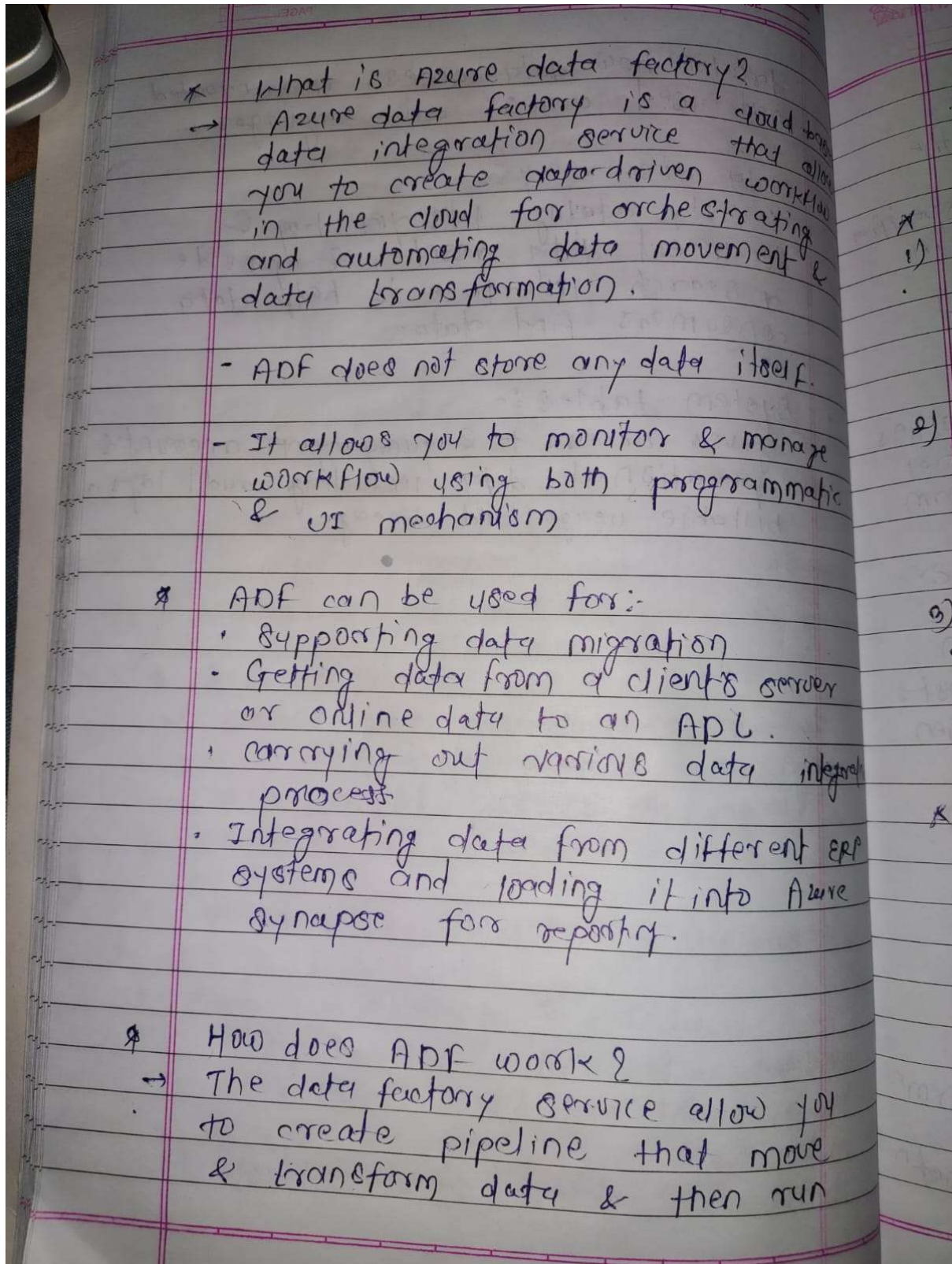


Name: Krushnakumar Patle

Email: [krishnapatle128@gmail.com](mailto:krishnapatle128@gmail.com)

Batch: Data Engineering Batch-1



the pipelines on a specified schedule (hourly, weekly, monthly etc).

\* ADF can perform in 3 steps:-

1) Step 1: Connect & collect  
Connect to all the required source of data & processing such as SQLS, SaaS, on-premise.

2) Step 2: Transform & enrich  
Once data present in centralized storage it is transformed using compute service such as HDInsight, Hadoop, Spark, Azure Databricks.

3) Step 3: Publish  
Deliver transformed data from the cloud to on-premise.

\* Azure Data factory key component:-  
It has 4 key component that work together to define input & output data, processing events, & the schedule & resources required to execute the desired data flow.

1) Dataset represent data structures within the data stores.

An input dataset represents the input



for an activity in the pipeline.  
An output dataset represents the output for the activity.  
ex. an azure blob dataset specifies the blob container & folder in the Azure blob storage from which the pipeline should read the data.

2) A pipeline is a group of activities. They are used to group activities into a unit that together perform a task.

A data factory may have one or more pipelines.

ex. a pipeline could contain a group of activities that ingest data from an Azure blob & then runs a Hive query on an HDInsight cluster to partition the data.

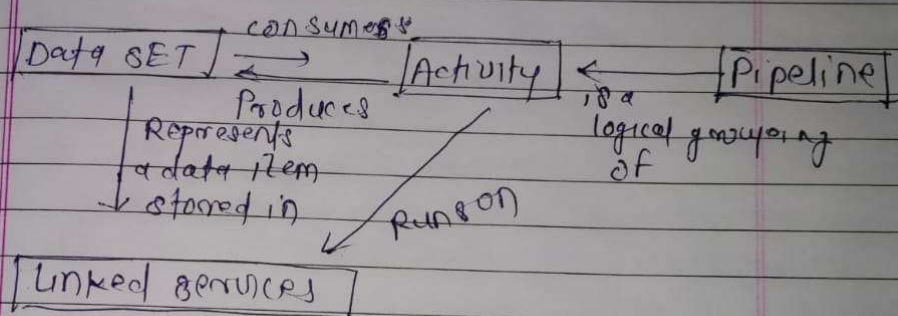
3) Activities define the actions to perform on your data :-

currently, Azure Data Factory supports two types of activities, data movements & data transformation.

4) Linked services define the information

needed for Azure data factory to connect to external resources :-

\* How ADF ~~work~~ component work together :-



you can use one of the following tools or APIs to create data pipeline in Azure data factory.

- Azure portal
- Visual studio
- Powershell
- .NET API
- REST API
- Azure resource management template

## Copy Activity:

### 1. Create a data factory

portal.azure.com/#create/Microsoft.DataFactory

Microsoft Azure

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 \* Azure subscription 1

Resource group \* rg-azuser1072\_mm1local-T96K8

Create new

Instance details

Name \* adfonhexa

Region \* East US

Version \* V2

Previous Next Review + create

Give feedback

portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2fsubscriptions%2f984f097c-963c-4eb6-a20d-839457ae9f08%2fresourceGroups%2frg-azuser1072...

Microsoft Azure

Home >

Microsoft.DataFactory-20240220171449 | Overview

Deployment

Search

Delete Cancel Redeploy Download Refresh

Overview

Inputs

Outputs

Template

✓ Your deployment is complete

Deployment name : Microsoft.DataFactory-20240220171449

Subscription : Azure subscription 1

Resource group : rg-azuser1072\_mm1local-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\_mm1local-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 >

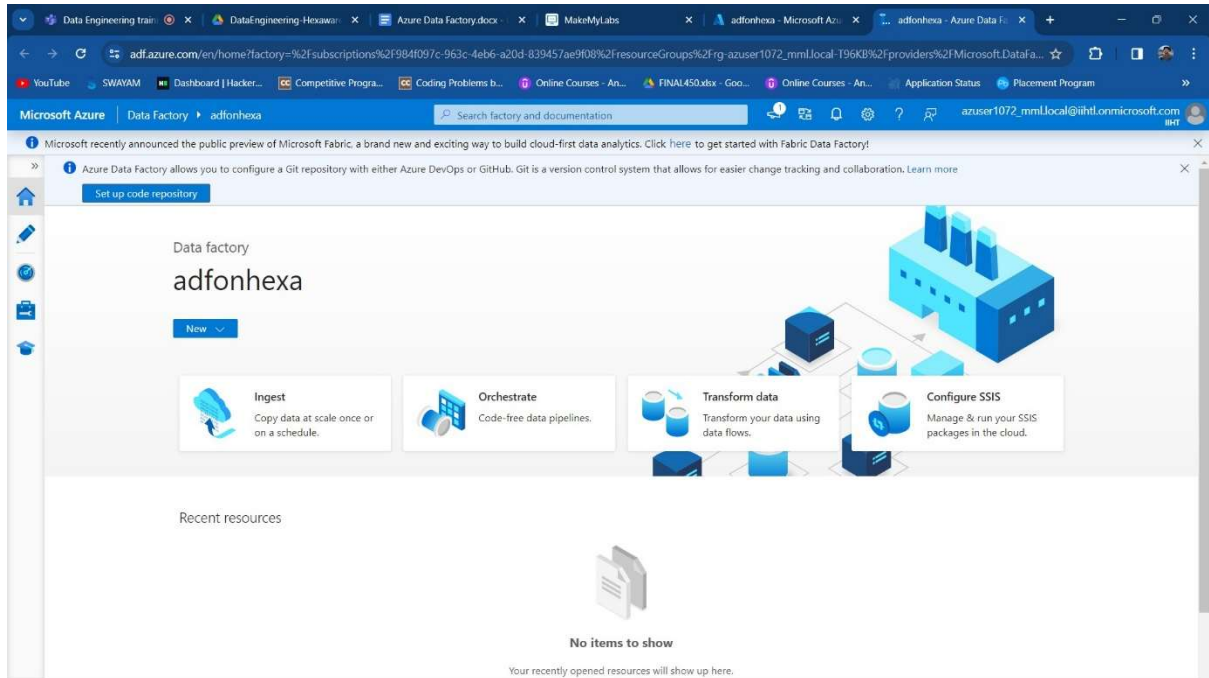
## Use the copy data tool to copy data

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

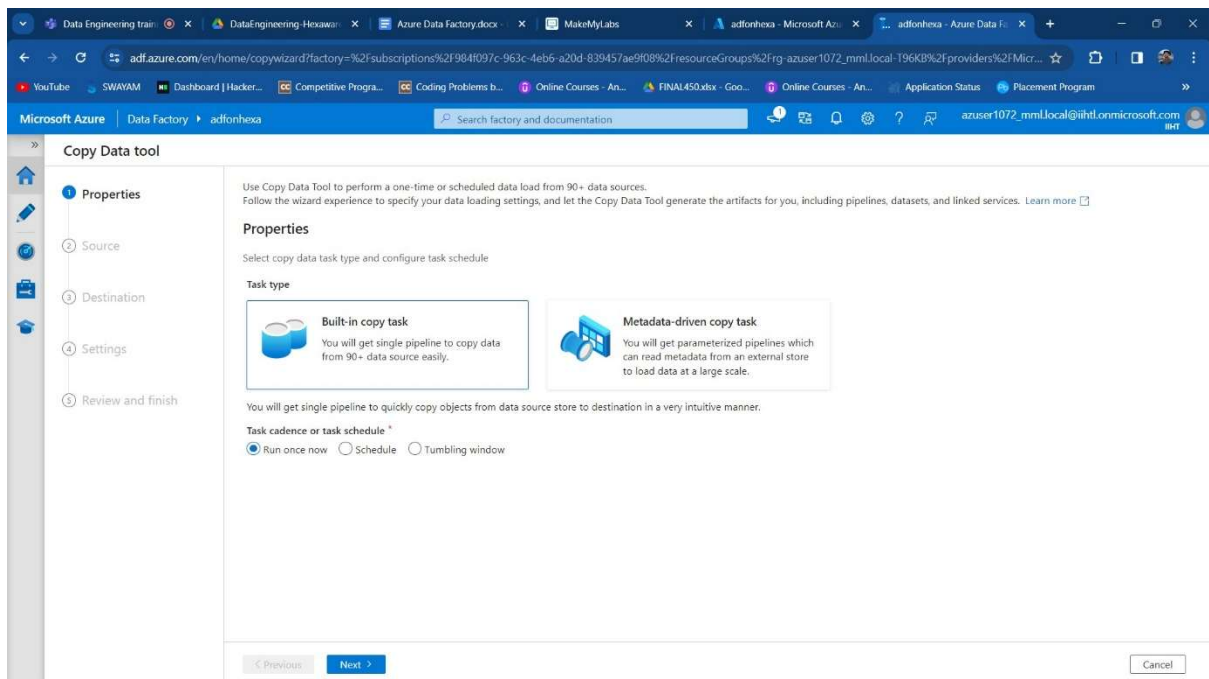


## Step 1: Start the copy data Tool

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

## Day 22 Assessment

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**.

The screenshot shows the 'Create a storage account' page in the Microsoft Azure portal. The page is divided into tabs: Basics, Advanced, Networking, Data protection, Encryption, Tags, and Review. The 'Basics' tab is active, showing the following fields:

- Resource group: rg-azuser1072\_mml-local-T96K8
- Storage account name: 1ststorageacc
- Region: (Asia Pacific) Central India
- Performance: Standard (Recommended for most scenarios (general-purpose v2 account))
- Redundancy: Geo-redundant storage (GRS)

At the bottom, there is a 'Review' button and a 'Next : Advanced >' button.

The screenshot shows the '1ststorageacc\_1708429817235 | Overview' page in the Microsoft Azure portal. The page displays the deployment details for the storage account. The deployment is complete, and a 'Deployment succeeded' notification is visible in the top right corner. The deployment details include:

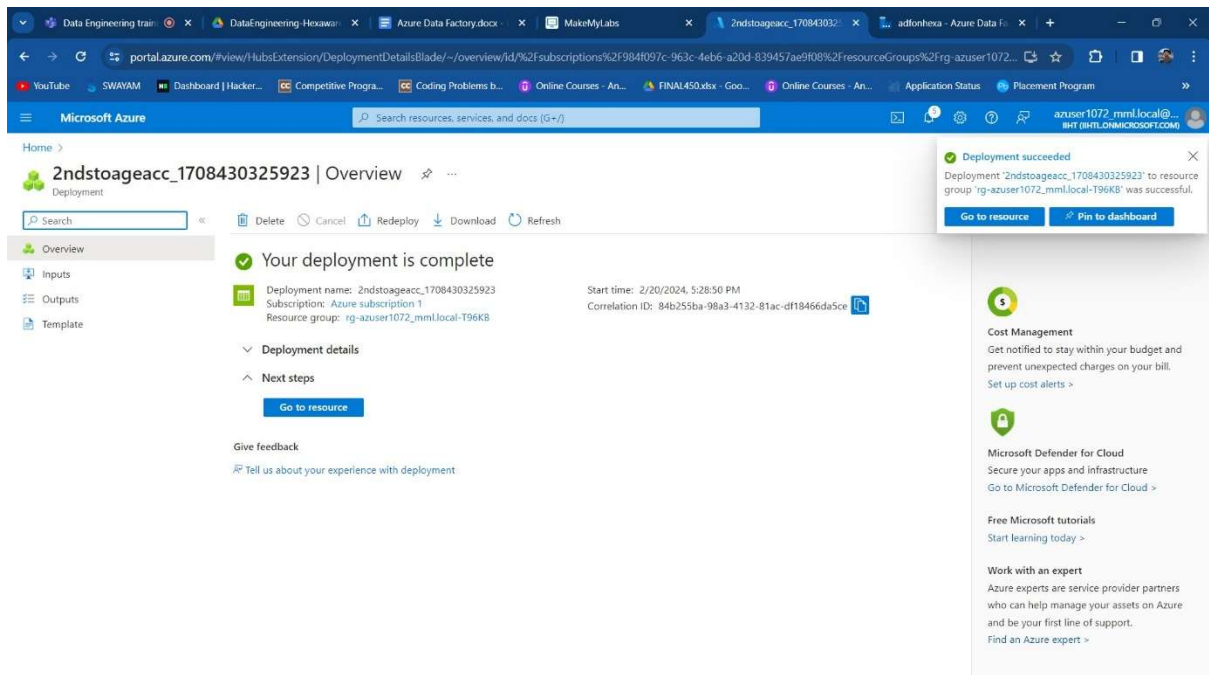
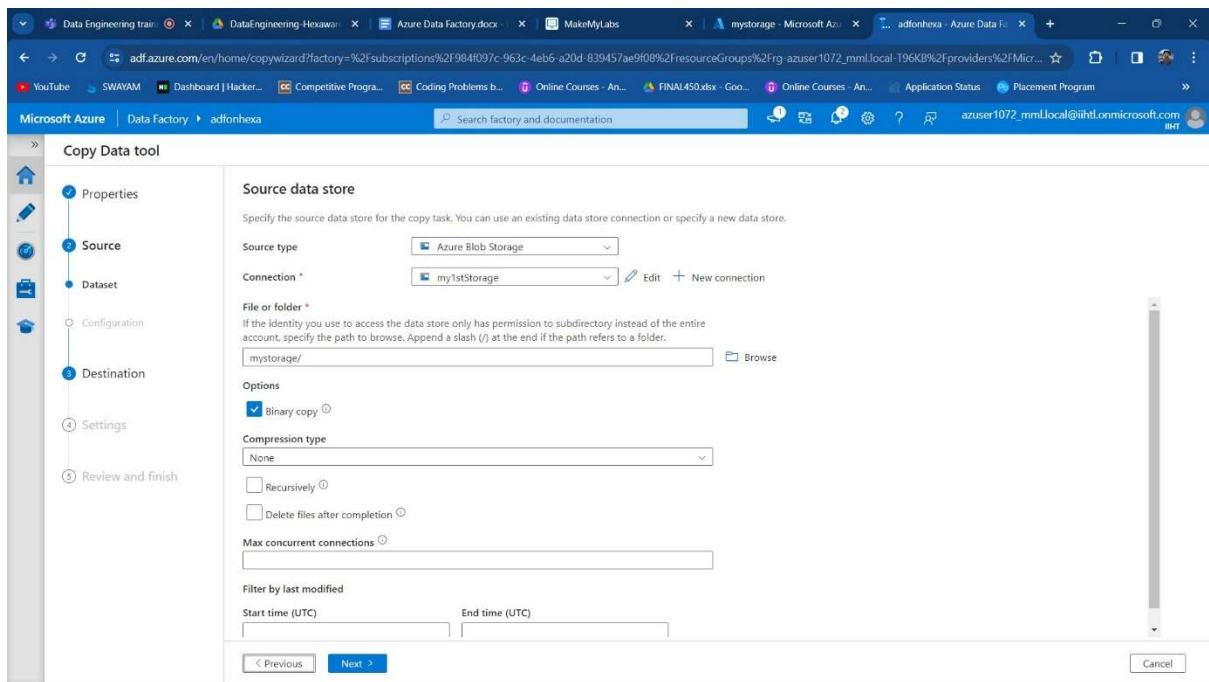
- Deployment name: 1ststorageacc\_1708429817235
- Subscription: Azure subscription 1
- Resource group: rg-azuser1072\_mml-local-T96K8
- Start time: 2/20/2024, 5:20:21 PM
- Correlation ID: 8be48fba-3021-4b10-b20-1e11c6a25eba

The page also includes a 'Go to resource' button and a 'Give feedback' link.

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**.

## Day 22 Assessment

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

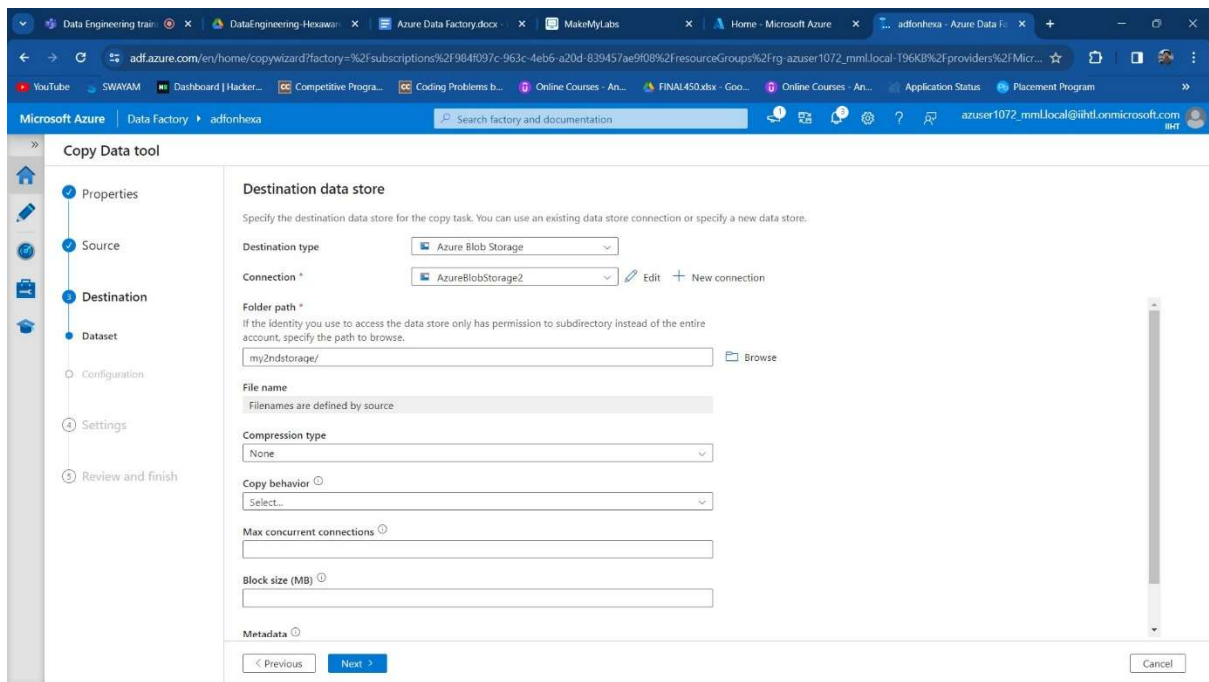


## Step 3: Complete destination configuration



## Day 22 Assessment

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.

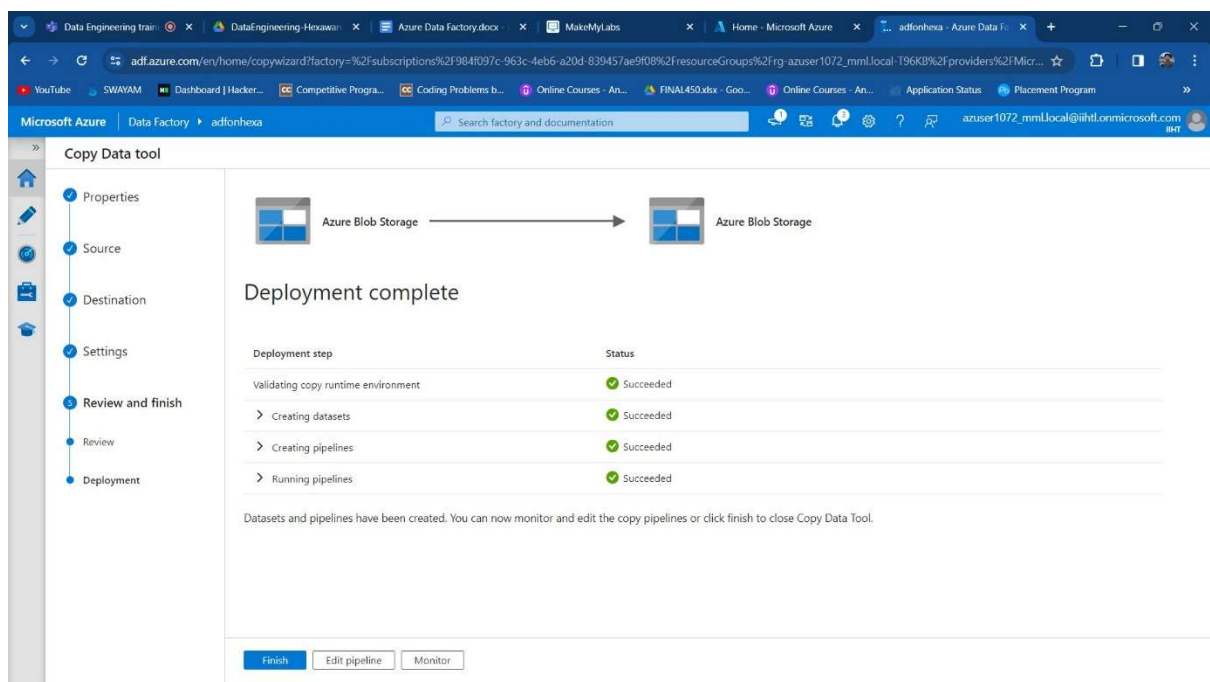
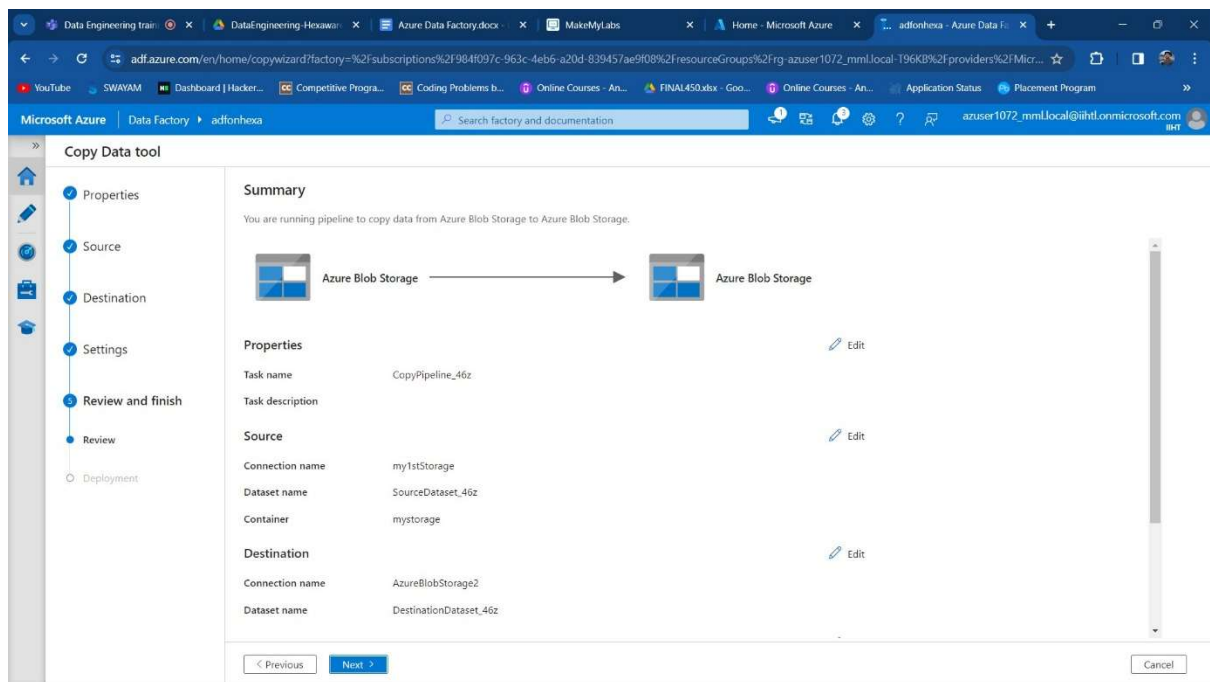


## Step 4: Review all settings and deployment

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.

## Day 22 Assessment



### 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.

## Day 22 Assessment

All pipeline runs > CopyPipeline\_46z - Activity runs

Rerun Cancel Refresh Update pipeline List Gantt

Copy data

Copy\_46z

Activity runs

Pipeline run ID: f7929b22-dd75-4bc5-9f57-0d9d28dbf53

All status Monitor in Azure Metrics Export to CSV

Showing 1 - 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Destination
Copy_46z	Succeeded	Copy data	2/20/2024, 5:39:55 PM	15s	AutoResolveIntegration		my2ndstorage//

Home > 2ndstoageacc | Containers >

my2ndstorage

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
student-dataset.csv	2/20/2024, 5:40:08 PM	Hot (Inferred)		Block blob	26.79 KiB	Available

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