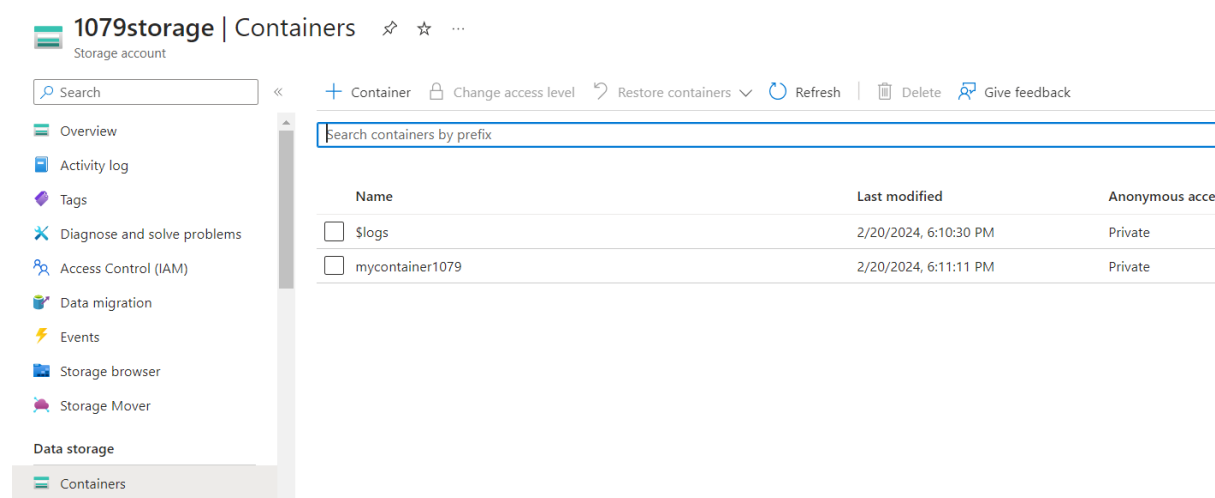**Akilesh K**

k.akilesh123@gmail.com

Data engineering - Batch 1

Date: 20-02-24

DAY 22 – Azure DataFactory

Transfer of data from blob storage account to another storage acount using Azure datafactory pipeline
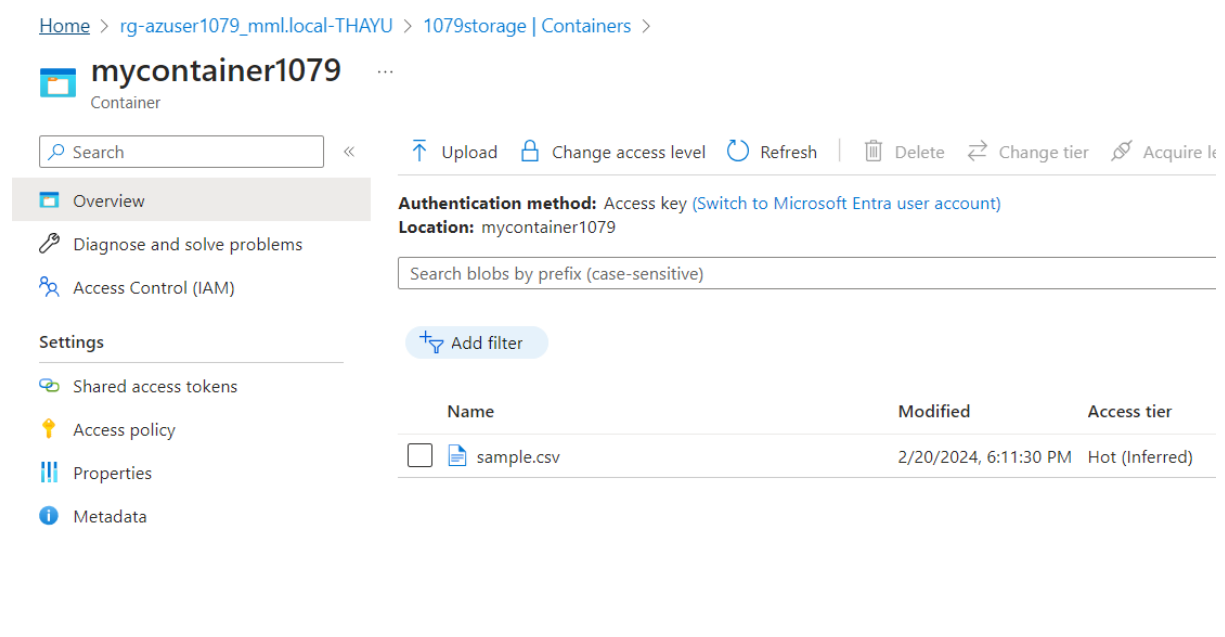
Creating a source storage account

Upload a sample file in the container

Create a destination storage account

Create a empty container

Create a azure datafactory

1079hexadf
Data factory (V2)

Search «

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

Settings

- Networking
- Managed identities
- Properties
- Locks

Getting started

- Quick start

Delete

∧ Essentials

Resource group (move)	: rg-azuser1079_mml.local-THAYU	Type	: Data factory (V2)
Status	: Succeeded	Getting started	: Quick start
Location	: East US		
Subscription (move)	: Azure subscription 1		
Subscription ID	: 984f097c-963c-4eb6-a20d-839457ae9f08		

Azure Data Factory Studio

Launch studio

Launch the studio and click ingest data

Microsoft Azure | Data Factory ▸ 1079hexadf | Search factory and documentation | azuser1079_mml.local@iihtl.

ⓘ Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!

ⓘ Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration. Learn more

Set up code repository

Data factory

1079hexadf

New ∨

Ingest
Copy data at scale once or on a schedule.

Orchestrate
Code-free data pipelines.

Transform data
Transform your data using data flows.

Configure SSIS
Manage & run your SSIS packages in the cloud.

Recent resources

Click built in copy task

Choose azure storage and create a new connection by mentioning the storage name

Specify the source container path

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source type   Azure Blob Storage

Connection *   1079storage   ✎ Edit   + New connection

If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse. Append a slash (/) at the end if the path refers to a folder.

mycontainer1079/   📁 Browse

Options

☐ Binary copy ⓘ

☑ Recursively ⓘ

☐ Enable partitions discovery ⓘ

Max concurrent connections ⓘ

Select the new storage account for destination datastore

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or spec

Destination type   Azure Blob Storage

Connection *   Select...   + New connection

New connection

Azure Blob Storage Learn more ⧉

Name *

1079newstorage

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime

Authentication type

Account key

Connection string   **Azure Key Vault**

Account selection method ⓘ

◉ From Azure subscription   ○ Enter manually

Azure subscription ⓘ

Azure subscription 1 (984f097c-963c-4eb6-a20d-839457ae9f08)

Storage account name *

1079newstorage   ↻

Mention the destination container path

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Destination type   Azure Blob Storage

Connection *   1079newstorage   ✏ Edit   + New connection

Folder path *
If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.

newcontainer1079/   🗀 Browse

File name

Copy behavior ⓘ

Select...

Max concurrent connections ⓘ

Create a pipeline name

Settings

Enter name and description for the copy data task, more options for data movement

Task name *   CopyPipeline_1079

Task description

Data consistency verification ⓘ   ☐

Fault tolerance ⓘ

Enable logging ⓘ   ☐

Enable staging ⓘ   ☐

> Advanced

Summary of the pipeline is provided

Summary

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

Azure Blob Storage → Azure Blob Storage

Properties   ✎ Edit

Task name	CopyPipeline_1079
Task description	

Source   ✎ Edit

Connection name	1079storage
Dataset name	SourceDataset_h86
Column delimiter	;
Row delimiter	
Escape character	\
Quote char	"

The pipeline gets executed

Azure Blob Storage → Azure Blob Storage

Deployment complete

Deployment step	Status
Validating copy runtime environment	✅ Succeeded
› Creating datasets	✅ Succeeded
› Creating pipelines	✅ Succeeded
› Running pipelines	✅ Succeeded

Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.

[Finish] [Edit pipeline] [Monitor]

The data gets transferred to destination blob storage

Home > rg-azuser1079_mml.local-THAYU > 1079newstorage | Containers >

newcontainer1079
Container

Overview	Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots
Diagnose and solve problems	
Access Control (IAM)	**Authentication method:** Access key (Switch to Microsoft Entra user account)
	Location: newcontainer1079

Settings

Shared access tokens

Access policy

Properties

Metadata

Search blobs by prefix (case-sensitive)

Add filter

Name	Modified	Access tier	Archive status	Blob type
sample.txt	2/20/2024, 6:21:49 PM	Hot (Inferred)		Block blob

The pipeline is executed successfully