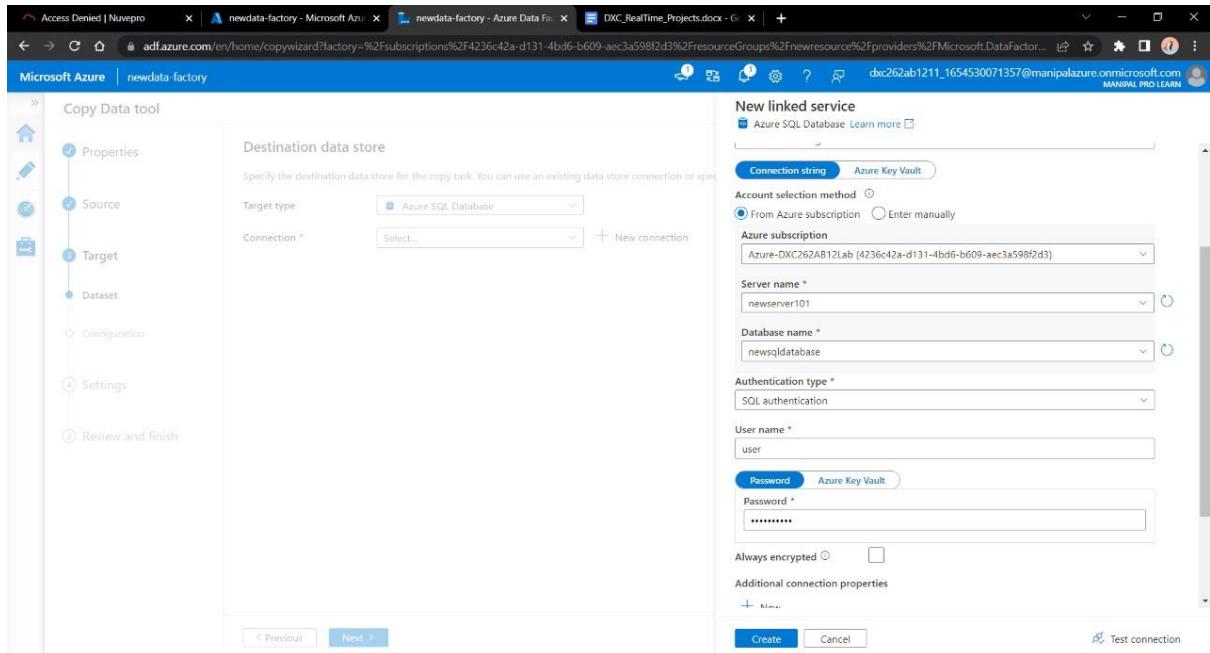
> Advanced

Compression type: None

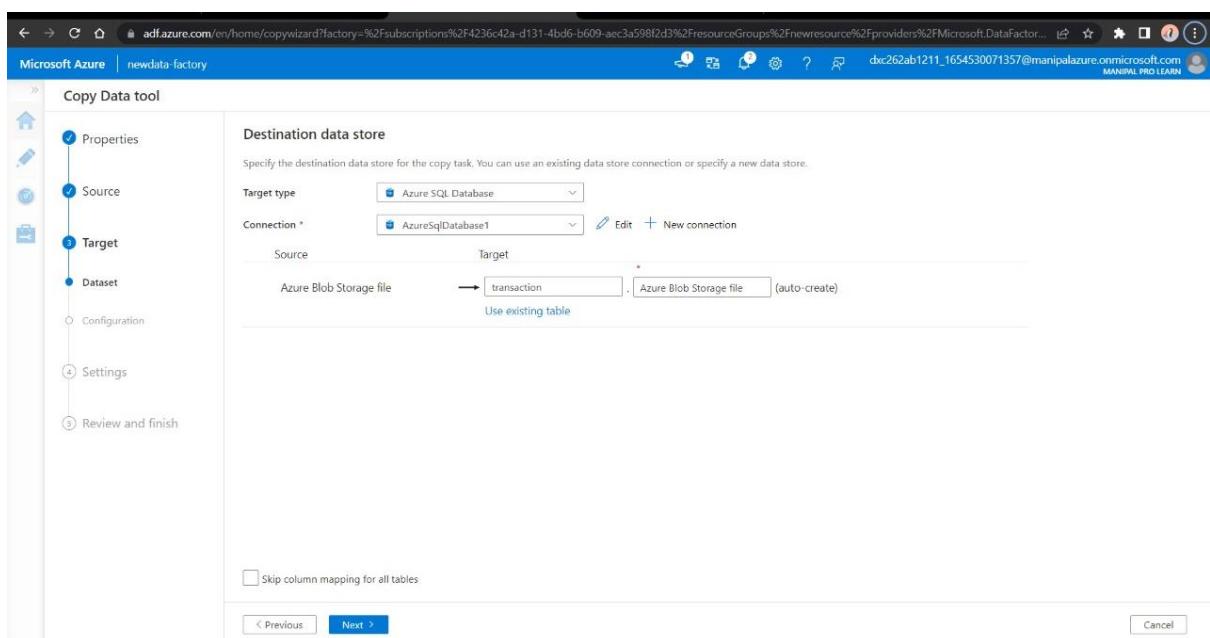
Additional columns: New

< Previous Next > Cancel

Step 5:-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 at the end stage of creating pipeline, click next.

The screenshot shows the 'Copy Data tool' settings page. On the left, a vertical navigation bar lists steps: Properties, Source, Target, Settings, Review and finish, and a final step. The 'Settings' step is currently selected. The main area is titled 'Settings' and contains the following fields:

- Task name ***: CopyPipeline_zbw
- Task description**: (empty text area)
- Data consistency verification**: (radio button)
- Fault tolerance**: (dropdown menu)
- Enable logging**: (checkbox)
- Enable staging**: (checkbox)
- Advanced**: (link)

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

Step 8:- Validate the data, connection, details and move forwards with Next

The screenshot shows the 'Copy Data tool' summary page. The left navigation bar shows steps: Properties, Source, Target, Settings, Review and finish, and Deployment. The 'Review and finish' step is selected. The main area is titled 'Summary' and displays the following information:

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

A diagram shows an arrow pointing from 'Azure Blob Storage' to 'Azure SQL Database'.

Properties section:

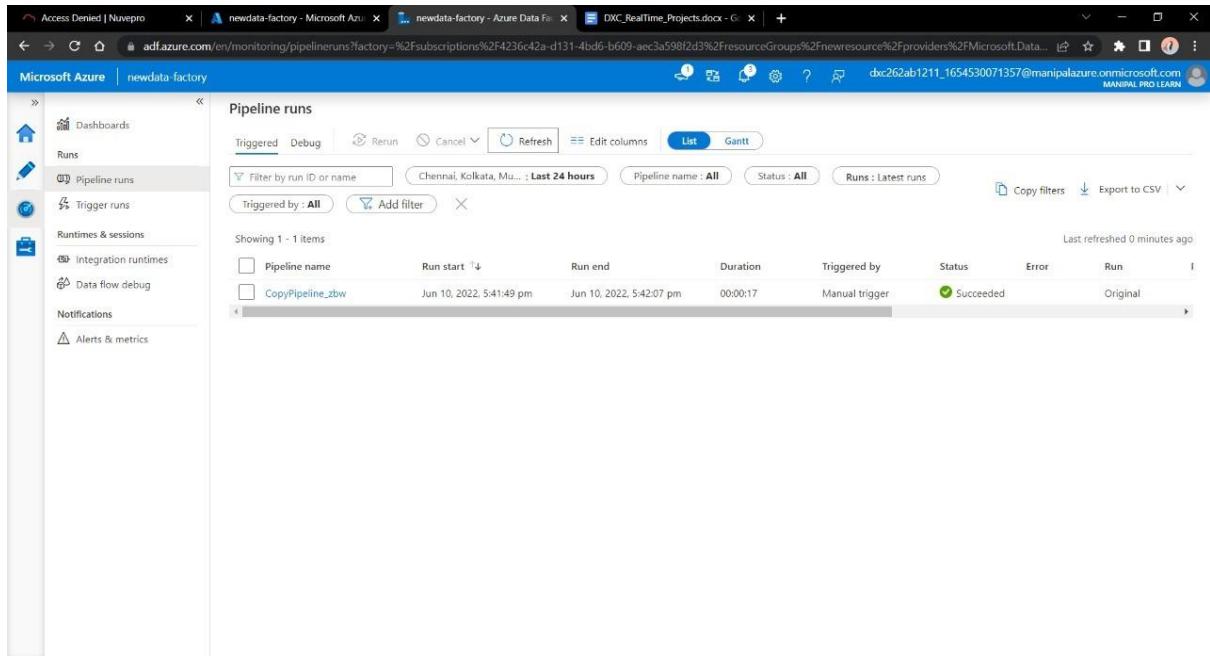
Task name	CopyPipeline_zbw
Task description	(empty)
Source	(Edit icon)

Source properties:

Connection name	AzureBlobStorage1
Dataset name	SourceDataset_zbw
Column delimiter	,
Escape character	\
Quote char	"
First row as header	true
File name	transactions.csv
Container	source

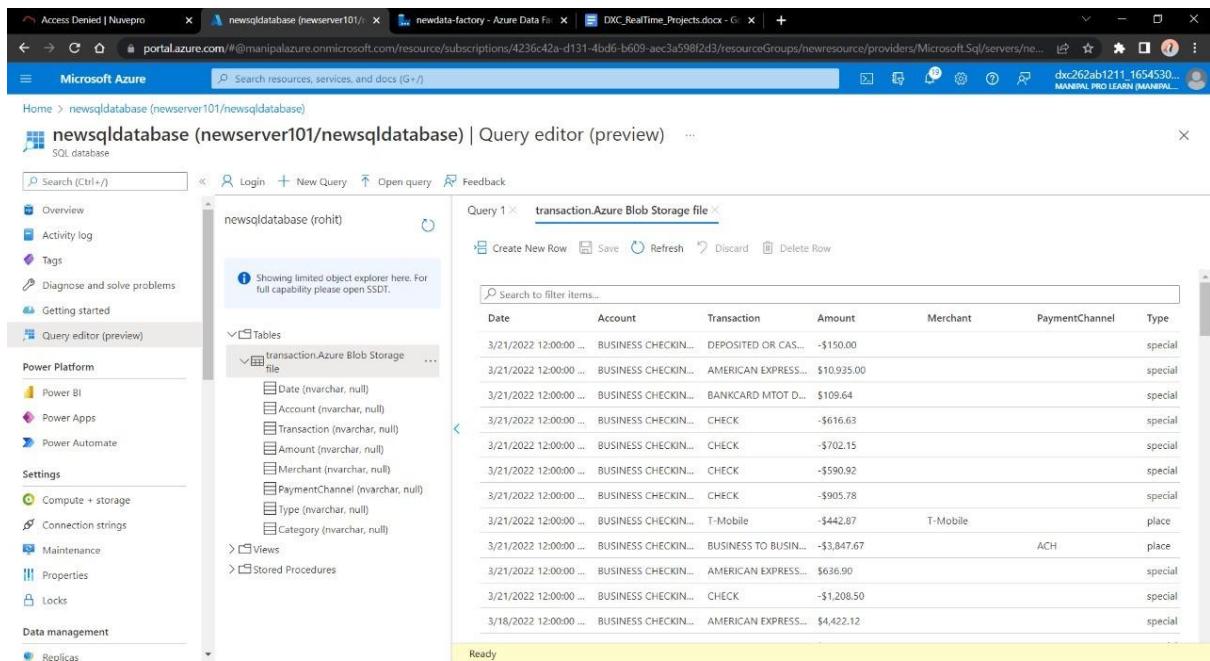
At the bottom are buttons for < Previous, Next >, and Cancel.

Step 9:-Trigger the pipeline manually and navigate towards monitor, Here we can see our pipeline has successfully ran once.



The screenshot shows the Microsoft Azure Pipeline runs monitor. The left sidebar includes options like Dashboards, Runs, Pipeline runs (which is selected), Trigger runs, Runtimes & sessions, Integration runtimes, Data flow debug, Notifications, and Alerts & metrics. The main area is titled 'Pipeline runs' and displays a table of recent runs. The table columns are Pipeline name, Run start, Run end, Duration, Triggered by, Status, Error, and Run. A single row is shown: 'CopyPipeline_zbw' with a run starting at Jun 10, 2022, 5:41:49 pm, ending at Jun 10, 2022, 5:42:07 pm, duration 00:00:17, triggered by 'Manual trigger', status 'Succeeded', and run type 'Original'. The table header indicates 'Showing 1 - 1 items' and 'Last refreshed 0 minutes ago'.

Step 10:-Navigate towards SQL database in order to check the data.



The screenshot shows the Microsoft Azure SQL database query editor. The left sidebar lists various database management options: Overview, Activity log, Tags, Diagnose and solve problems, Getting started, Query editor (preview) (which is selected), Power Platform (Power BI, Power Apps, Power Automate), Settings (Compute + storage, Connection strings, Maintenance, Properties, Locks), and Data management (Replicas). The main area is titled 'newsqldatabase (newserver101/newsqldatabase) | Query editor (preview)' and shows a table named 'transaction.Azure Blob Storage file'. The table has columns: Date, Account, Transaction, Amount, Merchant, PaymentChannel, and Type. The data shows multiple transactions from March 21, 2022, to March 18, 2022, involving various payment methods like BUSINESS CHECKIN, AMERICAN EXPRESS, BANKCARD, T-Mobile, and ACH, with amounts ranging from -\$150.00 to \$4,422.12.

Date	Account	Transaction	Amount	Merchant	PaymentChannel	Type
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	DEPOSITED QR 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

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

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

Name: SHAN

Reg No: DXC262AB1228

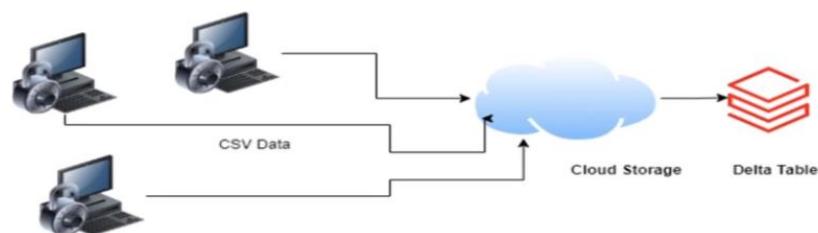
Project2 Name: AP Morgan Data Platform

Date: 10TH JUNE 2022

Project 2: AP Morgan Data Platform

Project 2 : AP Morgan

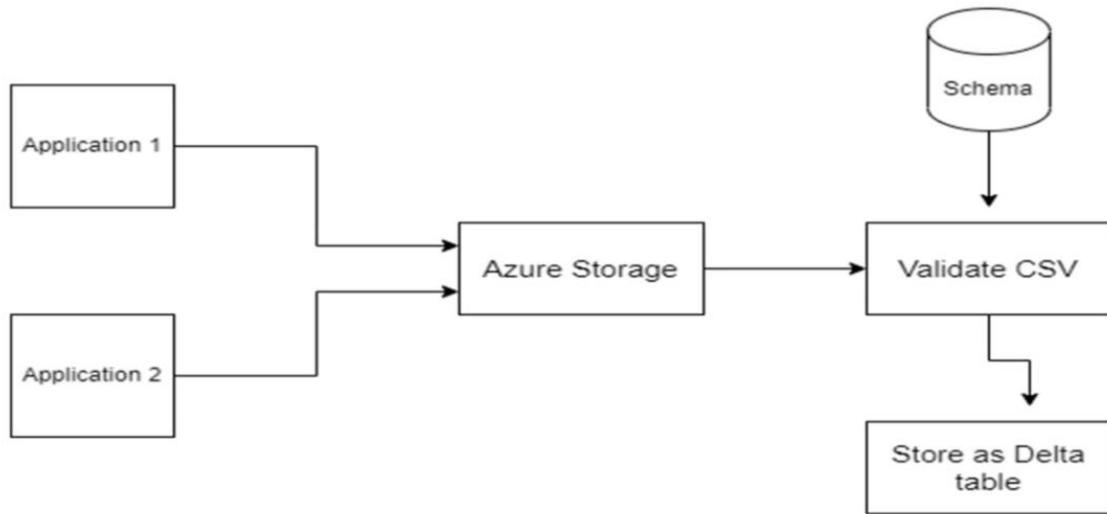
- Multiple Internal applications sends the data(huge size) in CSV format on daily basis in the cloud storage location. There are couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed data to be persisted as Delta table in Databricks for downstream system.



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

STEPS TO CREATE DATABRICKS:-

FIRST WE NEED TO OPEN MICROSOFT AZURE PORTAL THEN INSIDE THE PORTAL WE NEED TO FOLLOW THESE STEPS

STEP 1:- SEARCH FOR DATABRICKS IN THE SEARCH AREA AND CLICK ON IT.

The screenshot shows the Microsoft Azure portal search results for 'databricks'. The search bar at the top contains the query 'databricks'. Below the search bar, there are several categories: All, Services (26), Marketplace (9), Documentation (7), Resources (0), and Resource Groups (0). The 'All' category is selected. Under the 'Services' section, 'Azure Databricks' is highlighted with a yellow box. Other services listed include Azure Database Migration Projects, Azure Database Migration Services, Azure Database for MariaDB servers, and Azure Database for PostgreSQL servers. In the 'Marketplace' section, items like Infoworks for Databricks, Unravel for Azure Databricks, and NASH Video Analytics on Azure are shown. The 'Documentation' section includes links to 'What is Azure Databricks?' and 'Databricks File System (DBFS) - Azure Databricks'. At the bottom of the search results, there is a note: 'Searching all subscriptions.' and a 'Give feedback' button.

STEP 2:- CLICK ON CREATE.

The screenshot shows the Azure Databricks service list page. At the top, there are navigation links for Home > Azure Databricks. Below the header are various filtering options: 'Create' (highlighted in yellow), 'Manage view', 'Refresh', 'Export to CSV', 'Open query', and 'Assign tags'. There are also dropdown filters for 'Subscription == all', 'Resource group == all', 'Location == all', and 'Add filter'. Below the filters, there are sorting options: 'Name ↑', 'Type ↑↓', 'Resource group ↑↓', 'Location ↑↓', and 'Subscription ↑↓'. The main content area displays a large 'No azure databricks services to display' message with a small icon of three stacked bars. Below this message, there is a brief description: 'Unlock insights from all your data and build artificial intelligence (AI) solutions with Azure Databricks. Set up your Apache Spark environment in minutes, autoscale, and collaborate on shared projects in an interactive workspace.' A prominent blue button labeled 'Create azure databricks service' is centered at the bottom of the page, with a 'Learn more' link below it.

STEP 3:- IN CREATE AN AZURE DATABRICKS WORKSPACE FILL ALL THE PROJECT DETAILS AND CLICK NEXT.

The screenshot shows the 'Create an Azure Databricks workspace' wizard on the 'Basics' tab. The top navigation bar includes the Microsoft Azure logo and a search bar. The breadcrumb navigation shows 'Home > Azure Databricks > Create an Azure Databricks workspace'. The page title is 'Create an Azure Databricks workspace' with a '...' button. Below the title, there are tabs for 'Basics' (which is selected and highlighted in blue), 'Networking', 'Advanced', 'Tags', and 'Review + create'. The 'Project Details' section contains fields for 'Subscription *' (set to 'Azure-DXC262AB12Lab') and 'Resource group *' (set to 'dxcrg7781'). A 'Create new' link is available for creating a new resource group. The 'Instance Details' section includes fields for 'Workspace name *' (set to 'dxcdbricks231'), 'Region *' (set to 'East US'), and 'Pricing Tier *' (set to 'Trial (Premium - 14-Days Free DBUs)'). A red box highlights the 'Pricing Tier' field. At the bottom of the page are navigation buttons: 'Review + create' (highlighted in blue), '< Previous', and 'Next : Networking >' (highlighted in yellow).

STEP 4:- SKIP NETWORKING, ADVANCED, TAG AND THEN CLICK ON REVIEW+CREATE AND AFTER THE VALIDATION IS SUCCEEDED. CLICK ON CREATE.

The screenshot shows the Microsoft Azure portal interface for creating a Databricks workspace. At the top, there's a blue header bar with the Microsoft Azure logo and a search bar. Below the header, the breadcrumb navigation shows 'Home > Azure Databricks > Create an Azure Databricks workspace'. The main title is 'Create an Azure Databricks workspace'. A green banner at the top indicates 'Validation Succeeded'. Below the banner, there are tabs for 'Basics', 'Networking', 'Advanced', 'Tags', and 'Review + create', with 'Review + create' being underlined. Under the 'Basics' section, there are fields for 'Workspace name' (dxcdatalab231), 'Subscription' (Azure-DXC262AB12Lab), 'Resource group' (dxcrg7781), 'Region' (East US), and 'Pricing Tier' (trial). The 'Networking' section contains two options: 'Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)' and 'Deploy Azure Databricks workspace in your own Virtual Network (VNet)', both set to 'No'. The 'Advanced' section has one option, 'Enable Infrastructure Encryption', set to 'No'. At the bottom, there are three buttons: a large green 'Create' button, a smaller white '< Previous' button, and a link to 'Download a template for automation'.

AFTER DEPLOYMENT IS DONE THEN GO TO RESOURCES.

resourcegroup_bricks101 | Overview

Your deployment is complete

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

Start time: 6/11/2022, 10:26:04 PM
Correlation ID: 89c0e208-0dc0-4273-9799-a00365c96492

Deployment succeeded

Cost Management

Microsoft Defender for Cloud

Free Microsoft tutorials

Work with an expert

STEP 5:- AFTER CREATING CLICK ON LAUNCH WORKSPACE.

dxcdatabrick231

Overview

Status: Active

Resource group: dxcrg7781

Location: East US

Subscription: Azure-DXC262AB12Lab

Subscription ID: 378cc024-636d-4bed-8bde-fc66041e3b87

Tags (edit): Click here to add tags

Managed Resource Group: databricks-rg-dxcdatabrick231-sietltum6tjyo

URL: https://adb-7830094762102553.13.azuredatabricks.net

Pricing Tier: Trial (Premium - 14-Days Free DBUs)

Launch Workspace

STEP 6:-AFTER CLICKING ON LAUNCH WORKSPACE, DATABRICKS WILL OPEN IN NEW TAB.

The screenshot shows the Microsoft Azure Databricks 'Get started' page. The left sidebar includes links for Data Science & Engineering, Create, Workspace, Repos, Recents (with a red 'NEW' badge), Search, Data, Compute, Workflows, Partner Connect, Tasks Completed (0/3), Help, and Settings. The main content area has a 'Get started' heading with a sub-section 'Set up your workspace' containing 'Create a cluster', 'Ingest data.', and 'Invite your team'. Below this are 'Next steps' for 'Explore Notebook gallery' and 'Read documentation'. To the right, under 'Data Science & Engineering', there's a 'Notebook' section with a 'Create a notebook' link, a 'Partner Connect' section listing Fivetran, dbt, Tableau, Power BI, and a 'Recents' section which is currently empty. A 'Documentation' link is at the bottom right.

Practical Lab: Create Cluster in Azure Databricks

STEPS TO CREATE CLUSTER IN AZURE DATABRICKS:-

STEP 1:- IN DATABRICK CLICK ON COMPUTE.

The screenshot shows the Microsoft Azure Databricks portal interface. The top navigation bar includes 'Microsoft Azure' and 'Databricks'. On the right, it shows 'Portal dx26ab1237_165452996359' and a message 'Free trial ends in 14 days.' Below the navigation, there's a sidebar with various options: Data Science & Engineering (selected), Machine Learning, SQL, Recents, Search, Data, Compute (highlighted with a yellow box), and Workflows. The main content area has sections for 'Data import', 'Partner Connect', and 'Last viewed'. A message 'There are no recents yet' is displayed. On the right, there are sections for 'Release notes' (with links to Runtime release notes, Azure Databricks preview releases, Platform release notes, and More release notes) and 'Blog posts' (with links to Building ETL pipelines for the Live Tables, Streaming Windows Event Log Lakehouse, and Speed Up Streaming Queries Checkpointing).

STEP 2:-CLICK ON CREATE CLUSTER.

The screenshot shows the 'Compute' page in the Azure Databricks portal. At the top, there are tabs for 'All-purpose clusters' (selected), 'Job clusters', 'Pools', and 'Cluster policies'. Below the tabs, a green button labeled 'Create Cluster' is highlighted with a yellow box. To the right of the button are filters for 'Created by me' and 'Accessible by'. The main table lists cluster configurations with columns for Name, Policy, Runtime, Active memory, Active cores, and Active. A note at the bottom says 'Depending on your workloads we recommend different cluster configurations' and points to a guide for best practices.

STEP 3:- FILL DETAILS OF THE CLUSTER AS GIVEN IN THE PICTURE LIKE CLUSTER NAME, CLUSTER MODE- WHICH WE HAD TO CHOOSE SINGLE NODE. THEN CHOOSE DATABRICKS RUNTIME VERSION. AND SELECT AUTOPILOT AT 30 MIN. THEN CLICK ON CREATE CLUSTER.

The screenshot shows the 'Clusters / New Compute' section of the Microsoft Azure Databricks interface. A 'New Cluster' button is highlighted in blue at the top right. Below it, the 'Cluster name' field contains 'dxccluster2314'. The 'Cluster mode' dropdown is set to 'Single Node'. The 'Databricks runtime version' dropdown shows 'Runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1)'. A promotional message box states '50% promotional discount applied to Photon during preview'. Under 'Autopilot options', there is a checked checkbox for 'Terminate after 30 minutes of inactivity'. The 'Node type' dropdown shows 'Standard_DS3_v2' with '14 GB Memory, 4 Cores'. The 'DBU / hour' cost is listed as '0.75'. The 'Standard_DS3_v2' node type is highlighted in purple. At the bottom, a 'Create Cluster' button is visible.

STEP 4:- IT WILL LOAD AFTER SOME TIME. AND THEN CLUSTER WILL BE CREATED.

Microsoft Azure | Databricks

Clusters / dxcluster2314

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

Policy ? Unrestricted

Cluster mode ? Single Node

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

Use Photon Acceleration ? Preview

Autopilot options Terminate after 30 minutes of inactivity ?

Node type ? Standard_DS3_v2 14 GB Memory, 4 Cores

DBU / hour: 0.75 ? Standard_DS3_v2

Advanced options

1/3

Google Subscription Details | Nuvepro bricks101 - Microsoft Azure Create Cluster - Databricks AP-mergan - Azure Data Factory

Microsoft Azure | Databricks Clusters / newcluster

newcluster Configuration Notebooks (0) Libraries Event log Spark UI Driver logs Metrics Apps Spark cluster UI - Master ▾

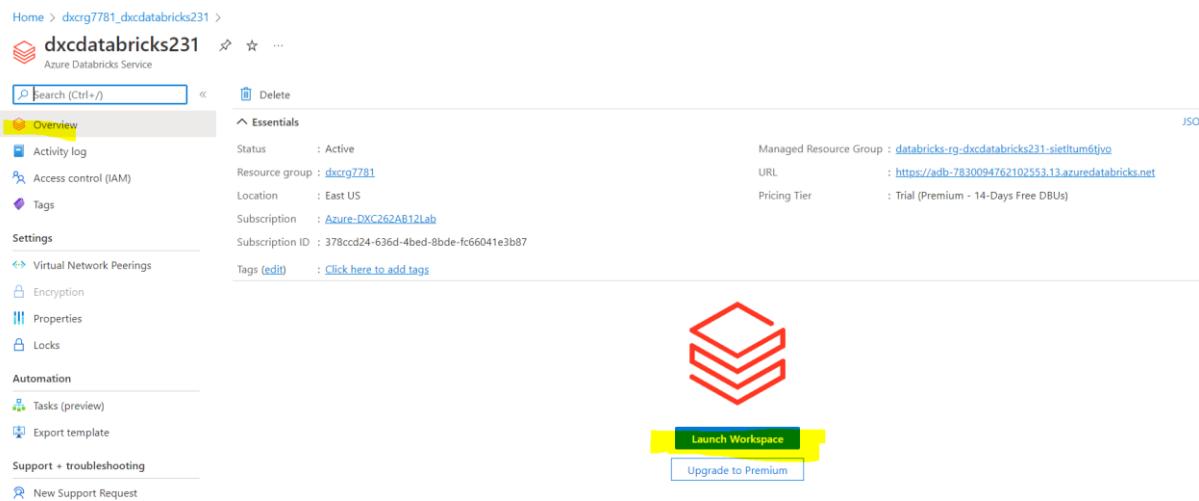
More ... Edit Terminate

No notebooks are attached to this cluster

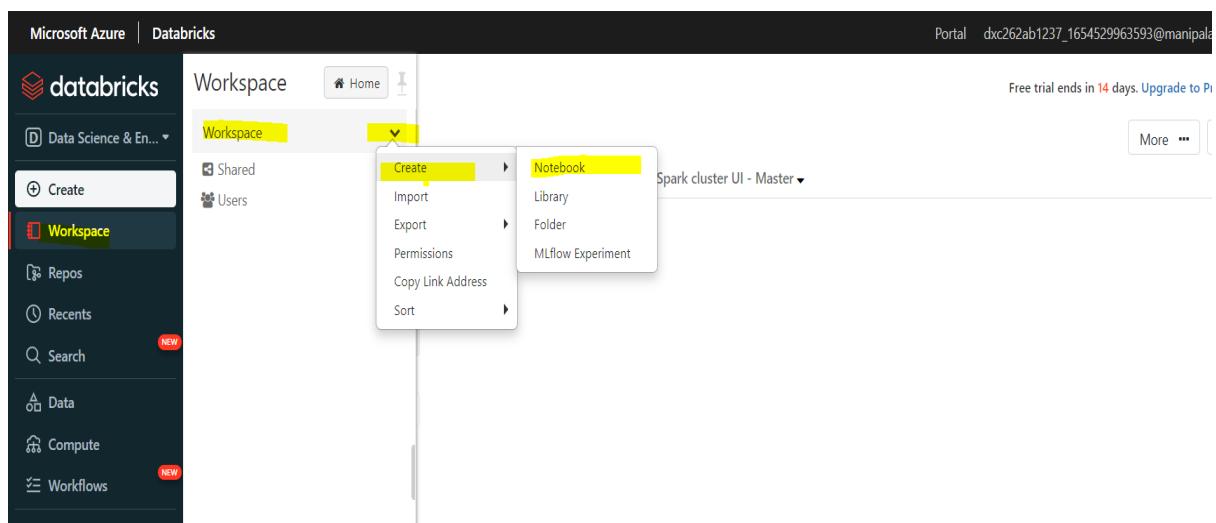
Practical Lab: Add notebook in Databricks and Implement the Business Logic.

STEPS TO CREATE NOTEBOOKS IN DATABRICKS:-

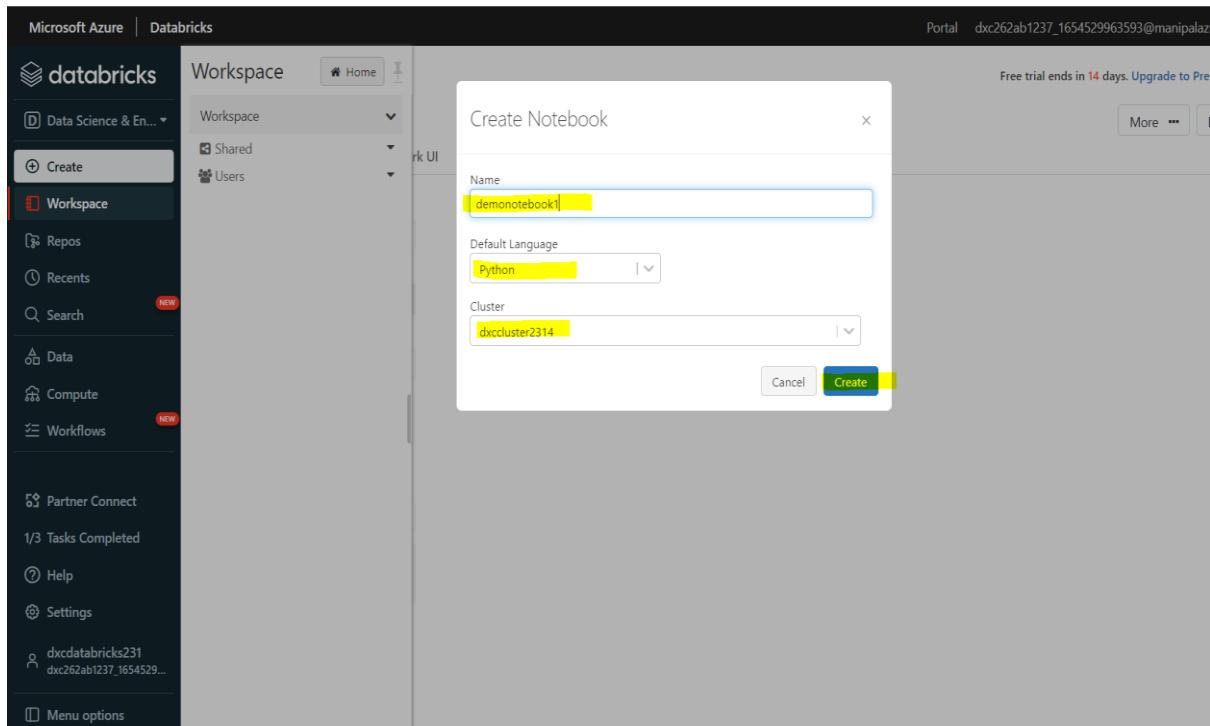
STEP 1:- AFTER CREATING CLICK ON LAUNCH WORKSPACE.



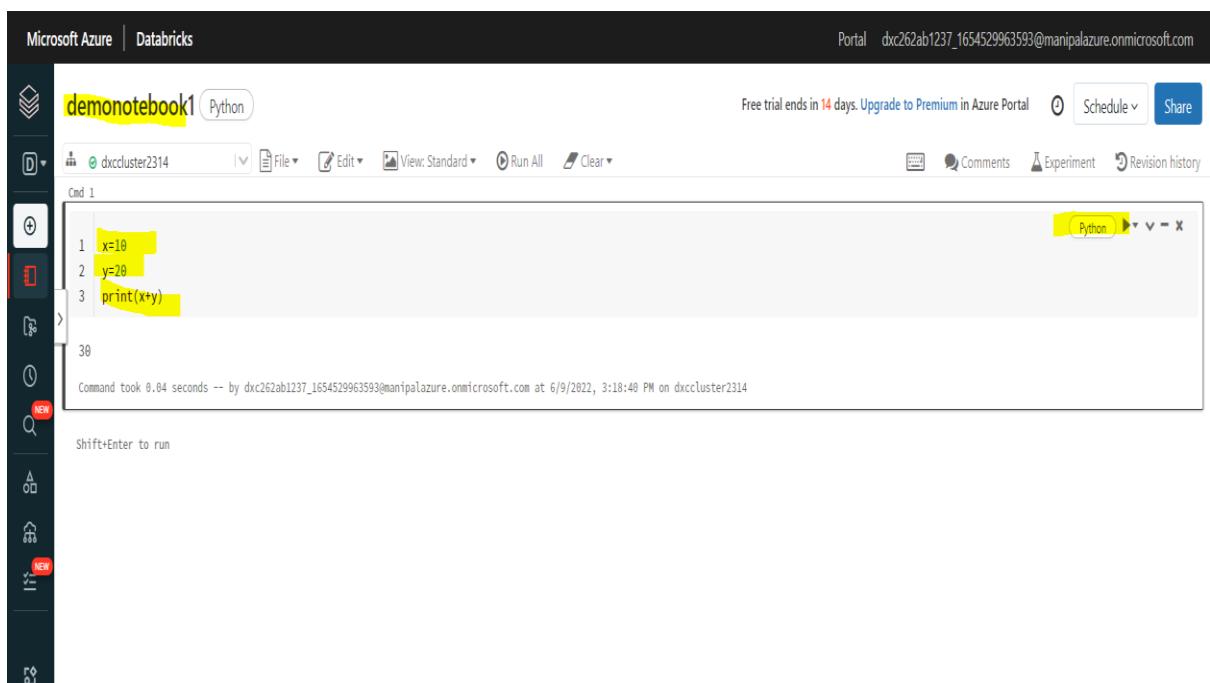
STEP 2:- CLICK ON WORKSPACE, THEN CLICK ON DROPDOWN MENU THEN CLICK ON CREATE AND SELECT NOTEBOOK.



STEP 3:- IN CREATE NOTEBOOK GIVE NAME, DEFAULT LANGUAGE AND CLUSTER NAME THEN CLICK ON CREATE.

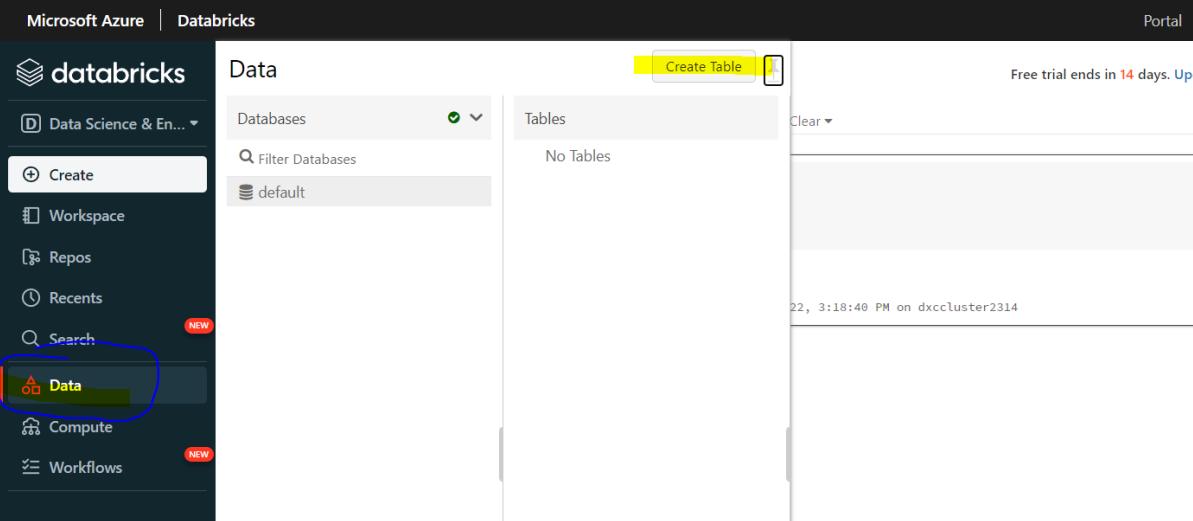


STEP 4:- THEN NOTEBOOK IS CREATED.



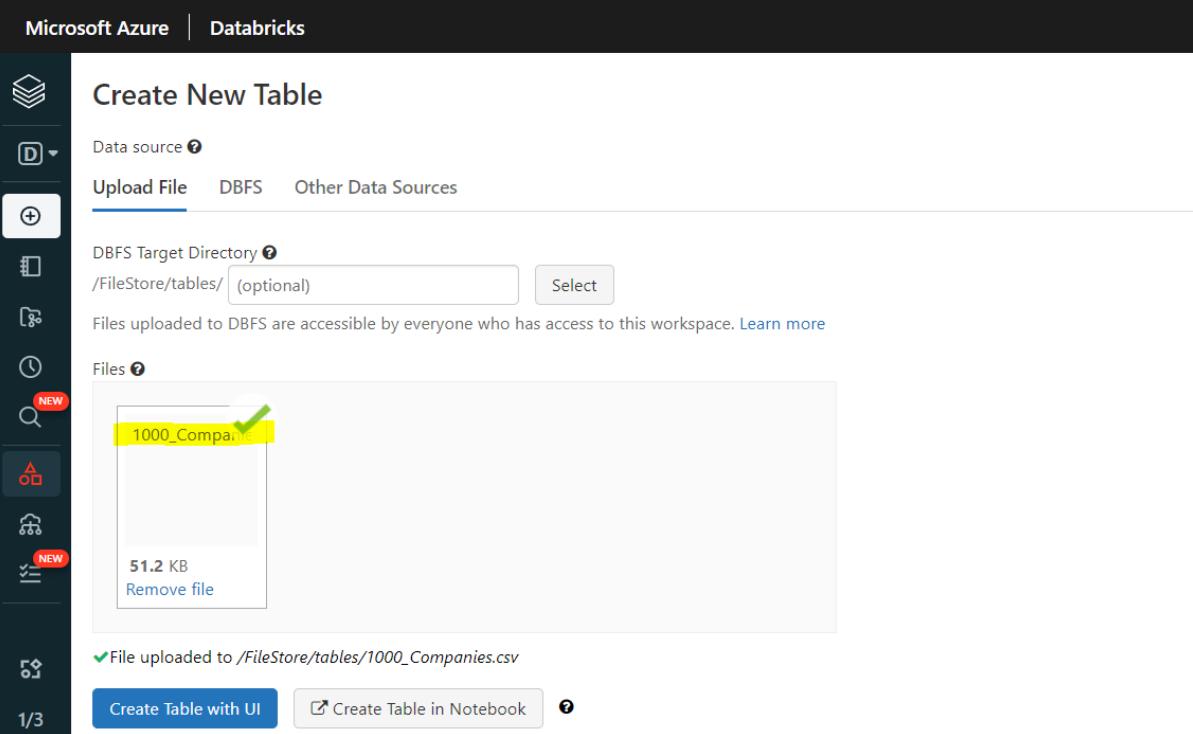
STEPS TO Implement the Business Logic AND INSERT DATA INTO THE DATABRICKS NOTEBOOK AND DISPLAY RESULTS:

STEP 5:- GO TO DATA, SELECT DATABASES THEN TABLES AND CLICK ON CREATE TABLE.



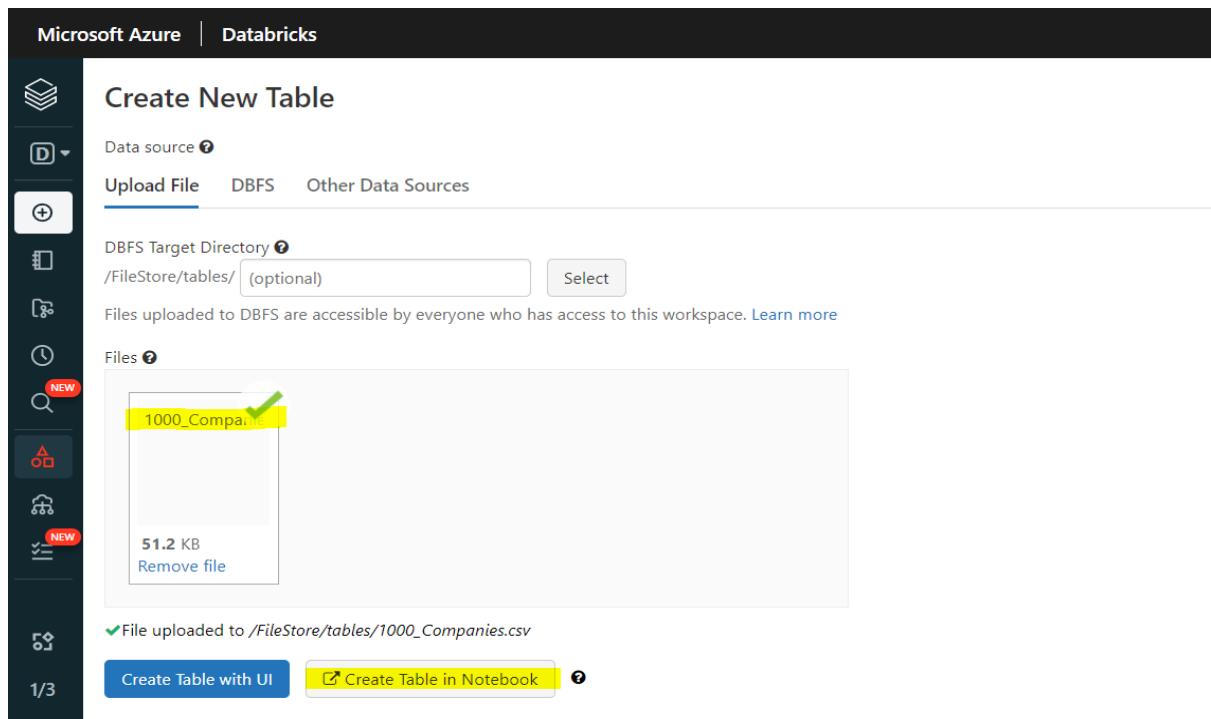
The screenshot shows the Microsoft Azure Databricks interface. On the left sidebar, the 'Data' option is highlighted with a blue oval. The main panel is titled 'Data' and shows 'Databases' selected. A 'Create Table' button is visible at the top right. The 'Tables' section shows 'No Tables' and a 'default' database. A message at the top right indicates a 'Free trial ends in 14 days. Upgrade'.

STEP 6:- IN UPLOAD FILES, SELECT FILES FROM THE DIRECTORY.



The screenshot shows the 'Create New Table' dialog in Databricks. The 'Upload File' tab is selected. A file named '1000_Companies.csv' is listed in the 'DBFS Target Directory' field. Below it, a note says 'Files uploaded to DBFS are accessible by everyone who has access to this workspace. Learn more'. At the bottom, a message confirms '✓ File uploaded to /FileStore/tables/1000_Companies.csv'. There are two buttons at the bottom: 'Create Table with UI' and 'Create Table in Notebook'.

STEP 7:- CLICK ON CREATE TABLE IN NOTEBOOK.



The screenshot shows the 'Create New Table' interface in Microsoft Azure Databricks. At the top, there's a navigation bar with 'Microsoft Azure' and 'Databricks'. Below it, a sidebar contains icons for file operations like upload, download, and search. The main area is titled 'Create New Table' and has tabs for 'Data source', 'Upload File', 'DBFS', and 'Other Data Sources'. Under 'Upload File', a file named '1000_Compan... (optional)' is listed with a size of '51.2 KB'. A 'Select' button is next to the file path. Below the file list, a note says 'Files uploaded to DBFS are accessible by everyone who has access to this workspace. Learn more'. A 'Files' section shows the uploaded file. At the bottom, there are two buttons: 'Create Table with UI' and 'Create Table in Notebook' (which is highlighted with a yellow box).

STEP 8:-YOU CAN SEE ALL THE DATA OF TABLE.



The screenshot shows a Databricks notebook titled '2022-06-09 - DBFS Example' in Python. The sidebar on the left includes icons for file operations and a search function. The notebook interface has a header with tabs for 'Detached' and 'Cmd 3'. Below the header, three code cells are visible:

- Cmd 3:**

```
1 # Create a view or table
2
3 temp_table_name = "Companies_csv"
4
5 df.createOrReplaceTempView(temp_table_name)
```
- Cmd 4:**

```
1 %sql
2
3 /* Query the created temp table in a SQL cell */
4
5 select * from `Companies_csv`
```
- Cmd 5:**

```
1 # With this registered as a temp view, it will only be available to this particular notebook. If you'
2 # table from the DataFrame.
3 # Once saved, this table will persist across cluster restarts as well as allow various users across d
4 # To do so, choose your table name and uncomment the bottom line.
5
6 permanent_table_name = "Companies_csv"
7
8 # df.write.format("parquet").saveAsTable(permanent_table_name)
```

STEP 9:- CLICK ON DROP DOWN MENU AND SELECT THE CLUSTER.

The screenshot shows a Microsoft Azure Databricks notebook titled "2022-06-09 - DBFS Example" in Python. The interface includes a left sidebar with various icons and a main workspace. At the top right, there is a message "Free trial ends in 14 days". A yellow box highlights the "Run All" button in the top navigation bar. The workspace contains an "Overview" section with text about creating and querying a DataFrame from DBFS, and a code cell (Cmd 2) containing Python code for reading a CSV file.

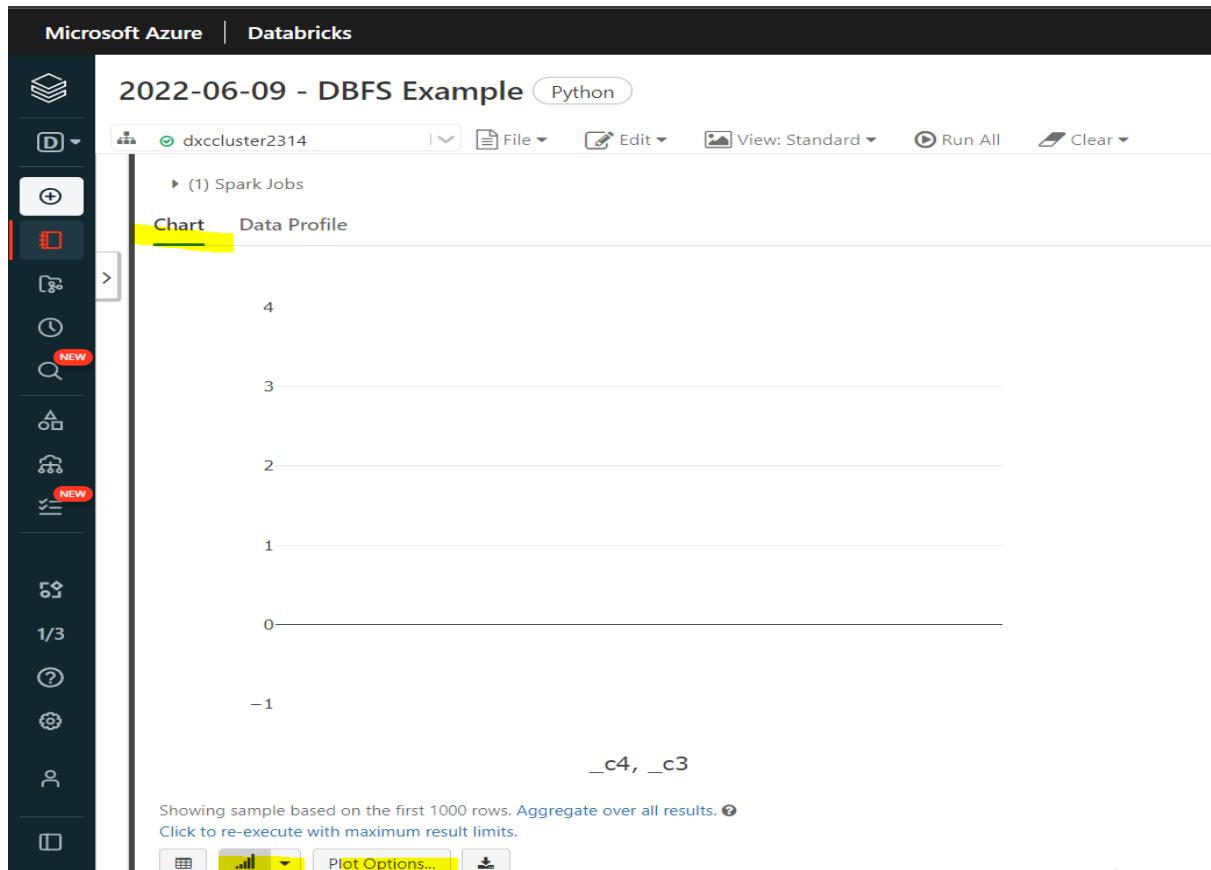
```
1 # File location and type
2 file_location = "/FileStore/tables/1000_Companies.csv"
3 file_type = "csv"
4
5 # CSV options
6 infer_schema = "false"
7 first_row_is_header = "false"
8 delimiter = ","
9
10 # The applied options are for CSV files. For other file types, these will be ignored.
11 df = spark.read.format(file_type) \
12     .option("inferSchema", infer_schema) \
13     .option("header", first_row_is_header) \
14     .option("sep", delimiter) \
15     .load(file_location)
```

STEP 10:- AFTER SELECTING CLUSTER CLICK ON RUN ALL COMMAND.

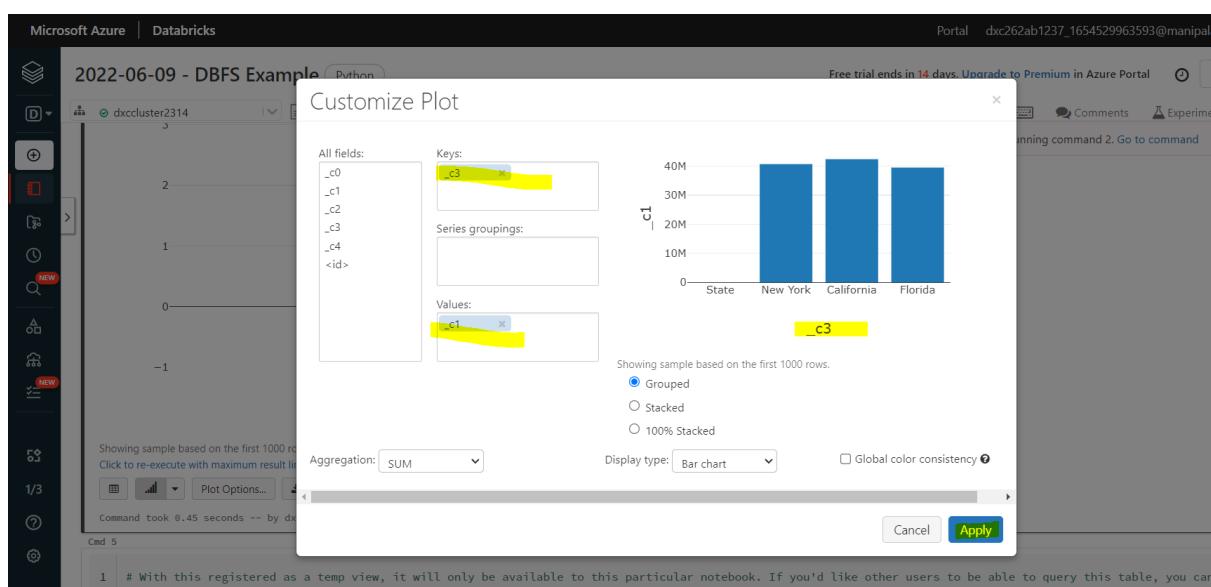
The screenshot shows the same Microsoft Azure Databricks notebook interface as before, but now the "Attached cluster" dropdown menu is open. A yellow box highlights the "Attached cluster:" label and the dropdown menu itself. The menu lists "dxccluster2314" with details: 14.00 GB | 4 Cores | DBR 10.4 LTS | Spark 3.2.1 | Scala 2.12. Other options in the menu include Detach, Restart Cluster, Detach & Re-attach, Spark UI, Driver logs, and Terminal. The workspace content remains the same as in Step 9.

```
1 # File location and type
2 file_location = "/FileStore/tables/1000_Companies.csv"
3 file_type = "csv"
4
5 # CSV options
6 infer_schema = "false"
7 first_row_is_header = "false"
8 delimiter = ","
9
10 # The applied options are for CSV files. For other file types, these will be ignored.
11 df = spark.read.format(file_type) \
12     .option("inferSchema", infer_schema) \
13     .option("header", first_row_is_header) \
14     .option("sep", delimiter) \
15     .load(file_location)
```

STEP 11:- HERE YOU CAN SEE THE DATA IN CHART FORMAT. CLICK ON PLOT OPTION.



STEP 12:- HERE YOU CAN CUSTOMIZE THE PLOT.



HERE IS ANOTHER METHOD

STEP 13:-IN CREATE NEW TABLE CLICK ON UPLOAD FILE AND UPLOAD THE FILE. AND THEN CLICK ON CREATE TABLE WITH UI. NOW CLICK ON PREVIEW TABLE.

The screenshot shows the 'Create New Table' page in the Databricks UI. On the left is a sidebar with various icons. The main area has a title 'Create New Table' with a yellow background. Below it, a 'Data source' section shows 'Upload File' (selected), 'DBFS', and 'Other Data Sources'. A 'DBFS Target Directory' input field contains '/FileStore/tables/' with '(optional)' text and a 'Select' button. A note says 'Files uploaded to DBFS are accessible by everyone who has access to this workspace.' Below this is a 'Files' section showing a file named '1000_Compa...' (partially visible) with a checkmark, circled in blue. It also shows '51.2 KB' and a 'Remove file' link. A message at the bottom says '✓ File uploaded to /FileStore/tables/1000_Companies-1.csv'. There are two buttons: 'Create Table with UI' (highlighted with a yellow background) and 'Create Table in Notebook'. Below these is a section titled 'Select a Cluster to Preview the Table' with a note 'Choose a cluster with which you will read and preview the data.' A 'Cluster' dropdown menu shows 'dxccluster2314' with a dropdown arrow. A 'Preview Table' button is at the bottom of this section.

STEP 14:-TABLE IS CREATED.

Microsoft Azure | Databricks

Create New Table

Preview Table

Specify Table Attributes

Table Name: companies_1_csv

Create in Database: default

File Type: CSV

Column Delimiter: ,

First row is header

Infer schema

Multi-line

Table Preview:

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
142107.34	91391.77	366168.42	Florida	166187.94
131876.9	99814.71	362861.36	New York	156991.12

Create Table

Microsoft Azure | Databricks

default.companies_1_csv | Refresh

dxcluster2314

Details History

Description:
Created at: 2022-06-09 10:23:09
Last modified: 2022-06-09 10:23:21
Partition columns:
Number of files: 1
Size: 40 kB

Schema:

col_name	data_type	comment
1	string	R_D_Spend
2	string	Administration
3	string	Marketing_Spend
4	string	State
5	string	Profit
6		# Partitioning
7		

Showing all 8 rows.

Sample Data:

	R_D_Spend	Administration	Marketing_Spend	State	Profit
1	165349.2	136897.8	471784.1	New York	192261.83
2	162597.7	151377.59	443898.53	California	191792.06
3	153441.51	101145.55	407934.54	Florida	191050.39
4	144372.41	118671.85	383199.62	New York	182901.99
5	142107.34	91391.77	366168.42	Florida	166187.94
6	131876.9	99814.71	362861.36	New York	156991.12
7	134615.46	147198.87	127716.82	California	156122.51

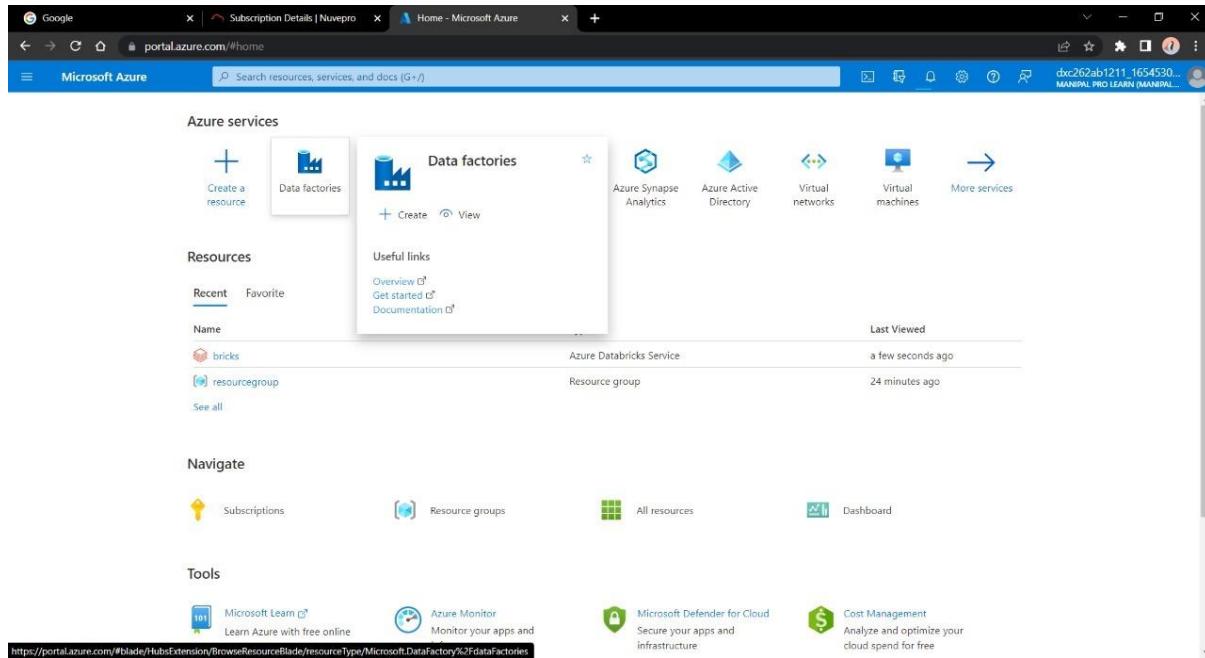
Showing all 1000 rows.

NOW OUR NOTEBOOK IS READY TO BE LINKED AND EXECUTED IN AZURE DATA FACTORY.

Practical Lab: Azure Data Factory for AP Morgan

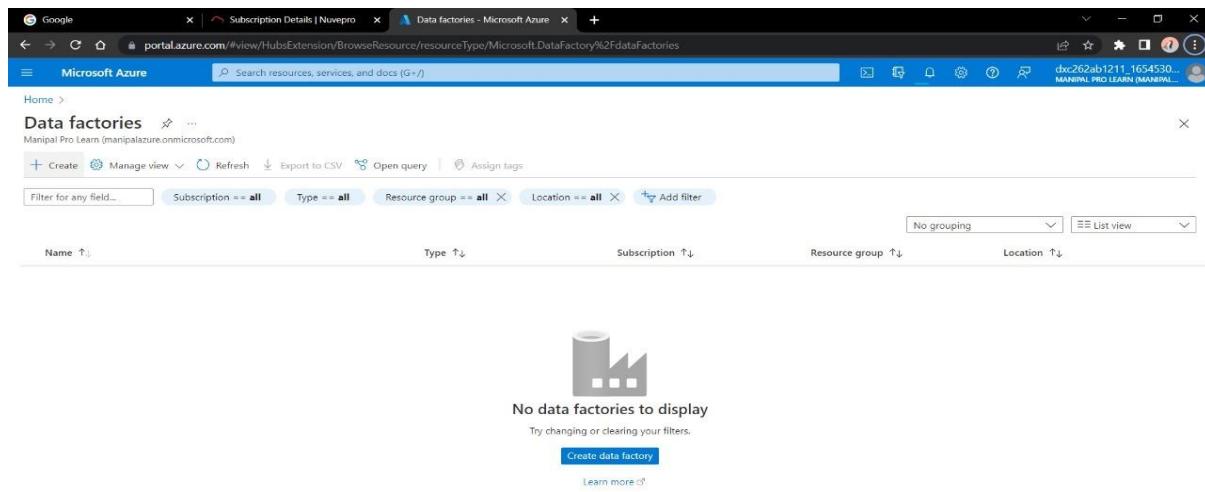
STEPS TO CREATE AZURE DATA FACTORY FOR AP MORGAN.

STEP 1:- GO TO AZURE HOME PAGE AND SEARCH FOR DATA FACTORY.



The screenshot shows the Microsoft Azure Home page. The 'Data factories' service is highlighted with a blue box. The left sidebar includes sections for 'Create a resource', 'Data factories', 'Resources' (Recent and Favorite), 'Useful links' (Overview, Get started, Documentation), 'Last Viewed' (Azure Databricks Service, Resource group), 'Navigate' (Subscriptions, Resource groups, All resources, Dashboard), and 'Tools' (Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, Cost Management). The URL in the address bar is <https://portal.azure.com/#blade/HubsExtension/BrowseResourceBlade/resourceType/Microsoft.DataFactory%2fdataFactories>.

STEPS 2: CREATE A NEW ACCOUNT ON AZURE DATA FACTORY.FOR THAT CLICK ON CREATE.



The screenshot shows the 'Data factories' blade. It displays a message 'No data factories to display' with the text 'Try changing or clearing your filters.' Below this are two buttons: 'Create data factory' and 'Learn more'. The top navigation bar shows the URL <https://portal.azure.com/#view/HubsExtension/BrowseResourceBlade/resourceType/Microsoft.DataFactory%2fdataFactories>.

[Give feedback](#)

STEP 3: Enter the details for the factory account. And fill all the project details.

The screenshot shows the Microsoft Azure portal with the URL portal.azure.com/#create/Microsoft.DataFactory. The page title is "Create Data Factory". The "Basics" tab is selected. The "Project details" section asks to select a subscription and resource group. The "Subscription" dropdown is set to "Azure-DXC262AB12Lab" and the "Resource group" dropdown is set to "resourcegroup". The "Instance details" section shows the "Name" as "AP-morgan", "Region" as "East US", and "Version" as "V2 (Recommended)". At the bottom, there are buttons for "Review + create", "< Previous", and "Next : Git configuration >".

STEP 4: VALIDATE AND DEPLOY THE DATA FACTORY. AND VALIDATING IS COMPLETED CLICK ON CREATE.

The screenshot shows the "Create Data Factory" wizard on the "Review + create" step. A message box indicates "Submitting deployment..." and "Submitting the deployment template for resource group 'resourcegroup'". The "Validation Passed" status is shown. The "Review + create" tab is selected. The "TERMS" section contains legal agreement text. The "Basics" section lists the selected values: Subscription "Azure-DXC262AB12Lab", Resource group "resourcegroup", Name "AP-morgan", Region "East US", and Version "V2 (Recommended)". The "Networking" section shows "Connect via" as "Public endpoint". At the bottom, there are buttons for "Create", "< Previous", "Next", and "Download a template for automation".

STEP 5: After the deployment is done, Navigate to “Go to resource”

The screenshot shows the Microsoft Azure portal with the URL <https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2fsubscriptions%2f4236c42a-d131-4bd6-b609-aec3a598f2d3%2fresourceGroups%2fresourcegroup%2fproviders%2fMicrosoft.DataFactory%2fFactories%2fAP-morgan>. The page title is "Microsoft.DataFactory-2022061121708 | Overview". The main content area displays a green checkmark icon and the message "Your deployment is complete". It shows deployment details: Deployment name: Microsoft.DataFactory-2022061121708, Subscription: Azure-DXC262AB12Lab, Resource group: resourcegroup. The start time is 6/11/2022, 10:18:22 PM, and the Correlation ID is aee1bae-0991-4da2-be42-e577d473e7dd. A "Go to resource" button is visible. To the right, there are promotional cards for Cost Management, Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

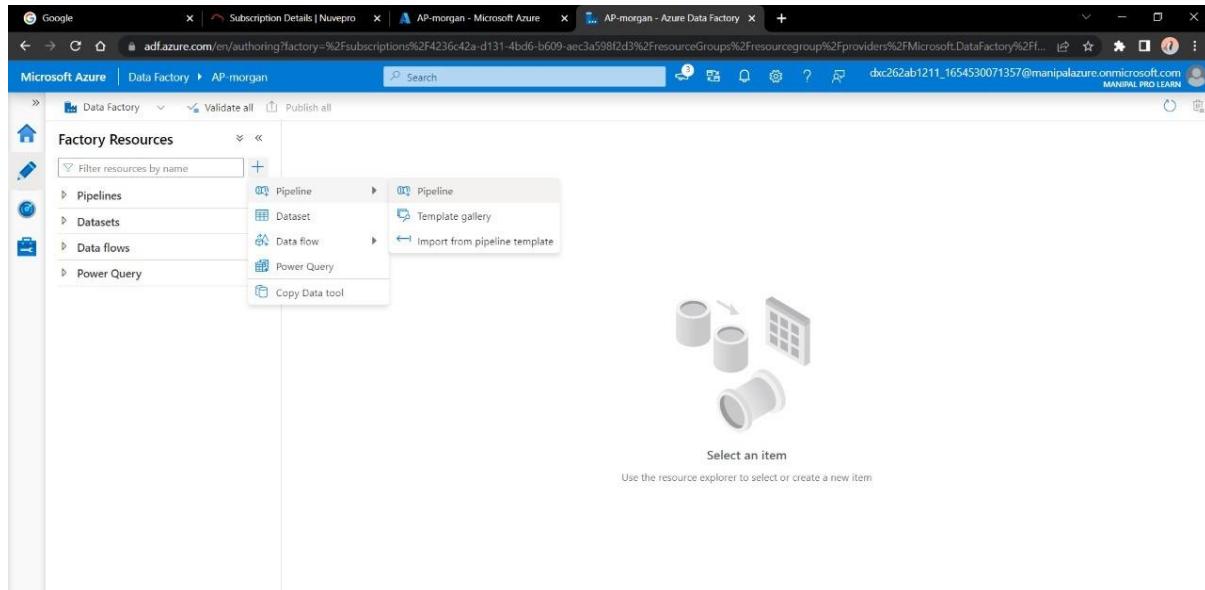
STEP 6: NOW OPEN THE DATA FACTORY STUDIO.

The screenshot shows the Microsoft Azure portal with the URL https://adf.azure.com/en/home?factory=%2fsubscriptions%2f4236c42a-d131-4bd6-b609-aec3a598f2d3%2fresourceGroups%2fresourcegroup%2fproviders%2fMicrosoft.DataFactory%2fFactories%2fAP-morgan#loginHint=dxc262ab1211_1654530071357@manipalazure.onmicrosoft.com. The page title is "AP-morgan - Microsoft Azure". The main content area shows the Data Factory studio interface with sections for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Networking, Properties, Locks, Quick start, Alerts, Metrics, Diagnostic settings, and Logs. It also includes Getting started, Monitoring, and PipelineRuns/ActivityRuns sections.

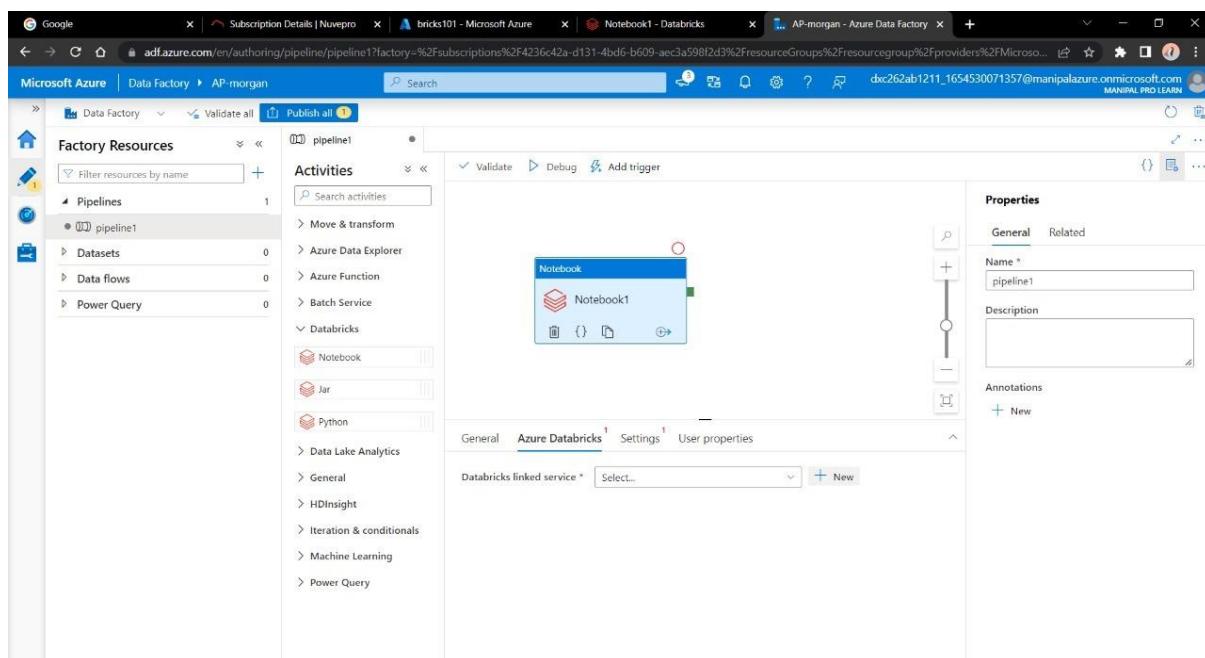
Practical Lab: Create Azure Databricks Linked Service in ADF.

WE NEED TO CREATE A PIPELINE THAT WILL CONNECT DATA BRICKS NOTEBOOK WITH DATA FACTORY.

STEP 1: CLICK ON THE ADD SIGN, GO TO PIPELINE AND SEECT PIPELINE.



STEP 2: Create a new Pipeline and drag and drop the notebook tab from Azure Databricks dropdown into the pipeline workspace.



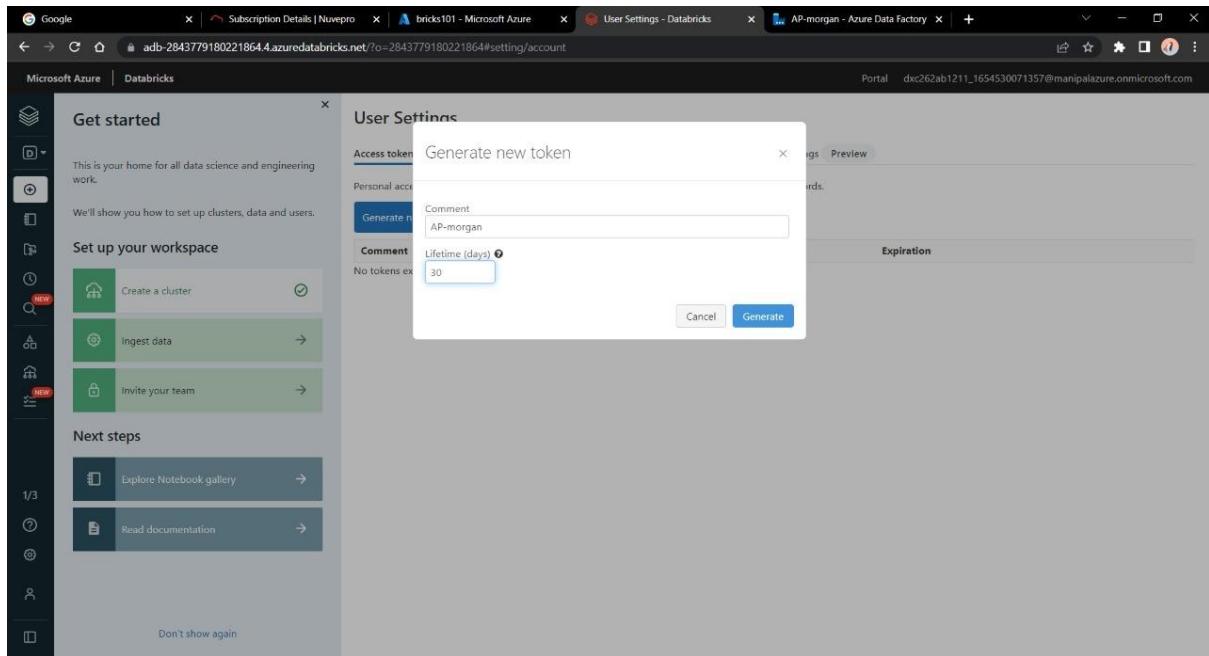
Step 3: Create a new Linked service for databricks.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a sidebar with 'Factory Resources' like Pipelines, Datasets, Data flows, and Power Query. In the center, under 'Activities', there's a 'Notebook' section. On the right, a modal window titled 'New linked service' is open, specifically for 'Azure Databricks'. It shows the configuration for creating a 'Databricks linked service'. The 'Account selection method' is set to 'From Azure subscription'. The 'Azure subscription' dropdown is set to 'Azure-DXC262AB12Lab (4236c42a-d131-4bd6-b609-aec3a598f2d3)'. The 'Databricks workspace' dropdown is set to 'bricks101'. Under 'Authentication type', 'Access Token' is selected, and the 'Access token' field contains a placeholder 'Access token'. There are tabs for 'General', 'Azure Databricks', 'Settings', and 'User properties'. At the bottom of the modal are 'Create' and 'Cancel' buttons, and a 'Test connection' link.

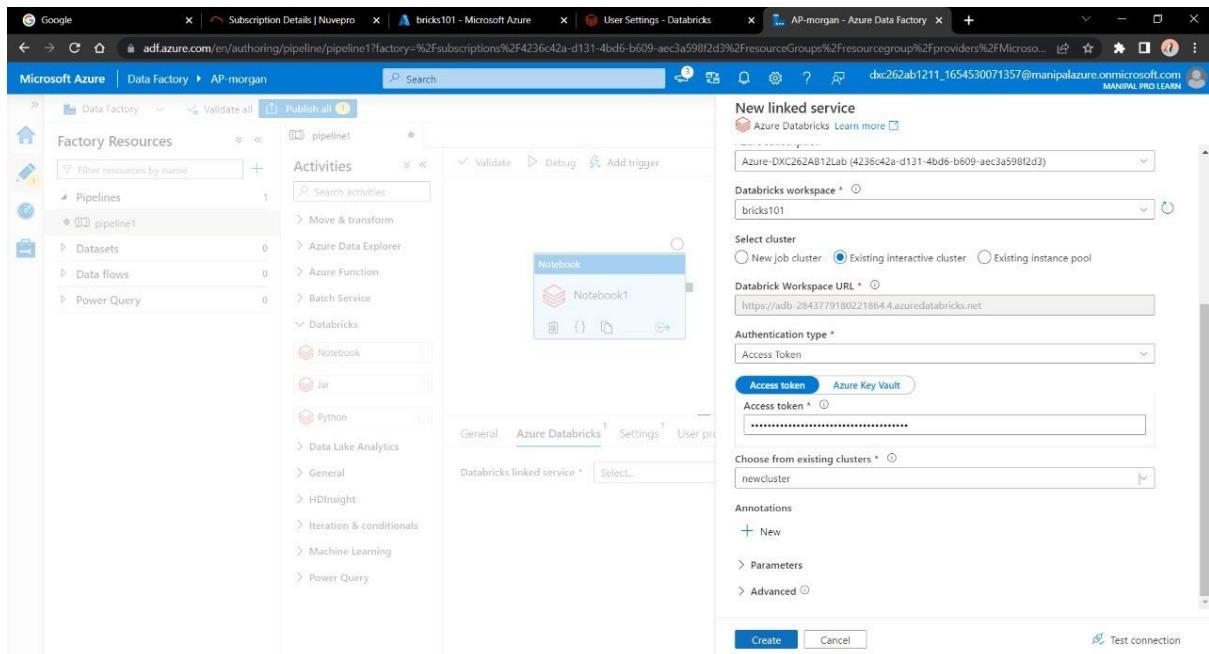
Step 4: Now, we need the access token of Data bricks account in order to access it. Navigate to Data bricks workspace and go to Users in settings.

The screenshot shows the Azure Databricks workspace interface. On the left, there's a sidebar with 'Data Science & Engineering', 'Create', 'Workspace', 'Repos', 'Recents', 'Search', 'Data', 'Compute', and 'Workflows'. A 'User Settings' dropdown is open, showing 'Admin Console' and 'Manage Account'. The main area is a 'Notebook1' tab with a Python kernel. The notebook contains a single command: 'print("Do some computation on data")'. The output shows the command ran successfully in 'newcluster' with a duration of 0.04 seconds. The URL of the browser is 'adb-2843779180221864.4.azuredatabricks.net/?o=2843779180221864#notebook/4202547851666467/command/4202547851666468'.

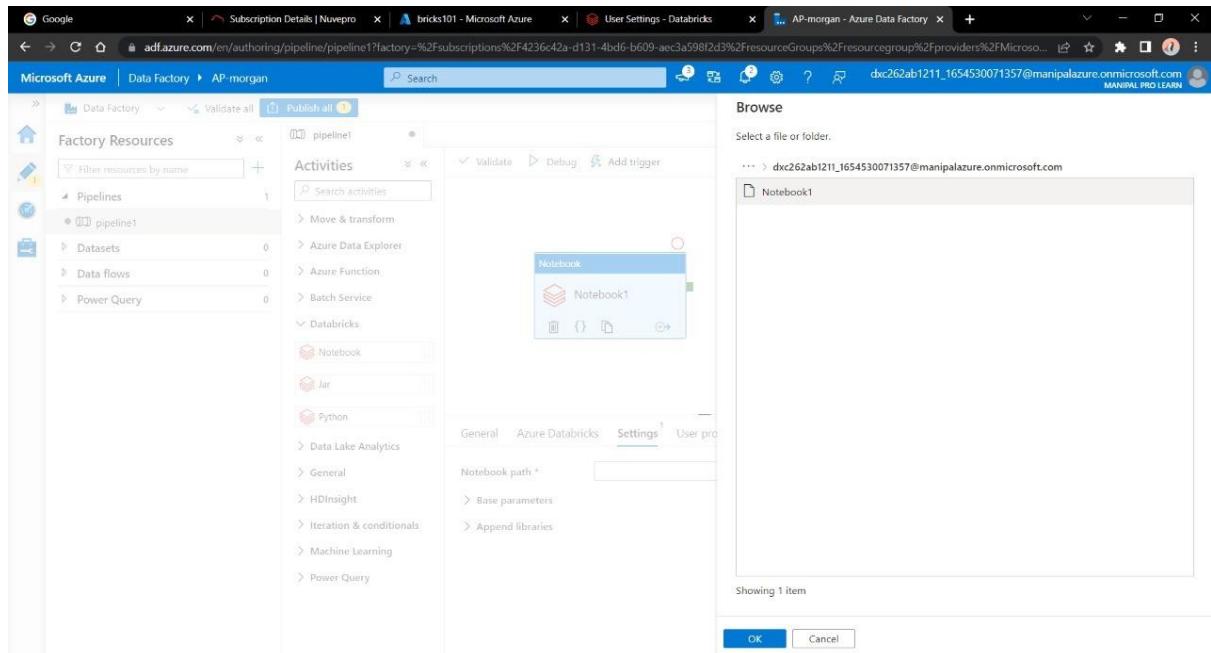
Step 5: Click on “Generate token” and mention a small quick description name for the token.



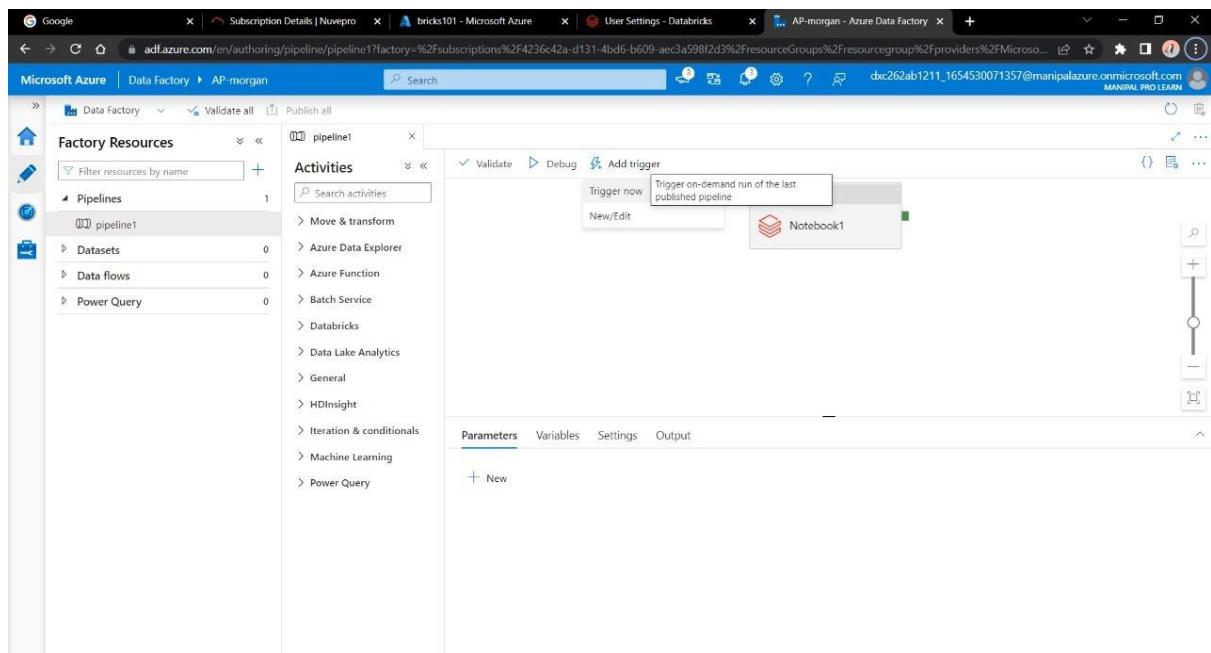
Step 6: After that a tab will open containing your access token, copy it and paste it in the azure factory data connection form.



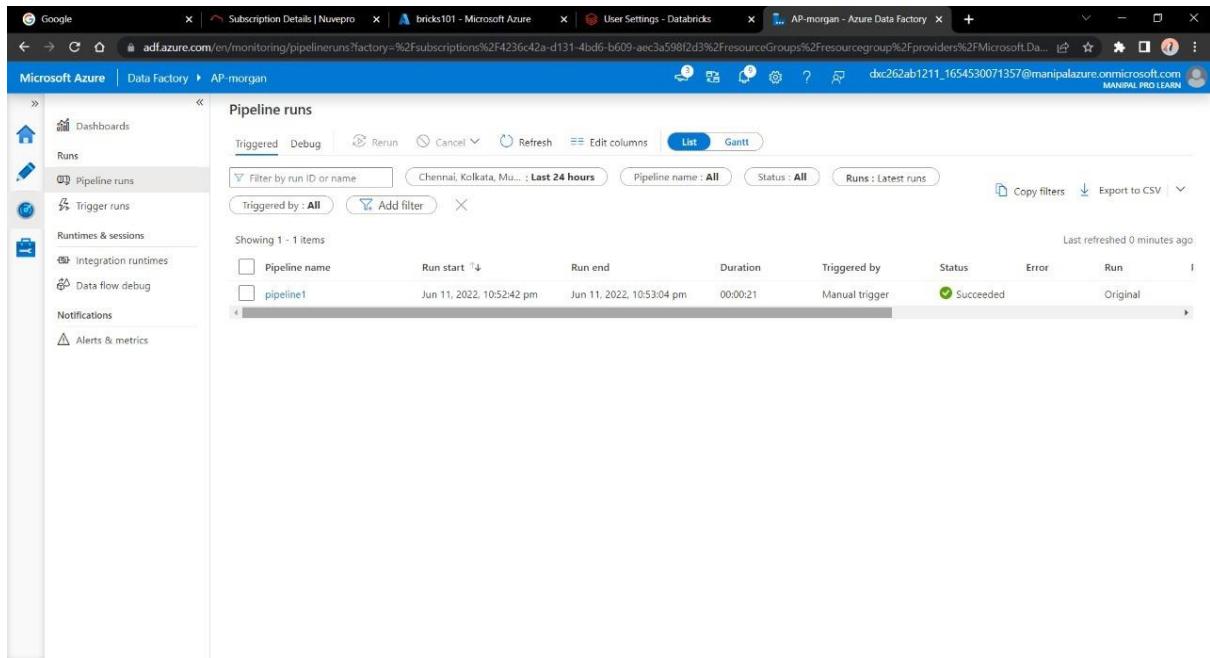
Step 7: Choose the right cluster and click on create.



Now choose the notebook in the pipeline notebook tab setting in order to trigger it. Publish the pipeline in order to trigger it, after publication is done click on “Trigger now”.



Step 8: Go to monitor window in order to check it's execution.



The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has a 'Pipeline runs' option selected. The main area is titled 'Pipeline runs' with tabs for 'List' and 'Gantt'. It displays a table of pipeline runs. The table includes columns for Pipeline name, Run start, Run end, Duration, Triggered by, Status, Error, and Run. One row is visible, showing 'pipeline1' with a run starting at Jun 11, 2022, 10:52:42 pm, ending at Jun 11, 2022, 10:53:04 pm, duration 00:00:21, triggered by 'Manual trigger', status 'Succeeded', and error 'None'. The status bar at the bottom right indicates 'Last refreshed 0 minutes ago'.

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error	Run
pipeline1	Jun 11, 2022, 10:52:42 pm	Jun 11, 2022, 10:53:04 pm	00:00:21	Manual trigger	Succeeded	None	Original

Here we can see we have successfully triggered a linked notebook of Data bricks from Data Factory.

Result: We had successfully linked and trigger Azure Databricks notebook using Data Factory.

Conclusion: Azure Data Factory linked with Azure Data bricks.