

Name: Krushnakumar Patle

Email: krishnapatle128@gmail.com

Batch: Data Engineering Batch-1

Azure Databricks Coding Challenge

Q3. Execute & explain, Azure data factory and its copy activity.

Azure Data Factory is a cloud-based data integration service that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation.

ADF does not store any data itself. It allows you to create data-driven workflows to orchestrate the movement of data between supported data stores and then process the data using compute services in other regions or in an on-premise environment. It also allows you to monitor and manage workflows using both programmatic and UI mechanisms.

ADF can be used for:

- Supporting data migrations
- Getting data from a client's server or online data to an Azure Data Lake
- Carrying out various data integration processes
- Integrating data from different ERP systems and loading it into Azure Synapse for reporting

Azure Data Factory pipelines (data-driven workflows) typically perform three steps.

Step 1: Connect and Collect

Connect to all the required sources of data and processing such as SaaS services, file shares, FTP, and web services. Then, move the data as needed to a centralized location for subsequent processing by using the Copy Activity in a data pipeline to move data from both on-premise and cloud source data stores to a centralization data store in the cloud for further analysis.

Step 2: Transform and Enrich

Once data is present in a centralized data store in the cloud, it is transformed using compute services such as HDInsight Hadoop, Spark, Azure Data Lake Analytics, and Machine Learning.

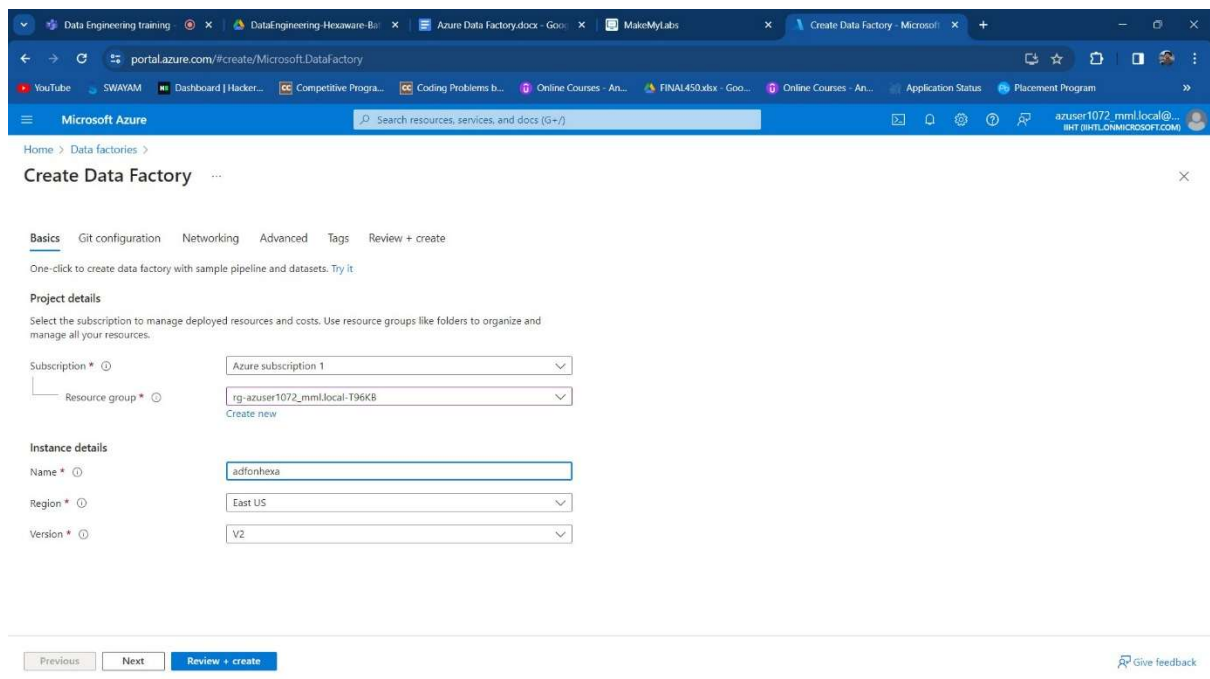
Step 3: Publish

Deliver transformed data from the cloud to on-premise sources like SQL Server or keep it in your cloud storage sources for consumption by BI and analytics tools and other applications.

Copy Activity:

Copy Activity is one of the activities available in Azure Data Factory. It is used to copy data from a source data store to a destination data store. It supports a wide range of source and destination data stores including Azure Blob Storage, Azure SQL Database, Azure Data Lake Storage, SQL Server, Amazon S3, Google Cloud Storage, etc. Copy Activity provides capabilities for efficient data movement, transformation, and monitoring, allowing you to move data reliably and efficiently across different environments.

1. Create a data factory



The screenshot shows the 'Create Data Factory' wizard in the Microsoft Azure portal. The 'Project details' section is active, showing fields for Subscription, Resource group, Instance details (Name, Region, Version), and navigation buttons (Previous, Next, Review + create).

Project details

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

Subscription *

Resource group *

Create new

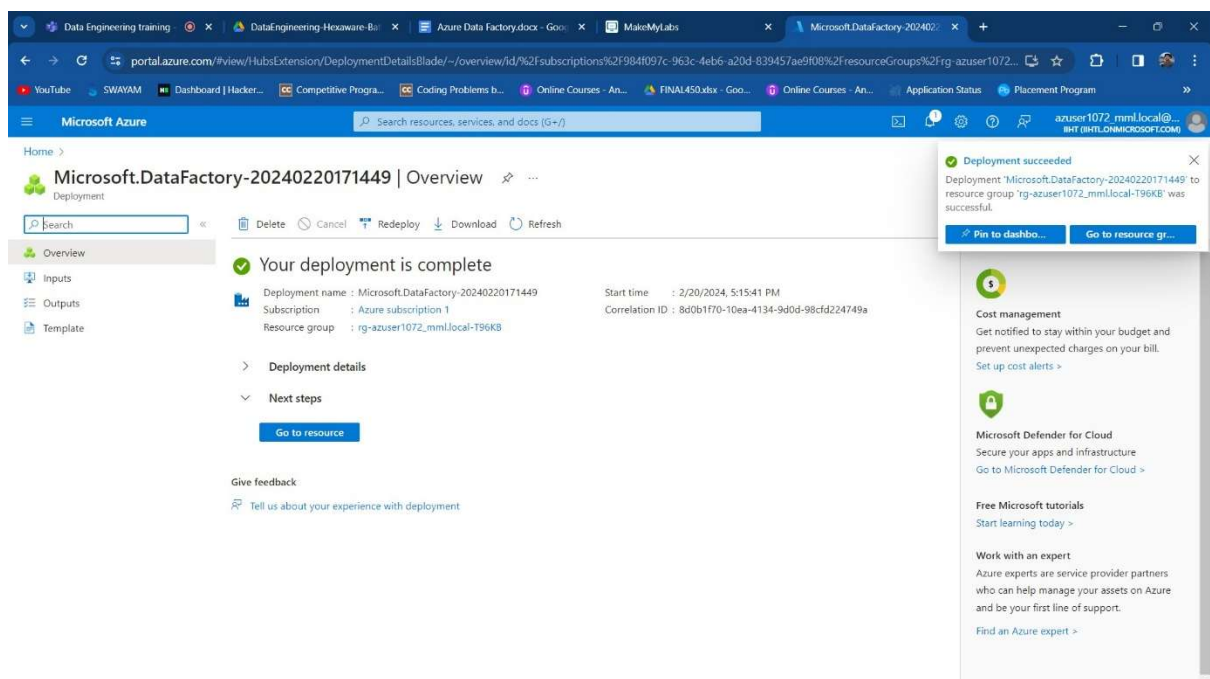
Instance details

Name *

Region *

Version *

Previous Next Review + create



The screenshot shows the 'Microsoft.DataFactory-20240220171449 | Overview' page in the Microsoft Azure portal. The page displays deployment details, next steps, and a sidebar with cost management and security recommendations.

Microsoft.DataFactory-20240220171449 | Overview

Deployment

Search

Overview

Inputs

Outputs

Template

Your deployment is complete

Deployment name : Microsoft.DataFactory-20240220171449

Subscription : Azure subscription 1

Resource group : rg-azuser1072_mm1.local-T96K8

Start time : 2/20/2024, 5:15:41 PM

Correlation ID : 8d0b1f70-10ea-4134-9d0d-98cf224749a

Deployment details

Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

Deployment succeeded

Deployment 'Microsoft.DataFactory-20240220171449' to resource group 'rg-azuser1072_mm1.local-T96K8' was successful.

Pin to dashbo... Go to resource gr...

Cost management

Get notified to stay within your budget and prevent unexpected charges on your bill.

Set up cost alerts >

Microsoft Defender for Cloud

Secure your apps and infrastructure

Go to Microsoft Defender for Cloud >

Free Microsoft tutorials

Start learning today >

Work with an expert

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.

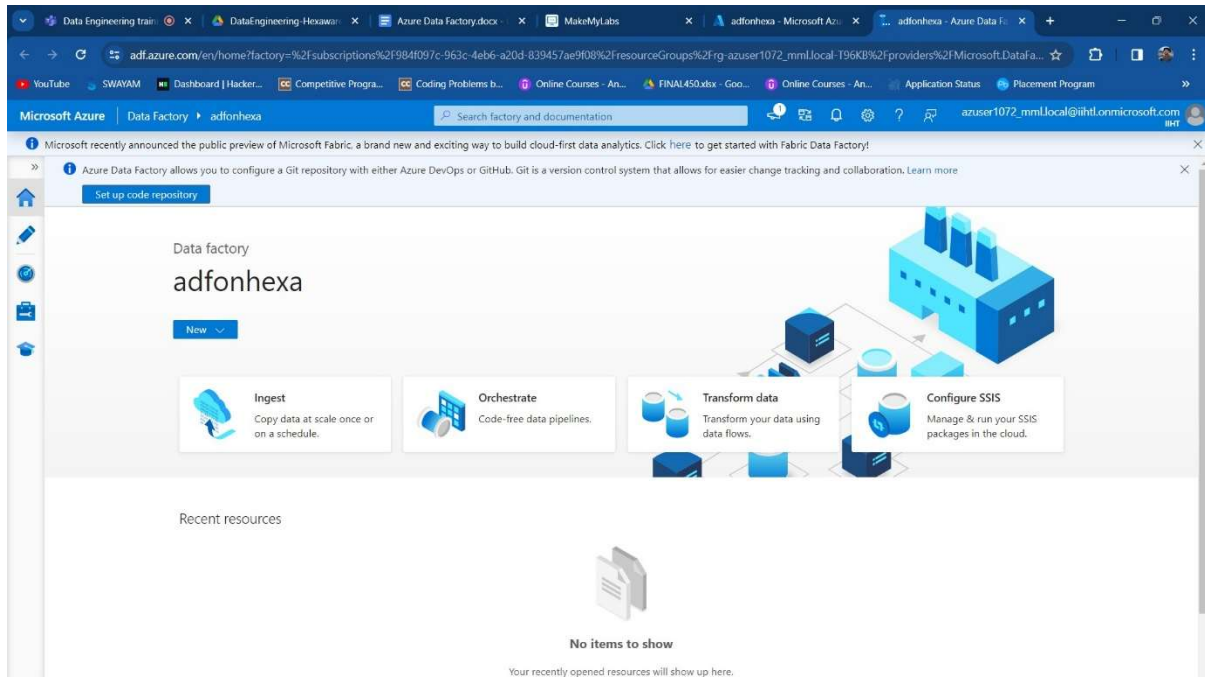
Find an Azure expert >

Use the copy data tool to copy data

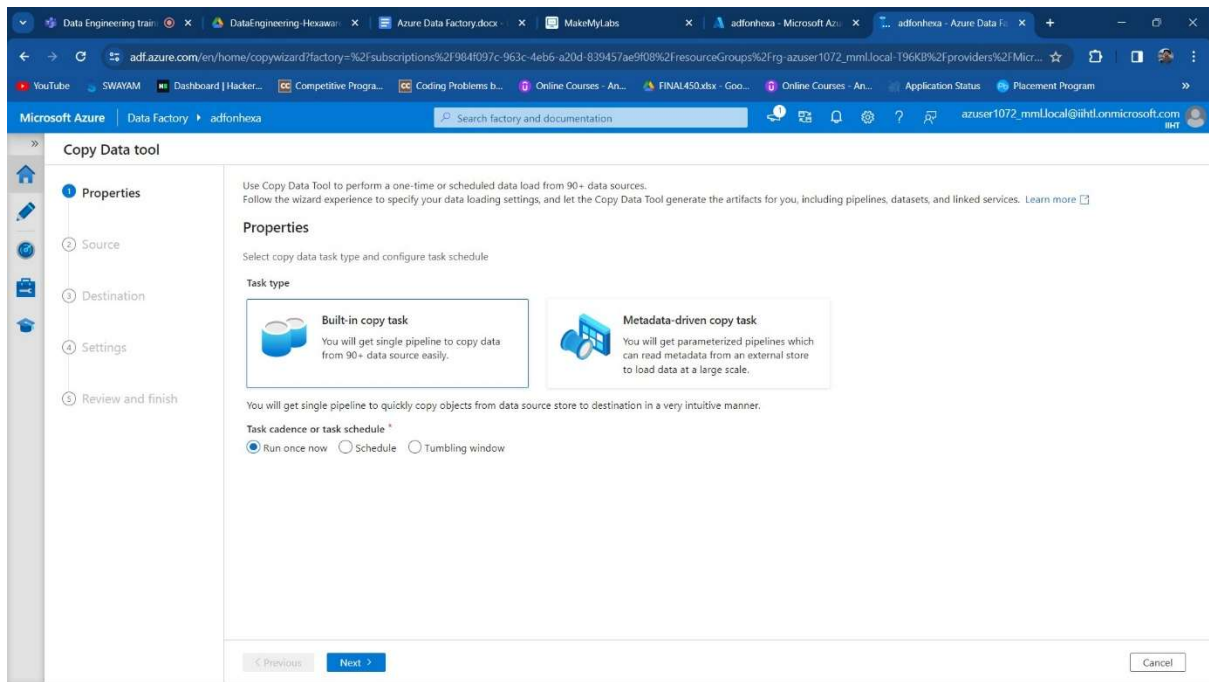
The steps below will walk you through how to easily copy data with the copy data tool in Azure Data Factory.

Step 1: Start the copy data Tool

1. On the home page of Azure Data Factory, select the **Ingest** tile to start the Copy Data tool.

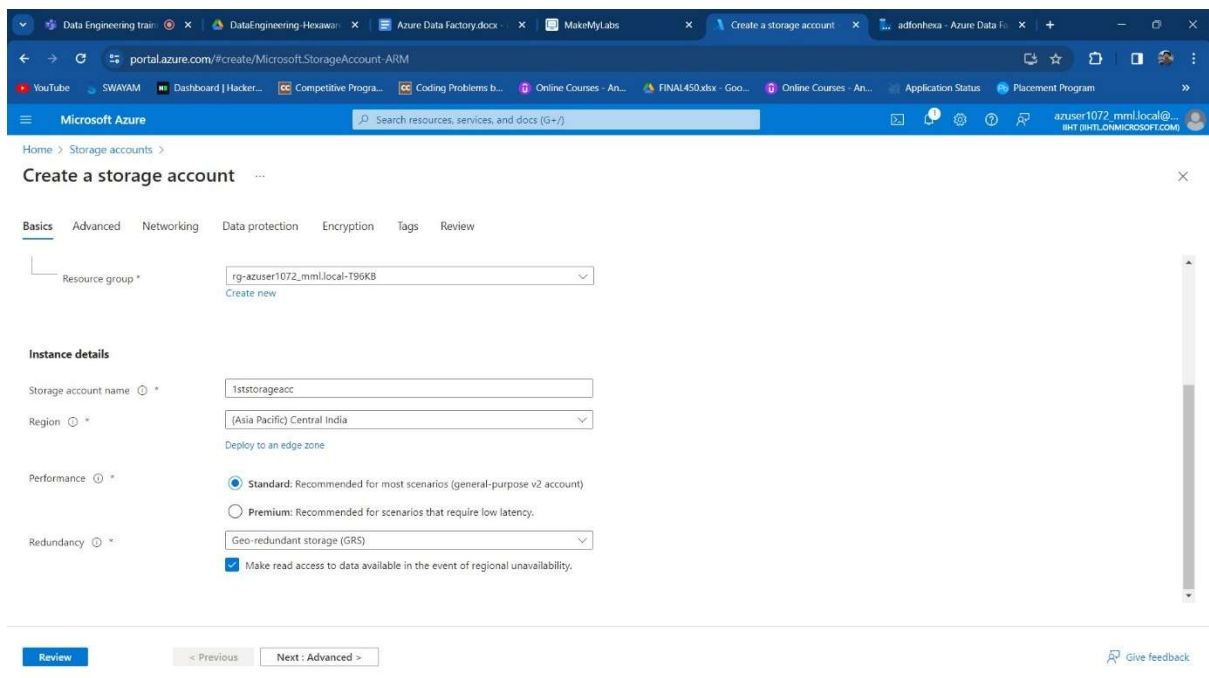


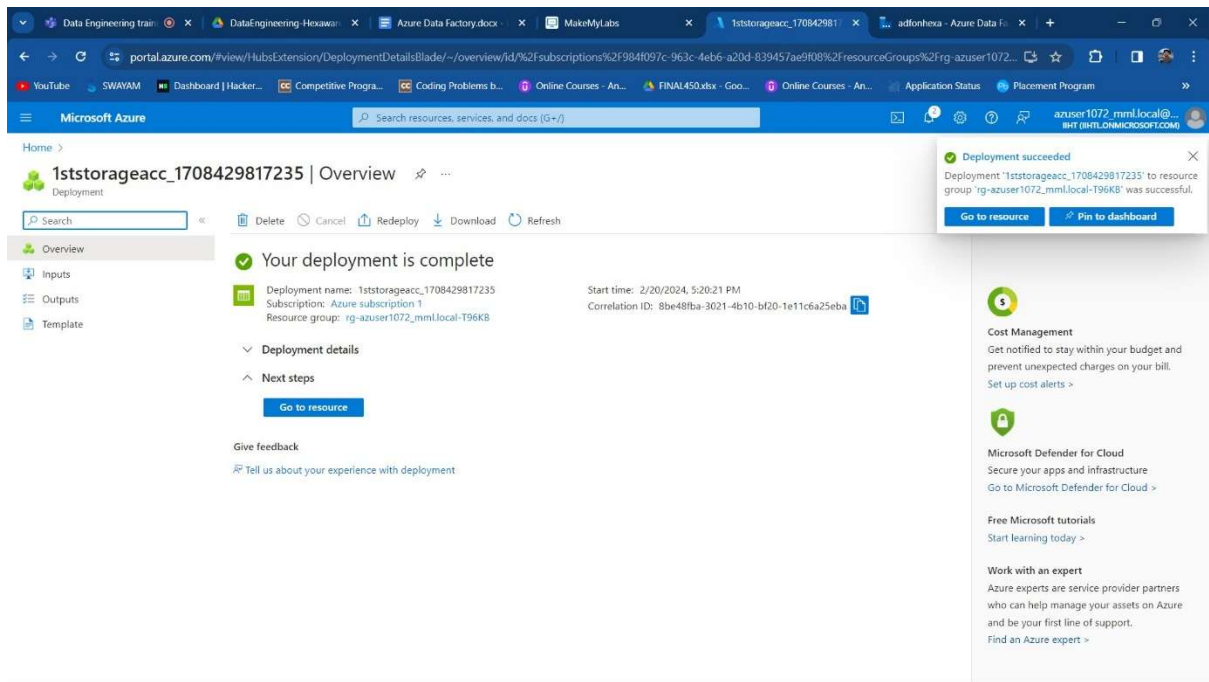
2. On the **Properties** page of the Copy Data tool, choose **Built-in copy task** under **Task type**, then select **Next**.



Step 2: Complete source configuration

1. Click + **Create new connection** to add a connection.
2. Select the linked service type that you want to create for the source connection. In this tutorial, we use **Azure Blob Storage**. Select it from the gallery, and then select **Continue**.

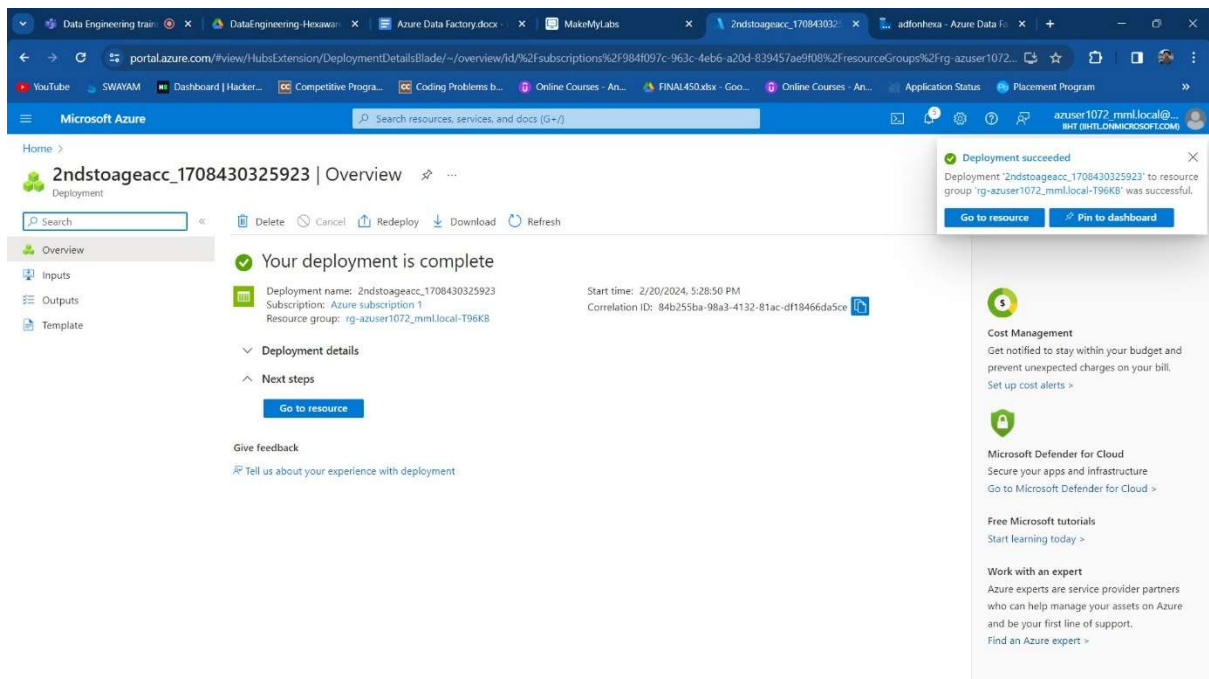
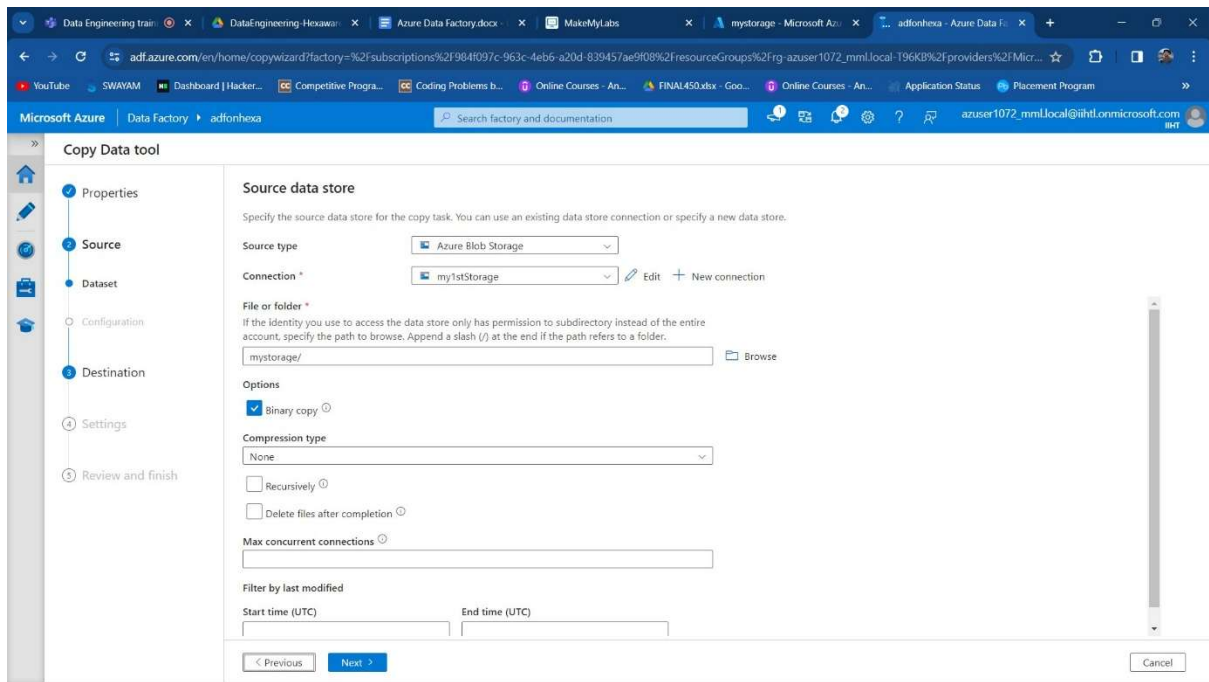




1. On the **New connection (Azure Blob Storage)** page, specify a name for your connection. Select your Azure subscription from the **Azure subscription** list and your storage account from the **Storage account name** list, test connection, and then select **Create**.

Select the newly created connection in the **Connection** block.

1. In the **File or folder** section, select **Browse** to navigate to the **mystorage/input** folder, select the file, and then click **OK**.
2. Select the **Binary copy** checkbox to copy file as-is, and then select **Next**.



Step 3: Complete destination configuration

1. Select the **AzureBlobStorage** connection that you created in the **Connection** block.
2. In the **Folder path** section, enter **my2ndstorage/output** for the folder path.

Copy Data tool

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:

Connection: [Edit](#) [+ New connection](#)

Folder path: [Browse](#)

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.

File name:

Compression type:

Copy behavior:

Max concurrent connections:

Block size (MB):

Metadata:

[< Previous](#) [Next >](#) [Cancel](#)

Step 4: Review all settings and deployment

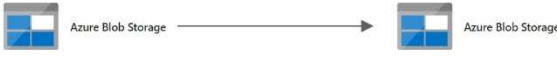
On the **Settings** page, specify a name for the pipeline and its description, then select **Next** to use other default configurations

1. On the **Summary** page, review all settings, and select **Next**.
2. On the **Deployment complete** page, select **Monitor** to monitor the pipeline that you created.

Copy Data tool

Summary

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



Properties [Edit](#)

Task name:

Task description:

Source [Edit](#)

Connection name:

Dataset name:

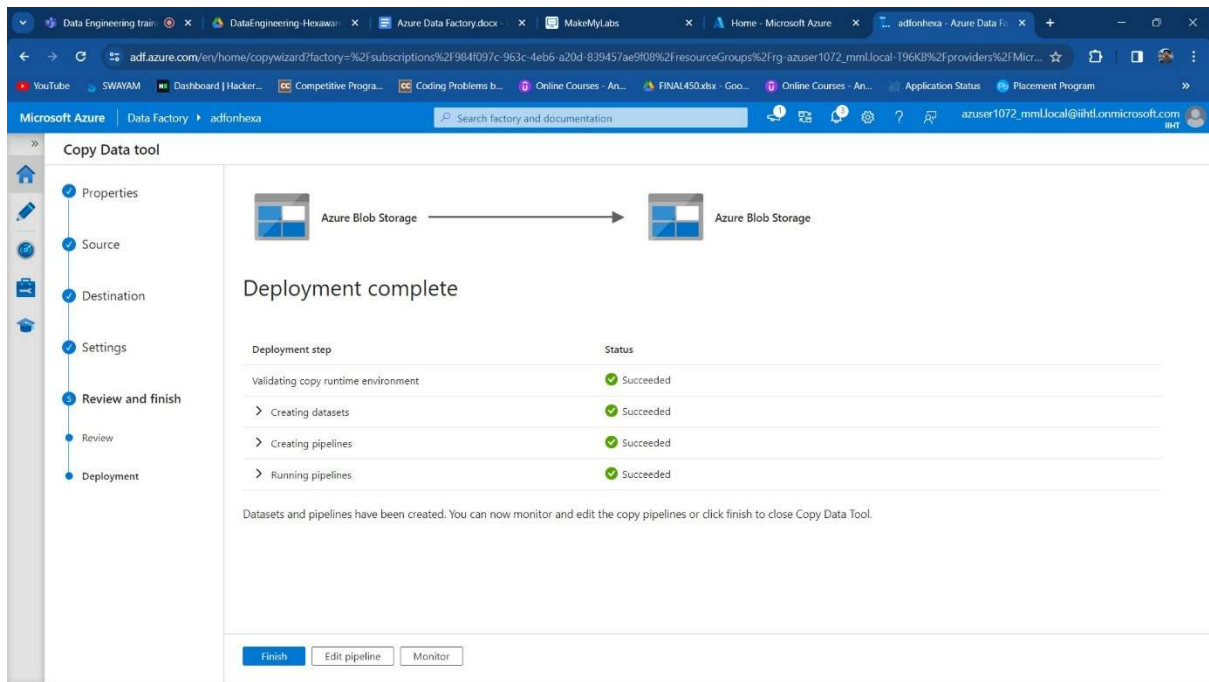
Container:

Destination [Edit](#)

Connection name:

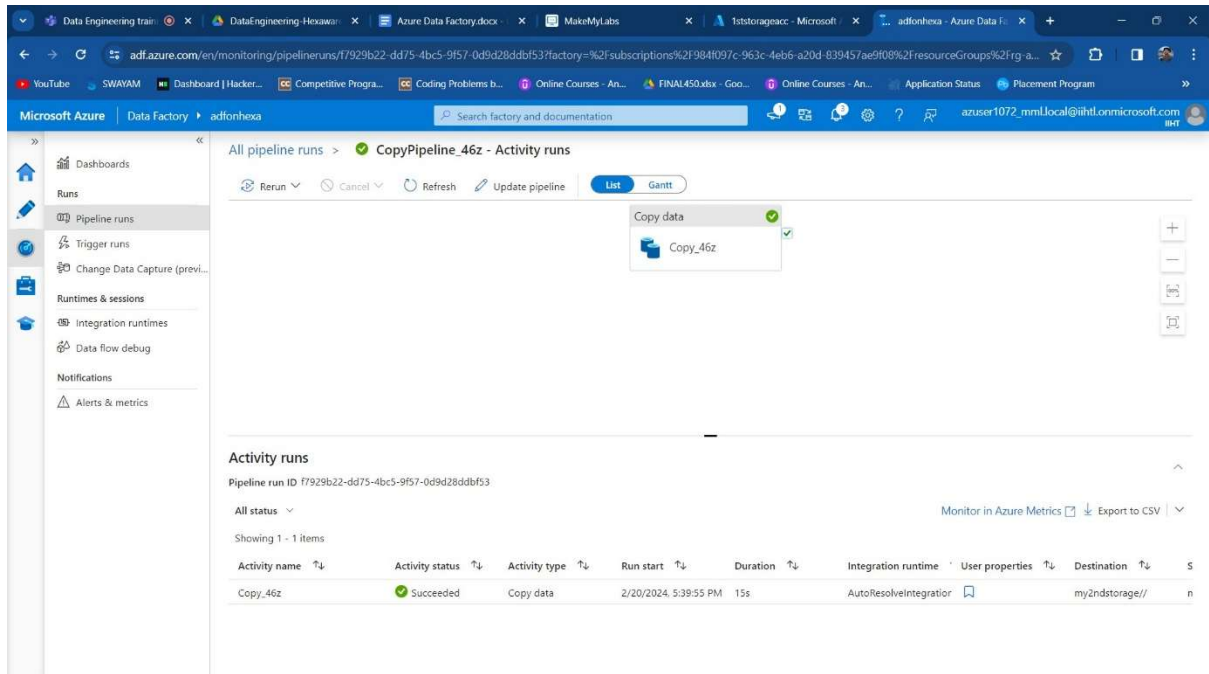
Dataset name:

[< Previous](#) [Next >](#) [Cancel](#)



Step 5: Monitor the running results

1. The application switches to the **Monitor** tab. You see the status of the pipeline on this tab. Select **Refresh** to refresh the list. Click the link under **Pipeline name** to view activity run details or rerun the pipeline.



my2ndstorage

Container

Search

UploadChange access levelRefreshDeleteChange tierAcquire leaseBreak leaseView snapshotsCreate snapshotGive feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Access policy

Properties

Metadata

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

Location: my2ndstorage

Search blobs by prefix (case-sensitive)

Show deleted blobs

Add filter

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state	
<input type="checkbox"/> student-dataset.csv	2/20/2024, 5:40:08 PM	Hot (Inferred)		Block blob	26.79 KiB	Available	***

https://portal.azure.com/#