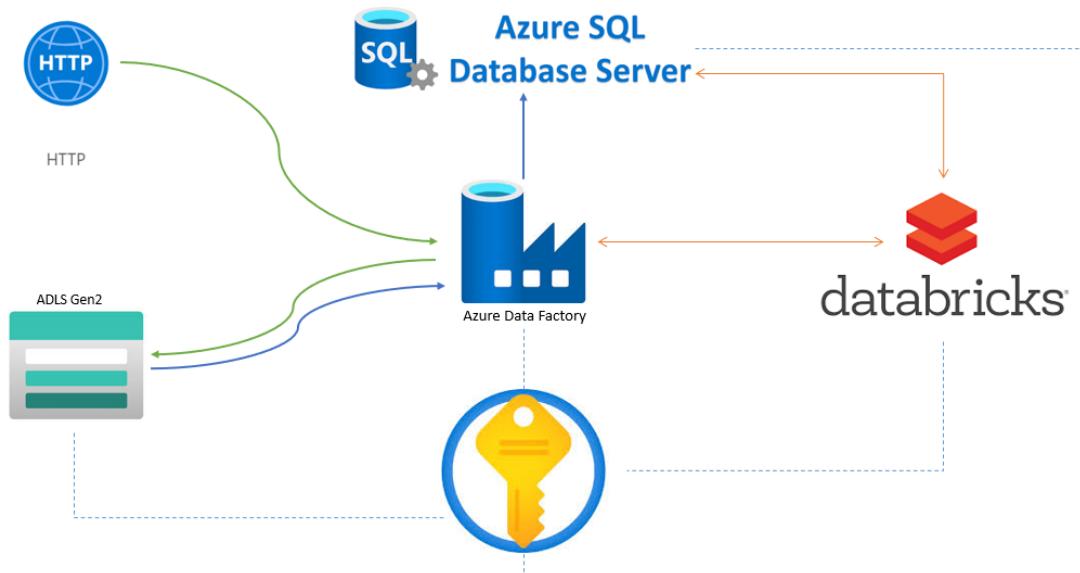


Capstone Project

Use Case:

1. Get Orders and Customer data from multiple sources and perform transformations.
Write the result data to the Azure SQL DB.
2. Secure all the credentials with Azure Key Vault.
3. Create a trigger event with type as “Storage Event” , So that pipeline will run automatically once we got the orders data in the container.
4. Get customer data from the external source link every time the pipeline runs.
5. Write back the transferred data to the Azure SQL DB, for further analysis.



Transformation Required

~Get the Total number of orders in each state.

Resources Required

=====

- >Storage Account
- >Azure Databricks
- >Azure DataFactory
- >Azure SQL DataBase
- >Azure Key Vault

Create a resource group ...

Basics Tags Review + create

Resource group - A container that holds related resources for an Azure solution. The resource group can include all the resources for the solution, or only those resources that you want to manage as a group. You decide how you want to allocate resources to resource groups based on what makes the most sense for your organization. [Learn more](#)

Project details

Subscription * ⓘ

Free Trial

Resource group * ⓘ

capstone-proj ✓

Resource details

Region * ⓘ

(Asia Pacific) Central India

[Review + create](#)

[< Previous](#)

[Next : Tags >](#)

Pin it to the Our own dashboard

The screenshot shows the Microsoft Azure dashboard with the 'Capstone' private dashboard selected. A 'Pin to dashboard' modal is open, allowing the user to pin the 'Capstone' dashboard to their own dashboard. The 'Existing' tab is selected, and the 'Type' dropdown is set to 'Private'. The 'Dashboard' dropdown also shows 'Capstone'. Below the modal, there is a 'Pin' button and a 'Cancel' button. To the right of the modal, the 'Notifications' panel is visible, displaying recent events: 'Resource group created' for 'capstone-proj' in 'Free Trial', 'Deleted resource group NetworkWatcherRG', and 'Deleted resource group databricks-rg-'.

Create an Storage account

The screenshot shows the 'Create a storage account' wizard. In the 'Instance details' section, the 'Storage account name' is set to 'capstonestorageacct', the 'Region' is '(Asia Pacific) Central India', and the 'Primary service' dropdown shows 'Select a primary service'. Under 'Performance', 'Standard' is selected as the recommended option for most scenarios. In the 'Redundancy' section, 'Locally-redundant storage (LRS)' is chosen. At the bottom of the screen, there are navigation buttons for 'Previous', 'Next', and 'Review + create', along with a 'Give feedback' link.

Make sure Hierarchical namespace enabled . -> to create it as a ADLS

The screenshot shows the 'Create a storage account' wizard. In the 'Basics' section, the subscription is 'Free Trial', resource group is 'capstone-proj', location is 'Central India', storage account name is 'capstonestorageacct', primary service is 'Standard', and replication is 'Locally-redundant storage (LRS)'. In the 'Advanced' section, 'Enable hierarchical namespace' is set to 'Enabled'. Other options like 'Enable SFTP' and 'Enable network file system v3' are disabled. At the bottom, there are 'Previous', 'Next', and 'Create' buttons, along with a 'Give feedback' link.

Create a Data Factory to run our pipeline.

The screenshot shows the 'Create Data Factory' wizard. Under 'Project details', the subscription is 'Free Trial' and the resource group is 'capstone-proj'. Under 'Instance details', the name is 'capstoneprojectdatafac', region is 'Central India', and version is 'V2'. At the bottom, there are 'Previous', 'Next', and 'Review + create' buttons, along with a 'Give feedback' link.

Create an Azure SQL DB

The screenshot shows the 'Create SQL Database' wizard. Under 'Database details', the database name is 'capstoneDB' and the server is 'Select a server'. At the bottom, there are 'Review + create' and 'Next : Networking >' buttons.

Create a new SQL server and attach it to our DB

The screenshot shows the 'Create SQL Database Server' step in the Azure portal. It's titled 'Create SQL Database Server'. Under 'Server details', the 'Server name' is set to 'capstonedbserver' and the 'Location' is '(Asia Pacific) Central India'. A note at the bottom says 'Azure Active Directory (Azure AD) is now Microsoft Entra ID. [Learn more](#)'. An 'OK' button is visible.

The screenshot continues the 'Create SQL Database Server' process. It asks for preferred authentication methods, with 'Use SQL authentication' selected. It also requires entering a 'Server admin login' ('capstoneProj') and 'Password' ('*****'). An 'OK' button is present.

The screenshot shows the 'Create SQL Database' step. It starts with 'Compute + storage' set to 'Basic' (1 GB storage). The 'Configure database' link is highlighted. In the 'Backup storage redundancy' section, 'Locally-redundant backup storage' is selected. Navigation buttons 'Review + create' and 'Next : Networking >' are at the bottom.

I have created an minimum level of resource , for our use case

The screenshot shows the Microsoft Azure portal interface for configuring a SQL database. At the top, there's a navigation bar with 'Microsoft Azure', 'Upgrade', a search bar, and user information. Below it, the URL 'Home > SQL databases >' and a 'Configure' button are visible. A feedback icon is present. The main area displays a 'SQL' icon and a 'Cost summary' section. The 'Service tier' is set to 'Basic (For less demanding workloads)'. The 'DTUs' selected are '5 (Basic)'. The 'Data max size (GB)' slider is at '1'. A 'Cost summary' table shows: Cost per DTU (in INR) 93.18, DTUs selected x 5, and ESTIMATED COST / MONTH 465.92 INR. An 'Apply' button is at the bottom.

Create an Databricks WorksSpace in Azure.

The screenshot shows the Microsoft Azure portal interface for creating a Databricks workspace. At the top, there's a navigation bar with 'Microsoft Azure', 'Upgrade', a search bar, and user information. Below it, the URL 'Dashboard > Azure Databricks >' and a 'Create an Azure Databricks workspace' button are visible. A 'Subscription *' dropdown is set to 'Free Trial' and a 'Resource group *' dropdown is set to 'capstone-proj'. A 'Create new' link is also visible. The 'Instance Details' section includes fields for 'Workspace name *' (set to 'capstone_Databricks'), 'Region *' (set to 'Central India'), and 'Pricing Tier *' (set to 'Premium (+ Role-based access controls)'). A note indicates that the recommended pricing tier was selected. Navigation buttons at the bottom include 'Review + create', '< Previous', and 'Next : Networking >'.

Home > Key vaults >

Create a key vault

The screenshot shows the Microsoft Azure portal interface for creating a key vault. At the top, there's a navigation bar with 'Microsoft Azure', 'Upgrade', a search bar, and user information. Below it, the URL 'Home > Key vaults >' and a 'Create a key vault' button are visible. A 'Project details' section notes that the subscription manages deployed resources and uses resource groups. A 'Subscription *' dropdown is set to 'Free Trial' and a 'Resource group *' dropdown is set to 'capstone-proj'. The 'Instance details' section includes fields for 'Key vault name *' (set to 'capstonevaultn'), 'Region *' (set to 'Central India'), and 'Pricing tier *' (set to 'Standard'). Navigation buttons at the bottom include 'Previous', 'Next', and 'Review + create'.

Below are all the required resources for our project

The screenshot shows the Microsoft Azure dashboard for the 'Capstone' project. The left sidebar lists resources under 'capstone-proj': capstoneprojectdatafac (Data factory V2), capstonestorageacct (Storage account), capstoneDB (SQL database), capstoneDBserver (SQL server), capstone_Databricks (Azure Databricks Service), and capstonevault (Key vault). The right side displays two deployment slots: 'capstonestorageacct_1' and 'Microsoft.DataFactory-'. Both slots show a green status icon with an upward arrow, indicating they are active or deploying.

Create an cluster with minimum requirement inside the databricks

The screenshot shows the Databricks Compute configuration page. The 'Compute' tab is selected. A new cluster named 'Nadarajan S's Cluster' is being created. The configuration includes: Access mode set to 'Single user access' (Single user), Performance settings (Runtime: 15.4 LTS (Scala 2.12, Spark 3.5.0), Use Photon Acceleration checked, Node type: Standard_D3_v2 (14 GB Memory, 4 Cores), and Terminate after 30 minutes of inactivity checked), and a 'Create compute' button at the bottom.

Set the IP in set server firewall option inside the sql DB. Then only we can able to access the DB.

The screenshot shows the Microsoft Azure SQL Database settings for 'capstoneDB'. The 'Overview' section is visible. Key details include: Resource group: 'capstone-proj', Server name: 'capstonedbserver.database.windows.net', Status: 'Online', Location: 'Central India', Subscription: 'Free Trial', and Subscription ID: 'a376f639-0268-48a9-84e5-2b741923ebe4'. The 'Set server firewall' button is highlighted in yellow.

Microsoft Azure Upgrade Search resources, services, and docs (G+)

Dashboard > capstoneDB (capstonedbserver/capstoneDB) > capstonedbserver

capstonedbserver | Networking

SQL server

Public access Private access Connectivity

Public network access

Public Endpoints allow access to this resource through the internet using a public IP address. An application or resource that is granted access with the following network rules still requires proper authorization to access this resource. [Learn more](#)

Public network access

Disable
 Selected networks

Connections from the IP addresses configured in the Firewall rules section below will have access to this database. By default, no public IP addresses are allowed. [Learn more](#)

Please save public network access value before adding new virtual networks.

Virtual networks

Allow virtual networks to connect to your resource using service endpoints. [Learn more](#)

Save **Discard**

Microsoft Azure Upgrade Search resources, services, and docs (G+)

Dashboard > capstoneDB (capstonedbserver/capstoneDB) > capstonedbserver

capstonedbserver | Networking

SQL server

Search

Overview Activity log Access control (IAM) Tags Quick start Diagnose and solve problems Settings Microsoft Entra ID SQL databases SQL elastic pools Properties Locks

Firewall rules

Allow certain public internet IP addresses to access your resource. [Learn more](#)

+ Add your client IPv4 address (117.193.27.114) + Add a firewall rule

| Rule name | Start IPv4 address | End IPv4 address |
|-----------------------------------|--------------------|------------------|
| ClientIPAddress_2025-1-12_8-30-34 | 117.193.27.114 | 117.193.27.114 |

Exceptions

Allow Azure services and resources to access this server

Save **Discard**

Click the “Allow Azure service ...” check box to give access to other azure services to access this SQL DB.

Microsoft Azure Upgrade Search resources, services, and docs (G+)

Dashboard > capstoneDB (capstonedbserver/capstoneDB) > capstonedbserver

capstonedbserver | Networking

SQL server

Search

Overview Activity log Access control (IAM) Tags Quick start Diagnose and solve problems Settings Microsoft Entra ID SQL databases SQL elastic pools Properties Locks

Firewall rules

Allow certain public internet IP addresses to access your resource. [Learn more](#)

+ Add your client IPv4 address (117.193.27.114) + Add a firewall rule

| Rule name | Start IPv4 address | End IPv4 address |
|-----------------------------------|--------------------|------------------|
| ClientIPAddress_2025-1-12_8-30-34 | 117.193.27.114 | 117.193.27.114 |

Exceptions

Allow Azure services and resources to access this server

Save **Discard**

The screenshot shows the Microsoft Azure SQL Database Query Editor. On the left, a sidebar menu includes 'Overview', 'Activity log', 'Tags', 'Diagnose and solve problems', and 'Query editor (preview)' (which is selected). The main area displays the 'Welcome to SQL Database Query Editor' page with a 'SQL' logo and a login form for 'capstoneProj'. To the right, a 'Notifications' panel lists three recent events: 'Successfully updated server exceptions', 'Successfully updated server firewall rules', and 'Successfully updated server public network access', all occurring 'a few seconds ago'. A 'More events in the activity log' link is at the top of the notifications.

Login to the DB and create orders table.

The screenshot shows the Microsoft Azure SQL Database Query Editor. The 'Query editor (preview)' section is active. In the center, a query window titled 'Query 1' contains the following T-SQL code:

```
1 create table orders(order_id int, order_date varchar(100),
2 customer_id int, order_status varchar(100));
```

The 'Messages' tab at the bottom indicates 'Query succeeded: Affected rows: 0'. The status bar at the bottom of the editor window shows 'Query succeeded | 0s'.

In Storage account give “Anonymous Access for Blob”. With that we can create an external link in our storage account.

The screenshot shows the Microsoft Azure Storage Account settings for 'capstonesorageacct'. The 'Configuration' section is highlighted in the sidebar. The main pane displays the following details:

- Resource group: capstone-proj
- Location: centralindia
- Subscription: Free Trial
- Subscription ID: a376f639-0268-48a9-84e5-2b741923eb4
- Disk state: Available
- Tags: (edit) Add tags

A 'JSON View' button is located in the top right corner of the main pane.

Microsoft Azure Upgrade Search resources, services, and docs (G+/-) Copilot nadarajan073@gmail.co... DEFAULT DIRECTORY

Dashboard > capstonestorageacct

capstonestorageacct | Configuration

Storage account

Save Discard Refresh Give feedback

The cost of your storage account depends on the usage and the options you choose below. [Learn more about storage pricing](#)

Account kind: StorageV2 (general purpose v2)

Performance: Standard

Secure transfer required: Enabled

Allow Blob anonymous access: Enabled

Allow storage account key access: Enabled

Resource sharing (CORS), SFTP, Advisor recommendations, Endpoints

Microsoft Azure Upgrade Search resources, services, and docs (G+/-) Copilot nadarajan073@gmail.co... DEFAULT DIRECTORY

Dashboard >

capstonestorageacct

Storage account

Search

Upload Open in Explorer Delete Move Refresh Open in mobile CLI / PS Feedback JSON View

Overview

Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser, Partner solutions, Data storage, Containers, File shares

Essentials

Resource group: capstone-proj, Location: centralindia, Subscription: Free Trial, Subscription ID: a376f639-0268-48a9-84e5-2b741923eb4, Disk state: Available, Tags: landing

Create landing container , where the orders data will be updated.

Microsoft Azure Upgrade Search resources, services, and docs (G+/-) Copilot nadarajan073@gmail.co... DEFAULT DIRECTORY

Dashboard > capstonestorageacct

capstonestorageacct | Containers

Storage account

Search

Container Change access level Restore containers Refresh Delete Give feedback

Containers

File shares, Queues, Tables

Containers

\$logs

New container

Name: landing, Anonymous access level: Private (no anonymous access)

Create Give feedback

Create customers container with anonymous access.

The screenshot shows the Microsoft Azure Storage account dashboard for the 'capstonestorageacct' storage account. On the left sidebar, under 'Containers', there is a list of existing containers: '\$logs' and 'landing'. A new container named 'customers' is being created on the right. The 'Anonymous access level' dropdown is set to 'Container (anonymous read access for containers and blobs)'. A warning message states: 'All container and blob data can be read by anonymous request. Clients can enumerate blobs within the container by anonymous request, but cannot enumerate containers within the storage account. Anonymous access bypasses Access Control List (ACL) settings.' A 'Create' button is at the bottom right of the container creation form.

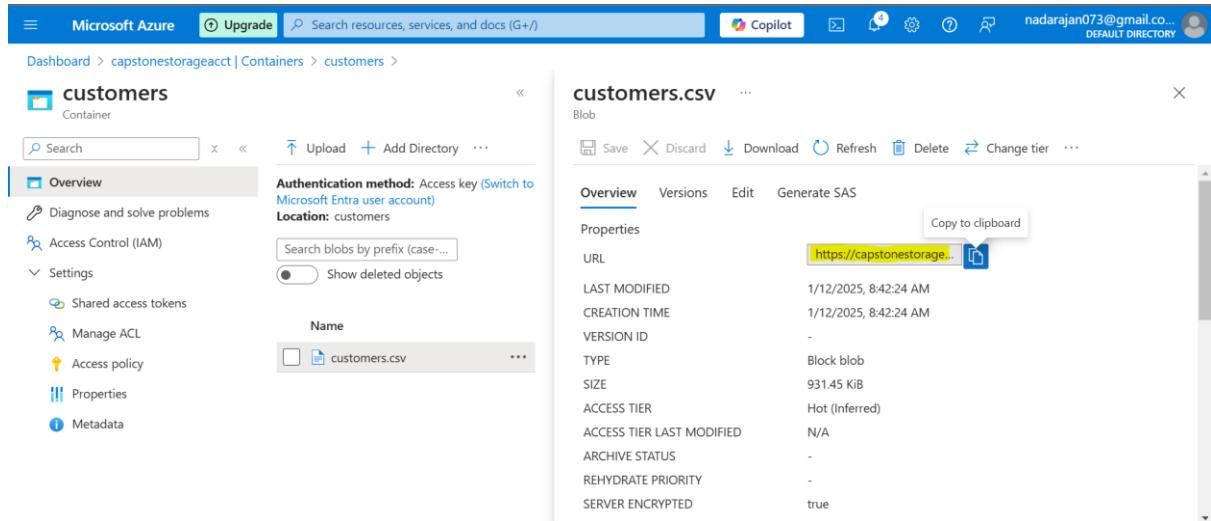
Upload the below customer's data.

The screenshot shows the 'customers' container within the 'capstonestorageacct' storage account. The 'Overview' tab is selected. On the right, an 'Upload blob' dialog is open, showing a file named 'customers.csv' selected for upload. The 'Overwrite if files already exist' checkbox is checked. A 'Upload' button is at the bottom right of the dialog.

The screenshot shows a Microsoft Excel spreadsheet titled 'customer_id' containing 20 rows of customer data. The columns are labeled: customer, customer_fname, customer_lname, username, password, address, city, state, pincode. The data includes various names like Richard Hernandez, Mary Barrett, Ann Smith, etc., along with their addresses and other details.

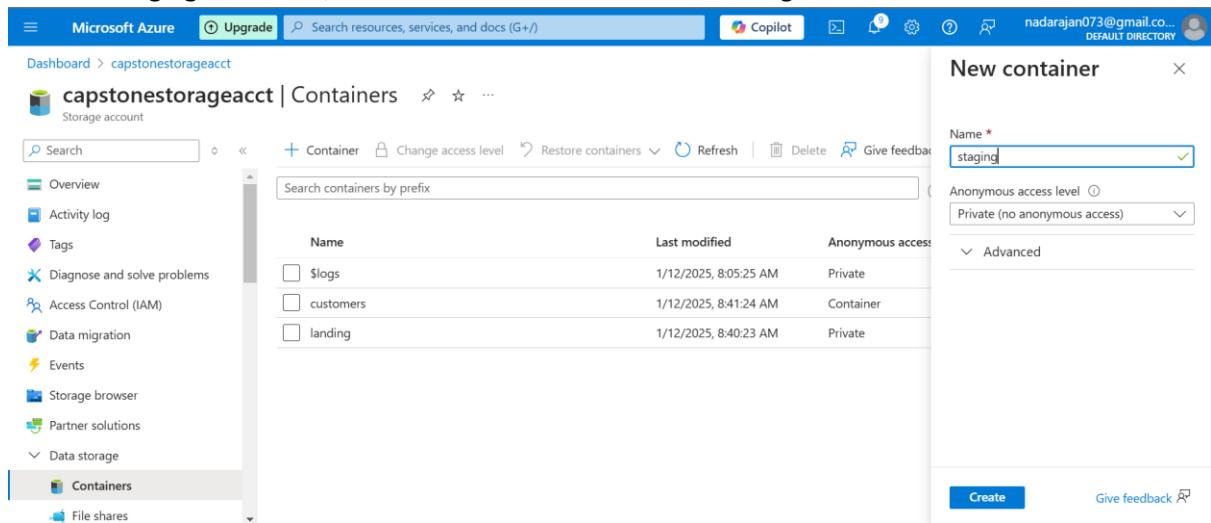
| | A | B | C | D | E | F | G | H | I | J | K |
|----|----------|----------------|----------------|------------|------------|-----------------------------|---------------|-------|---------|---|---|
| 1 | customer | customer_fname | customer_lname | username | password | address | city | state | pincode | | |
| 2 | 1 | Richard | Hernandez | XXXXXXXXXX | XXXXXXXXXX | 6303 Heather Plaza | Brownsville | TX | 78521 | | |
| 3 | 2 | Mary | Barrett | XXXXXXXXXX | XXXXXXXXXX | 9526 Noble Embers Ridge | Littleton | CO | 80126 | | |
| 4 | 3 | Ann | Smith | XXXXXXXXXX | XXXXXXXXXX | 3422 Blue Pioneer Bend | Caguas | PR | 725 | | |
| 5 | 4 | Mary | Jones | XXXXXXXXXX | XXXXXXXXXX | 8324 Little Common | San Marcos | CA | 92069 | | |
| 6 | 5 | Robert | Hudson | XXXXXXXXXX | XXXXXXXXXX | 10 Crystal River Mall | Caguas | PR | 725 | | |
| 7 | 6 | Mary | Smith | XXXXXXXXXX | XXXXXXXXXX | 3151 Sleepy Quail Promenade | Passaic | NJ | 7055 | | |
| 8 | 7 | Melissa | Wilcox | XXXXXXXXXX | XXXXXXXXXX | 9453 High Concession | Caguas | PR | 725 | | |
| 9 | 8 | Megan | Smith | XXXXXXXXXX | XXXXXXXXXX | 3047 Foggy Forest Plaza | Lawrence | MA | 1841 | | |
| 10 | 9 | Mary | Perez | XXXXXXXXXX | XXXXXXXXXX | 3616 Quaking Street | Caguas | PR | 725 | | |
| 11 | 10 | Melissa | Smith | XXXXXXXXXX | XXXXXXXXXX | 8598 Harvest Beacon Plaza | Stafford | VA | 22554 | | |
| 12 | 11 | Mary | Huffman | XXXXXXXXXX | XXXXXXXXXX | 3169 Stony Woods | Caguas | PR | 725 | | |
| 13 | 12 | Christopher | Smith | XXXXXXXXXX | XXXXXXXXXX | 5594 Jagged Embers By-pass | San Antonio | TX | 78227 | | |
| 14 | 13 | Mary | Baldwin | XXXXXXXXXX | XXXXXXXXXX | 7922 Iron Oak Gardens | Caguas | PR | 725 | | |
| 15 | 14 | Katherine | Smith | XXXXXXXXXX | XXXXXXXXXX | 5666 Hazy Pony Square | Pico Rivera | CA | 90660 | | |
| 16 | 15 | Jane | Luna | XXXXXXXXXX | XXXXXXXXXX | 673 Burning Glen | Fontana | CA | 92336 | | |
| 17 | 16 | Tiffany | Smith | XXXXXXXXXX | XXXXXXXXXX | 6651 Iron Port | Caguas | PR | 725 | | |
| 18 | 17 | Mary | Robinson | XXXXXXXXXX | XXXXXXXXXX | 1325 Noble Pike | Taylor | MI | 48180 | | |
| 19 | 18 | Robert | Smith | XXXXXXXXXX | XXXXXXXXXX | 2734 Hazy Butterfly Circle | Martinez | CA | 94553 | | |
| 20 | 19 | Stephanie | Mitchell | XXXXXXXXXX | XXXXXXXXXX | 3543 Red Treasure Bay | Caguas | PR | 725 | | |
| 21 | 20 | Mary | Ellis | XXXXXXXXXX | XXXXXXXXXX | 4703 Old Route | West New York | NJ | 7093 | | |

Open the file , then we can get the external link , with that we can get this data from anywhere later.



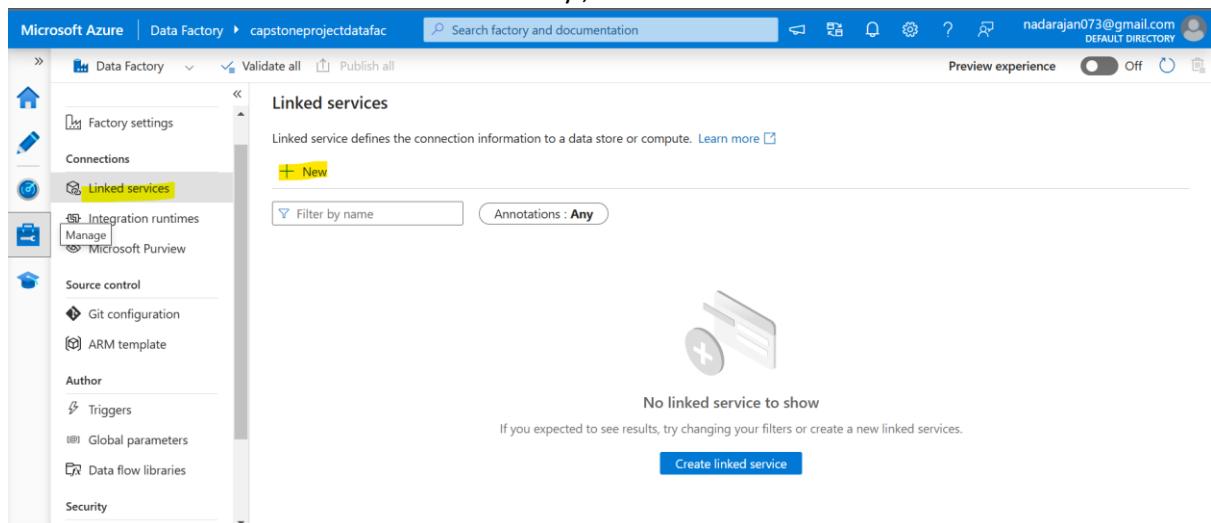
The screenshot shows the Microsoft Azure Storage Explorer interface. On the left, the 'customers' container is selected under the 'Containers' section. In the center, a blob named 'customers.csv' is listed. On the right, the properties for this blob are displayed, including its URL, which is highlighted with a yellow box: <https://capstonestorage...>. Other properties shown include Last Modified (1/12/2025, 8:42:24 AM), Creation Time (1/12/2025, 8:42:24 AM), Type (Block blob), Size (931.45 KB), and Access Tier (Hot (Inferred)).

Create a staging container , so that we can save the data that we get from the external link.



The screenshot shows the Microsoft Azure Storage Explorer interface. On the left, the 'Containers' section is selected under the 'Storage account' section. A new container named 'staging' is being created on the right. The 'Name' field is filled with 'staging'. Under 'Anonymous access level', 'Private (no anonymous access)' is selected. A 'Create' button is visible at the bottom right.

Now create a linked service's in the data factory , so that we can connect to the data sources.



The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Linked services' section is selected under the 'Data Factory' section. A message states 'No linked service to show'. Below it, a button says 'Create linked service'. The central area displays a list of existing linked services, which is currently empty.

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

Validate all Publish all

Linked services

Linked service defines the connection information to a data store

+ New Filter by name Annotations : Any

Azure Data Lake Storage Gen2 Magento (Legacy)

If you expected to see

New linked service

Data store Compute

All Azure Database File Generic protocol

Continue Cancel

This screenshot shows the 'Linked services' blade in the Microsoft Azure Data Factory portal. The left sidebar shows various navigation items like 'Factory settings', 'Connections', 'Source control', etc. The main area displays a list of available linked services, categorized by type (Data store, Compute). Under 'Data store', 'Azure Data Lake Storage Gen2' and 'Magento (Legacy)' are listed. A search bar at the top allows filtering by name and applying annotations.

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

Validate all Publish all

Linked services

Linked service defines the connection information to a data store

+ New Filter by name Annotations : Any

Azure Data Lake Storage Gen2 Learn more

Name * ls_adls

Description

Connect via integration runtime * AutoResolveIntegrationRuntime

Authentication type Account key

Account selection method From Azure subscription Enter manually

Azure subscription

Create Back Test connection Cancel

This screenshot shows the 'New linked service' configuration page for 'Azure Data Lake Storage Gen2'. It includes fields for 'Name' (ls_adls), 'Description', 'Connect via integration runtime' (set to 'AutoResolveIntegrationRuntime'), 'Authentication type' (set to 'Account key'), and 'Account selection method' (set to 'From Azure subscription'). The 'Create' and 'Back' buttons are at the bottom.

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

Validate all Publish all

Linked services

Linked service defines the connection information to a data store

+ New Filter by name Annotations : Any

Azure Data Lake Storage Gen2 Learn more

Account selection method From Azure subscription Enter manually

Azure subscription

Storage account name * capstonestorageacct

Test connection To linked service To file path

Annotations + New

Parameters

Connection successful

Create Back Test connection Cancel

This screenshot shows the 'New linked service' configuration page for 'Azure Data Lake Storage Gen2' after a successful connection test. The 'Storage account name' field is set to 'capstonestorageacct'. The 'Test connection' status is shown as 'Connection successful'. The 'Create' and 'Back' buttons are at the bottom.

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

New linked service

Data store Compute

http

All Azure Database File Generic protocol

HTTP

ls_adls

Continue Cancel

This screenshot shows the 'New linked service' dialog in the Microsoft Azure Data Factory interface. The left sidebar is visible with 'Data Factory' selected. The main area shows a list of 'Linked services' with one item, 'ls_adls', listed. On the right, a detailed view of the 'ls_adls' entry is shown, specifically the 'HTTP' type. A 'Continue' button is at the bottom.

Give the Base URL.

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

New linked service

HTTP Learn more

Name * ls_externalSource

Description

Connect via integration runtime * AutoResolveIntegrationRuntime

Base URL * https://capstonestorageacct.blob.core.windows.net

⚠ Information will be sent to the URL specified. Please ensure you trust the URL entered.

Server certificate validation Enable Disable

Authentication type * Anonymous

Create Back Test connection Cancel

This screenshot shows the 'New linked service' dialog with the 'Name' field set to 'ls_externalSource'. The 'Base URL' field contains 'https://capstonestorageacct.blob.core.windows.net'. The 'Authentication type' dropdown is set to 'Anonymous'. A 'Test connection' button is visible at the bottom.

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

New linked service

HTTP Learn more

⚠ Information will be sent to the URL specified. Please ensure you trust the URL entered.

Server certificate validation Enable Disable

Authentication type * Anonymous

Auth headers New

Annotations New

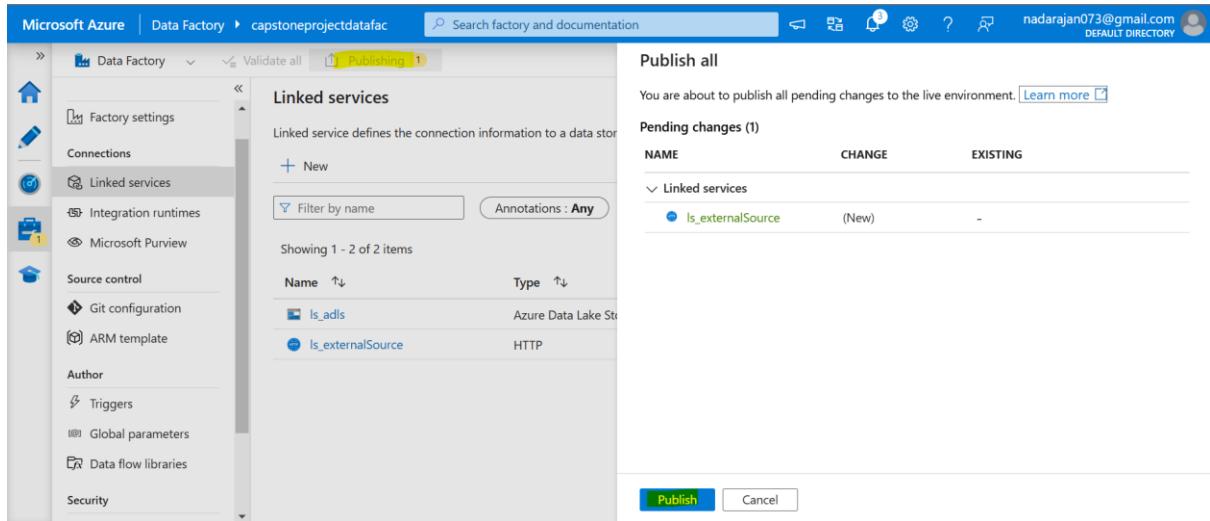
Parameters Advanced

Connection successful

Create Back Test connection Cancel

This screenshot shows the 'New linked service' dialog with a green checkmark next to 'Connection successful'. The 'Authentication type' is set to 'Anonymous'. A 'Test connection' button is visible at the bottom.

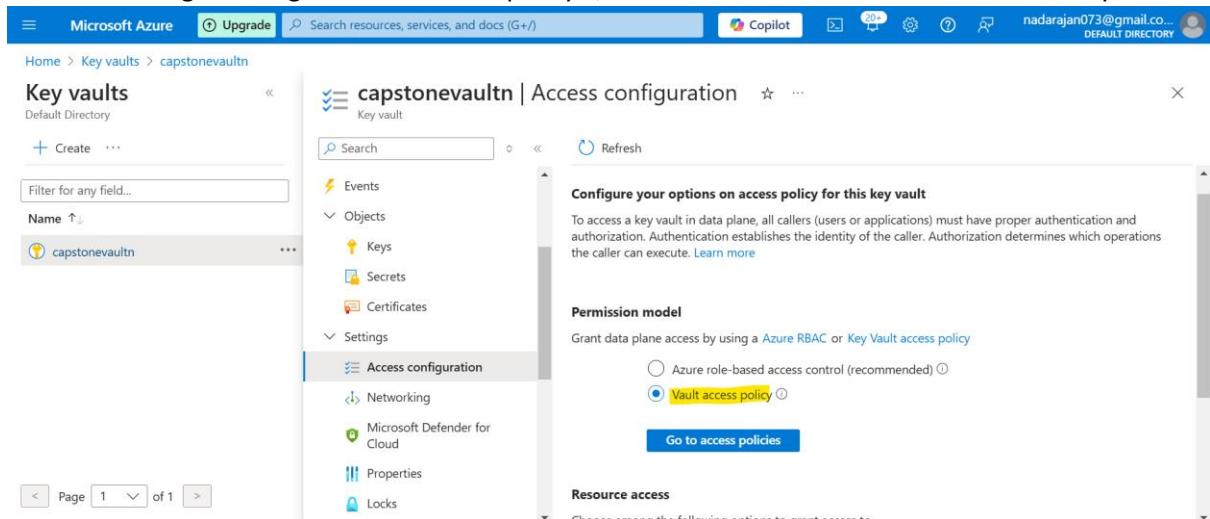
We have created Linked Services for External Link and the ADLS gen2. We need to publish all to save all the changes.



The screenshot shows the Microsoft Azure Data Factory interface. On the left, a sidebar lists various options: Home, Data Factory, Factory settings, Connections (selected), Linked services, Integration runtimes, Microsoft Purview, Source control, Git configuration, ARM template, Author, Triggers, Global parameters, Data flow libraries, and Security. In the main area, under 'Linked services', there are two items listed: 'ls_adls' (Azure Data Lake Storage) and 'ls_externalSource' (HTTP). A 'Pending changes (1)' section shows a single entry: 'ls_externalSource' with a status of '(New)'. At the bottom right, there are 'Publish' and 'Cancel' buttons.

For all the credentials we are using Azure Key vault in this project.
Create an azure key vault , and assign the “Vault Secret Officer” role to us.

In Access configuration give “Vault access policy” , So that other resources can access the key vault.



The screenshot shows the Microsoft Azure Key vaults interface. On the left, a sidebar lists 'Key vaults' (Default Directory) and a search bar. A table lists one item: 'capstonevaultn'. On the right, the 'capstonevaultn | Access configuration' page is shown. It has a sidebar with 'Events', 'Objects' (Keys, Secrets, Certificates), 'Settings' (Access configuration selected), 'Networking', 'Microsoft Defender for Cloud', 'Properties', and 'Locks'. The main content area is titled 'Configure your options on access policy for this key vault'. It explains that to access a key vault in data plane, authentication and authorization are required. It shows a 'Permission model' section with a radio button for 'Vault access policy' (which is selected) and 'Azure role-based access control (recommended)'. Below this is a 'Go to access policies' button. At the bottom, there's a 'Resource access' section with a note: 'Choose among the following options to grant access to'.

Microsoft Azure Upgrade Search resources, services, and docs (G+)

nadarajan073@gmail.co... DEFAULT DIRECTORY

Dashboard > capstonevaultn | Access control (IAM) Key vault

Add role assignment

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Access policies Events Objects Keys Secrets Certificates Settings

Search Add Download role assignments Edit columns Refresh Delete Feedback

Add role assignment Check access Role assignments Roles Deny assignments Classic administrators

My access View my level of access to this resource. View my access

Check access Review the level of access a user, group, service principal, or managed identity has to this resource. Learn more Check access

Grant access to this resource View access to this resource View deny assignments

Grant access to resources by assigning a role. Learn more View the role assignments that grant access to this and other resources. Learn more View the role assignments that have been denied access to specific actions at this scope.

Microsoft Azure Upgrade Search resources, services, and docs (G+)

nadarajan073@gmail.co... DEFAULT DIRECTORY

Dashboard > capstonevaultn | Access control (IAM) >

Add role assignment

Job function roles Privileged administrator roles

Grant access to Azure resources based on job function, such as the ability to create virtual machines.

key vault secret Type : All Category : All

| Name ↑ | Description ↑ | Type ↑↓ | Category ↑↓ | Details |
|-------------------------------------|---|-------------|-------------|---------|
| Key Vault Administrator | Perform all data plane operations on a key vault and all objects in it, including certificates, keys, and secrets. Cannot read sensitive values. | BuiltinRole | Security | View |
| Key Vault Data Access Administrator | Manage access to Azure Key Vault by adding or removing role assignments for the Key Vault. | BuiltinRole | None | View |
| Key Vault Reader | Read metadata of key vaults and its certificates, keys, and secrets. Cannot read sensitive values. | BuiltinRole | Security | View |
| Key Vault Secrets Officer | Perform any action on the secrets of a key vault, except manage permissions. Only works for key vaults that use the 'Azure role-based access contr... | BuiltinRole | Security | View |
| Key Vault Secrets User | Read secret contents. Only works for key vaults that use the 'Azure role-based access contr... | BuiltinRole | Security | View |

Showing 1 - 5 of 5 results.

Review + assign Previous Next Feedback

Microsoft Azure Upgrade Search resources, services, and docs (G+)

nadarajan073@gmail.co... DEFAULT DIRECTORY

Dashboard > capstonevaultn | Access control (IAM) >

Add role assignment

Role Members Conditions Review + assign

Selected role Key Vault Secrets Officer

Assign access to User, group, or service principal Managed identity

Members + Select members

| Name | Object ID | Type |
|---------------------|-----------|------|
| No members selected | | |

Description Optional

Review + assign Previous Next

Select members

Search by name or email address

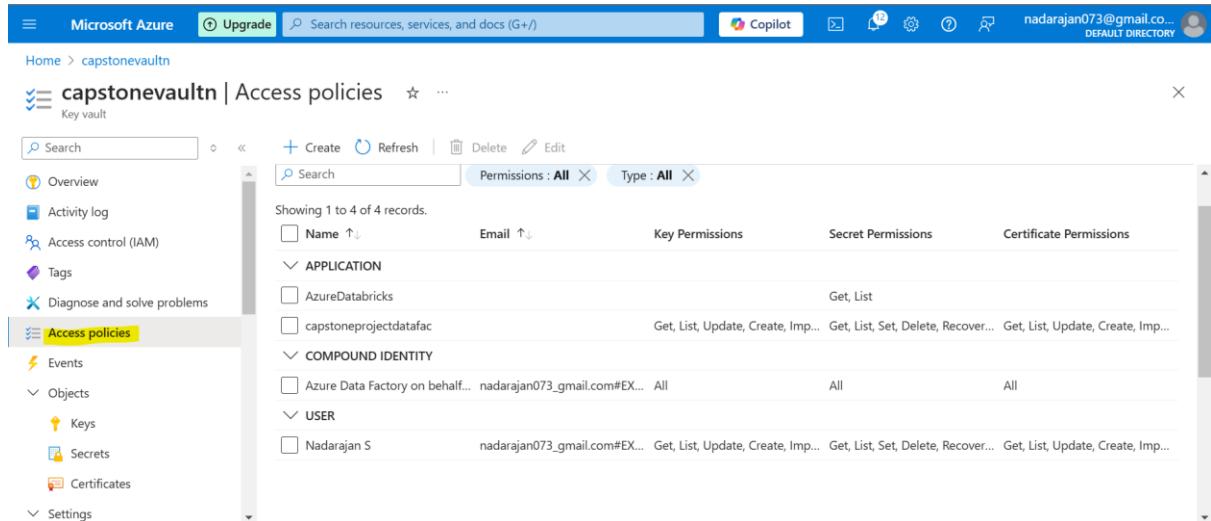
Nadarajan S(Guest)
nadarajan073_gmail.com#EXT#@nadarajan073gmail.onmicrosoft.com

Selected members:

Nadarajan S(Guest)
nadarajan073_gmail.com#EXT#@nadarajan073gmail.onmicrosoft.co...

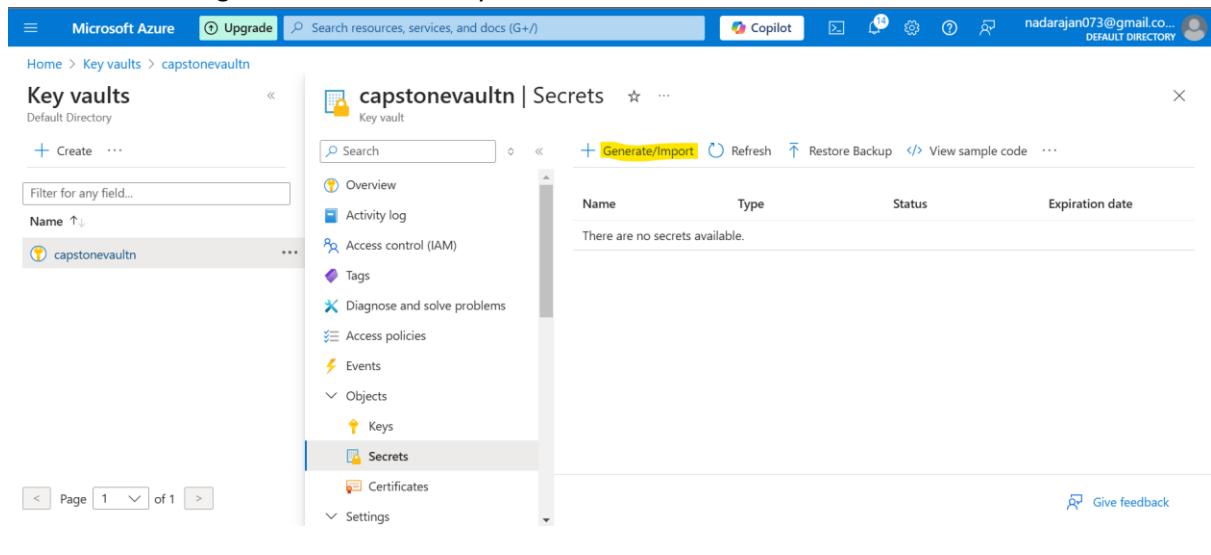
Select Close

Give the required access to all the resources that are going to use this key vault.

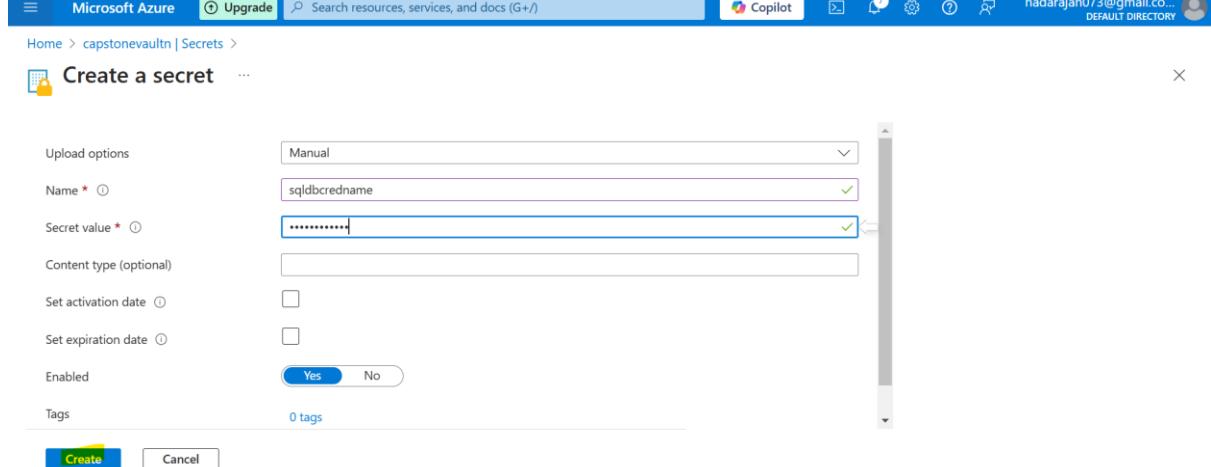


The screenshot shows the 'Access policies' section of the Azure Key Vault. It lists four entries under 'APPLICATION' and 'COMPOUND IDENTITY'. The first entry for 'AzureDatabricks' has 'Key Permissions' set to 'Get, List'. The second entry for 'capstoneprojectdatafac' has 'Key Permissions' set to 'Get, List, Update, Create, Imp...' and 'Secret Permissions' set to 'Get, List, Set, Delete, Recover...'. The third entry for 'Azure Data Factory on behalf of...' has 'Key Permissions' set to 'All' and 'Secret Permissions' set to 'All'. The fourth entry for 'Nadarajan S' has 'Key Permissions' set to 'Get, List, Update, Create, Imp...' and 'Secret Permissions' set to 'Get, List, Set, Delete, Recover...'. The sidebar on the left shows options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, and Access policies (which is selected).

Now Start creating the secrets in the key vault.



The screenshot shows the 'Secrets' section of the Azure Key Vault. It displays a table with one row: 'There are no secrets available.' The sidebar on the left shows options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, and Secrets (which is selected). A yellow box highlights the '+ Generate/Import' button at the top of the main content area.



The screenshot shows the 'Create a secret' form. The 'Name' field is set to 'sqldbcredname' and the 'Secret value' field contains '.....'. The 'Enabled' switch is set to 'Yes'. At the bottom, there are 'Create' and 'Cancel' buttons. The sidebar on the left shows options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, and Secrets (which is selected).

Microsoft Azure | [Upgrade](#) | Search resources, services, and docs (G+)

Home > capstonevaultn | Secrets > Create a secret

Upload options: Manual

Name *: sqldbcredpass

Secret value *: *****

Content type (optional):

Set activation date:

Set expiration date:

Enabled: Yes

Tags: 0 tags

Create **Cancel**

To access the key vault in databricks , we need to create an scope in databricks . So that only we can able to access the key vault secrets.

First of all generate a token for our databricks workspace.

Welcome to Databricks

Search data, notebooks, recents, and more... CTRL + P

Set up your workspace

Get started

Recents, Favorites, Popular, Mosaic AI, What's new

Untitled Notebook 2025-01-12 08:47:31 /Users/nadarajan073@gmail.com 16 seconds ago Notebook

Nadarajan S nadarajan073@gmail.com

Settings

Azure Portal

Privacy Policy

Previews

Send feedback

Log out

<https://adb-3262179368725353.13.azuredatabricks.net/settings/user/profile?g=3262179368725353>

Settings

Developer

Manage your development settings

Access tokens

Set up secure authentication to Databricks API using access tokens **Manage**

SQL query snippets

Configure SQL query snippets. Note: SQL query snippets will be moving to the SQL editor dropdown menu. **SQL Editor**

Editor settings

General

Spark tips

Enrich notebook error stack traces by displaying high-level "error hints" which explain the cause of errors from external services. **On**

The screenshot shows the Databricks Settings page under User settings > Developer > Access tokens. A green box highlights the "Generate new token" button. Below it, a table lists tokens with columns for Comment, Creation, and Expiration. A note states: "Personal access tokens can be used for secure authentication to the [Databricks API](#) instead of passwords."

Add this secret in the key vault secret.

The screenshot shows the "Create a secret" page in Azure Key Vault. The "Name" field is set to "databrickstoken" and the "Secret value" field contains a redacted token. Other fields include "Content type (optional)", "Set activation date", "Set expiration date", "Enabled" (set to Yes), and "Tags". Buttons for "Create" and "Cancel" are at the bottom.

We can get the secret for the storage account in the “Access Keys” option inside the storage account.

The screenshot shows the "Access keys" page for the storage account "capstonestorageacct". The "key1" key is selected. It shows the key was last rotated on 1/12/2025 and provides a "Copy to clipboard" button. A "Hide" button is also present. The "Connection string" is shown as a redacted string. The left sidebar lists other storage account options like Tables, Networking, Shared access signature, Encryption, and Data protection.

Copy the storage account secret and create a secret in key vault.

The screenshot shows the 'Create a secret' dialog in the Microsoft Azure Key Vault interface. The 'Name' field is set to 'storageaccountkey'. The 'Secret value' field contains a masked password. The 'Enabled' switch is set to 'Yes'. At the bottom, there are 'Create' and 'Cancel' buttons.

These are all the all available secrets that we have created in our key vault.

The screenshot shows the 'Secrets' list in the Microsoft Azure Key Vault interface. It displays four secrets: 'databrickstoken', 'sqldbcredname', 'sqldbcredpass', and 'storageaccountkey', all of which are enabled. A message at the top indicates that 'The secret 'databrickstoken' has been successfully created.'

Meanwhile turn on the cluster and create an Note book.

The screenshot shows the Databricks Compute interface. Under the 'Compute' tab, it displays 'Nadarajan S's Cluster'. Configuration settings include 'Policy' (Unrestricted), 'Access mode' (Single user access), and 'Databricks Runtime Version' (15.4 LTS). The 'Performance' section shows 'Node type' as 'Standard_D3_v2' with '14 GB Memory, 4 Cores'.

The screenshot shows the Microsoft Azure Databricks workspace. On the left, the sidebar includes options like 'New', 'Workspace' (selected), 'Recents', 'Catalog', 'Workflows', 'Compute', 'SQL', 'SQL Editor', 'Queries', 'Dashboards', 'Genie', 'Alerts', 'Query History', 'SQL Warehouses', and 'Data Engineering'. The main area is titled 'Capstone_Notebook' and 'Python'. It contains a single code cell with the placeholder text 'Start typing or generate with AI (Ctrl + I)...'. The status bar at the bottom indicates 'Last edit was now'.

Create an linked service for the key vault.

The screenshot shows the 'New linked service' dialog in Azure Data Factory. The 'Data store' tab is selected. In the search bar, 'key' is typed. Below the search bar, there are tabs for 'All', 'Azure', 'Database', 'File', and 'Generic protocol'. A card for 'Azure Key Vault' is displayed, featuring a key icon and the text 'Azure Key Vault'. At the bottom of the dialog are 'Continue' and 'Cancel' buttons.

The screenshot shows the 'New linked service' dialog in Azure Data Factory, specifically for an 'Azure Key Vault' connection. The 'Name' field is set to 'ls_key_vault'. The 'Description' field is empty. Under 'Azure key vault selection method', the radio button 'From Azure subscription' is selected. The 'Azure subscription' dropdown shows 'Free Trial (a376f639-0268-48a9-84e5-2b741923eb4)'. The 'Azure key vault name' dropdown shows 'capstonevault'. At the bottom, a green checkmark indicates 'Connection successful', and there are 'Create', 'Back', 'Test connection', and 'Cancel' buttons.

Create a linked service for our Azure SQL Database with key vault credentials.

The screenshot shows the 'New linked service' dialog in the Microsoft Azure Data Factory interface. On the left, the navigation menu is visible with 'Linked services' selected. The main area shows a grid of data store options under the 'Data store' tab. The 'Azure SQL Database' option is highlighted. Below the grid, there are 'Continue' and 'Cancel' buttons.

The screenshot shows the 'New linked service' dialog with the 'Name' field set to 'ls_sql_db'. The 'Account selection method' section is expanded, showing the 'From Azure subscription' radio button selected. Other options like 'Legacy' and 'Enter manually' are also shown. At the bottom, there are 'Create', 'Back', 'Test connection', and 'Cancel' buttons.

The screenshot shows the 'New linked service' dialog with detailed connection settings for an Azure SQL Database. The 'Account selection method' is set to 'From Azure subscription'. The 'Azure subscription' dropdown shows 'Free Trial (a376f639-0268-48a9-84e5-2b741923ebc4)'. The 'Server name' is set to 'capstonebserver', 'Database name' to 'capstoneDB', and 'Authentication type' to 'SQL authentication'. The 'User name' field contains 'capstoneProj'. At the bottom, there are 'Create', 'Back', 'Test connection', and 'Cancel' buttons.

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

Validate all Publish all

Linked services

Linked service defines the connection information to a data store

+ New Filter by name Annotations: Any

Showing 1 - 3 of 3 items

| Name | Type |
|-------------------|-------------------------|
| ls_adls | Azure Data Lake Storage |
| ls_externalSource | HTTP |
| ls_key_vault | Azure Key Vault |

New linked service

Azure SQL Database Learn more

AKV linked service * ls_key_vault

Secret name * sqldbcredpass

Secret version Latest version

Always encrypted

Encrypt Mandatory

Connection successful Test connection Cancel

Create an linked service for our databricks , in the Compute tab.

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

Validate all Publish all

Linked services

Linked service defines the connection information to a data store

+ New Filter by name Annotations: Any

Showing 1 - 4 of 4 items

| Name | Type |
|-------------------|-------------------------|
| ls_adls | Azure Data Lake Storage |
| ls_externalSource | HTTP |
| ls_key_vault | Azure Key Vault |
| ls_sql_db | Azure SQL Database |

New linked service

Data store Compute

databri

All Azure Compute

Azure Databricks

Continue Cancel

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

Validate all Publish all

Linked services

Linked service defines the connection information to a data store

+ New Filter by name Annotations: Any

Showing 1 - 4 of 4 items

| Name | Type |
|-------------------|-------------------------|
| ls_adls | Azure Data Lake Storage |
| ls_externalSource | HTTP |
| ls_key_vault | Azure Key Vault |
| ls_sql_db | Azure SQL Database |

New linked service

Azure Databricks Learn more

Name * ls_databricks

Description

Connect via integration runtime * AutoResolveIntegrationRuntime

Account selection method * From Azure subscription Enter manually

Azure subscription * Free Trial (a376f639-0268-48a9-84e5-2b741923be4)

Databricks workspace * ranstone Databricks

Create Back Test connection Cancel

Login to the databricks workspace by using the token which we have already generated before,
Use azure Key vault.

Here we have used “Existing interaction cluster” , since our usecase was small and so it was a demo purpose. Usually it will be a job Clusters only.

The screenshot shows the 'Linked services' configuration page in Microsoft Azure Data Factory. The left sidebar lists various options like General, Factory settings, Connections, and Linked services. The 'Connections' section is currently selected. On the right, there's a form for creating a new linked service. The 'Type' dropdown is set to 'Azure Databricks'. Under 'Select cluster', the radio button for 'Existing interactive cluster' is selected. The 'Databrick Workspace URL' field contains the URL 'https://adb-3262179368725353.13.azure.databricks.net'. The 'Authentication type' dropdown is set to 'Access Token', and the 'Access token' tab is selected. The 'AKV linked service' dropdown is set to 'ls_key_vault', and the 'Secret name' dropdown is set to 'databrickstoken'. At the bottom, there are 'Create', 'Back', 'Test connection', and 'Cancel' buttons.

This screenshot is similar to the previous one but shows a different selection in the 'Choose from existing clusters' dropdown. Instead of 'Nadarajan S's Cluster', it now shows 'Nadarajan S's Cluster' again, indicating a selection loop or a different step in the process. The rest of the configuration remains the same, with 'ls_key_vault' selected as the AKV linked service and 'databrickstoken' as the secret name.

Select the cluster name that we have already created.

The screenshot shows the final step of the linked service creation process. The 'Choose from existing clusters' dropdown now correctly displays 'Nadarajan S's Cluster'. A green checkmark icon and the text 'Connection successful' are visible at the bottom right, indicating that the connection has been successfully tested. The other fields remain the same as in the previous screenshots.

These are all the linked services that we have created for our project.

Microsoft Azure | Data Factory | capstoneprojectdatafac | Search factory and documentation

Validate all Publish all

Preview experience Off

General

Factory settings

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Global parameters

Data flow libraries

Linked services

Linked service defines the connection information to a data store or compute. Learn more

+ New

Filter by name Annotations : Any

Showing 1 - 5 of 5 items

| Name | Type | Related | Annotations |
|-------------------|------------------------------|---------|-------------|
| ls_adls | Azure Data Lake Storage Gen2 | 0 | |
| ls_databricks | Azure Databricks | 0 | |
| ls_externalSource | HTTP | 0 | |
| ls_key_vault | Azure Key Vault | 2 | |
| ls_sql_db | Azure SQL Database | 0 | |

Upload the orders data to the landings container.

Microsoft Azure | capstonestorageacct | Containers | Search resources, services, and docs (G+)

Home > capstonestorageacct | Containers >

landing Container

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: landing

Search blobs by prefix (case-sensitive)

| Name | Modified |
|------------|----------|
| No results | |

Upload blob

1 file(s) selected: orders_new.csv

Drag and drop files here or Browse for files

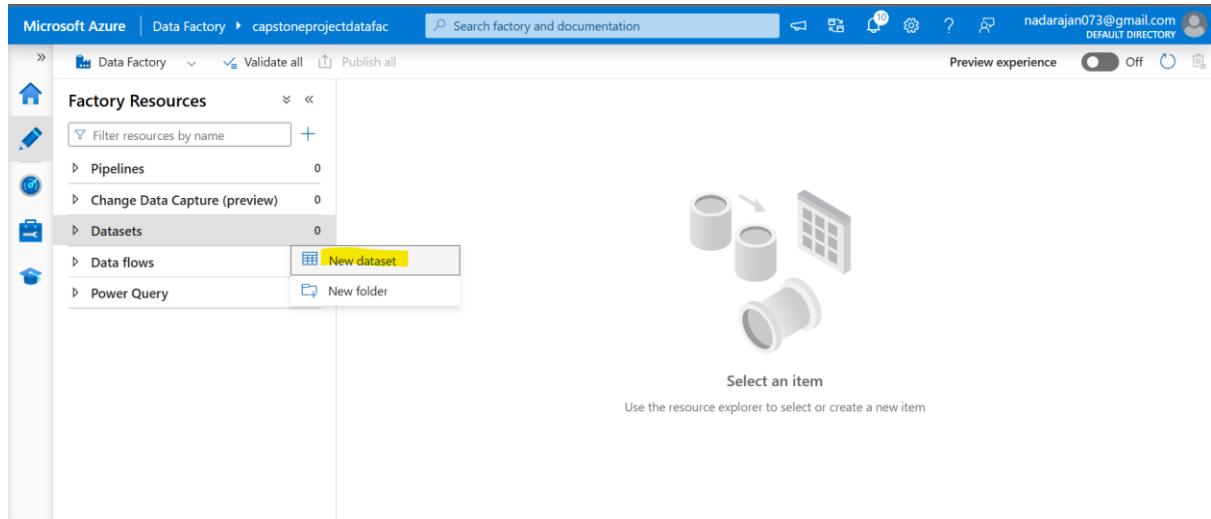
Overwrite if files already exist

Advanced

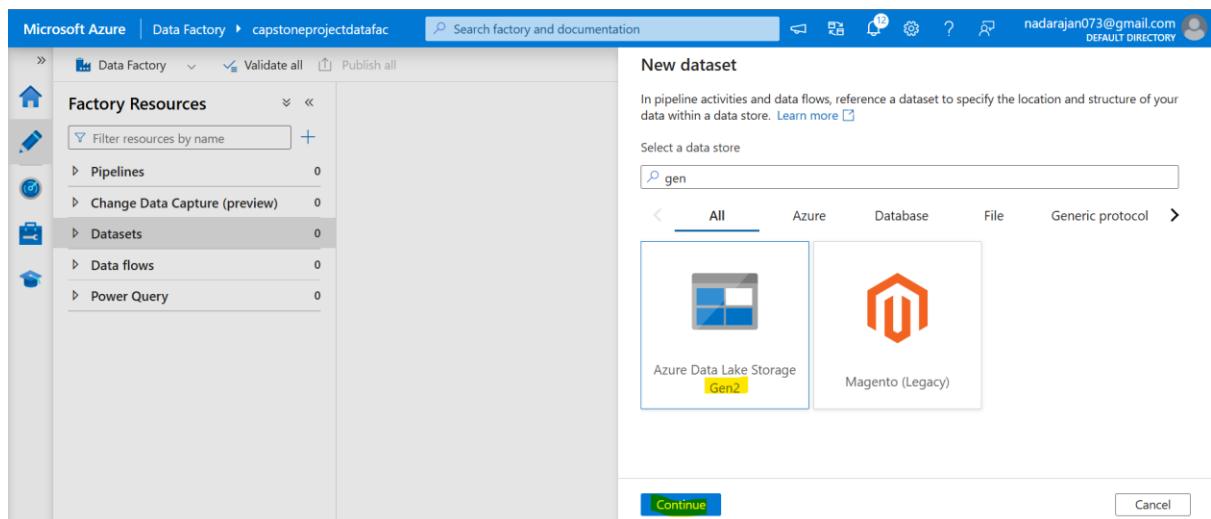
Upload

Give feedback

Create a dataset for the ADLS and our Azure SQL DB

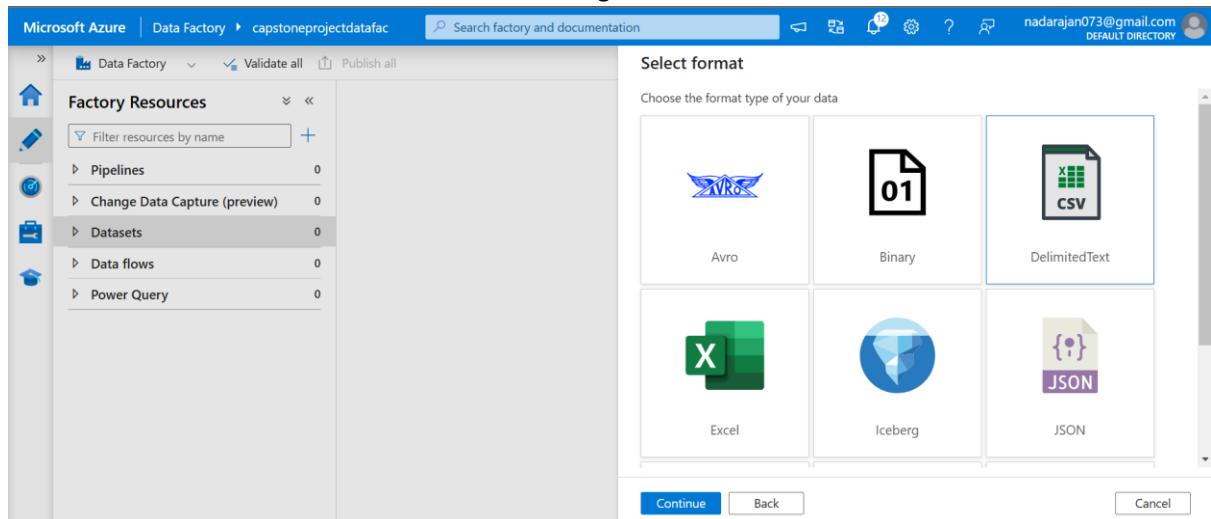


The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a sidebar titled 'Factory Resources' with icons for Home, Pipelines, Change Data Capture (preview), Datasets, Data flows, and Power Query. Under 'Datasets', the option 'New dataset...' is highlighted with a yellow box. The main area has a large icon of two cylinders and a grid, with the text 'Select an item' and 'Use the resource explorer to select or create a new item'.



This screenshot shows the 'New dataset' dialog. The left sidebar is identical to the previous one. The main area has a heading 'New dataset' and a sub-instruction 'In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)'. Below this is a search bar with 'gen' typed in, and a tab navigation bar with 'All' selected, followed by 'Azure', 'Database', 'File', and 'Generic protocol'. Two data store options are listed: 'Azure Data Lake Storage Gen2' and 'Magento (Legacy)'. At the bottom are 'Continue' and 'Cancel' buttons.

Our data was in “csv” format . So that we are using Delimited Text here.



This screenshot shows the 'Select format' dialog. The left sidebar is consistent. The main area has a heading 'Select format' and a sub-instruction 'Choose the format type of your data'. A 3x2 grid of icons represents different formats: Avro (blue icon), Binary (white document with '01'), DelimitedText (green document with 'CSV'), Excel (green icon with 'X'), Iceberg (blue icon with a funnel), and JSON (purple icon with curly braces). The 'DelimitedText' option is highlighted with a yellow box. At the bottom are 'Continue', 'Back', and 'Cancel' buttons.

Select the file path.

Microsoft Azure | Data Factory | capstoneprojectdatafac | Search factory and documentation | nadarajan073@gmail.com | DEFAULT DIRECTORY

Set properties

Name: ds_adls

Linked service: ls_adls

File path: File system / Directory / File name !

First row as header

Import schema:

From connection/store From sample file None

OK Back Cancel

Microsoft Azure | Data Factory | capstoneprojectdatafac | Search factory and documentation | nadarajan073@gmail.com | DEFAULT DIRECTORY

Browse

Select a file or folder.

Root folder > landing

orders_new.csv

Showing 1 item

OK Cancel

Microsoft Azure | Data Factory | capstoneprojectdatafac | Search factory and documentation | nadarajan073@gmail.com | DEFAULT DIRECTORY

Set properties

Name: ds_adls

Linked service: ls_adls

File path: landing / Directory / orders_new.csv !

First row as header

Import schema:

From connection/store From sample file None

OK Back Cancel

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar is open, showing 'Pipelines' (0), 'Change Data Capture (preview)' (0), 'Datasets' (1), 'Data flows' (0), and 'Power Query' (0). The 'Datasets' section is expanded, and 'ds_adls' is selected. In the main pane, a dataset named 'ds_adls' is displayed, which is a 'DelimitedText' type with a 'CSV' icon. Below it, the 'Connection' tab is selected, showing a 'Linked service' dropdown set to 'ls_adls', a 'File path' of 'landing / Directory / orders_new.csv', and other settings like 'Compression type' (No compression), 'Column delimiter' (Comma (,),) and 'Row delimiter' (Default (\r,\n, or \r\n)). A green checkmark indicates a 'Connection successful'. The top navigation bar includes 'Search factory and documentation', 'Validate all', 'Publish all (1)', and 'Preview experience' toggle.

Now create a dataset for azure SQL DB.

This screenshot shows the 'New dataset' dialog in Microsoft Azure Data Factory. The left side of the screen shows the same 'Factory Resources' and dataset configuration as the previous screenshot. The right side is the 'New dataset' dialog. It starts with the message 'In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store.' Below this is a 'Select a data store' section with a search bar containing 'sql'. A grid of data stores is shown, with 'Azure SQL Database' highlighted. Other options include 'Azure Database for PostgreSQL' and 'Azure SQL Database Managed Instance'. At the bottom of the dialog are 'Continue' and 'Cancel' buttons.

Select the Table.

This screenshot shows the 'Set properties' dialog for creating a new dataset named 'ds_sqldb'. The 'Name' field is filled with 'ds_sqldb'. The 'Linked service' dropdown is set to 'ls_sqldb'. The 'Table name' dropdown is set to 'dbo.orders'. Under 'Import schema', the radio button is selected for 'From connection/store'. The bottom of the dialog has 'OK', 'Back', and 'Cancel' buttons. The left side of the screen shows the familiar 'Factory Resources' and dataset configuration.

Pulish all the changes.

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar lists 'Factory Resources' including 'Pipelines', 'Change Data Capture (preview)', 'Datasets', 'Data flows', and 'Power Query'. Under 'Datasets', 'ds_adls' and 'ds_sql_db' are listed. The main panel shows 'ds_adls' and 'ds_sql_db' under 'Azure SQL Database'. The 'Connection' tab is selected, showing 'ls_sql_db' as the linked service, which has a 'Connection successful' status. The 'Table' dropdown is set to 'dbo.orders'. A message box at the top right says 'Publishing completed' and 'Successfully published'.

Create a new pipeline in authors tab.

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar lists 'Factory Resources' including 'Pipelines', 'Change Data Capture (preview)', 'Datasets', 'Data flows', and 'Power Query'. The 'Pipelines' section is expanded, showing 'New pipeline' highlighted with a yellow box. The main panel shows 'ds_adls' and 'ds_sql_db' under 'Azure SQL Database'. The 'Connection' tab is selected, showing 'ls_sql_db' as the linked service, which has a 'Connection successful' status. The 'Table' dropdown is set to 'dbo.orders'.

Bring the CopyTo activity to our pipeline.

The screenshot shows the Microsoft Azure Data Factory Pipeline Editor. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (pipeline1), 'Datasets' (ds_adls, ds_sql_db), 'Data flows', and 'Power Query'. The main area shows the 'Activities' pane with 'Move and transform' expanded, showing 'Copy data', 'Data flow', and 'Synapse'. A 'Copy data' activity is selected and highlighted in the list. Below the list are tabs for 'Parameters', 'Variables', 'Settings', and 'Output', with a '+ New' button.

Select the source as the dataset that we have already created for the ADLS orders data.

The screenshot shows the 'Copy data' activity configuration for the 'pl_capstone' pipeline. The 'Source' tab is selected. Under 'Source dataset', 'ds_adls' is chosen. Under 'File path type', 'File path in dataset' is selected. There are fields for 'Start time (UTC)' and 'End time (UTC)'. The 'Recursively' and 'Enable partitions discovery' options are checked. The 'Mapping' tab is visible at the top right.

Select the Sink as the dataset that we have created for the Azure SQL DB.

The screenshot shows the 'Copy data' activity configuration for the 'pl_capstone' pipeline. The 'Sink' tab is selected. Under 'Sink dataset', 'ds_sql_db' is chosen. Under 'Write behavior', 'Insert' is selected. Under 'Table option', 'Use existing' is selected. The 'Mapping' tab is visible at the top right.

The screenshot shows the Microsoft Azure Data Factory interface. In the left sidebar, under 'Move and transform', the 'Copy data' option is selected. In the main workspace, a 'Copy data' activity is highlighted in a blue box. Below it, the 'Mapping' tab is active in the ribbon. The mapping table shows the following data:

| Source | Type | Destination | Type |
|--------------|------------|--------------|-------------|
| order_id | abc String | order_id | 123 int |
| order_date | abc String | order_date | abc varchar |
| customer_id | abc String | customer_id | 123 int |
| order_status | abc String | order_status | abc varchar |

Import the schema , and verify the mappers in source and sink.

The screenshot shows the Microsoft Azure Data Factory interface with the 'Mapping' tab selected. The 'Import schemas' button is visible, and the previously defined mapping table is displayed:

| Source | Type | Destination | Type |
|--------------|------------|--------------|-------------|
| order_id | abc String | order_id | 123 int |
| order_date | abc String | order_date | abc varchar |
| customer_id | abc String | customer_id | 123 int |
| order_status | abc String | order_status | abc varchar |

The screenshot shows the Microsoft Azure Data Factory interface with a message box indicating 'Publishing completed' and 'Successfully published'. The 'Copy data' activity is still highlighted, and the mapping table remains the same as in the previous screenshots.

After publish the changes , try to debug the pipeline.

The screenshot shows the Microsoft Azure Data Factory interface. In the center, a pipeline named 'pl_capstone' is displayed. Under the 'Activities' section, a 'Copy data' task named 'Copy orders data' is shown with a green checkmark indicating success. Below this, the 'Output' tab displays the pipeline run ID (437e91f0-c34a-4b45-827e-e3a0801e9ccd) and the pipeline status as 'In progress'. A table below lists the activity details: Activity name: Copy orders data, Status: Succeeded, Activity type: Copy data, Run start: 1/12/2025, 12:05:50 PM, Duration: 34s, Integration runtime: AutoResolveIntegration.

If everything is worked fine we can able to see the data in our Azure SQL table.

The screenshot shows the Microsoft Azure SQL database query editor. On the left, a sidebar provides navigation options like Overview, Activity log, Tags, and Query editor (preview). The main area shows a database named 'capstoneDB' with a table named 'orders'. A query is run: 'select * from [dbo].[orders]'. The results table shows two rows of data:

| order_id | order_date | customer_id | order_status |
|----------|-----------------------|-------------|-----------------|
| 1 | 2013-07-25 00:00:00.0 | 11599 | CLOSED |
| 2 | 2013-07-25 00:00:00.0 | 256 | PENDING_PAYMENT |

A message at the bottom indicates 'Query succeeded | 1s'.

Now create a dataset for our external source.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar is open, showing 'Pipelines' (1), 'Datasets' (2), 'Data flows' (0), and 'Power Query' (0). In the center, a 'New dataset' dialog is open. It asks to 'Select a data store' and shows an 'HTTP' option under the 'All' tab. Other tabs include 'Azure', 'Database', 'File', and 'Generic protocol'. At the bottom, there are 'Continue' and 'Cancel' buttons.

Select the file format as Delimited text.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists 'Pipelines' (1), 'Datasets' (2), and other resources. In the center, the 'Activities' pane shows various options like 'Move and transform', 'Synapse', etc. A modal window titled 'Select format' is open, displaying icons for different file formats: Avro, Binary, DelimitedText (highlighted in yellow), Excel, JSON, and ORC. At the bottom of the modal are 'Continue', 'Back', and 'Cancel' buttons.

Give the Relative URL.

The screenshot shows the 'Set properties' dialog for a dataset named 'ds_externalSource'. It includes fields for 'Name' (ds_externalSource), 'Linked service' (ls_externalSource), 'Relative URL' (customers/customers.csv), 'First row as header' (checked), 'Import schema' (From connection/store selected), 'Request method' (GET), and 'Additional headers'. At the bottom are 'OK', 'Back', and 'Cancel' buttons.

The screenshot shows the 'Connection' tab for the 'ds_externalSource' dataset. It displays the 'Connection' section with 'Linked service' set to 'ls_externalSource' (with a green checkmark for 'Connection successful'), 'Base URL' as 'https://capstonestorageacct.blob.core.wi...', 'Relative URL' as 'customers/customers.csv', 'Compression type' as 'No compression', and 'Column delimiter' as 'Comma ()'. There are tabs for 'Schema' and 'Parameters' at the top of this section. At the bottom right are 'Test connection', 'Edit', 'New', and 'Learn more' buttons.

Now create another dataset for our ADLS gen 2 , We are going to fetch the data from external source and copy it into ADLS gen 2 "Staging" container.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, there's a sidebar with icons for Home, Pipelines, Datasets, Data flows, and Power Query. The main area shows a list of datasets under 'Factory Resources'. A new dataset named 'ds_adls' is being created. The 'Connection' tab is selected, showing 'Linked service' set to 'ls_adl' and 'File path' set to 'landing'. The 'Format' tab is also visible. On the right, a search bar says 'Search factory and documentation' and a user 'nadarajan073@gmail.com' is logged in. Below the search bar, there's a 'New dataset' section with a sub-section for 'Select a data store' where 'gen' is selected. Buttons for 'All', 'Azure', 'Database', 'File', and 'Generic protocol' are shown. Below that, two options are listed: 'Azure Data Lake Storage Gen2' (selected) and 'Magento (Legacy)'. At the bottom right are 'Continue', 'Cancel', and 'Back' buttons.

This screenshot continues the dataset creation process. The 'Format' tab is selected in the ds_adls dataset configuration. A grid of format options is shown: Avro, Binary, CSV (selected), Excel, Iceberg, and JSON. Each option has a small icon and a label. The 'Continue' button is at the bottom of the configuration pane.

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

Validate all Publish all

Factory Resources <>

- Pipelines 1
 - pl_capstone
- Change Data Capture (preview) 0
- Datasets 3
 - ds_adls
 - ds_externalSource
 - ds_sql_db
 - Data flows 0
 - Power Query 0

ds_adls x ds_sql_db

DelimitedText ds_adls

CSV

Connection Schema Parameters

Linked service * ls_adls

File path landing

Compression type No comp

Column delimiter (,) Comma (,

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

Browse

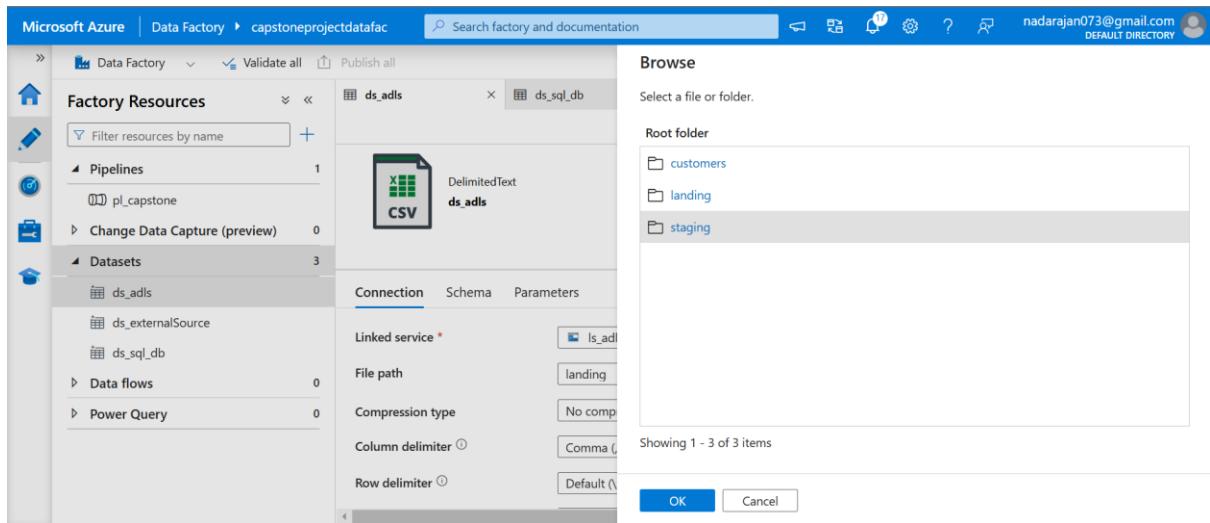
Select a file or folder.

Root folder

- customers
- landing
- staging

Showing 1 - 3 of 3 items

OK Cancel



Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

Validate all Publish all

Factory Resources <>

- Pipelines 1
 - pl_capstone
- Change Data Capture (preview) 0
- Datasets 3
 - ds_adls
 - ds_externalSource
 - ds_sql_db
 - Data flows 0
 - Power Query 0

ds_adls x ds_sql_db

DelimitedText ds_adls

CSV

Set properties

Name ds_adls_customers

Linked service * ls_adls

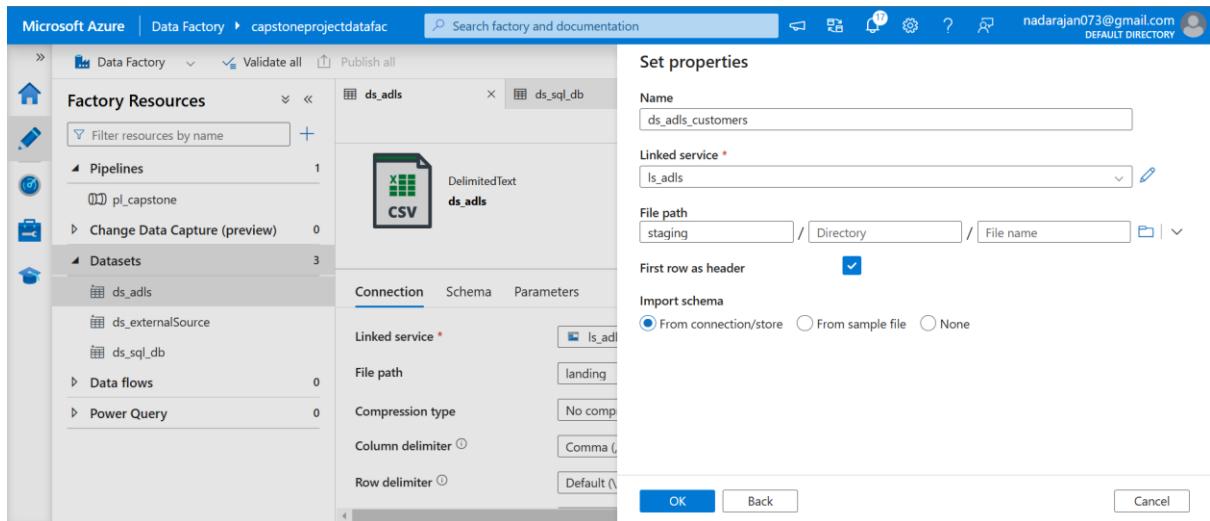
File path staging / Directory / File name

First row as header

Import schema

From connection/store From sample file None

OK Back Cancel



Publish all the changes.

Microsoft Azure | Data Factory > capstoneprojectdatafac | Search factory and documentation

Validate all Publish all

Factory Resources <>

- Pipelines 1
 - pl_capstone
- Change Data Capture (preview) 0
- Datasets 4
 - ds_adls
 - ds_adls_customers
 - ds_externalSource
 - ds_sql_db
 - Data flows 0
 - Power Query 0

ds_adls x ds_sql_db x pl_capstone x ds_adls

DelimitedText ds_adls_customers

CSV

Connection Schema Parameters

Linked service * ls_adls Test connection

File path staging / Directory / customers.csv

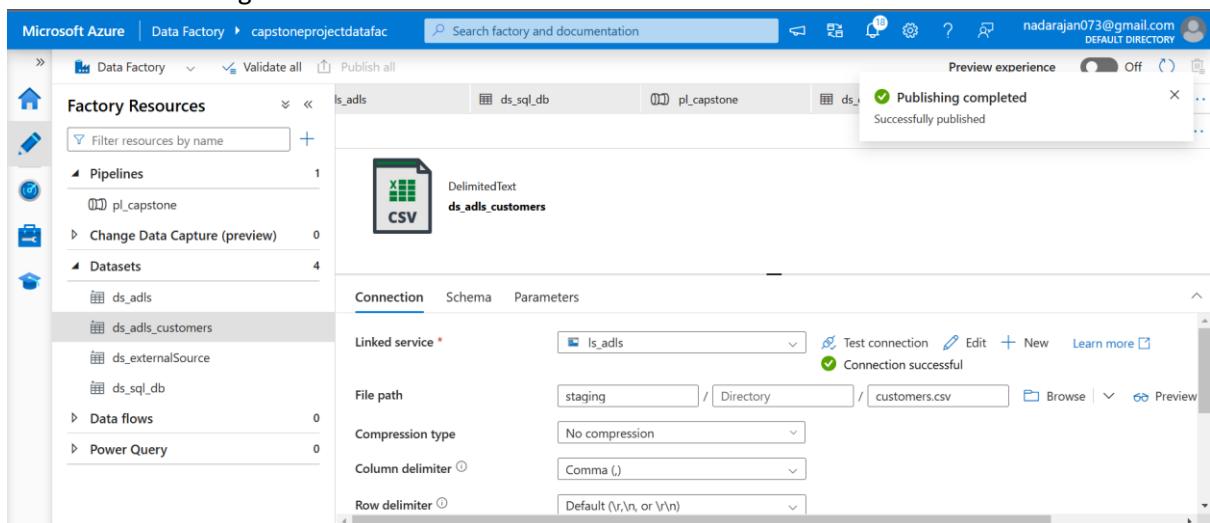
Compression type No compression

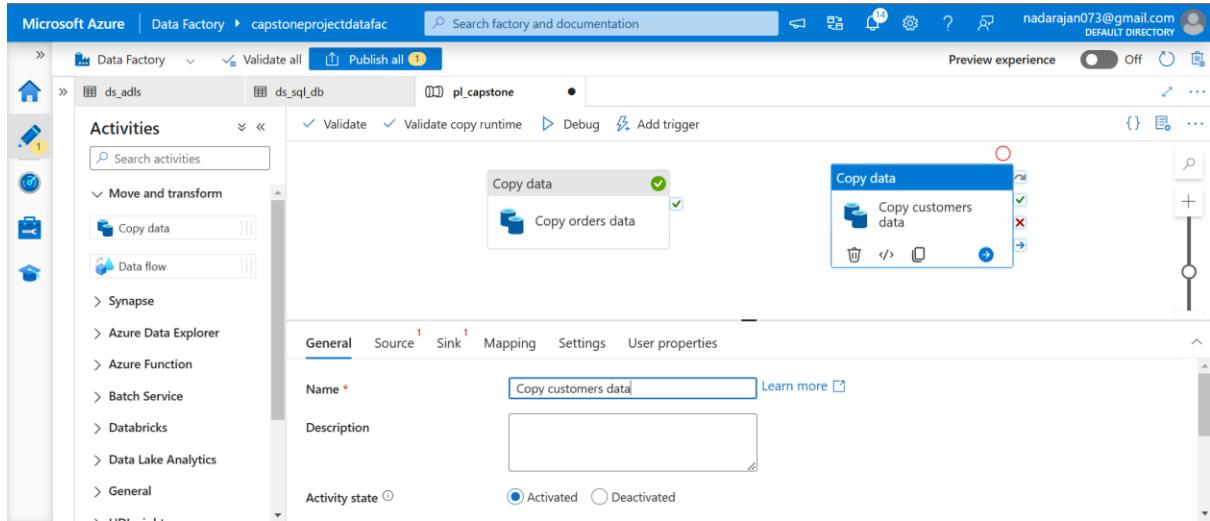
Column delimiter (,) Comma (,

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

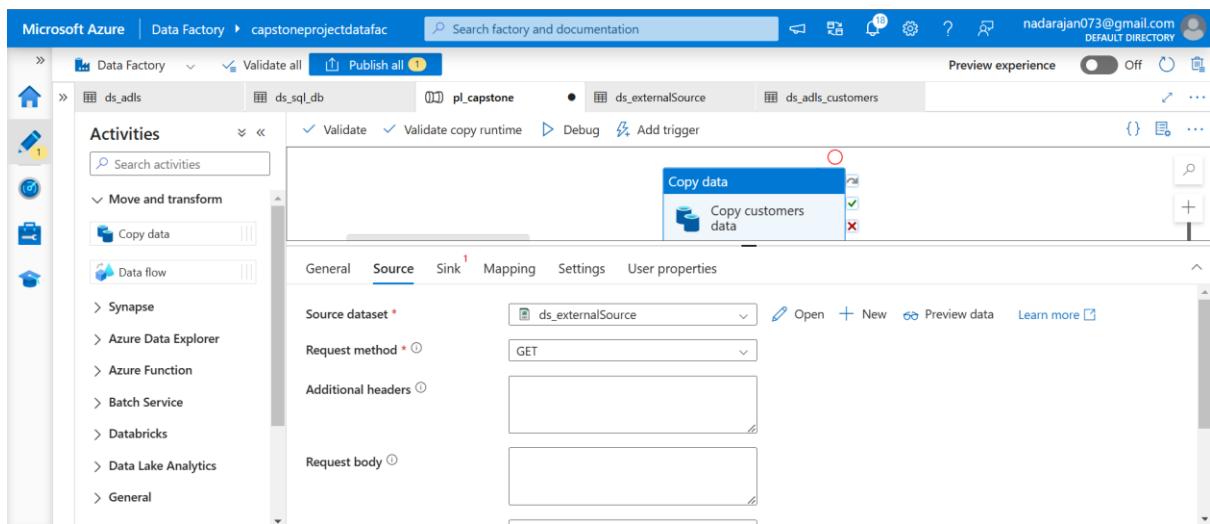
Preview experience Off

✓ Publishing completed Successfully published





Select the source as the external source dataset and the sink as the ADLS gen 2 staging container dataset.



The screenshot shows the Microsoft Azure Data Factory interface. In the center, a 'Copy data' activity is selected. The 'Sink' tab is active, showing the destination dataset as 'ds_adls_customers'. Below this, 'Copy behavior' is set to 'Select...', and 'Max concurrent connections' and 'Block size (MB)' are configured. The 'Mapping' tab is visible at the bottom.

Verify the mappers schema .

The screenshot shows the 'Mapping' tab for the 'Copy data' activity. It lists 10 columns from the source dataset, each mapped to a corresponding column in the sink dataset 'ds_adls_customers'. All columns are of type 'String'.

| Source | Type | Destination | Type |
|----------------|--------|----------------|--------|
| customer_id | String | customer_id | String |
| customer_fname | String | customer_fname | String |
| customer_lname | String | customer_lname | String |
| username | String | username | String |
| password | String | password | String |
| address | String | address | String |
| city | String | city | String |
| state | String | state | String |
| pincode | String | pincode | String |

The screenshot shows the Microsoft Azure Data Factory interface after publishing. A message in the top right corner indicates 'Publishing completed' and 'Successfully published'. The 'Copy data' activity is now listed in the activities list under 'Move and transform'.

Link both copy to activities to run one after another.

The screenshot shows the Microsoft Azure Data Factory pipeline editor. A pipeline named 'pl_capstone' is displayed, containing two 'Copy data' activities. The first activity, 'Copy orders data', has a green checkmark indicating success. An arrow points from its output to the second activity, 'Copy customers data', which also has a green checkmark. The pipeline status is 'Succeeded'. Below the pipeline, a table lists the run details for both activities:

| Activity name | Activity status | Activity type | Run start | Duration | Integration runtime |
|---------------------|-----------------|---------------|------------------------|----------|------------------------|
| Copy customers data | Succeeded | Copy data | 1/12/2025, 12:23:43 PM | 12s | AutoResolveIntegration |
| Copy orders data | Succeeded | Copy data | 1/12/2025, 12:23:43 PM | 32s | AutoResolveIntegration |

If there is no issue then , we can able to see the customers data in the staging container.

The screenshot shows the Microsoft Azure Storage Explorer interface. A container named 'staging' is selected. Inside, there is a single blob file named 'customers.csv'. The blob details are as follows:

| Name | Modified | Access tier | Archive status | Blob type | Size |
|---------------|-------------------------|----------------|----------------|------------|------|
| customers.csv | 1/12/2025, 12:23:54 ... | Hot (Inferred) | | Block blob | 1.13 |

<https://portal.azure.com/#>

Our cluster might be turned off by this time. Turn on the cluster again.

The screenshot shows the Databricks Compute page. On the left sidebar, under the 'Compute' section, the 'Compute' option is selected. The main area displays a table titled 'Compute' with the following columns: State, Name, Policy, Runtime, Active ... (with ellipsis), Active c..., Active D..., Source, Creator, Notebo..., and a small icon. A single row is present in the table, labeled 'Nadarajan S's Cluster'. The table has a header row with sorting icons for State and Name. At the bottom right of the table, there are navigation buttons for 'Previous', 'Next', and '20 / page'.

To create a scope in the databricks , enter "#secrets/createScope" at last of the databricks URL homepage

The screenshot shows the Databricks homepage. The URL in the browser address bar is `https://adb-3262179368725353.13.azuredatabricks.net/?o=3262179368725353#secrets/createScope`. The page features a 'Welcome to Databricks' header, a search bar, and a 'Set up your workspace' callout with a 'Get started' button. Below this, there are tabs for 'Recents', 'Favorites', 'Popular', 'Mosaic AI', and 'What's new'. A recent item, 'Capstone Notebook' by 'Users/nadarajan073@gmail.com', is listed with a timestamp of '58 minutes ago' and a 'Notebook' link.

We will navigate to the create scope page inside the databricks.

We can get the DNS name and the Resource Id , In the “Properties” option inside the keyvault.

The screenshot shows two side-by-side browser windows. The top window is the 'Key vaults' blade in Microsoft Azure, specifically for the 'capstonevaultn' key vault. It displays properties like Name (capstonevaultn), Sku (Standard), Location (centralindia), and various URLs and IDs. The 'Vault URI' is highlighted with a yellow box, showing the URL <https://capstonevaultn.vault.azure.net/>. The bottom window is the 'Create Secret Scope' dialog in Databricks. It shows the 'Scope Name' field set to 'capstonescope', the 'Manage Principal' dropdown set to 'All workspace users', and the 'Azure Key Vault' section with the 'DNS Name' set to <https://capstonevaultn.vault.azure.net/> and 'Resource ID' set to </subscriptions/a376f639-0268-48a9-84e5-2b741923eb4/resourceGroups/capstone>.

We need to turn the DBFS setting to view the folders in the DBFS system.

The screenshot shows the Databricks interface. On the left is the navigation sidebar with options like Workspace, Recents, Catalog, Workflows, Compute, SQL, and Data Engineering. The main area is titled 'Welcome to Databricks' and shows a 'Set up your workspace' guide with a 'Get started' button. Below it are tabs for Recents, Favorites, Popular, Mosaic AI, and What's new. A recent notebook entry is listed: 'Untitled Notebook 2025-01-12 08:47:31' by 'nadarajan073@gmail.com'. The top right corner shows the user profile 'Nadarajan S' and the email 'nadarajan073@gmail.com'. A vertical menu on the far right includes 'Settings', 'Azure Portal', 'Privacy Policy', 'Previews', 'Send feedback', and 'Log out'.

Create another table in the Azure SQL DB to save the results to it.

By following the below approach , we can get connection to the Azure SQL database. Through that we can create the DF and write out derived results also to it.

```

01:21 PM (1s)
1

dbServer = 'capstonedbserver'
dbName = 'capstonedb'
dbPort = '1433'
dbUser = dbutils.secrets.get(scope="newScope", key="sqlDbCredName")
dbPassword = dbutils.secrets.get(scope="newScope", key="sqlDbCredPass")

connectionURL = 'jdbc:sqlserver://{}.{}.database.windows.net:{};database={};user={};'.format(dbServer, dbPort, dbName,
dbUser)

connectionProperties = {'password': dbPassword, 'driver': 'com.microsoft.sqlserver.jdbc.SQLServerDriver'}

orders_df = spark.read.jdbc(url=connectionURL, table='dbo.orders', properties=connectionProperties)

orders_df: pyspark.sql.dataframe.DataFrame = [order_id: integer, order_date: string ... 2 more fields]

```

Now start writing the transformations for our usecase in the databricks.

Create Mount Points.

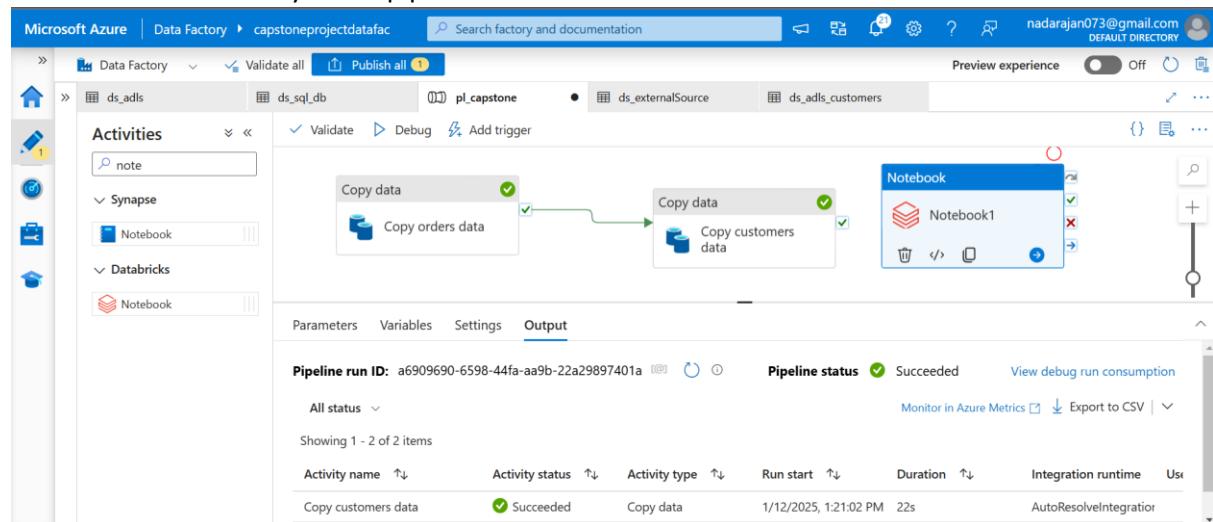
Create the dataframes for both the data's and perform required join operations.

Save the result to the Azure SQL database again.

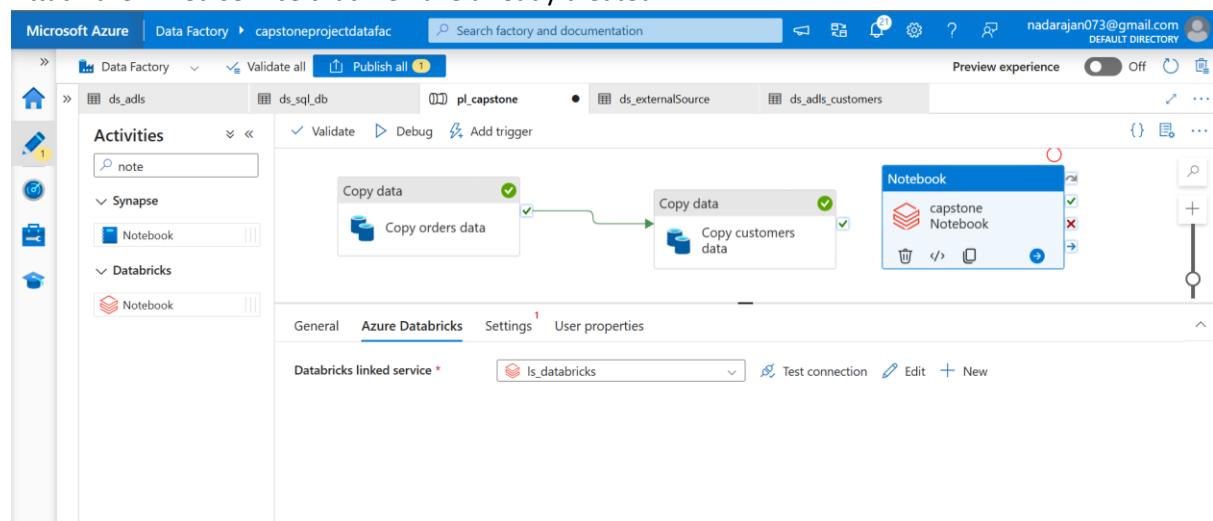
Please get the databricks code in the below drive.

https://drive.google.com/drive/folders/1_dAVqM4PzQik9Ou2pTrfPzFazHbc4hBM?usp=sharing

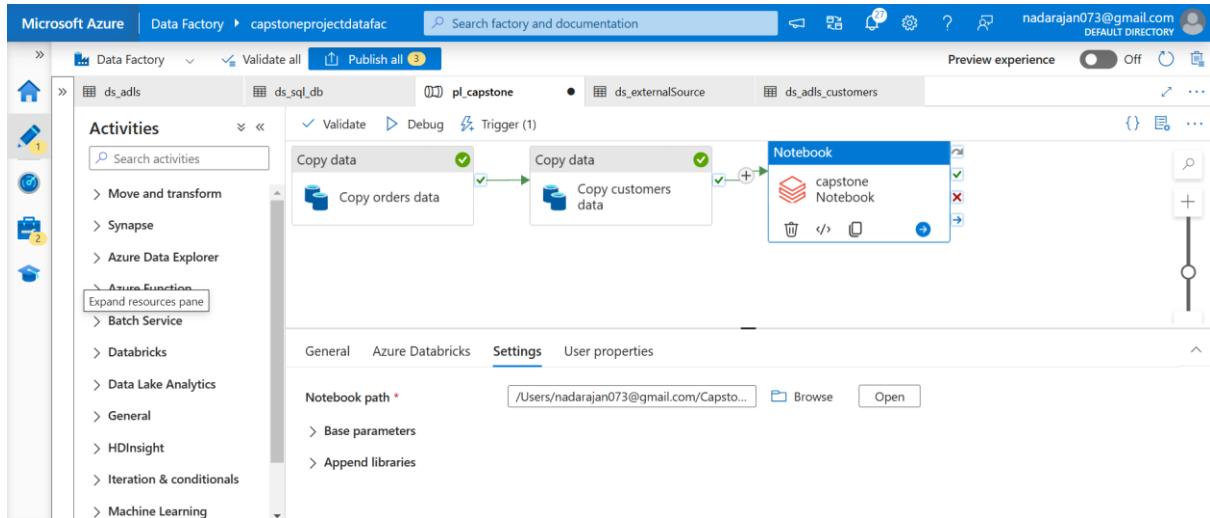
Add a Databricks activity in our pipeline.



Attach the linked service that we have already created.



Link the notebook activity to our copy to activity. So that once the copy to activity completed our notebook will be start running.



Create a trigger to run our pipeline. Type of the trigger should be storage event.

So that our pipeline will be start running once the orders_new data arrived to our container.

The screenshot shows the 'Add trigger' dialog for the 'Copy orders data' activity. The dialog is titled 'New trigger' and contains fields for 'Name' (set to 'Capstone_trigger'), 'Description', 'Type' (set to 'Storage events'), 'Account selection method' (radio button selected for 'From Azure subscription'), 'Azure subscription' (set to 'Free Trial (a376f639-0268-48a9-84e5-2b741923ebc4)'), and 'Storage account name' (set to 'capstonestorageacct'). At the bottom are 'Continue' and 'Cancel' buttons.

The screenshot shows the 'Triggers' page in the Microsoft Azure Data Factory portal. The 'Triggers' section lists one item: 'capstone_t...'. On the right, the 'Edit trigger' dialog is open, showing configuration for an 'Azure subscription' (selected 'Free Trial (a376f639-0268-48a9-84e5-2b741923ebc4)'), 'Storage account name' ('capstonestorageacct'), 'Container name' ('landing'), and 'Event' type ('Blob created'). The dialog also includes fields for 'Blob path begins with' and 'Blob path ends with'. At the bottom are 'Continue' and 'Cancel' buttons.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, a sidebar navigation menu includes General, Factory settings, Connections, Integration runtimes, Microsoft Purview, Source control, Author (with Triggers selected), Global parameters, and Data flow libraries. The main content area is titled 'Triggers' and contains a table with one item: 'capstone_trigger' (Type: Storage events). To the right, a 'Data preview' section displays a filtered list of blobs from the 'landing' container, ending in '.csv'. One blob, 'orders_new.csv', is shown in the preview table.

orders data will be looks like this.

The screenshot shows a Microsoft Excel spreadsheet titled 'orders_new'. The table has four columns: 'order_id', 'order_date', 'customer_id', and 'order_status'. The data consists of 20 rows, each containing a unique order ID, a timestamp for the order date, a customer ID, and a status. The statuses include 'CLOSED', 'PENDING_PAYMENT', 'COMPLETE', and 'PROCESSING'.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|----|----------|------------|-------------|-----------------|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | order_id | order_date | customer_id | order_status | | | | | | | | | | | | |
| 2 | 1 | 00:00:0 | 11599 | CLOSED | | | | | | | | | | | | |
| 3 | 2 | 00:00:0 | 256 | PENDING_PAYMENT | | | | | | | | | | | | |
| 4 | 3 | 00:00:0 | 12111 | COMPLETE | | | | | | | | | | | | |
| 5 | 4 | 00:00:0 | 8827 | CLOSED | | | | | | | | | | | | |
| 6 | 5 | 00:00:0 | 11318 | COMPLETE | | | | | | | | | | | | |
| 7 | 6 | 00:00:0 | 7130 | COMPLETE | | | | | | | | | | | | |
| 8 | 7 | 00:00:0 | 4530 | COMPLETE | | | | | | | | | | | | |
| 9 | 8 | 00:00:0 | 2911 | PROCESSING | | | | | | | | | | | | |
| 10 | 9 | 00:00:0 | 5657 | PENDING_PAYMENT | | | | | | | | | | | | |
| 11 | 10 | 00:00:0 | 5648 | PENDING_PAYMENT | | | | | | | | | | | | |
| 12 | 11 | 00:00:0 | 918 | PAYMENT_REVIEW | | | | | | | | | | | | |
| 13 | 12 | 00:00:0 | 1837 | CLOSED | | | | | | | | | | | | |
| 14 | 13 | 00:00:0 | 9149 | PENDING_PAYMENT | | | | | | | | | | | | |
| 15 | 14 | 00:00:0 | 9842 | PROCESSING | | | | | | | | | | | | |
| 16 | 15 | 00:00:0 | 2568 | COMPLETE | | | | | | | | | | | | |
| 17 | 16 | 00:00:0 | 7276 | PENDING_PAYMENT | | | | | | | | | | | | |
| 18 | 17 | 00:00:0 | 2667 | COMPLETE | | | | | | | | | | | | |
| 19 | 18 | 00:00:0 | 1205 | CLOSED | | | | | | | | | | | | |
| 20 | 19 | 00:00:0 | 9488 | PENDING_PAYMENT | | | | | | | | | | | | |
| 21 | 20 | 00:00:0 | 9198 | PROCESSING | | | | | | | | | | | | |

Exit the notebook once the transformations are completed.

The screenshot shows a Jupyter Notebook cell with the following Scala code:

```

Run cell (Ctrl + Enter)
▶ 01:45 PM (2s) 6

result.write.jdbc(url = connectionURL , table = "dbo.aggregateResult" , mode = "overwrite", properties =
connectionProperties)
dbutils.notebook.exit("Success")

▶ (5) Spark Jobs
Notebook exited: Success

```

Below the cell, keyboard shortcuts are listed:

- [Shift+Enter] to run and move to next cell
- [Ctrl+Shift+P] to open the command palette
- [Esc H] to see all keyboard shortcuts

Trigger was paused right now. Start the trigger now.

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has a 'Triggers' section selected. A table lists one item: 'capstone_trig' (Type: Storage events, Status: Stopped). The top navigation bar includes 'Validate all', 'Publish all', and a search bar for 'Search factory and documentation'. The top right shows the user's email (nadarajan073@gmail.com) and a 'DEFAULT DIRECTORY' indicator.

To the pipeline ,upload the orders data to landing container.

The screenshot shows the Azure Storage Blob Container 'landing'. The left sidebar has an 'Overview' section selected. The main area shows a table with no results. On the right, an 'Upload blob' dialog is open, showing a file 'orders_new.csv' selected for upload. The top navigation bar includes 'Upgrade', 'Search resources, services, and docs (G+)', 'Copilot', and the user's email (nadarajan073@gmail.com).

The screenshot shows the same Azure Storage Blob Container 'landing' after the file has been uploaded. The 'Upload blob' dialog is closed, and a success message 'Successfully uploaded blob(s)' is displayed. The main area shows the uploaded file 'orders_new.csv' in the table. The top navigation bar includes 'Upgrade', 'Search resources, services, and docs (G+)', 'Copilot', and the user's email (nadarajan073@gmail.com).

Trigger will start the pipeline once the file got uploaded.

The screenshot shows the Microsoft Azure Data Factory interface. The top navigation bar includes 'Microsoft Azure', 'Data Factory', 'capstoneprojectdatafac', 'Search factory and documentation', and user information 'nadarajan073@gmail.com DEFAULT DIRECTORY'. The main area is titled 'Pipeline runs' with tabs for 'Triggered' (selected), 'Debug', 'Rerun', 'Cancel options', 'Refresh', 'Edit columns', 'List' (selected), and 'Gantt'. Filter options include 'Filter by run ID or name: Chennai, Kolkata, Mu...', 'Last 24 hours', 'Pipeline name: All', 'Status: All', 'Runs: Latest runs', 'Triggered by: All', and 'Add filter'. A message 'Showing 1 - 1 items' is displayed. The table lists one run: 'pl_capstone' started at '1/12/2025, 2:52:52 PM' with a duration of '3s', triggered by 'capstone_trigger', and is currently 'In progress'.

The screenshot shows the Microsoft Azure Data Factory interface. The top navigation bar includes 'Microsoft Azure', 'Data Factory', 'capstoneprojectdatafac', 'Search factory and documentation', and user information 'nadarajan073@gmail.com DEFAULT DIRECTORY'. The left sidebar shows 'Runs' selected, with sub-options: 'Pipeline runs' (selected), 'Trigger runs', 'Change Data Capture (previ...)', 'Runtimes & sessions', 'Integration runtimes', 'Data flow debug', 'Notifications', and 'Alerts & metrics'. The main area is titled 'All pipeline runs > pl_capstone - Activity runs'. It displays a Gantt chart with three activities: 'Copy data' (Copy orders data) → 'Copy data' (Copy customers data) → 'Notebook' (capstone Notebook). Below the chart, it says 'Pipeline run ID e296bbfe-bb4c-4270-9796-ec18c400b7c3'. The table below shows 'Activity runs' with three entries: 'capstone Notebook' (In progress, Notebook, 1/12/2025, 2:53:43 PM, 6s), 'Copy customers data' (Succeeded, Copy data, 1/12/2025, 2:53:29 PM, 13s, AutoResolveIntegratio...), and 'Copy orders data' (Succeeded, Copy data, 1/12/2025, 2:52:54 PM, 34s, AutoResolveIntegratio...).

The screenshot shows the Microsoft Azure Data Factory interface. The top navigation bar includes 'Microsoft Azure', 'Data Factory', 'capstoneprojectdatafac', 'Search factory and documentation', and user information 'nadarajan073@gmail.com DEFAULT DIRECTORY'. The left sidebar shows 'Runs' selected, with sub-options: 'Pipeline runs' (selected), 'Trigger runs', 'Change Data Capture (previ...)', 'Runtimes & sessions', 'Integration runtimes', 'Data flow debug', 'Notifications', and 'Alerts & metrics'. The main area is titled 'All pipeline runs > pl_capstone - Activity runs'. It displays a Gantt chart with three activities: 'Copy data' (Copy orders data) → 'Copy data' (Copy customers data) → 'Notebook' (capstone Notebook). A modal window titled 'Details' is open over the first activity, showing 'Monitor real-time execution in Azure Databricks', 'Duration: 00:00:16', and 'Run page URL: https://adb-3262179368725353.13.azuredatabricks.net/?o=3262179368725353#job/187366531180929/run/5790118'. The table below shows 'Activity runs' with three entries: 'capstone Notebook' (In progress, Notebook, 1/12/2025, 2:53:43 PM, 35s), 'Copy customers data' (Succeeded, Copy data, 1/12/2025, 2:53:29 PM, 13s, AutoResolveIntegratio...), and 'Copy orders data' (Succeeded, Copy data, 1/12/2025, 2:52:54 PM, 34s, AutoResolveIntegratio...).

Output

```

Joined_df.createOrReplaceTempView("Joined_df")
result = spark.sql("select city, count(order_id) as Total_orders from Joined_df
group by city order by Total_orders desc")
result.write.jdbc(url = connectionURL , table = "dbo.aggregateResult" , mode =
"overwrite", properties = connectionProperties)
dutils.notebook.exit("Success")

```

Notebook exited: Success

Task run

Details

- Job ID: 329570619836830
- Task run ID: 165159066062771
- Run as: Nadarajan S
- Started: Jan 12, 2025, 03:10 PM
- Ended: Jan 12, 2025, 03:11 PM

All pipeline runs > pl_capstone - Activity runs

Pipeline run ID: 99f1090c-e31f-4b96-a73e-9d9570cb7f45

| Activity name | Activity status | Activity type | Run start | Duration | Integration runtime | User proc |
|---------------------|-----------------|---------------|-----------------------|----------|------------------------|-----------|
| capstone Notebook | Succeeded | Notebook | 1/12/2025, 3:10:42 PM | 34s | AutoResolveIntegration | |
| Copy customers data | Succeeded | Copy data | 1/12/2025, 3:10:30 PM | 12s | AutoResolveIntegration | |
| Copy orders data | Succeeded | Copy data | 1/12/2025, 3:09:56 PM | 33s | AutoResolveIntegration | |

Once the pipeline run completed successfully , we can see the customers data in the staging container.

staging

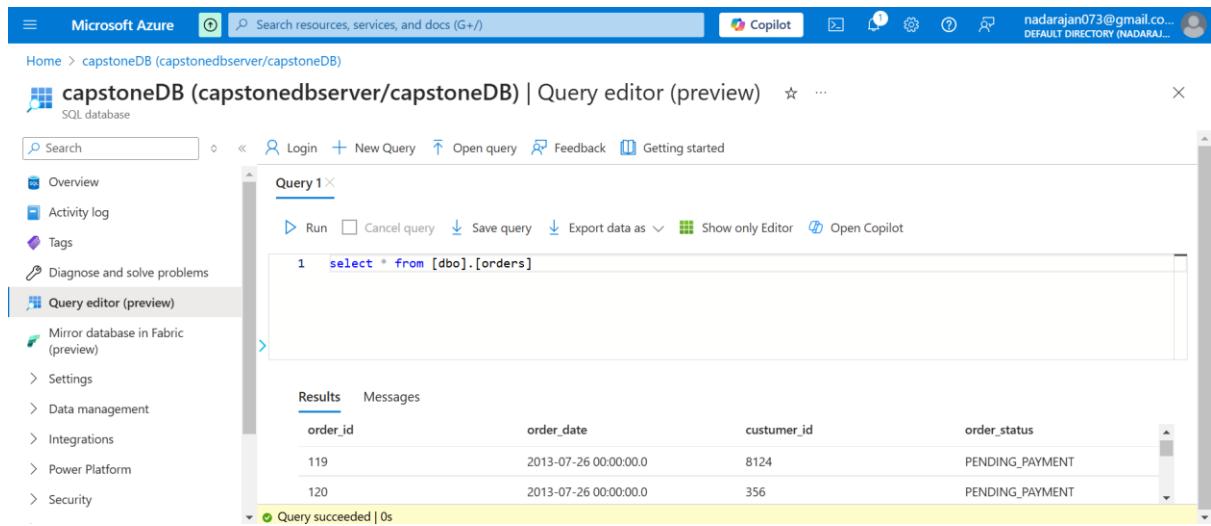
Overview

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: staging

| Name | Modified | Access tier | Archive status | Blob type | Size |
|---------------|-----------------------|----------------|----------------|------------|------|
| customers.csv | 1/12/2025, 3:10:40 PM | Hot (Inferred) | | Block blob | 1.13 |

once the pipeline runs successfully , we can see the orders data and the result data in the Azure SQL DB



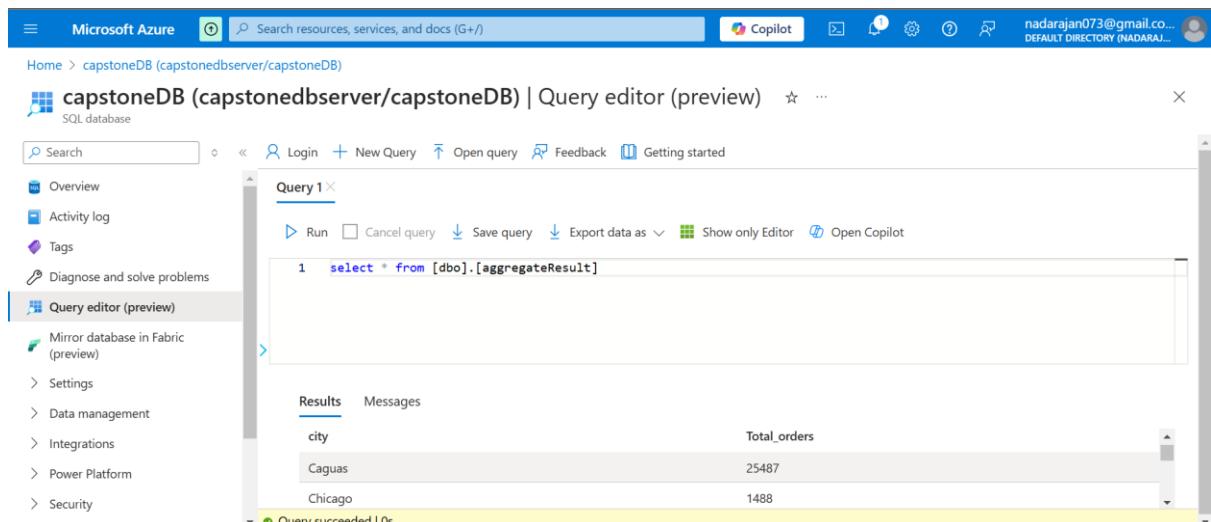
The screenshot shows the Microsoft Azure portal interface for a SQL database named 'capstoneDB'. The left sidebar has 'Query editor (preview)' selected. The main area displays a query window with the following content:

```
1 select * from [dbo].[orders]
```

The results pane shows the following data:

| order_id | order_date | customer_id | order_status |
|----------|---------------------|-------------|-----------------|
| 119 | 2013-07-26 00:00:00 | 8124 | PENDING_PAYMENT |
| 120 | 2013-07-26 00:00:00 | 356 | PENDING_PAYMENT |

A message at the bottom indicates: 'Query succeeded | 0s'



The screenshot shows the Microsoft Azure portal interface for a SQL database named 'capstoneDB'. The left sidebar has 'Query editor (preview)' selected. The main area displays a query window with the following content:

```
1 select * from [dbo].[aggregateResult]
```

The results pane shows the following data:

| city | Total_orders |
|---------|--------------|
| Caguas | 25487 |
| Chicago | 1488 |

A message at the bottom indicates: 'Query succeeded | 0s'