

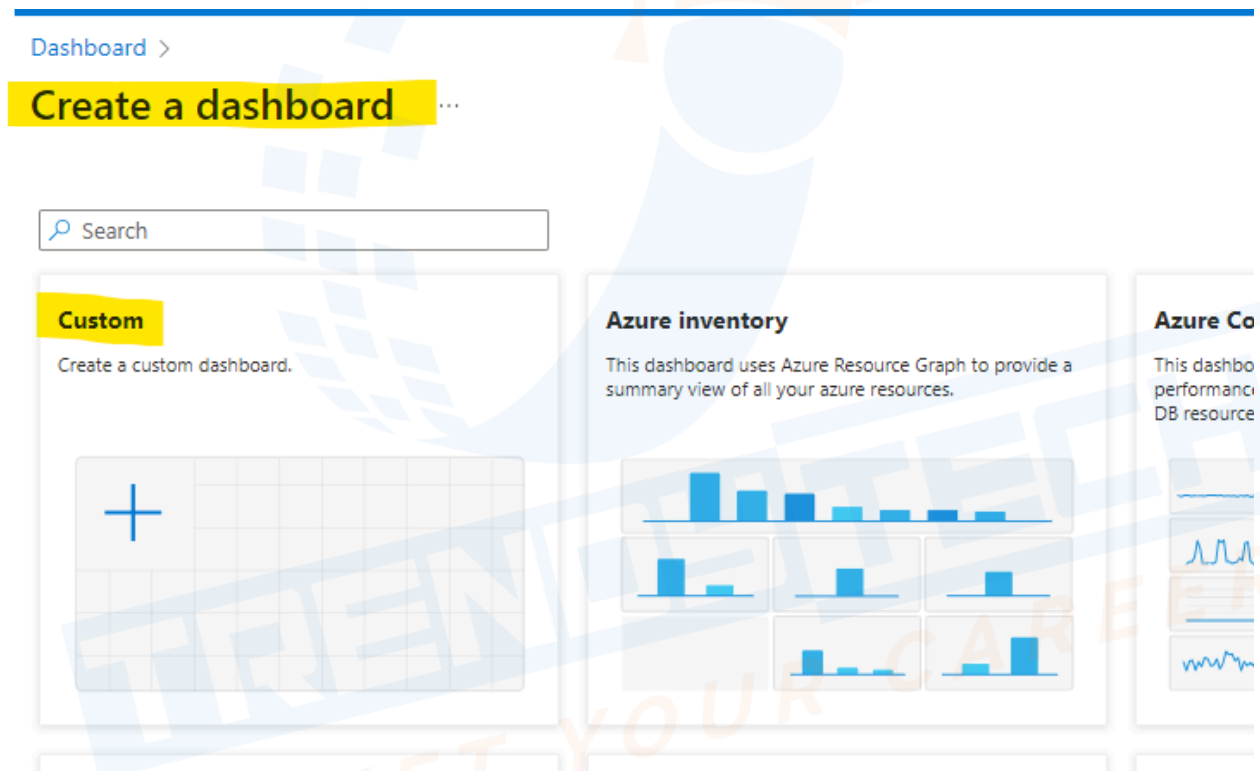
Use case 2 - Ingesting Data from Amazon S3 to Azure ADLS Gen2

Ingestion:

1. Create a Resource Group and pin it to the dashboard.

Create a Dashboard and a Resource Group for the project to organize the resources related to the project at one place.

To create dashboard: Click on Dashboard => Create => Custom =>



To create Resource group (trendytech-rg):

[Home](#) > [Resource groups](#) >

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 * ⓘ Pay-As-You-Go (136f20b5-00f7-4eb5-a52e-0843e7ad1034) ▼
Resource group * ⓘ Trendytech-rg ✓

Resource details

Region * ⓘ (Asia Pacific) Central India ▼

[Review + create](#)

[< Previous](#)

[Next : Tags >](#)

[Home](#) > [Resource groups](#) >

Create a resource group ...

✓ Validation passed.

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

Basics

Subscription Pay-As-You-Go
Resource group Trendytech-rg
Region Central India

Tags

None

[Create](#)

[< Previous](#)

[Next >](#)

[Download a template for automation](#)

2. Create Storage Accounts a Normal Blob Storage Account (trendytechsa101)

[Home](#) > [Storage accounts](#) >

Create a storage account

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

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more about Azure storage accounts](#)

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *

Pay-As-You-Go (136f20b5-00f7-4eb5-a52e-0843e7ad1034)

Resource group *

Trendytech-rg

[Create new](#)

Instance details

Storage account name *

trendytechsa101

Region *

(Asia Pacific) Central India

[Deploy to an Azure Extended Zone](#)

Performance *

☒ Standard: Recommended for most scenarios (general-purpose v2 account)
 ☐ Premium: Recommended for scenarios that require low latency.

Redundancy *

Locally-redundant storage (LRS)

Previous

Next

Review + create

[Home](#) > [Storage accounts](#) >

Create a storage account

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

[View automation template](#)

Basics

Subscription	Pay-As-You-Go
Resource group	Trendytech-rg
Location	Central India
Storage account name	trendytechsa101
Performance	Standard
Replication	Locally-redundant storage (LRS)

Advanced

Enable hierarchical namespace	Disabled
-------------------------------	----------

Previous

Next

Create

Click on review and create and then create the storage account.

Note: Allow the “blob anonymous access” for the storage account

The screenshot shows the 'Configuration' page for the storage account 'trendytechsa101'. The left sidebar lists various settings, with 'Configuration' highlighted. The main area displays the following settings:


- Account kind:** StorageV2 (general purpose v2)
- Performance:** Standard (selected), Premium
- Secure transfer required:** Disabled, Enabled (selected)
- Allow Blob anonymous access:** Disabled, Enabled (selected). A note states: "Some blobs may become anonymously readable."
- Allow storage account key access:** Disabled, Enabled (selected)
- Allow recommended upper limit for shared access signature (SAS) expiry interval:** Disabled (selected), Enabled

Now create the folder `order_input` and upload the file `orders.csv` in it.

The 'New container' dialog box is shown with the following details:

- Name:** ordersinput (with a green checkmark)
- Anonymous access level:** Container (anonymous read access for containers and blobs) (with a green checkmark)
- Warning:** 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.
- Advanced:** A collapsed section indicated by a downward arrow.
- Create:** A green button at the bottom left.
- Give feedback:** A link with a speech bubble icon at the bottom right.

[Home](#) > [trendytechsa101](#) | Containers >

 **ordersinput**

Container

◊ <<

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Upload

Change access level

Refresh

Delete


Change tier

Acquire lease

Break lease

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: ordersinput

+ Add filter

	Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/>	 orders.csv	5/24/2024, 3:52:04 PM	Hot (Inferred)		Block blob

3. ADLS Gen2 Storage Account (with hierarchical namespace enabled)(trendytechadls101)

[Home](#) > [Storage accounts](#) >

Create a storage account

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *

Pay-As-You-Go (136f20b5-00f7-4eb5-a52e-0843e7ad1034)

Resource group *

Trendytech-rg

Create new

Instance details

Storage account name * ⓘ

trendytechadls101

Region * ⓘ

(Asia Pacific) Central India

Deploy to an Azure Extended Zone

Performance * ⓘ

☒ Standard: Recommended for most scenarios (general-purpose v2 account)

☐ Premium: Recommended for scenarios that require low latency.

Redundancy * ⓘ

Locally-redundant storage (LRS)

Previous

Next

Review + create

In advance tab, select option “**Enable hierarchical namespace**”, refer attached screenshot

The screenshot shows the 'Advanced' tab of the Azure Storage account configuration page. The 'Security' section includes options for 'Require secure transfer for REST API operations' (checked), 'Allow enabling anonymous access on individual containers' (unchecked), 'Enable storage account key access' (checked), and 'Default to Microsoft Entra authorization in the Azure portal' (unchecked). The 'Minimum TLS version' is set to 'Version 1.2' and 'Permitted scope for copy operations (preview)' is set to 'From any storage account'. The 'Hierarchical Namespace' section has a description and the 'Enable hierarchical namespace' checkbox is checked. At the bottom, there are 'Previous', 'Next', and 'Review + create' buttons.

Click on review and create and then create the storage account.

Note: Allow the “blob anonymous access” for the storage account

The screenshot shows the 'Configuration' tab of the Azure Storage account configuration page. The left sidebar lists various settings categories, with 'Configuration' selected. The main panel shows the 'Account kind' as 'StorageV2 (general purpose v2)' and 'Performance' as 'Standard'. The 'Secure transfer required' is set to 'Enabled'. The 'Allow Blob anonymous access' checkbox is checked, with a note stating 'Some blobs may become anonymously readable.' The 'Allow storage account key access' is also set to 'Enabled'. The 'Allow recommended upper limit for shared access signature (SAS) expiry interval' is set to 'Disabled'.

Create the container “landing” in the adls gen2 storage and in landing create the folders orders and order_items

New container

Name *

landing

Anonymous access level ⓘ

Container (anonymous read access for containers and blobs)

⚠ 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.

Advanced

Create

Give feedback

Home > trendytechadls101 | Containers >

landing

Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

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

Location: landing

Search blobs by prefix (case-sensitive)

Name	Modified	Access tier	Archive status	Blob type
order_items				
orders				

4. Create an Azure Data Factory within a Resource Group (trendytech-df-1001)

Home > Data factories >

Create Data Factory ...

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

One-click to create data factory with sample pipeline and datasets. [Try it](#)

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 * ⓘ

[Previous](#) [Next](#) [Review + create](#)

Home > Data factories >

Create Data Factory ...

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

[View automation template](#)

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

Basics

Subscription	Pay-As-You-Go
Resource group	Trendytech-rg
Name	trendytech-df-1001
Region	East US
Version	V2

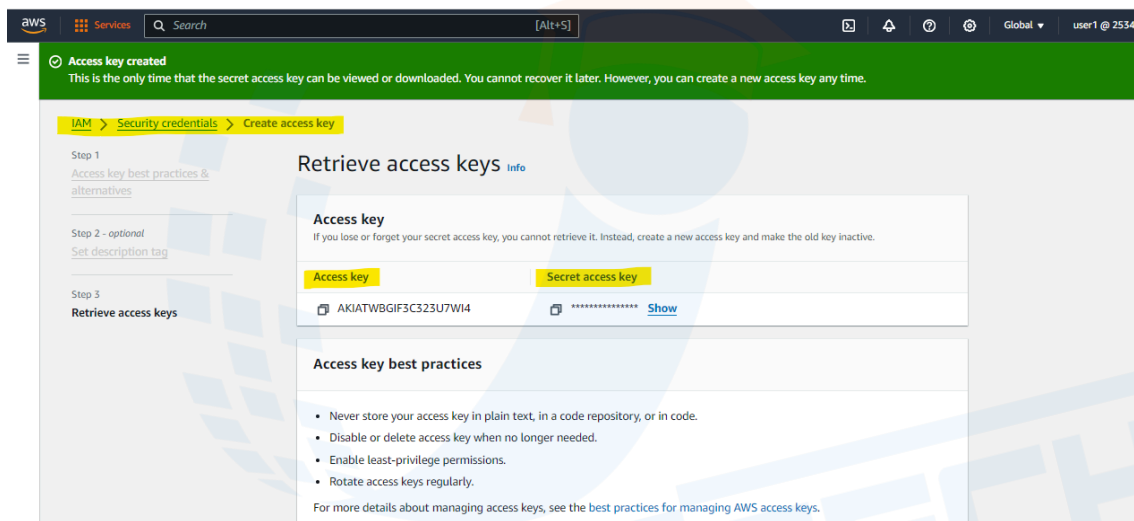
[Previous](#) [Next](#) [Create](#)

5. Create an Amazon Web Services Account with S3 Storage

First create the IAM user and then login to AWS console using that user and create access key.

To create an access key you can refer to this document.

https://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_access-keys.html



Create a bucket (trendytech-azure) within the S3 Storage and add the order_items.csv file to the bucket.

While creating the s3 bucket, deselect the option "Block public access" and keep other settings as it is and create the bucket.

Create bucket [Info](#)

Buckets are containers for data stored in S3.

General configuration

AWS Region

US East (N. Virginia) us-east-1

Bucket type [Info](#)



General purpose

Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.



Directory - New

Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.

Bucket name [Info](#)

trendytech-azure

Bucket name must be unique within the global namespace and follow the bucket naming rules. [See rules for bucket naming](#)

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

[Choose bucket](#)

Object Ownership [Info](#)

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.



ACLs disabled (recommended)

All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.



ACLs enabled

Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

Object Ownership

Bucket owner enforced

Block Public Access settings for this bucket

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. [Learn more](#)



Block all public access

Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.



Block public access to buckets and objects granted through new access control lists (ACLs)

S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.

Then upload the file `order_items.csv`.

Add tag

Default encryption [Info](#)
Server-side encryption is automatically applied to new objects stored in this bucket.

Encryption type | [Info](#)

- ☒ Server-side encryption with Amazon S3 managed keys (SSE-S3)
- ☐ Server-side encryption with AWS Key Management Service keys (SSE-KMS)
- ☐ Dual-layer server-side encryption with AWS Key Management Service keys (DSSE-KMS)
Secure your objects with two separate layers of encryption. For details on pricing, see [DSSE-KMS pricing](#) on the [Storage](#) tab of the [Amazon S3 pricing page](#).

Bucket Key
Using an S3 Bucket Key for SSE-KMS reduces encryption costs by lowering calls to AWS KMS. S3 Bucket Keys aren't supported for DSSE-KMS. [Learn more](#)

- ☐ Disable
- ☒ Enable

► **Advanced settings**

After creating the bucket, you can upload files and folders to the bucket, and configure additional bucket settings.

[Cancel](#) [Create bucket](#)

And once it is created upload file `order_items.csv` in it.

Amazon S3 > Buckets > trendytech-azure-new1

trendytech-azure-new1 [Info](#)

[Objects](#) | [Properties](#) | [Permissions](#) | [Metrics](#) | [Management](#) | [Access Points](#)

Objects (1) [Info](#)

[Copy](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#) [Upload](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	order_items.csv	csv	May 24, 2024, 15:11:19 (UTC+05:30)	5.2 MB	Standard

6. Create a Key vault (trendytech-keyvault1)

Create a key vault

Project details

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

Subscription *	Pay-As-You-Go (136f20b5-00f7-4eb5-a52e-0843e7ad1034) ▼
Resource group *	Trendytech-rg ▼

[Create new](#)

Instance details

Key vault name * ⓘ	trendytech-keyvault1 ✓
Region *	East US ▼
Pricing tier * ⓘ	Standard ▼

Recovery options

Soft delete protection will automatically be enabled on this key vault. This feature allows you to recover or permanently delete a key vault and secrets for the duration of the retention period. This protection applies to the key vault and the secrets stored within the key vault.

[Previous](#) [Next](#) [Review + create](#)

Note: In Access configuration select option “vault access policy”

[Basics](#) [Access configuration](#) [Networking](#) [Tags](#) [Review + create](#)

Configure data plane access for this key vault

To access a key vault in data plane, all callers (users or applications) must have proper authentication

Permission model

Grant data plane access by using a [Azure RBAC](#) or [Key Vault access policy](#)

- ☐ Azure role-based access control (recommended) ⓘ
- ☒ Vault access policy ⓘ

Resource access

- ☐ Azure Virtual Machines for deployment ⓘ
- ☐ Azure Resource Manager for template deployment ⓘ
- ☐ Azure Disk Encryption for volume encryption ⓘ

[Previous](#) [Next](#) [Review + create](#)

[Home](#) > [Key vaults](#) >

Create a key vault

Basics Access configuration Networking Tags **Review + create**

Review + create

Basics

Subscription	Pay-As-You-Go
Resource group	Trendytech-rg
Key vault name	trendytech-keyvault1
Region	East US
Pricing tier	Standard
Soft-delete	Enabled
Purge protection during retention period	Disabled
Days to retain deleted vaults	90 days

Access configuration

Azure Virtual Machines for deployment	Disabled
---------------------------------------	----------

Previous

Next

Create

And click on the option “create”.

a. Store access key and access key secret in keyvault

In the key vault select the option “Generate/import”.

Home > trendytech-keyvault1 | Overview > trendytech-keyvault1 | Secrets >



Create a secret

...

Upload options

Manual

Name * ⓘ

AWS-s3k-access-key

Secret value * ⓘ

.....

Content type (optional)

Set activation date ⓘ

☐

Set expiration date ⓘ

☐

Enabled

Yes

No

Tags

0 tags

Create

Cancel

Home > trendytech-keyvault1 | Overview > trendytech-keyvault1



trendytech-keyvault1 | Secrets ☆ ...

Key vault

Search

+ Generate/import



Refresh



Restore Backup



View sample code



Manage deleted secrets

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Access policies

Events

> Objects

> Settings

> Monitoring

> Automation

> Help

Name

Type

Status

There are no secrets available.

First we will store the access key

Home > trendytech-keyvault1 | Overview > trendytech-keyvault1 | Secrets >



Create a secret ...

Upload options	<input type="text" value="Manual"/>
Name * ⓘ	<input type="text" value="AWS-s3k-access-key"/>
Secret value * ⓘ	<input type="password" value="....."/>
Content type (optional)	<input type="text"/>
Set activation date ⓘ	<input type="checkbox"/>
Set expiration date ⓘ	<input type="checkbox"/>
Enabled	<input checked="" type="radio"/> Yes <input type="radio"/> No
Tags	0 tags

Create

Cancel

And click on option create to create an access key.

Home > trendytech-keyvault1 | Overview > trendytech-keyvault1 | Secrets >



Create a secret ...

Upload options	<input type="text" value="Manual"/>
Name * ⓘ	<input type="text" value="AWS-s3-access-secret"/>
Secret value * ⓘ	<input type="password" value="....."/>
Content type (optional)	<input type="text"/>
Set activation date ⓘ	<input type="checkbox"/>
Set expiration date ⓘ	<input type="checkbox"/>
Enabled	<input checked="" type="radio"/> Yes <input type="radio"/> No
Tags	0 tags

Create

Cancel

Home > trendytech-keyvault1 | Overview > trendytech-keyvault1

trendytech-keyvault1 | Secrets

Key vault

Search

+ Generate/import Refresh Restore Backup View sample code Manage deleted secrets

The secret 'AWS-s3-access-secret' has been successfully created.

Name	Type	Status	Expiration date
AWS-s3-access-secret		✓ Enabled	
AWS-s3-access-key		✓ Enabled	

Creating the secret 'AWS-s3-access-secret'.
The secret 'AWS-s3-access-secret' has been successfully created.

In Key vault grant access permission to the Data Factory service principal under the access policies.

trendytech-keyvault1 | Access policies

Key vault

Search

+ Create Refresh Delete Edit

Access policies enable you to have fine grained control over access to vault items. [Learn more](#)

Search Permissions: All Type: All

Showing 1 to 1 of 1 records.

<input type="checkbox"/>	Name ↑↓	Email ↑↓	Key Permissions	Secret Permissions	Certificate
<input type="checkbox"/>	USER				
<input type="checkbox"/>	Arati Hatti	aratishatti15_gmail.com#...	Get, List, Update, Create, ...	Get, List, Set, Delete, Rec...	Get, List, U

Home > Key vaults > trendytech-keyvault1 | Access policies

Create an access policy

trendytech-keyvault1

1 Permissions 2 Principal 3 Application (optional) 4 Review + create

Configure from a template

Key, Secret, & Certificate Management

Key permissions	Secret permissions	Certificate permissions
Key Management Operations	Secret Management Operations	Certificate Management Operations
<input checked="" type="checkbox"/> Select all	<input checked="" type="checkbox"/> Select all	<input checked="" type="checkbox"/> Select all
<input checked="" type="checkbox"/> Get	<input checked="" type="checkbox"/> Get	<input checked="" type="checkbox"/> Get
<input checked="" type="checkbox"/> List	<input checked="" type="checkbox"/> List	<input checked="" type="checkbox"/> List
<input checked="" type="checkbox"/> Update	<input checked="" type="checkbox"/> Set	<input checked="" type="checkbox"/> Update
<input checked="" type="checkbox"/> Create	<input checked="" type="checkbox"/> Delete	<input checked="" type="checkbox"/> Create
<input checked="" type="checkbox"/> Import	<input checked="" type="checkbox"/> Recover	<input checked="" type="checkbox"/> Import

Previous Next

✓ Permissions **2 Principal** ③ Application (optional) ④ Review + create

Only 1 principal can be assigned per access policy.
Use the new embedded experience to select a principal. The previous popup experience can be accessed here. [Select a principal](#)

🔍

trendytech


✕



trendytech-df-1001
aa95f711-f99f-45f2-b357-88a884c867b4

▲
▼

Selected item



trendytech-df-1001
aa95f711-f99f-45f2-b357-88a884c867b4

Previous

Next

[Home](#) > [Key vaults](#) > [trendytech-keyvault1](#) | [Access policies](#) >

Create an access policy ...

trendytech-keyvault1

✓ Permissions ✓ Principal ✓ Application (optional) **4 Review + create**

Key Permissions

Key Management Operations	All selected
Cryptographic Operations	None selected
Privileged Key Operations	None selected
Rotation Policy Operations	All selected

Secret Permissions

Secret Management Operations	All selected
Privileged Secret Operations	None selected

Certificate Permissions

Certificate Management Operations	All selected
-----------------------------------	--------------

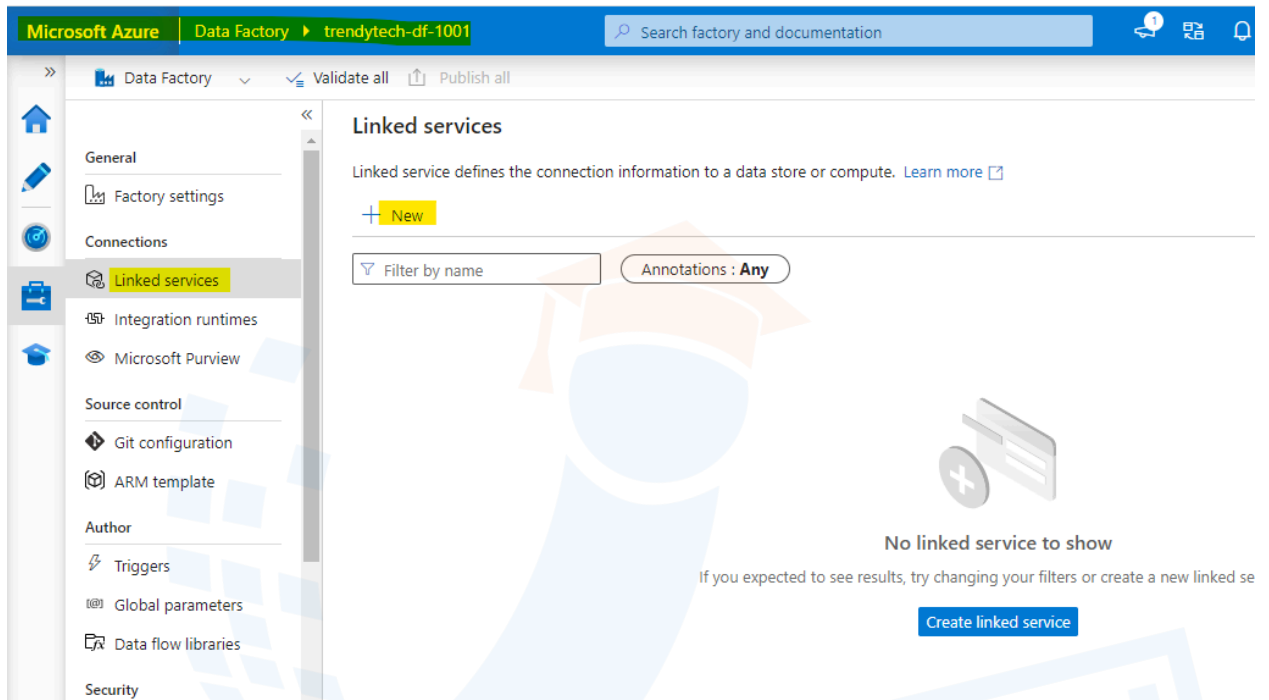
Previous

Create

Launch Azure Studio and Create Linked Services:

Linked service - For key vault , AWS s3, Datalake, blob


Go to Monitor => Linked Service => New



We will create linked service for Blob storage:

TRENDY TECH
UPLIFT YOUR CAREER!

New linked service

 Azure Blob Storage [Learn more](#)

Connect via integration runtime * ⓘ
AutoResolveIntegrationRuntime

Authentication type
Account key

Connection string Azure Key Vault

Account selection method ⓘ
☒ From Azure subscription ☐ Enter manually

Azure subscription ⓘ
Pay-As-You-Go (136f20b5-00f7-4eb5-a52e-0843e7ad1034)

Storage account name *
trendytechsa101


Additional connection properties
[+ New](#)

Create [Back](#)

✓ Connection successful
[Test connection](#) [Cancel](#)

We will create linked service for Azure Data Lake Storage gen2:

New linked service

 Azure Data Lake Storage Gen2 [Learn more](#)

Connect via integration runtime * ⓘ
AutoResolveIntegrationRuntime

Authentication type
Account key

Account selection method ⓘ
☒ From Azure subscription ☐ Enter manually

Azure subscription ⓘ
Pay-As-You-Go (136f20b5-00f7-4eb5-a52e-0843e7ad1034)

Storage account name *
trendytechadls101

Test connection ⓘ
☒ To linked service ☐ To file path


Annotations

Create [Back](#)

✓ Connection successful
[Test connection](#) [Cancel](#)


We will create linked service for Key Vault:

New linked service

 Azure Key Vault

Azure key vault selection method ⓘ
☒ From Azure subscription ☐ Enter manually

Azure subscription ⓘ
Pay-As-You-Go (136f20b5-00f7-4eb5-a52e-0843e7ad1034) ▼

Azure key vault name *
trendytech-keyvault1 ▼ 

[Edit key vault](#)

Authentication method
System Assigned Managed Identity ▼


Managed identity name: **trendytech-df-1001**
Managed identity object ID: **5caa1289-9bb6-4abe-b6d7-333100e65fee**
Grant Data Factory service managed identity access to your Azure Key Vault. [Learn more](#) ⓘ


Test connection
☒ To linked service ☐ To secret

Annotations

Create

Back


 Connection successful

 Test connection

Cancel

We will create linked service for Amazon S3:

New linked service

 Amazon S3 [Learn more](#) ⓘ

Name *
ls_aws_s3


Description

Connect via integration runtime * ⓘ
AutoResolveIntegrationRuntime ▼

Authentication type
Access key ▼

Access key ID

Azure Key Vault

AKV linked service * ⓘ
ls_trendytech_kv ▼ 

Authentication type

Access key

Access key ID Azure Key Vault

AKV linked service * ⓘ

ls_trendytech_kv

Secret name * ⓘ

AWS-s3k-access-key

☐ Edit

Secret version ⓘ

Latest version

☐ Edit

Secret access key Azure Key Vault

AKV linked service * ⓘ

ls_trendytech_kv

Secret name * ⓘ

AWS-s3-access-secret

☐ Edit

Secret version ⓘ

Latest version

☐ Edit

Service URL ⓘ

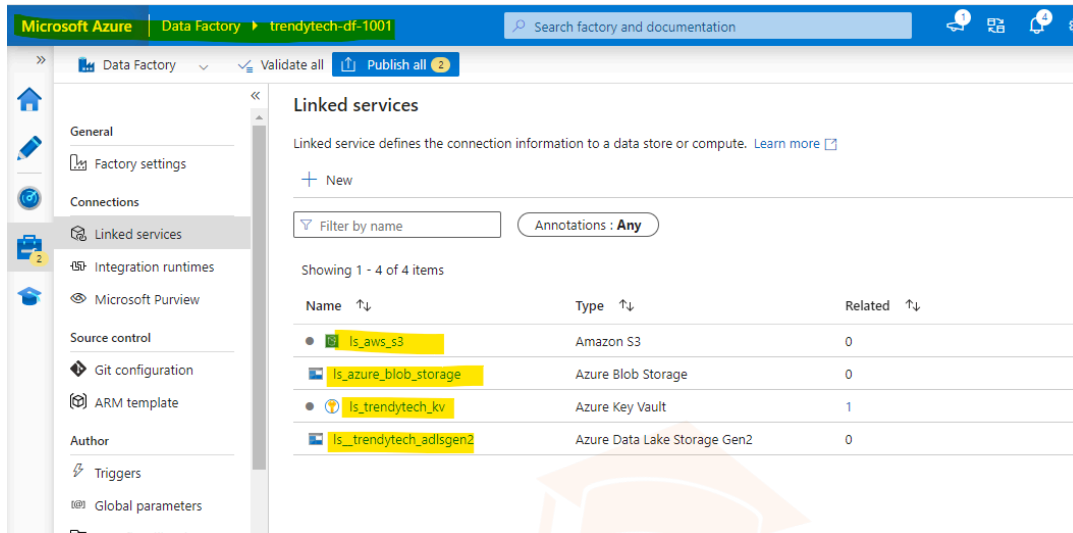
https://s3.amazonaws.com

Test connection ⓘ

✔ Connection successful

Create Back Test connection Cancel

These are all the link services for all the resources that we have have created



Create Datasets:

Source: Amazon S3, blob


1. Creating dataset for orders_items.csv in s3:

New dataset


In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store


All
Azure
Database
File
Generic protocol
NoSQL
Services and apps



Amazon S3



Amazon S3 Compatible



Oracle Cloud Storage (S3 API)

Continue

Cancel

Select option "delimited text"

Set properties

Name

Linked service *

File path
 / /

First row as header ☒

Import schema
☒ From connection/store ☐ From sample file ☐ None

2. For orders.csv file in blob storage

Click on New dataset => select "Azure Blob Storage" => delimited text file =>

Set properties

Name

Linked service *

File path
 / /

First row as header ☒

Import schema
☒ From connection/store ☐ From sample file ☐ None

Sink: ADLS Gen2 (landing folder - orders, order_items)

For orders:

Set properties

Name
ds_orders_adls_gen2_sink

Linked service *
ls_trendytech_adlsgen2

File path
landing / orders / File name

First row as header ☒

Import schema
☒ From connection/store ☐ From sample file ☐ None

OK Back Cancel

For order_items:

Set properties

Name
ds_order_items_adlsgen2_sink

Linked service *
ls_trendytech_adlsgen2

File path
landing / order_items / File name

First row as header ☒

Import schema
☒ From connection/store ☐ From sample file ☐ None

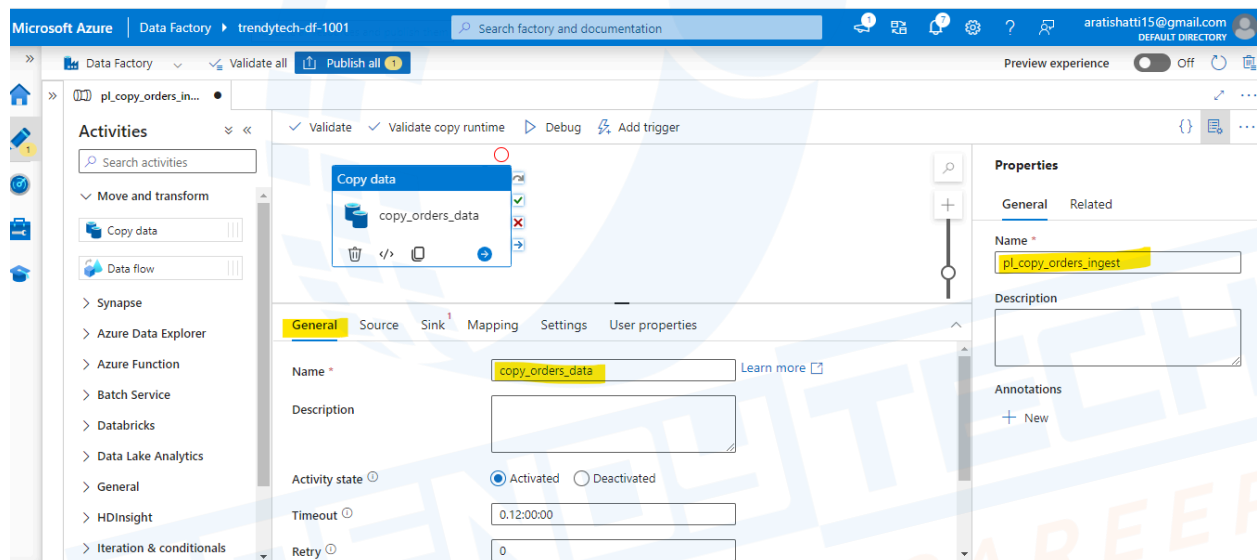
OK Back Cancel

Note: Publish all the resources that you have created.

Now we will create pipelines for ingestion, processing and execution.

1. For copying the order.csv from s3 => adlsgen2 (Debug and verify it)





Add the copy data activity and set source and sink dataset as shown below.



✓ Validate ✓ Validate copy runtime ▶ Debug ⚡ Add trigger

Copy data

copy_orders_data



General

Source

Sink





Mapping

Settings

User properties

Source dataset *

ds_trendytechsa_orders



File path type

☒ File path in dataset ☐ Prefix ☐ Wildcard file path ☐ List of files ⓘ

Filter by last modified ⓘ

Start time (UTC)

End time (UTC)

Properties


General Related

Name *

pl_copy_orders_ingest

Description





Annotations

 New

✓ Validate ✓ Validate copy runtime ▶ Debug ⚡ Add trigger

Copy data

copy_orders_data



General

Source

Sink




Mapping

Settings

User properties

Sink dataset *

ds_orders_adls_gen2_sink



Copy behavior ⓘ

Select...

Max concurrent connections ⓘ

Block size (MB) ⓘ

Properties


General Related

Name *

pl_copy_orders_ingest

Description

Annotations

 New

Now click on Debug and check the pipeline

✓ Validate ▶ Debug ⚡ Add trigger

Copy data ✓
copy_orders_data ✓

Properties
General Related

Name *
pl_copy_orders_ingest

Description

Annotations
+ New

Parameters Variables Settings **Output**

Pipeline run ID: 2d73a75a-d640-43cc-b693-906f29b66955 ⓘ ⌂ ⓘ
Pipeline status ✓ Succeeded View debug run consumption
All status ▾ Monitor in Azure Metrics ⓘ Export to CSV ▾
Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration
copy_orders_data	✓ Succeeded	Copy data	5/24/2024, 4:47:49 PM	15s

2. For copying the order_items.csv from blob => adlsgen2
(Debug and verify it)

Add the copy data activity and set source and sink dataset as shown below.

✓ Validate ✓ Validate copy runtime ▶ Debug ⚡ Add trigger

Copy data
copy order items

Properties
General Related

Name *
pl_copy_order_items

Description

Annotations
+ New

General Source 1 Sink 1 Mapping Settings User properties

Name *
copy order items Learn more ⓘ

Description

Activity state ⓘ
☒ Activated ☐ Deactivated

Timeout ⓘ
0.12:00:00

✓ Validate ✓ Validate copy runtime ▶ Debug ⚡ Add trigger

Copy data

copy order items

General **Source** Sink¹ Mapping Settings User properties

Source dataset * ds_order_items_s3_source

Open + New Preview data Learn more

File path type ☒ File path in dataset ☐ Prefix ☐ Wildcard file path ☐ List of files ⓘ

Filter by last modified ⓘ

Start time (UTC) End time (UTC)

Properties

General Related

Name * pl_copy_order_items

Description

Annotations + New

Now click on Debug and check the pipeline

✓ Validate ✓ Validate copy runtime ▶ Debug ⚡ Add trigger

Copy data

copy order items

General Source **Sink** Mapping Settings User properties

Sink dataset * ds_order_items_adlsgen2_sink

Open + New Learn more

Copy behavior ⓘ Select...

Max concurrent connections ⓘ

Block size (MB) ⓘ

Properties

General Related

Name * pl_copy_order_items

Description

Annotations + New

Now click on Debug and check the pipeline

The screenshot shows the Azure Data Factory (ADF) console. On the left, the 'Activities' pane is open, showing a search bar and a list of activities under 'Move and transform', including 'Copy data' and 'Data flow'. The main canvas displays a pipeline named 'pl_copy_order_items' with a single 'Copy data' activity. The activity is configured with a source and a sink, and its status is 'Succeeded'. Below the canvas, the 'Output' tab is selected, showing the 'Pipeline run ID' (690811f4-7ee2-4163-9e77-dd06a127c8c9) and the 'Pipeline status' (Succeeded). A table below shows the activity details:

Activity name	Activity status	Activity type	Run start	Duration
copy order items	Succeeded	Copy data	5/24/2024, 4:54:28 PM	20s

3. Create new pipeline for processing the data

But before creating a pipeline we will create a Data Flow.

For processing data using Data Flow

In Adls gen 2 create container “**output**” and segregate data into 3 categories:

High value orders (>500)

Low value orders (<=500)

Erroneous orders (no order amount).

Create 3 folders for the above categories in the output folder of the storage container as “**high_value_orders**”, “**low_value_orders**”, “**erroneous**”.

Home > trendytechadls101

trendytechadls101 Containers

Storage account

Search

+ Container Change access level Restore containers Refresh Delete Give

Search containers by prefix

Name	Last modified
<input type="checkbox"/> \$logs	5/24/2024, 2:17:01 PM
<input type="checkbox"/> landing	5/24/2024, 4:09:12 PM

New container

Name *
output

Anonymous access level
Container (anonymous read access for containers and blobs)

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.

Advanced

Create Give feedback

Click on the Add directory and add the folder high_value, low value and erroneous as shown.

Home > trendytechadls101 Containers >

output Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease

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

Search blobs by prefix (case-sensitive)

Name	Modified	Access tier	Archive status	Blob type
<input type="checkbox"/> erroneous				
<input type="checkbox"/> high_value				
<input type="checkbox"/> low_value				

Now create a New Data Flow:

-Add Source for order_item dataset (in adls gen2) and import the schema (*.csv).

Note: for order_item dataset also import the schema as shown below.

Microsoft Azure | Data Factory | trendytech-df-1001

Search factory and documentation

Validate all Publish all

Factory Resources

Filter resources by name

- Pipelines
 - pl_copy_order_items
 - pl_copy_orders_ingest
- Change Data Capture (preview) 0
- Datasets 4
 - ds_order_items_adlsngen2_sink
 - ds_order_items_s3_source**
 - ds_orders_adls_gen2_sink
 - ds_trendytechsa_orders
- Data flows 1
 - dataflow1
- Power Query 0

dataflow1

DelimitedText
ds_order_items_s3_source

Connection Schema Parameters

Import schema Clear

Column name	Type
order_item_id	String
order_item_order_id	String
order_item_product_id	String
order_item_quantity	String
order_item_subtotal	String
order_item_product_price	String

Preview the data to check if it is in the desired form. Change data types if required (e.g., order_item_quantity from string to integer, order_item_subtotal and order_item_product_price to float). Refer the below screenshot

dataflow1

Validate Data flow debug Debug Settings

orderitems

Source settings Source options **Projection** Optimize Inspect

Define default format Detect data type Import projection Reset schema

Column name	Type	Format
order_item_id	abc string	Specify format
order_item_order_id	abc string	Specify format
order_item_product_id	abc string	Specify format
order_item_quantity	123 integer	Specify format
order_item_subtotal	1.2f float	Specify format
order_item_product_price	1.2f float	Specify format

And preview the data

Source settings	Source options	Projection	Optimize	Inspect	Data preview	
-----------------	----------------	------------	----------	---------	--------------	--

Number of rows	INSERT 100	UPDATE 0	DELETE 0	UPSERT 0	LOOKUP 0
----------------	------------	----------	----------	----------	----------

Refresh	Typecast	Modify	Map drifted	Statistics	Remove	Export to CSV
---------	----------	--------	-------------	------------	--------	---------------

order_item_id	order_item_order...	order_item_prod...	order_item_quan...	order_item_subt...
1	1	957	1	299.98
2	2	1073	1	199.99
3	2	502	5	250.0
4	2	403	1	129.99
5	4	897	2	49.98
6	4	365	5	299.95
7	4	502	3	150.0

Turn on the Debug mode to check if the processing is as per the requirement.

- Use aggregate transformation activity to calculate the subtotal of respective order items.

Group by column "order_item_order_id"

Aggregate settings	Optimize	Inspect	Data preview	Previous	Next
--------------------	----------	---------	--------------	----------	------

Output stream name *	orderitemaggregate	Learn more
Description	Add aggregate columns	Reset
Incoming stream *	orderitems	
<div>Group by</div> <div>Aggregates</div>		
Columns	Name as	
abc order_item_order_id	order_item_order_id	+

Aggregate settings

Optimize

Inspect

Data preview

Previous

Next

Output stream name *

orderitemaggregate

Learn more

Description

Aggregating data by 'order_item_order_id' producing columns 'order_item_subtotal'

Reset

Incoming stream *

orderitems

Group by

Aggregates

Grouped by: order_item_order_id

+ Add

Clone

Delete

Open expression builder

☐ Column

☐ Expression

☐ order_item_subtotal

sum(order_item_subtotal)

1.2

+

Data Factory

Dataflow expression builder

orderitemaggregate

Aggregate Columns

+ Create new

123 order_item_subtotal

Column name *

order_item_subtotal

Expression

sum(order_item_subtotal)

+ - * / || && ! ^ == === <=>

Expression elements

Expression values

All

Filter by keyword

Save and finish

Cancel

Clear contents

Click on save and finish and Click on Data preview as shown below.

Aggregate settings

Optimize

Inspect

Data preview

← Previous

Next →

Number of rows

+

INSERT

100

*

UPDATE

0

✗

DELETE

0

+

UPSERT

0

🔍

LOOKUP

0

✗

ERROR

0

TOTAL

341

↻ Refresh

▼

Typecast

▼

🔧 Modify

▼

📄 Map drifted

📊 Statistics

✕ Remove

↓ Export to CSV

▼

↑↓

order_item_order_id

abc

↑↓

order_item_subtotal

1.2

↑↓

+

1

299.9800109863281

+

2

579.9800109863281

+

4

699.8500099182129

+

5

1129.8600387573242

+

7

579.9200134277344

+

8

729.8400115966797

+

9

599.9600067138672

+

10

651.920015335083

+

11

919.7899932861328

- Add Source for orders dataset and import the schema (*.csv).
Preview the data to check if it is in the desired form.

Again import the schema of orders dataset as shown below.

Data Factory

Validate all

Publish all

Factory Resources

Filter resources by name

Pipelines

2

pl_copy_order_items

pl_copy_orders_ingest

Change Data Capture (preview)

0

Datasets

4

ds_order_items_adlsgen2_sink

ds_order_items_s3_source

ds_orders_adls_gen2_sink

ds_trendytechsa_orders

Data flows

1

dataflow1

Power Query

0

DelimitedText

ds_orders_adls_gen2_sink

Connection

Schema

Parameters

Import schema

Clear

Column name	Type
order_id	String
order_date	String
order_customer_id	String
order_status	String

Now add the source in Data flow.

Source settings Source options Projection Optimize Inspect Data preview ●

Output stream name * [Learn more](#)

Description [Reset](#)

Source type * Dataset Inline

Dataset * [Test connection](#) [Open](#) [New](#) Connection successful

Options ☒ Allow schema drift ⓘ ☐ Infer drifted column types ⓘ ☐ Validate schema ⓘ

Import the schema and preview the data

Source settings Source options **Projection** Optimize Inspect Data preview ●

[Define default format](#) [Detect data type](#) [Import projection](#) [Reset schema](#)

Column name	Type	Format
order_id	abc string	Specify format
order_date	abc string	Specify format
order_customer_id	abc string	Specify format
order_status	abc string	Specify format

-Use the transformation Join to perform a full outer join for orders and order_item datasets on the condition order_item_id == order_id.

dataflow1 ds_order_items_adls... pl_copy_order_items ds_order_items_s3

✓ Validate Data flow debug Debug Settings

orderitems Import data from ds_order_items_adlsgen2_sink

orderitemaggregate Aggregating data by 'order_item_order_id' producing columns 'order_item_subtotal'

orders Import data from ds_orders_adls_gen2_sink

join1 Columns: 2 total

Join settings Optimize Inspect Data preview

Description Full outer join on 'orderitemaggregate' and 'orders' Reset

Left stream * orderitemaggregate

Right stream * orders

Join type * Full outer Inner Left outer Right outer Custom (cross)

Use fuzzy matching ☐

Join conditions * Left: orderitemaggregate's column Right: orders's column

abc order_item_order_id == abc order_id + -

Click on the option “data preview” and preview the data

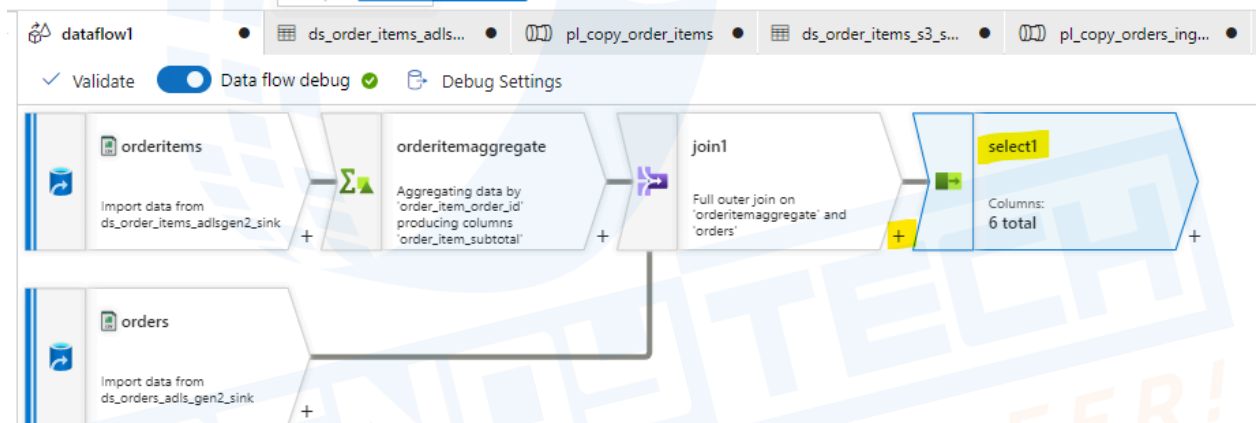
Join settings	Optimize	Inspect	Data preview	
---------------	----------	---------	---------------------	--

Number of rows	INSERT 100	UPDATE 0	DELETE 0	UPSERT 0	LOOKUP 0	ERROR 0	TOTAL 1
----------------	------------	----------	----------	----------	----------	---------	---------

Refresh | Typecast | Modify | Map drifted | Statistics | Remove | Export to CSV

↑↓	order_item_...	abc ↑↓	order_item_...	1,2 ↑↓	order_id	abc ↑↓	order_date	abc ↑↓	order_custo...	abc ↑↓	order_status	abc ↑↓
+	1		299.9800109...		1		2013-07-25 ...		11599		CLOSED	
+	10		651.9200153...		10		2013-07-25 ...		5648		PENDING PA...	
+	100		549.9400100...		100		2013-07-25 ...		12131		PROCESSING	
+	101		899.9400253...		101		2013-07-25 ...		5116		CLOSED	
+	NULL		NULL		102		2013-07-25 ...		8027		COMPLETE	
+	103		829.9200286...		103		2013-07-25 ...		12256		PROCESSING	

-Use the transformation Select to remove duplicate columns like order_item_order_id and rename the columns if required.



We will change the column name `order_customer_id` to `customer_id` and will rename the column `order_item_order_id` as shown below.

dataflow1 ds_order_items_adls... pl_copy_order_items ds_order_items_s3_s... pl_copy_orders_ing...

✓ Validate Data flow debug ✓ Debug Settings

Select settings Optimize Inspect Data preview ● Previous Next

columns 'order_id, order_date, customer_id, order_status, order_amount'

Incoming stream * join1

Options

- ✓ Skip duplicate input columns ⓘ
- ✓ Skip duplicate output columns ⓘ

Input columns *

☐ Auto mapping ⓘ Reset + Add mapping Delete 5 mappings: 1 column(s) from the inputs left unmapped

<input type="checkbox"/>	join1's column		Name as	
<input type="checkbox"/>	abc order_id	→	order_id	+ ⋈
<input type="checkbox"/>	abc order_date	→	order_date	+ ⋈
<input type="checkbox"/>	abc order_customer_id	→	customer_id	+ ⋈
<input type="checkbox"/>	abc order_status	→	order_status	+ ⋈
<input type="checkbox"/>	1.2 order_item_subtotal	→	order_amount	+ ⋈

Click on the option “Data Preview”

Select settings Optimize Inspect Data preview ●

Number of rows + INSERT 100 UPDATE 0 DELETE 0 UPSERT 0 LOOKUP 0 ERROR 0 TOTAL 104

Refresh | Typecast | Modify | Map drifted | Statistics | Remove | Export to CSV |

↑↓	order_id	abc	↑↓	order_date	abc	↑↓	customer_id	abc	↑↓	order_status	abc	↑↓	order_amount	1.2	↑↓
+	1			2013-07-25 00:0...			11599			CLOSED			299.9800109863...		
+	10			2013-07-25 00:0...			5648			PENDING PAYM...			651.9200153350...		
+	100			2013-07-25 00:0...			12131			PROCESSING			549.9400100708...		
+	101			2013-07-25 00:0...			5116			CLOSED			899.9400253295...		
+	102			2013-07-25 00:0...			8027			COMPLETE			NULL		

-Use the transformation Conditional Split to segregate the data into high value, low value, and erroneous.

dataflow1 • ds_order_items_adls... • pl_copy_order_items • ds_order_items_s3_s... • pl_copy_orders_ing... • ds_ord

✓ Validate Data flow debug ✓ Debug Settings

Conditional split settings Optimize Inspect Data preview

Output stream name * conditionalsplit [Learn more](#)

Description Conditionally distributing the data in order_amount groups, based on columns '1' [Reset](#)

Incoming stream * selectfromjoin

Split on ☒ First matching condition ☐ All matching conditions

Stream names	Condition
highvalue	order_amount > 500
lowvalue	order_amount <= 500
erroneous	Rows that do not meet any condition will use this output stream



-Write the outputs (high value, low value, erroneous) to their respective sinks in the output folder of ADLS GEN2.

Create dataset for storing “**high_value_orders**”, “**low_value_orders**”, “**erroneous**” data in adls gen2, refer the below screenshot.

Set properties

Name

ds_order_highvalue

Linked service *

ls_trendytech_adlsgen2

File path

output

/ high_value

/ File name

First row as header



Import schema

☒ From connection/store

☐ From sample file

☐ None

OK

Back

Cancel

Set properties

Name

ds_lowvalue

Linked service *

ls_trendytech_adlsgen2

File path

output

/ low_value

/ File name

First row as header



Import schema

☒ From connection/store

☐ From sample file

☐ None

OK

Back

Cancel

Set properties

Name

ds_orders_erroneous

Linked service *

ls_trendytech_adlsgen2

File path

output

/ erroneous

/ File name

First row as header



Import schema

☒ From connection/store ☐ From sample file ☐ None

OK

Back

Cancel

Note: Use single partition option while storing the data

TRENDYTECH
UPLIFT YOUR CAREER!

dataflow1 • ds_order_items_ads... • pl_copy_order_items • ds_order_items_s3_s...

✓ Validate Data flow debug ✓ Debug Settings

join1 Full outer join on 'orderitemaggregate' and 'orders'

selectfromjoin Renaming join1 to selectfromjoin with columns 'order_id,'

highvalue Conditionally distributing the data in order_amount groups...

sink1 Columns: 5 total

Sink Settings Errors Mapping Optimize Inspect ...

Incoming stream * conditionalsplit@highvalue

Sink type * Dataset Inline Cache

Dataset * ds_order_highvalue

✓ Connection successful

Test connection Open + New

Skip line count

Sink Settings Errors Mapping Optimize Inspect ...

! This sink currently has Single partition set in Optimize. This will make your data flow execution longer. The recommended setting is Use current partitioning.

Partition option * ☐ Use current partitioning ☒ Single partition ☐ Set partitioning

Similarly we will set sink for low value and erroneous values also.

dataflow1

ds_order_items_adls...

pl_copy_order_items

ds_order_items_s3_s...

✓ Validate

☒ Data flow debug ✓

Debug Settings

lowvalue

Conditionally distributing the data in order_amount groups.

+

sink2

Columns: 5 total

Sink

Settings

Errors

Mapping

Optimize

Inspect

...

Incoming stream *

conditionalsplit@lowvalue

Sink type *

Dataset

Inline

Cache

Dataset *

ds_lowvalue

✓ Connection successful

Test connection

Open

New

Skip line count

Options


☒ Allow schema drift ⓘ

TRENDY TECH

UPLIFT YOUR CAREER!

dataflow1 • ds_order_items_adls... • pl_copy_order_items • ds_order_items_s3_s...

✓ Validate ☒ Data flow debug ✓ Debug Settings



Sink Settings

Incoming stream * conditionalsplit@erroneous

Sink type * Dataset Inline Cache

Dataset * ds_orders_erroneous

✓ Connection successful

Test connection Open + New

Skip line count

Options ☒ Allow schema drift ⓘ

Publish all the changes.

Create pipeline and add this dataflow in it and debug that pipeline

Data Factory • Validate all • Publish all 1

dataflow1 • pl_process_orders...

✓ Validate ▶ Debug Add trigger ☒ Data flow debug ✓

Activities

- Move and transform
- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

Properties

General Related

Name * pl_process_orders_data

Description

Annotations + New

Data flow dataflow1

Parameters Variables Settings Output

Pipeline run ID: 537a07a5-ba90-40be-a712-375ec7a64006

Pipeline status ✓ Succeeded View debug run consumption

All status

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration
dataflow1	✓ Succeeded	Data flow	5/24/2024, 8:38:36 PM	15m 51s

4. For Populating the SQL Database for Reporting Team:

Create and Deploy an Azure SQL Database

Home >

Azure SQL


Default Directory

+ Create ⌚ Reservations ⚙️ Manage view ↕ Refresh ⬇ Export to CSV 🔗 Open query | 🏷 Assign tags 🗑 Delete

Filter for any field... Subscription equals all Resource group equals all X Location equals all X + Add filter

Showing 0 to 0 of 0 records.

Name ↑↓	Resour... ↑↓	Service tier ↑↓	Resource group ↑↓	Locat
---------	--------------	-----------------	-------------------	-------



No Azure SQL resources to display

Try changing or clearing your filters.

[Create Azure SQL resource](#)

Home > Azure SQL >


Select SQL deployment option

Microsoft

[Feedback](#)

How do you plan to use the service?

✓ SQL Database Hyperscale: Low price, high scalability, and best feature set. [Learn more](#)




SQL databases

Best for modern cloud applications. Hyperscale and serverless options are available.

Resource type
Single database

[Create](#) [Show details](#)




SQL managed instances

Best for most migrations to the cloud. Lift-and-shift ready.

Resource type
Single instance

[Create](#) [Show details](#)



SQL virtual machines

Best for migrations and applications requiring OS-level access. Lift-and-shift ready.

Image

[Create](#) [Show details](#) ☐ High availability

Create SQL Database ...

Microsoft

Project details

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

Subscription * ⓘ

Pay-As-You-Go (136f20b5-00f7-4eb5-a52e-0843e7ad1034) ▼

Resource group * ⓘ

Trendytech-rg ▼

[Create new](#)

Database details

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

Database name *

retaildb ✓

Server * ⓘ

Select a server ▼



[Create new](#)



TRENDYTECH
UPLIFT YOUR CAREER!

[Home](#) > [Azure SQL](#) > [Select SQL deployment option](#) > [Create SQL Database](#) >

Create SQL Database Server ...

Microsoft

 Azure Active Directory (Azure AD) is now Microsoft Entra ID. [Learn more](#) 

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 Microsoft Entra authentication [Learn more](#)  using an existing Microsoft Entra user, group, or application as Microsoft Entra admin [Learn more](#) , or select both SQL and Microsoft Entra authentication.

Authentication method

- ☐ Use Microsoft Entra-only authentication
- ☐ Use both SQL and Microsoft Entra authentication
- ☒ Use SQL authentication

Server admin login *

trendytech



Password *



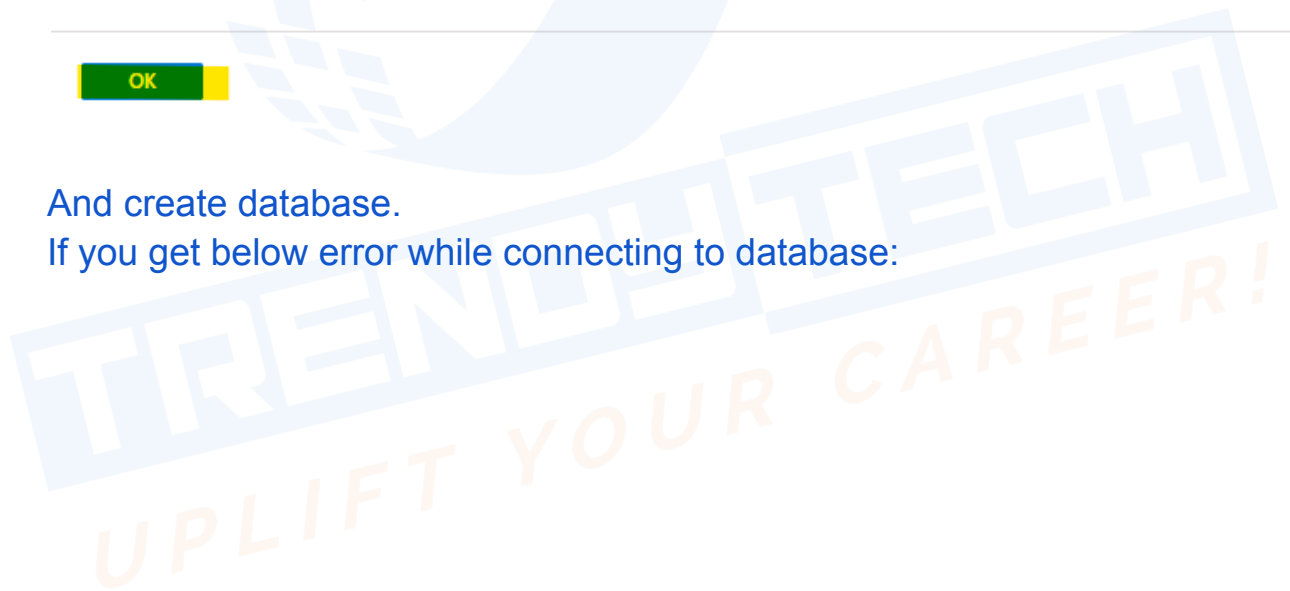
Confirm password *



OK

And create database.

If you get below error while connecting to database:



Error details

Error code [SqlDeniedPublicAccess](#)

Details

Cannot connect to SQL Database:

'trendytechserver.database.windows.net', Database:

'retaildb', Reason: Connection was denied since Deny

Public Network Access is set to Yes. To connect to this

server, 1. If you persist public network access disabled,

please use Managed Virtual Network IR and create

private endpoint. <https://docs.microsoft.com/en-us/azure/data-factory/managed-virtual-network-private-endpoint>; <https://docs.microsoft.com/en-us/azure/data-factory/tutorial-copy-data-portal-private>; 2. Otherwise you can enable public network



How helpful or unhelpful was this error message?



Follow below steps:

Go to “Overview” => Properties => Networking

Home >

retaildb (trendytechserver/retaildb) SQL database

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

Overview Activity log Tags Diagnose and solve problems Query editor (preview) Settings Data management Integrations Power Platform Security Intelligent Performance Monitoring Automation Help

Location : East US Pricing tier : Free - General Purpose - Serverless

Subscription (move) : Pay-As-You-Go Overage billing : Disabled

Subscription ID : 136f20b5-00f7-4eb5-a52e-0843e7ad1034 Free monthly vCore am... : 100,000 vCore seconds remaining

Earliest restore point : 2024-05-24 15:45 UTC

Tags (edit) : Add tags

Getting started Monitoring **Properties** Features Notifications (0) Integrations Tutorials

Compute + storage

Service tier	General Purpose
Compute tier	Serverless
vCores	1 vCore
Max storage	32 GB

Availability

Zone redundancy	Disabled
Replication	0 Replicas
Availability Zone	NoPreference

Backups

Networking

Public access	Enabled
Firewall rules	1 firewall rule
Virtual networks	0 virtual network service endpoints
Private access	0 private endpoint connections

Connections


Primary endpoint	trendytechserver.database.windows.net
------------------	---------------------------------------


Authentication


Authentication method	SQL
SQL admin	trendytech


trendytechserver | Networking ☆ ...

SQL server


 Feedback


 Overview

 Activity log

 Access control (IAM)

 Tags


 Quick start


 Diagnose and solve problems


> Settings

> Data management

> Security

 Networking

 Microsoft Defender for Cloud

 Transparent data encryption

Public access

Private access

Connectivity


Public network access


Public Endpoints allow access to this resource through the internet using a public IP address to access this resource. [Learn more](#)

Public network access

☐ Disable

☒ Selected networks

 Connections from the IP addresses configure [more](#)

 Please save public network access value befo

Virtual networks

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

+ Add a virtual network rule

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

+ Add a virtual network rule

Rule	Virtual network	Subnet	Address range	Endpoint status	Resource group	Subscription	State
------	-----------------	--------	---------------	-----------------	----------------	--------------	-------

Firewall rules

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

+ Add your client IPv4 address (49.207.194.137) + Add a firewall rule

Rule name	Start IPv4 address	End IPv4 address
-----------	--------------------	------------------


ClientIPAddress_2024-5-24_21-11-20

49.207.194.137

49.207.194.137



Exceptions

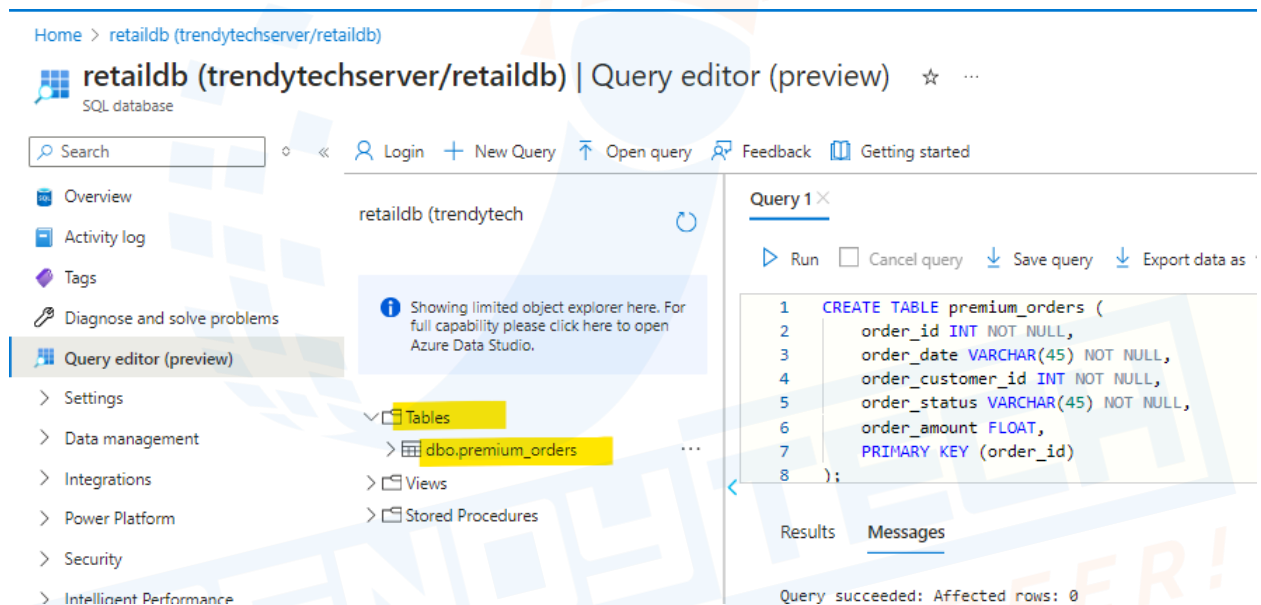
☒ Allow Azure services and resources to access this server 

Save

Discard



In database create table using below schema

```
CREATE TABLE premium_orders (  
    order_id INT NOT NULL,  
    order_date VARCHAR(45) NOT NULL,  
    order_customer_id INT NOT NULL,  
    order_status VARCHAR(45) NOT NULL,  
    order_amount FLOAT,  
    PRIMARY KEY (order_id)  
);
```



Create Linked Services:
Pointing to SQL DB


New linked service

 Azure SQL Database [Learn more](#) 

Account selection method 

☒ From Azure subscription ☐ Enter manually

Azure subscription

Pay-As-You-Go (136f20b5-00f7-4eb5-a52e-0843e7ad1034) 

Server name *

trendytechserver 



Database name *

retaildb 






Authentication type *

SQL authentication 

User name *


trendytech

 Password  Azure Key Vault

 Connection successful

Create

Back

 Test connection

Cancel

Pointing to ADLS GEN2 where the high value order output is stored.(already created)

Create Datasets:

Pointing to the high value order data. (already created)

Pointing to the table in SQL DB.

Set properties

Name

Linked service *

Table name

☐ Enter manually

Import schema
☒ From connection/store ☐ None

Create a Pipeline (copy pipeline adls to sql):

pl_adls_to_sql

Activities

Search activities

Move and transform

- Copy data
- Data flow

Synapse

- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General

Copy data

Copy data1

General Source Sink Mapping Settings

Name *

Copy data1

Learn more

Description

Properties

General Related

Name *

pl_adls_to_sql

Description

Annotations

+ New

The screenshot displays the Azure Data Factory (ADF) interface for configuring a Copy data activity. The activity is named 'Copy data1' and is part of a pipeline named 'pl_adls_to_sql'.

Source Tab Configuration:

- Source dataset:** ds_order_highvalue
- File path type:** Wildcard file path (selected)
- Wildcard paths:** output / high_value / *
- Filter by last modified:** Start time (UTC) and End time (UTC) are empty.
- Recursively:** Checked (indicated by a blue checkmark).

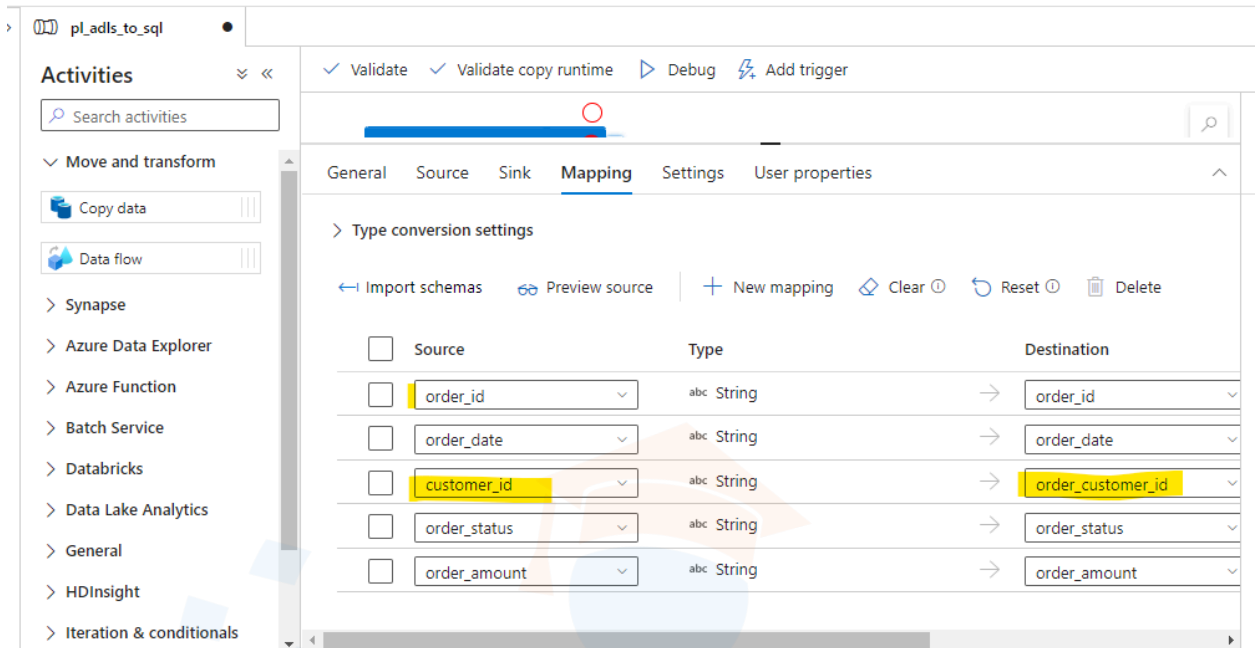
Sink Tab Configuration:

- Sink dataset:** ds_sql_retaildb_orders
- Write behavior:** Insert (selected)
- Bulk insert table lock:** No (selected)
- Table option:** None (selected)
- Pre-copy script:** (Empty text box)

Properties Pane:

- Name:** pl_adls_to_sql
- Description:** (Empty text box)
- Annotations:** + New

And define the mapping as shown below.



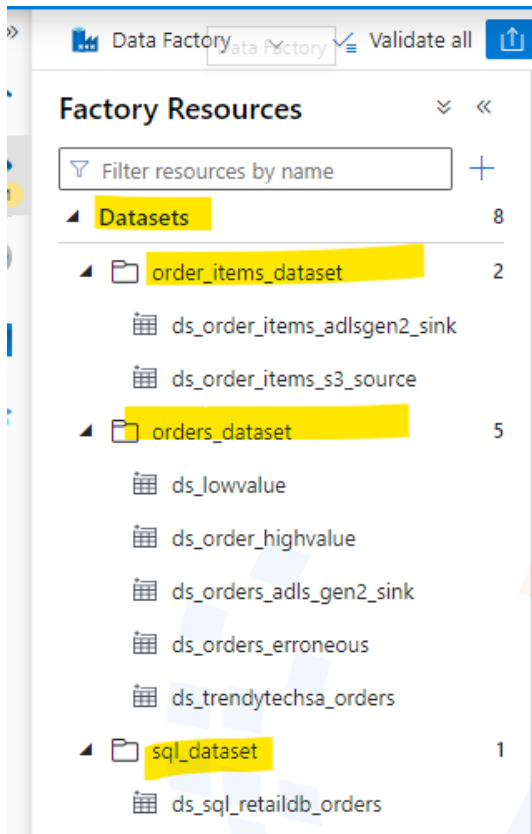
Push data from ADLS GEN2 (Source) to Azure SQL DB (Sink).

Debug the pipeline.

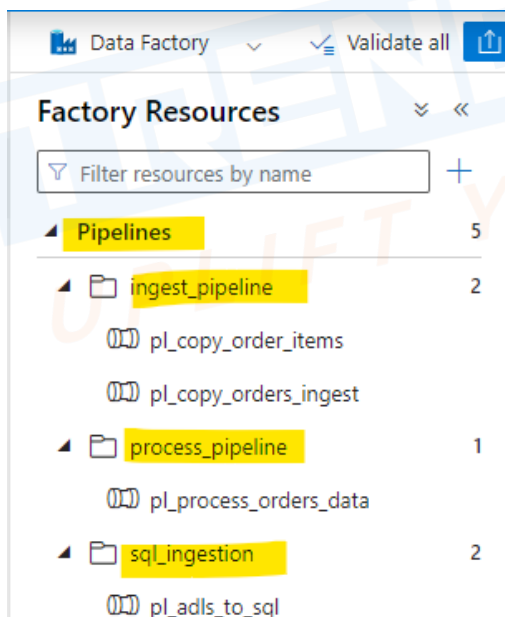
Note: remove the data from all the folders and truncate table also before executing the final pipeline.

Organize Datasets and Pipelines into Folders:

1. Create folders for datasets and add the datasets to their respective folders (orders dataset, order_items dataset, sql).



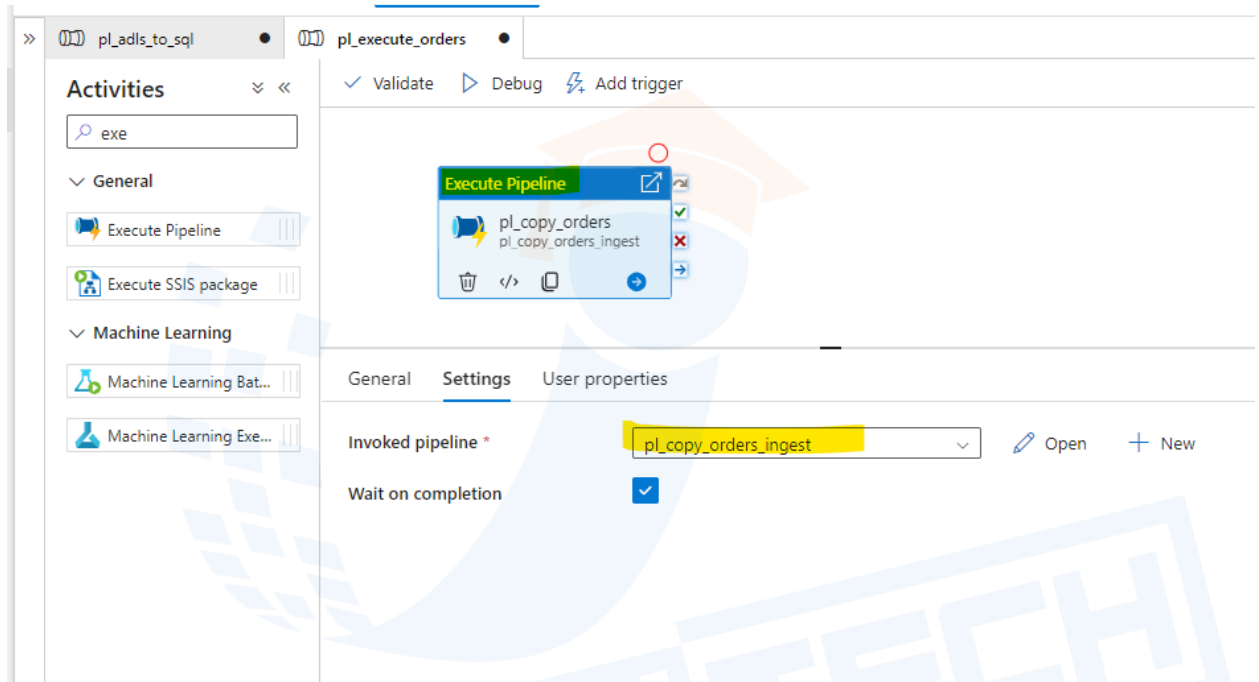
2. Create folders for pipelines and add the pipelines associated to specific activities to their respective folders (ingest, process, sql ingestion).



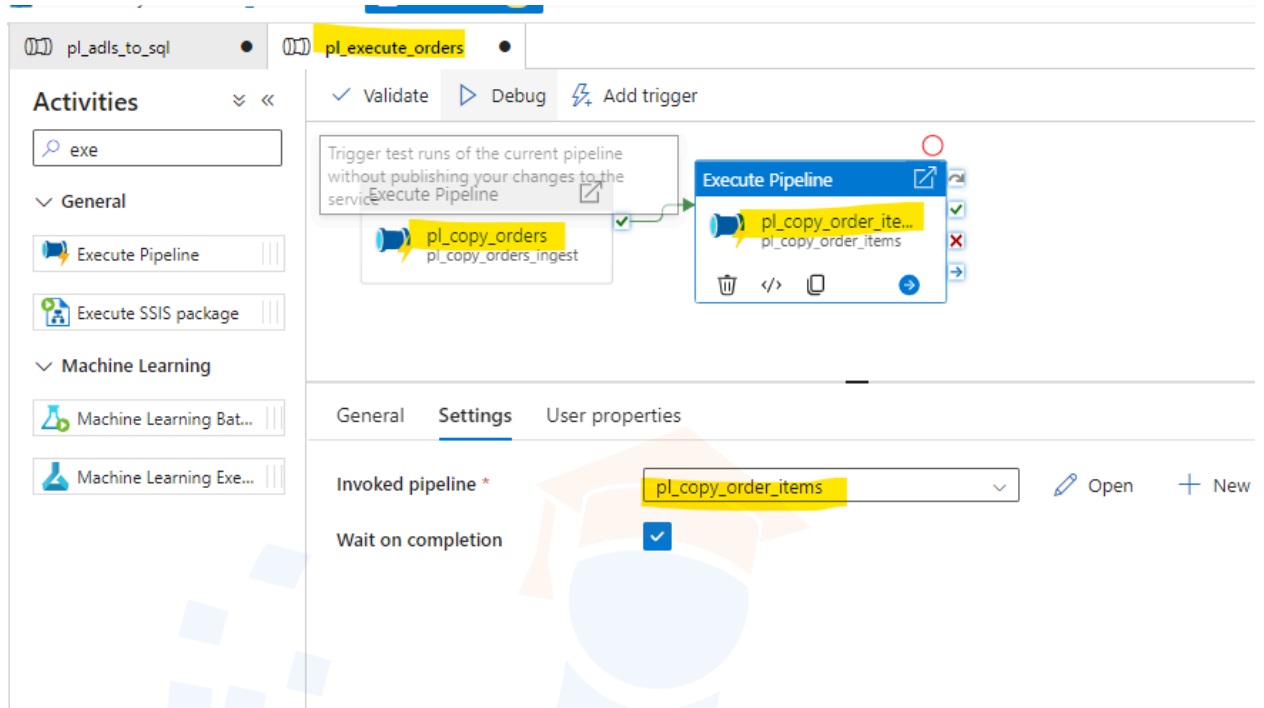
3. Create an Execute Pipeline Activity:

Chain all the pipelines to get executed in a specific order.

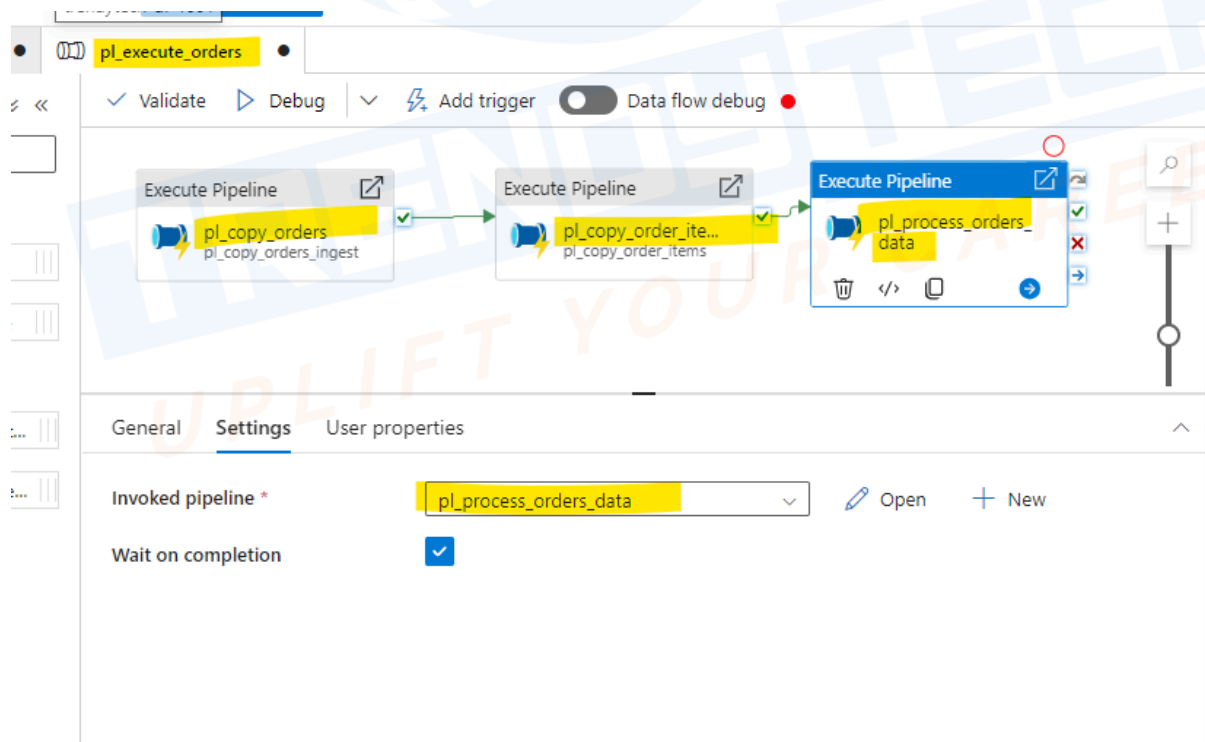
We will first add the pipeline `pl_copy_orders` using the option execution pipeline.



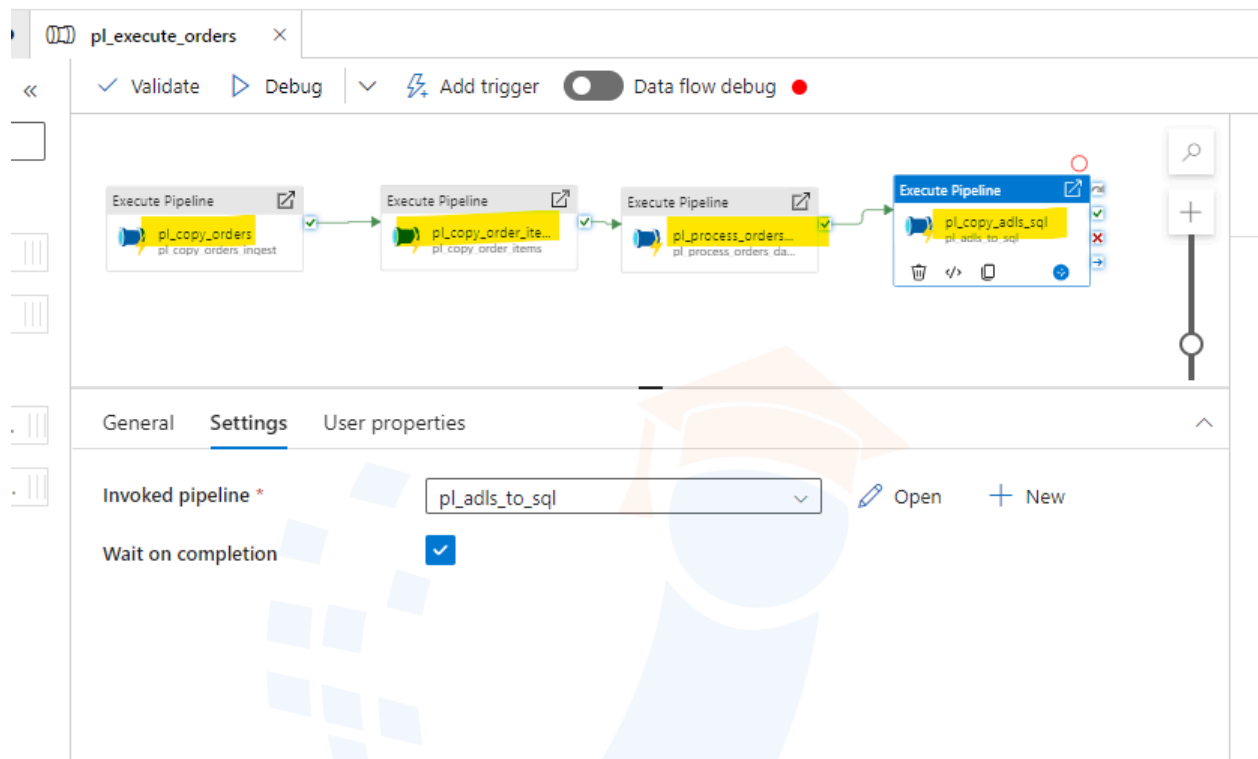
Now we will add the pipeline “`pl_copy_order_items`” as shown below using the option execution pipeline.



We will add the pipeline `pl_process_orders_data` as shown below using the option execution pipeline.



Our final pipeline will be as shown below.



4. Create a Trigger:

Ingest and process data with Storage Event Type (the trigger will get initiated whenever a new data file arrives at the blob storage as mentioned in the storage account and container name fields while creating the trigger).

New trigger

Name *
tr_orders_pipeline

Description

Type *
Storage events

Account selection method * ⓘ
☒ From Azure subscription ☐ Enter manually

Azure subscription ⓘ
Pay-As-You-Go (136f20b5-00f7-4eb5-a52e-0843e7ad1034)

Storage account name * ⓘ
trendytechsa101

Container name * ⓘ
ordersinput

Blob path begins with ⓘ

Blob path ends with ⓘ

Event * ⓘ
☒ Blob created ☐ Blob deleted

Ignore empty blobs * ⓘ
☒ Yes ☐ No

Annotations
+ New

Start trigger ⓘ
☒ Start trigger on creation

Continue Cancel

Note: After creating new pipelines in Azure Data Factory, be sure to publish these changes to make them active and available for use in your data workflows.

5. Attach the Created Trigger to the Execute Pipeline:

Trigger the pipeline execution without any manual intervention whenever a new file gets added to the blob storage.

If you get below error

Register Azure Event Grid resource provider to your subscription before creating an event trigger.
[Learn more here](#)

Follow this document:

<https://learn.microsoft.com/en-us/azure/azure-resource-manager/management/resource-providers-and-types#azure-portal>

Monitor the trigger and pipeline under the monitor tab.

The image shows two screenshots of the Microsoft Azure Data Factory interface. The top screenshot displays the 'Trigger runs' tab, showing a single trigger run named 'tr' with a status of 'Succeeded'. The bottom screenshot displays the 'Pipeline runs' tab, showing a list of pipeline runs, including 'pl_adls_to_sql', 'pl_process_orders_data', 'pl_copy_order_items', 'pl_copy_orders_ingest', and 'pl_execute_orders', all with a status of 'Succeeded'.

Trigger runs

Trigger name	Trigger type	Trigger time	Status	Pipelines	Run
tr	Storage events tri...	5/25/2024, 1:50:36 AM	Succeeded	1	Original

Pipeline runs

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Run
pl_adls_to_sql	5/25/2024, 2:00:13 AM	5/25/2024, 2:00:29 AM	16s	7b3ebddf-2b21-4d0...	Succeeded	Orig
pl_process_orders_data	5/25/2024, 1:51:41 AM	5/25/2024, 2:00:12 AM	8m 32s	91d853b3-bfae-4cc0...	Succeeded	Orig
pl_copy_order_items	5/25/2024, 1:51:20 AM	5/25/2024, 1:51:40 AM	20s	d9ea117c-c5e4-4b3...	Succeeded	Orig
pl_copy_orders_ingest	5/25/2024, 1:50:37 AM	5/25/2024, 1:51:20 AM	43s	c6912c70-98ca-4847...	Succeeded	Orig
pl_execute_orders	5/25/2024, 1:50:36 AM	5/25/2024, 2:00:30 AM	9m 55s	tr	Succeeded	Orig

Note: At the end delete all the resources that you have created.