

Azure Coding Challenge

-Sarthak Niranjana Kulkarni (Maverick)

- sarthakkul2311@gmail.com

- (+91) 93256 02791

10/12/2024 (Tuesday)

Copy activity from datalake to datalake in azure datafactory: -

1. Opening Data Factory

- Create Data Factory

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](#)

Activate Windows
Go to Settings to activate Windows. [Give feedback](#)

- Then launch Azure Data Factory Studio

Microsoft Azure Search resources, services, and docs (G+/I) Copilot azuser2371_mmllocal@... TECHADEMY LEARNING SOLUTIONS

Home > Data factories >

Data factories

Techademy Learning Solutions Private Limited (tec...

+ Create Manage view

Filter for any field...

Name ↑↓

Sarthak-data-factory1

Sarthak-data-factory1 Data factory (V2)

Search Delete

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Settings Getting started Monitoring Automation Help

Status Succeeded

Location East US

Subscription (move) [MML Learners](#)

Subscription ID 2a3c6418-97b9-4d96-a24b-2c2d7633d375

Getting started [Quick start](#)

Azure Data Factory Studio

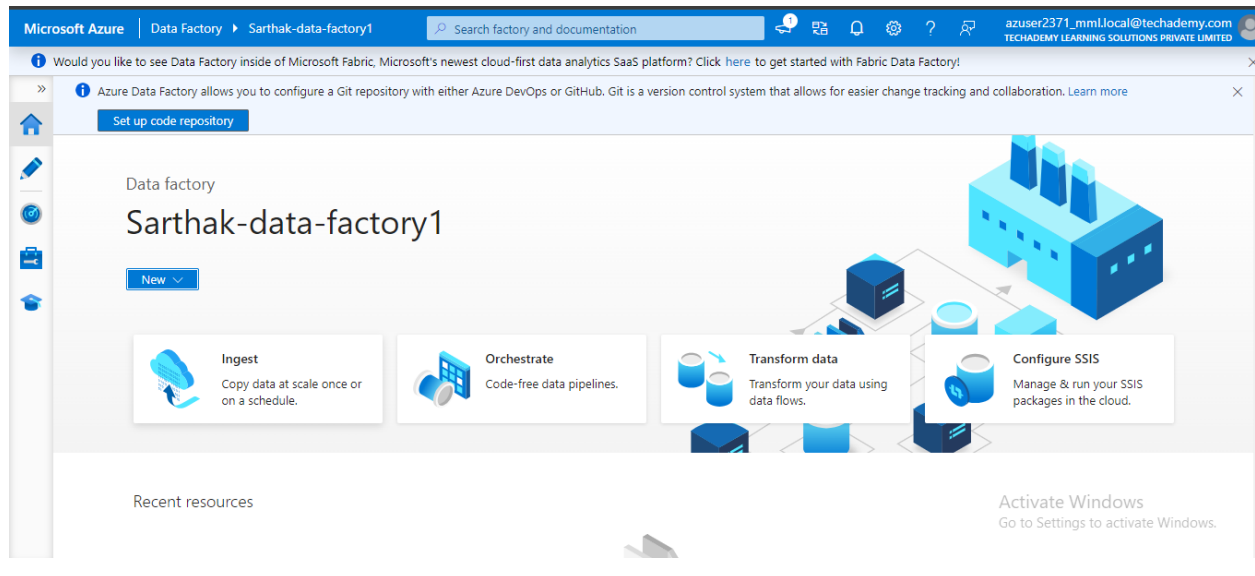
[Launch studio](#)

Activate Windows
Go to Settings to activate Windows.

Page 1 of 1

2. Ingesting the data

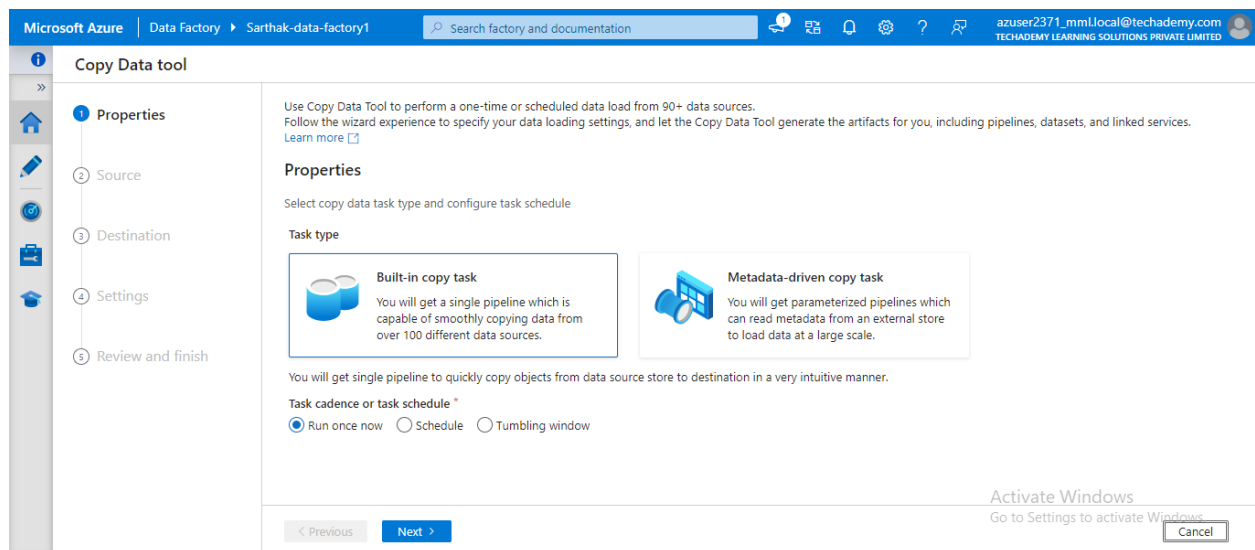
- Click on Ingest



3. Setting up Copy data tool

- Properties

Select your property and proceed



- Source → Dataset

Select your source file and check binary copy checkbox

Microsoft Azure | Data Factory | Sarthak-data-factory1 | Search factory and documentation | azuser2371_mmllocal@techademy.com | TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Copy Data tool

- 1 Properties
- 2 Source**
- 3 Dataset
- 4 Configuration
- 5 Destination
- 6 Settings
- 7 Review and finish

Source data store

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

Source type: All

Connection: AzureDataLakeStorage1 | Edit | + New connection

File or folder

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.

container123/loan.csv | Browse

Options

☒ Binary copy

Compression type: None

☒ Recursively

☐ Delete files after completion

< Previous | Next >

Activate Windows | Go to Settings to activate Windows | Cancel

- Source → Dataset → Configuration

Select your desired configurations and click on next

Microsoft Azure | Data Factory | Sarthak-data-factory1 | Search factory and documentation | azuser2371_mmllocal@techademy.com | TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Copy Data tool

- 1 Properties
- 2 Source
- 3 Dataset
- 4 Configuration**
- 5 Destination
- 6 Settings
- 7 Review and finish

File format settings

File format: DelimitedText | Detect text format | Preview data

Column delimiter: Comma (,) | Edit

Row delimiter: Default (\r\n or \n) | Edit

☒ First row as header

> Advanced

Compression type: No compression

Additional columns

< Previous | Next >

Activate Windows | Go to Settings to activate Windows | Cancel

- Destination → Dataset

Create a new connection of Data Lake Storage

Microsoft Azure | Data Factory | Sarthak-data-factory1 | Search factory and documentation | azuser2371_mm1local@techademy.com | TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Copy Data tool

Properties
Source
Destination
Dataset
Configuration
Settings
Review and finish

Destination data store

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

Destination type: Azure Data Lake Storage Gen2

Connection: AzureDataLakeStorage1

Folder path
If the identity you use to access the data store only has permission to subdi account, specify the path to browse.

File name

Copy behavior: Select...

Max concurrent connections: 1

< Previous Next >

New connection

Azure Data Lake Storage Gen2 Learn more

Name: AzureDataLakeStorage2

Description

Connect via integration runtime: AutoResolveIntegrationRuntime

Authentication type: Account key

Account selection method: From Azure subscription Enter manually

Azure subscription: MML Learners (2a3c6418-97b9-4d96-a24b-2c2d7633d375)

Storage account name

Create Cancel

Activate Windows
Go to Settings to activate Windows.
Test connection

Select destination path and give a name to the file

Microsoft Azure | Data Factory | Sarthak-data-factory1 | Search factory and documentation | azuser2371_mm1local@techademy.com | TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Copy Data tool

Properties
Source
Destination
Dataset
Configuration
Settings
Review and finish

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 Data Lake Storage Gen2

Connection: AzureDataLakeStorage2 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.

sarthak-container/sarthak-directory/ Browse

File name
Coding-challenge-copy-file

Copy behavior: Select...

Max concurrent connections: 1

< Previous Next >

Activate Windows
Go to Settings to activate Windows.
Cancel

- Destination → Dataset → Configuration

Select your desired configurations and click on next

Microsoft Azure | Data Factory | Sarthak-data-factory1 | Search factory and documentation | azuser2371_mmllocal@techademy.com | TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Copy Data tool

- Properties
- Source
- Destination
- Dataset
- Configuration
- Settings
- Review and finish

File format settings

File format: DelimitedText

Column delimiter: Comma (,)

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

☒ Add header to file

> Advanced

Compression type: No compression

Max rows per file:

< Previous | Next >

Activate Windows
Go to Settings to activate Windows
Cancel

- Settings

Give some name to the task and proceed

Microsoft Azure | Data Factory | Sarthak-data-factory1 | Search factory and documentation | azuser2371_mmllocal@techademy.com | TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Copy Data tool

- Properties
- Source
- Destination
- Settings
- Review and finish

Settings

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

Task name: Copy-Coding-challenge-pipeline

Task description:

Data consistency verification: ☐

Fault tolerance:

Enable logging: ☐

Enable staging: ☐

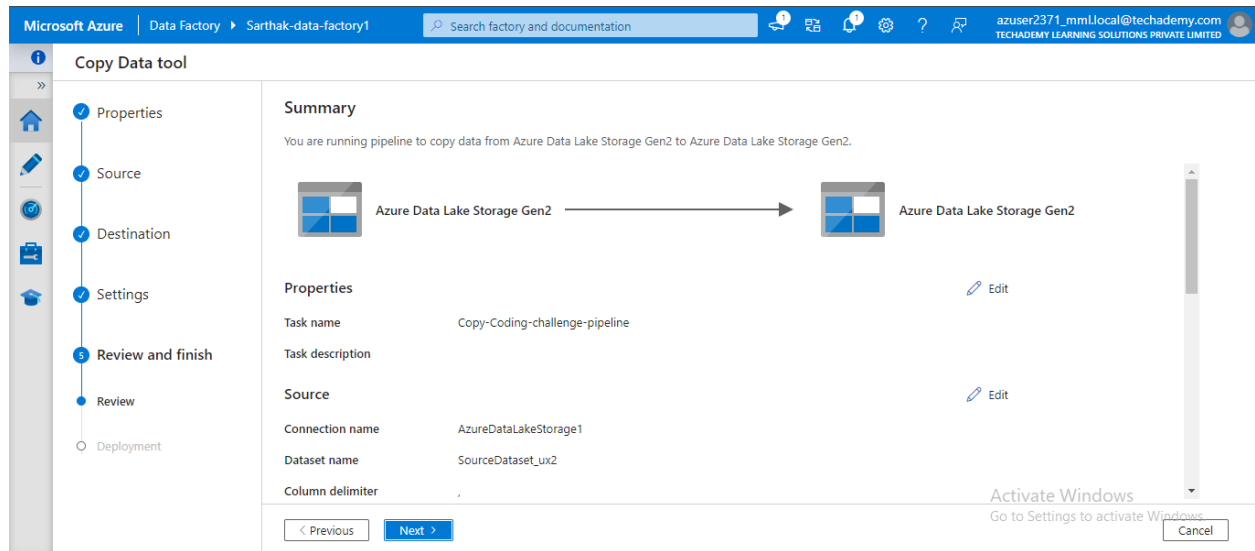
> Advanced

< Previous | Next >

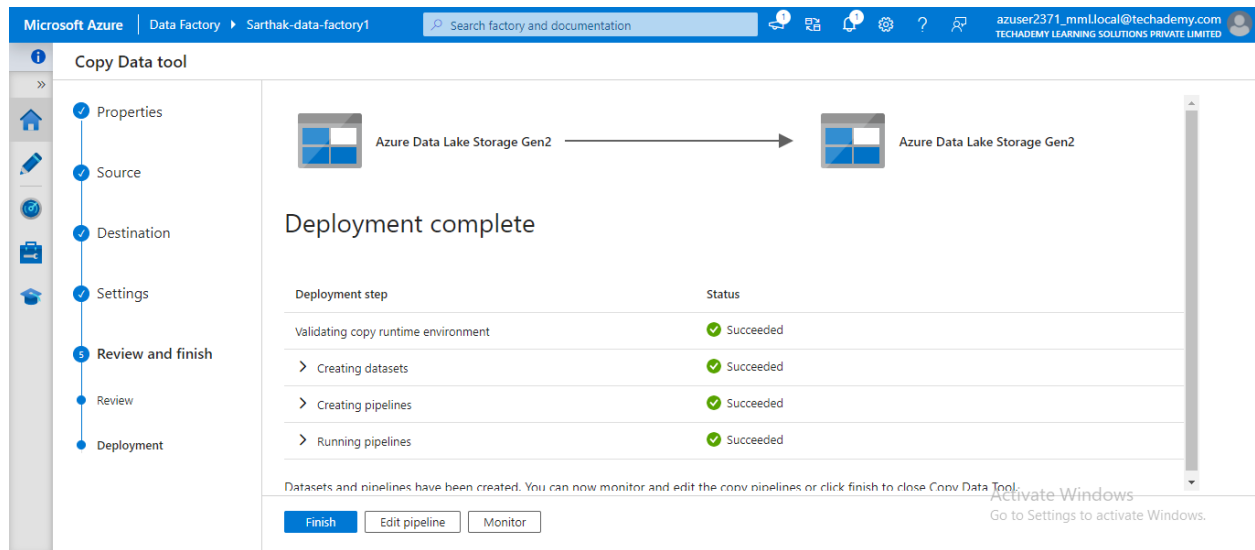
Activate Windows
Go to Settings to activate Windows
Cancel

4. Deploying the Copy Data Pipeline

- Review all the configuration and click on next



- This window would ensure the deployment of Copy data Pipeline



5. Validating the Pipeline

Microsoft Azure | Data Factory | Sarhak-data-factory1 | Search factory and documentation | azuser2371_mml.local@techademy.com | TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Copy-Coding-chall... x

Activities

Search activities

Move and transform

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

Copy data

Sarthak-coding-challenge

General

Name * Sarhak-coding-challenge

Description

Activity state ☒ Activated ☐ Deactivated

Timeout 0.12:00:00

Pipeline validation output

Your pipeline has been validated.
No errors were found.

Activate Windows
Go to Settings to activate Windows.

Close

6. Debugging the pipeline

- Debugging process would start wait for some time to complete

Microsoft Azure | Data Factory | Sarhak-data-factory1 | Search factory and documentation | azuser2371_mml.local@techademy.com | TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Copy-Coding-chall... x

Activities

Search activities

Move and transform

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Copy data

Sarthak-coding-challenge

Parameters

Variables

Settings

Output

Pipeline run ID: ba19c50d-c657-4437-a6c9-498c52d6a1ff

Pipeline status In progress

All status

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
Sarthak-coding-challenge	Queued	Copy data	12/10/2024, 5:07:59 PM	6s		

Microsoft Azure | Data Factory | Sarhak-data-factory1 | Search factory and documentation | azuser2371_mml.local@techademy.com | TECHADEMY LEARNING SOLUTIONS PRIVATE LIMITED

Copy-Coding-chall... x

Activities

Search activities

Move and transform

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Copy data

Sarthak-coding-challenge

Parameters

Variables

Settings

Output

Pipeline run ID: ba19c50d-c657-4437-a6c9-498c52d6a1ff

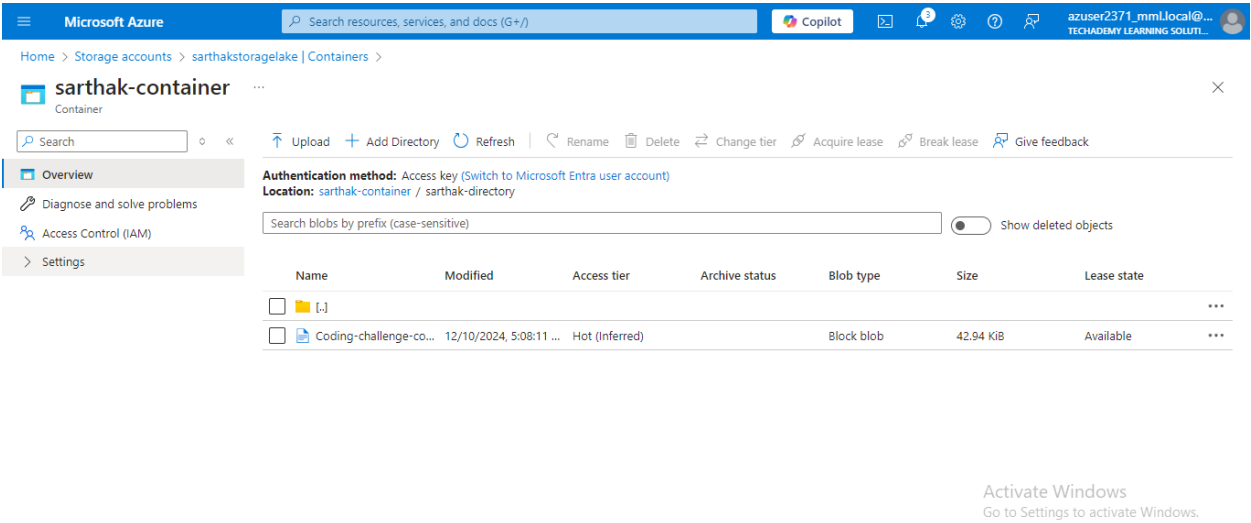
Pipeline status Succeeded

All status

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties
Sarthak-coding-challenge	Succeeded	Copy data	12/10/2024, 5:07:59 PM	19s		

7. Verify the whole process by visiting Data Lake storage account



Data Factory: Azure Data Factory is a cloud-based data integration service that enables you to create, schedule, and orchestrate data workflows. It supports data movement and transformation across a wide range of sources, making it ideal for ETL processes and data automation.

Copy Data Pipeline: The Copy Data Pipeline in Azure Data Factory is used to move data from one source to a destination efficiently. It supports diverse data formats and can handle complex transformations during the data transfer process.