

Coding Challenge-4.3

Azure Databricks

Name-Ajay Chaudhary

Batch-Data Engineering(Batch-1)

Execute & explain, Azure datafactory and its copy activity.

Azure Data Factory is a cloud-based data integration service that allows you to create, schedule, and orchestrate data workflows at scale. It enables you to build data pipelines to move data between various sources and destinations, transform data, and monitor the entire process.

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.

Execution of Copy Activity in Microsoft Azure-

Copy Activity:

1. Create a data factory

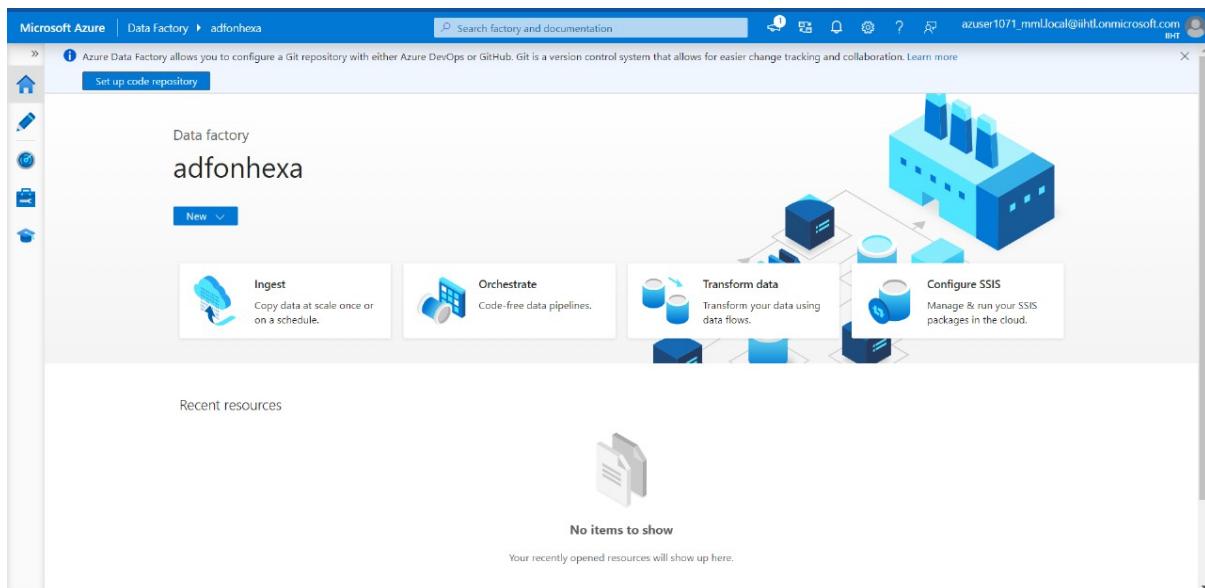
The screenshot shows the 'Create Data Factory' wizard in the Azure portal. The 'Basics' tab is selected. At the top, there's a breadcrumb navigation: Home > Data factories > Create Data Factory. Below the tabs are several informational links: 'One-click to create data factory with sample pipeline and datasets. Try it', 'Project details', and 'Instance details'. Under 'Project details', the 'Subscription' dropdown is set to 'Azure subscription 1' and the 'Resource group' dropdown is set to 'rg-azuser1072_mml.local-T96K8'. Under 'Instance details', the 'Name' field is 'adfonhexa', 'Region' is 'East US', and 'Version' is 'V2'. At the bottom of the form are 'Previous' and 'Next' buttons, and a prominent blue 'Review + create' button.

Use the copy data tool to copy data

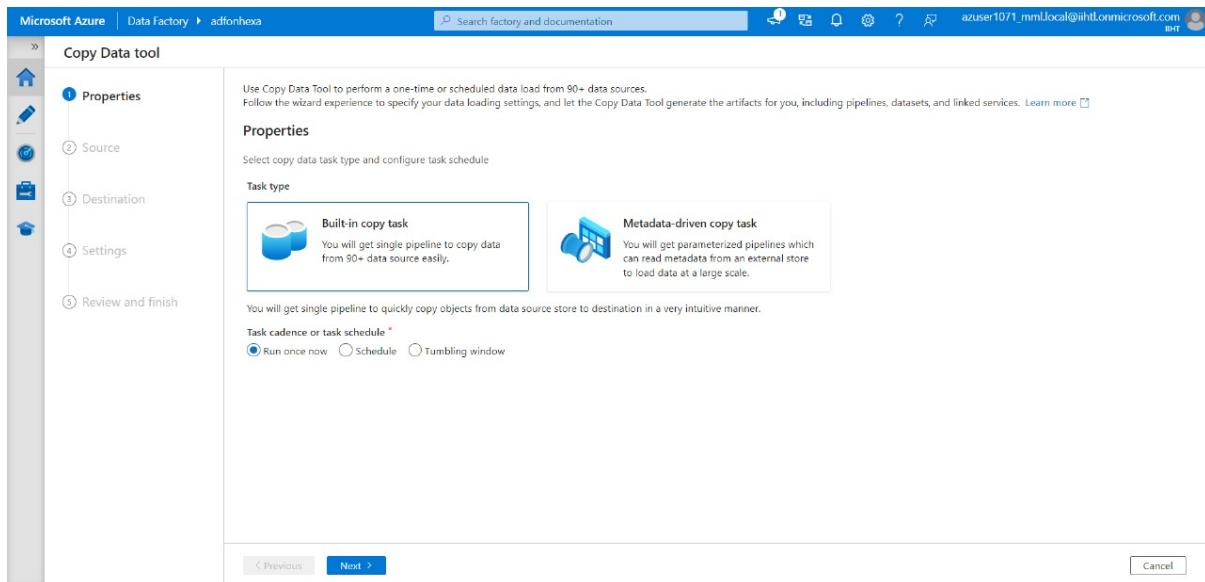
The steps below will walk you through how to easily copy data with the copy data tool in Azure Data Factory.

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**.

Microsoft Azure | Data Factory > adfonhexa

Search factory and documentation

New connection

Azure Blob Storage Learn more

Authentication type

Account key

Connection string Azure Key Vault

Account selection method

From Azure subscription Enter manually

Azure subscription

Select all

Storage account name

my2ndstorageacc

Additional connection properties

+ New

Test connection

To linked service To file path

Annotations

+ New

> Parameters

Connection successful

Test connection

Copy Data tool

Properties

Source

Destination

Dataset

Configuration

Settings

Review and finish

Previous Next Create Cancel

Microsoft Azure

Search resources, services, and docs (G+)

azuser1071_mmlLocal@... IHT (IHT.ONMICROSOFT.COM)

my1ststorageacc_1708493754076 | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

Deployment succeeded

Deployment 'my1ststorageacc_1708493754076' to resource group 'databricks-rg-learn-dbx-1071' was successful.

Go to resource Pin to dashboard

Overview

Inputs

Outputs

Template

Your deployment is complete

Deployment name: my1ststorageacc_1708493754076

Subscription: Azure subscription 1

Resource group: databricks-rg-learn-dbx-1071

Start time: 2/21/2024, 11:05:58 AM

Correlation ID: 6b20d461-fbfe-4017-805b-985552342ea4

Deployment details

Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

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 >

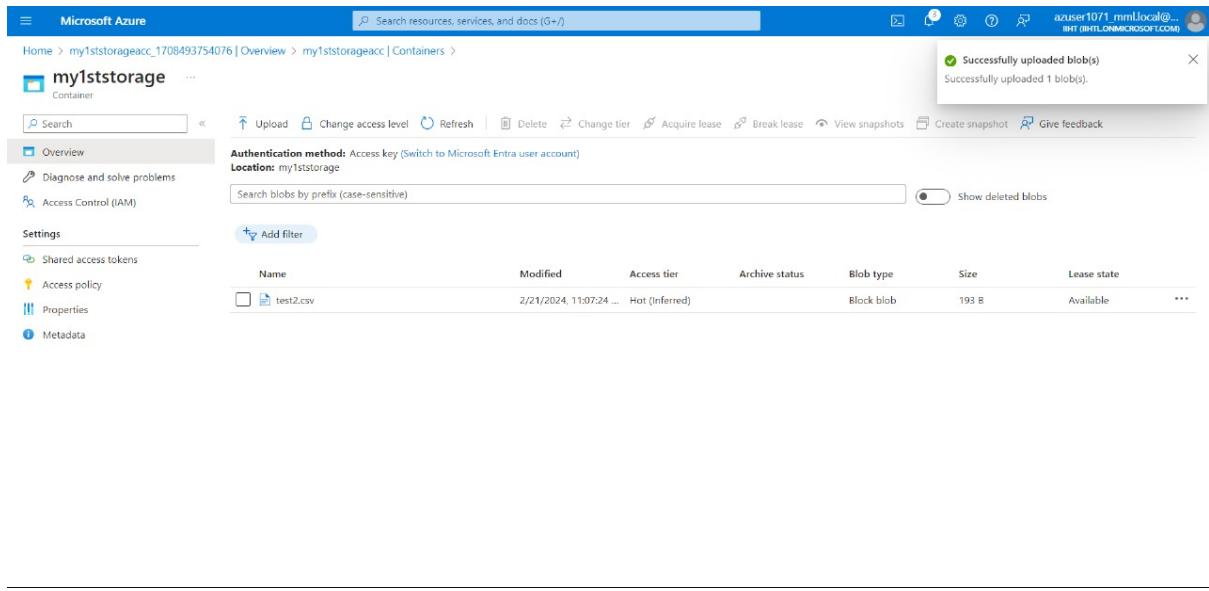
Free Microsoft tutorials

Start learning today >

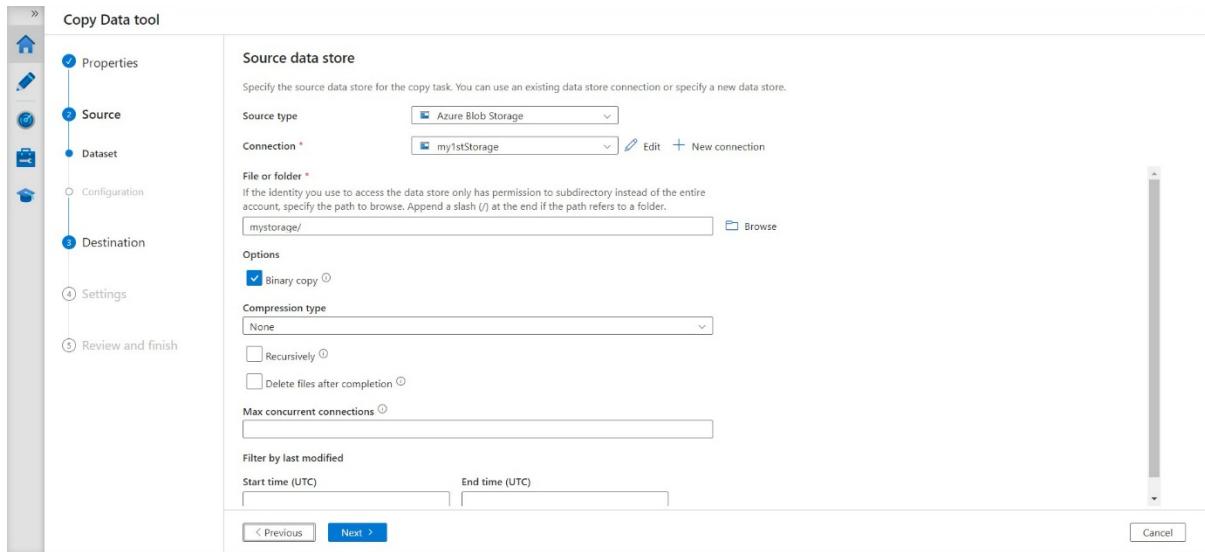
Work with an expert

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 >



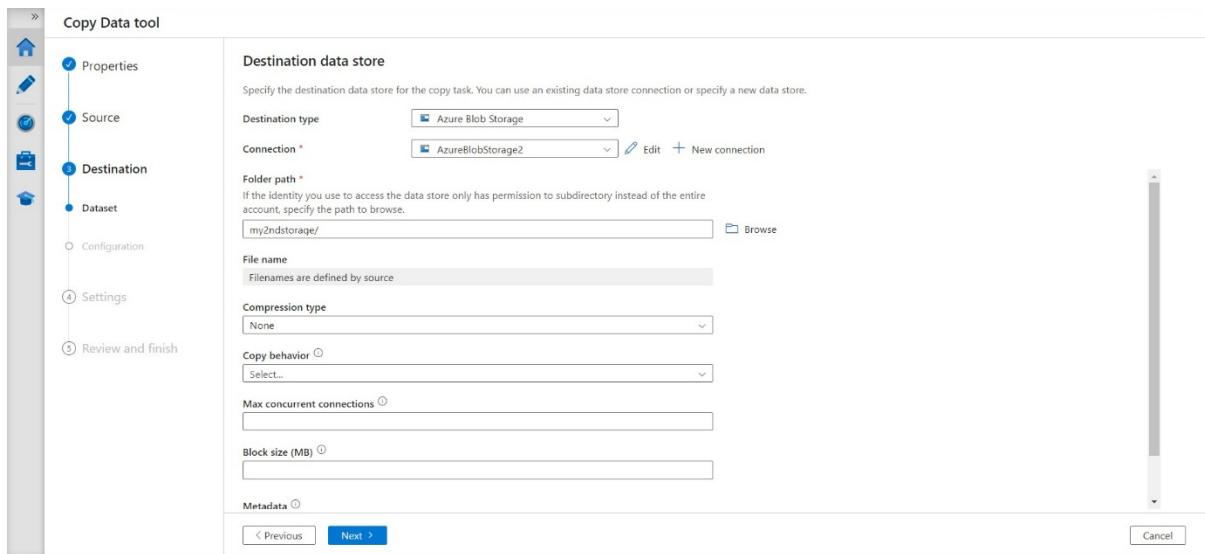
1. On the **New connection (Azure Blob Storage)** page, specify a name for your connection. Select your Azure subscription from the **Azure subscription** list and your storage account from the **Storage account name** list, test connection, and then select **Create**.
1. Select the newly created connection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



The screenshot shows the Microsoft Azure portal's deployment overview page. The deployment name is 'my2ndstorageacc_1708493889574'. The status bar indicates 'Deployment succeeded'. Deployment details include a start time of 2/21/2024, 11:08:13 AM, and a correlation ID of a30331f4-47d1-4d20-b41b-21ad8bf532a3. A 'Go to resource' button is visible.

Step 3: Complete destination configuration

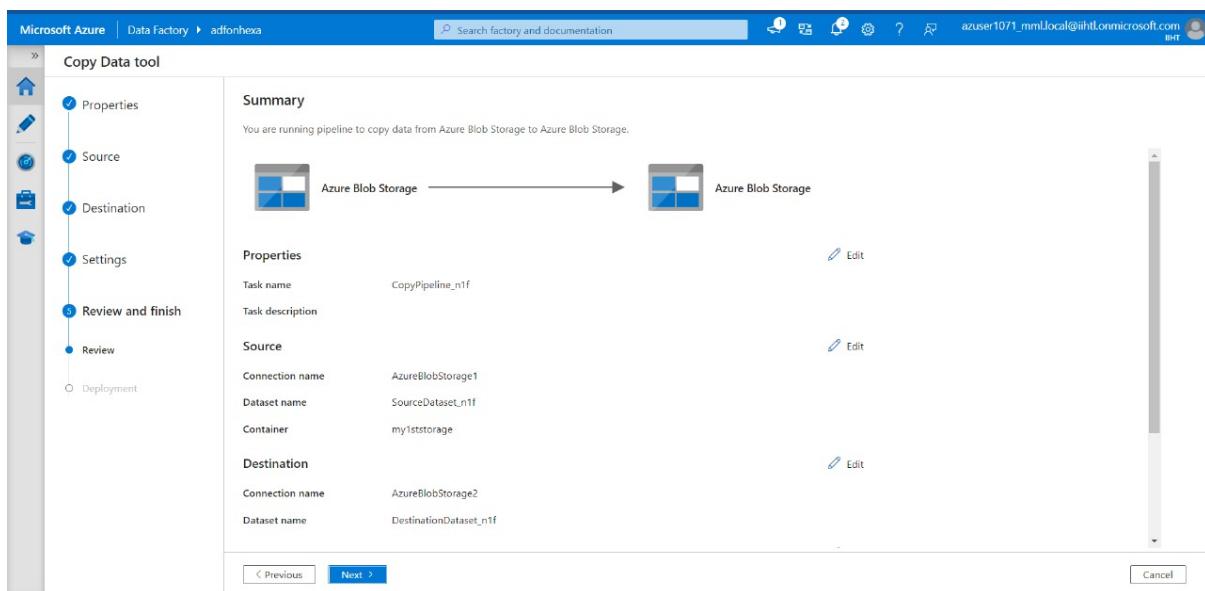
1. Select the **AzureBlobStorage** connection that you created in the **Connection** block.
2. In the **Folder path** section, enter **my2ndstorage/output** for the folder path.

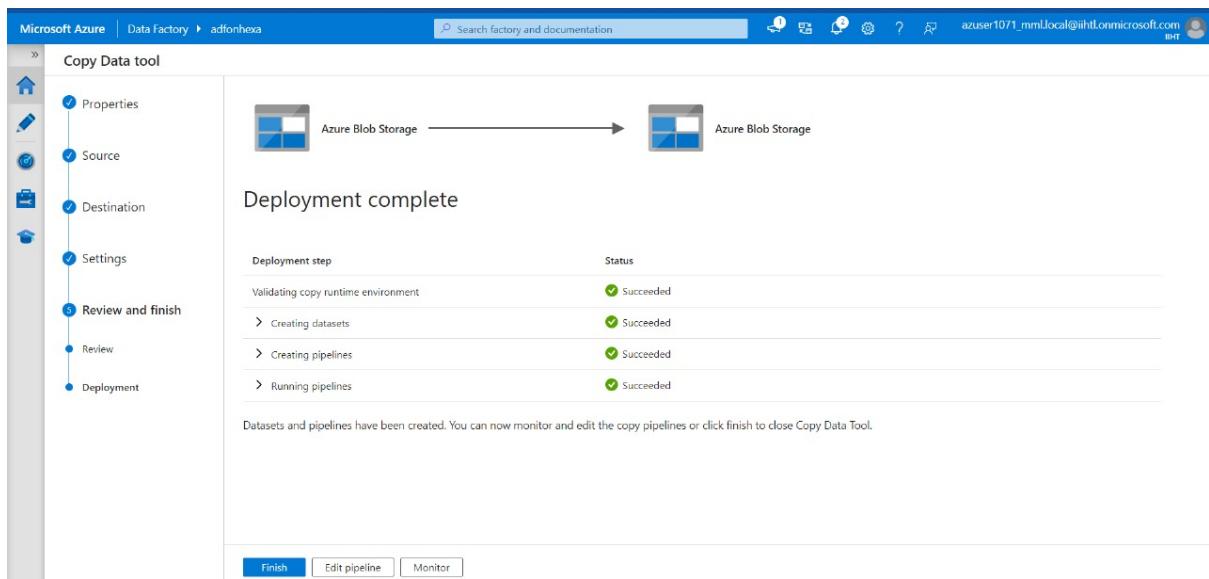


Step 4: Review all settings and deployment

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.





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.

The screenshot shows the Microsoft Azure Data Factory Monitor tab. The left sidebar includes links for Dashboards, Runs, Pipeline runs (which is selected and highlighted in grey), Trigger runs, Change Data Capture, Runtimes & sessions, Integration runtimes, Data flow debug, Notifications, and Alerts & metrics. The main area shows 'All pipeline runs > CopyPipeline_46z - Activity runs'. It features a toolbar with 'Rerun', 'Cancel', 'Refresh', 'Update pipeline', 'List' (selected), and 'Gantt'. Below the toolbar, a table titled 'Activity runs' lists one item: 'Copy data' for 'Copy_46z'. The table columns are: Activity name, Activity status, Activity type, Run start, Duration, Integration runtime, User properties, Destination, and \$\$. The status for the activity is 'Succeeded'. At the bottom right of the table, there are buttons for 'Monitor in Azure Metrics' and 'Export to CSV'.

Microsoft Azure

Search resources, services, and docs (G+)

azuser1071.mmllocal@... IHT (BTHLONMICROSOFT.COM)

Home > my2ndstorageacc_1708493889574 | Overview > my2ndstorageacc | Containers >

my2ndstorage Container

Search Upload Change access level Delete Acquire lease View snapshots Give feedback

Overview Diagnose and solve problems Access Control (IAM)

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
test2.csv	2/21/2024, 11:12:34 ...	Hot (Inferred)		Block blob	193 B	Available

Settings Shared access tokens Access policy Properties Metadata