

Name: Pradip Bochare

Azure Databricks Coding Assessment Question 3

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

❖ **Azure Data Factory**

- 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.

❖ **ADF Use cases**

- 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

❖ **How does ADF work**

- The Data Factory service allows you to create data pipelines that move and transform data and then run the pipelines on a specified schedule.
- 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 in Azure Data Factory copies data from a source data store to a sink data store.
- Azure supports various data stores such as source or sink data stores like Azure Blob storage, Azure Cosmos DB (DocumentDB API), Azure Data Lake Store, Oracle, Cassandra, etc.

➤ Creating two storage containers first is source container and another is destination container

- Source Container

The screenshot shows the 'Create a storage account' wizard in the Microsoft Azure portal, specifically the 'Basics' tab. The page is titled 'Create a storage account' with a close button (X) in the top right corner. Below the title are tabs for 'Basics', 'Advanced', 'Networking', 'Data protection', 'Encryption', 'Tags', and 'Review'. The 'Project details' section instructs the user to select a subscription and a resource group. The 'Subscription' dropdown is set to 'Azure subscription 1', and the 'Resource group' dropdown is set to 'rg-azuser1077_mml.local-WFLY', with a 'Create new' link below it. The 'Instance details' section includes the 'Storage account name' (hexadeb1077source), 'Region' (Asia Pacific Central India), and 'Performance' (Standard: Recommended for most scenarios (general-purpose v2 account)). At the bottom, there are 'Review', '< Previous', and 'Next: Advanced >' buttons, along with a 'Give feedback' link.

Microsoft Azure

Search resources, services, and docs (G+/I)

Home > Storage accounts >

Create a storage account

Basics Advanced Networking Data protection Encryption Tags Review

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription * Azure subscription 1

Resource group * rg-azuser1077_mml.local-WFLY

[Create new](#)

Instance details

Storage account name * hexadeb1077source

Region * (Asia Pacific) Central India

[Deploy to an edge zone](#)

Performance * ☒ Standard: Recommended for most scenarios (general-purpose v2 account)

Review < Previous Next: Advanced > Give feedback

The screenshot shows the 'Overview' page for the deployment 'hexadeb1077source_1708495277286' in the Microsoft Azure portal. The page title is 'hexadeb1077source_1708495277286 | Overview' with a close button (X) in the top right corner. Below the title is a 'Deployment' section with a search bar and buttons for 'Delete', 'Cancel', 'Redeploy', 'Download', and 'Refresh'. The 'Overview' tab is selected, showing a green checkmark and the message 'Your deployment is complete'. The deployment details include the name 'hexadeb1077source_1708495277286', subscription 'Azure subscription 1', resource group 'rg-azuser1077_mml.local-WFLY', start time '21/02/2024, 11:31:21', and correlation ID '3a39f1c9-d76b-4b2e-8d9d-888d75041b74'. There are sections for 'Deployment details', 'Next steps' (with a 'Go to resource' button), and 'Give feedback'. The right sidebar contains links for 'Cost Management', 'Microsoft Defender for Cloud', 'Free Microsoft tutorials', and 'Work with an expert'.

Microsoft Azure

Search resources, services, and docs (G+/I)

Home >

hexadeb1077source_1708495277286 | Overview

Deployment

Search < Delete Cancel Redeploy Download Refresh

Overview

☒ Your deployment is complete

Deployment name: hexadeb1077source_1708495277286 Start time: 21/02/2024, 11:31:21

Subscription: Azure subscription 1 Correlation ID: 3a39f1c9-d76b-4b2e-8d9d-888d75041b74

Resource group: rg-azuser1077_mml.local-WFLY

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 >](#)

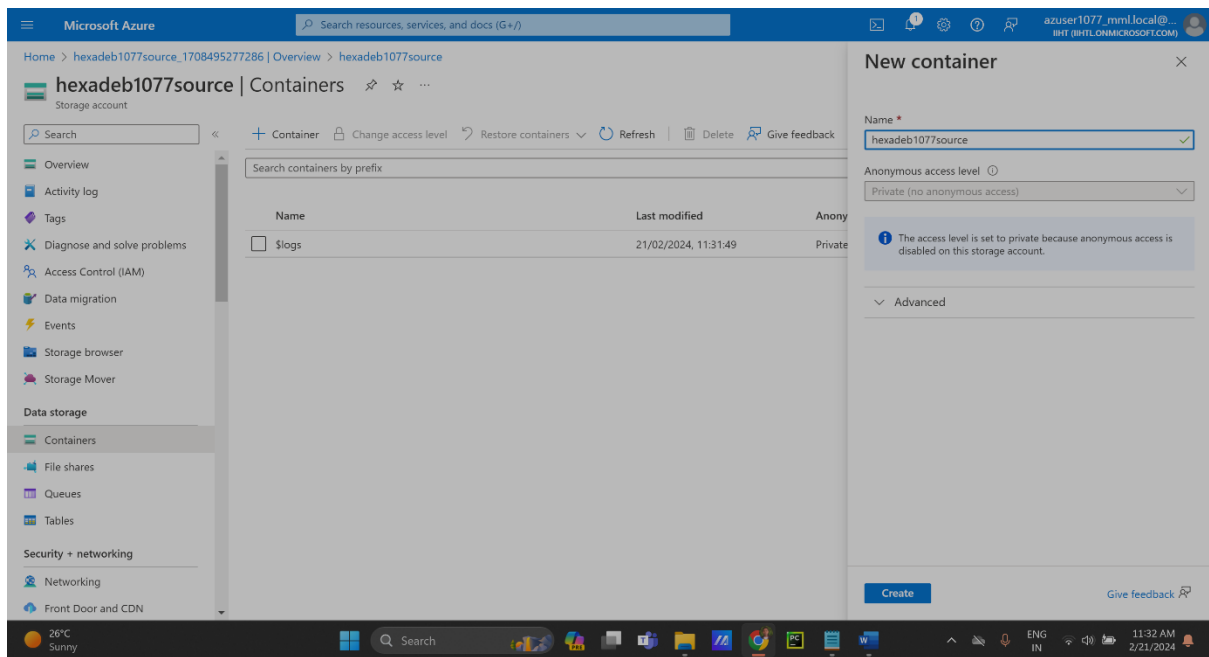
Free Microsoft tutorials

[Start learning today >](#)

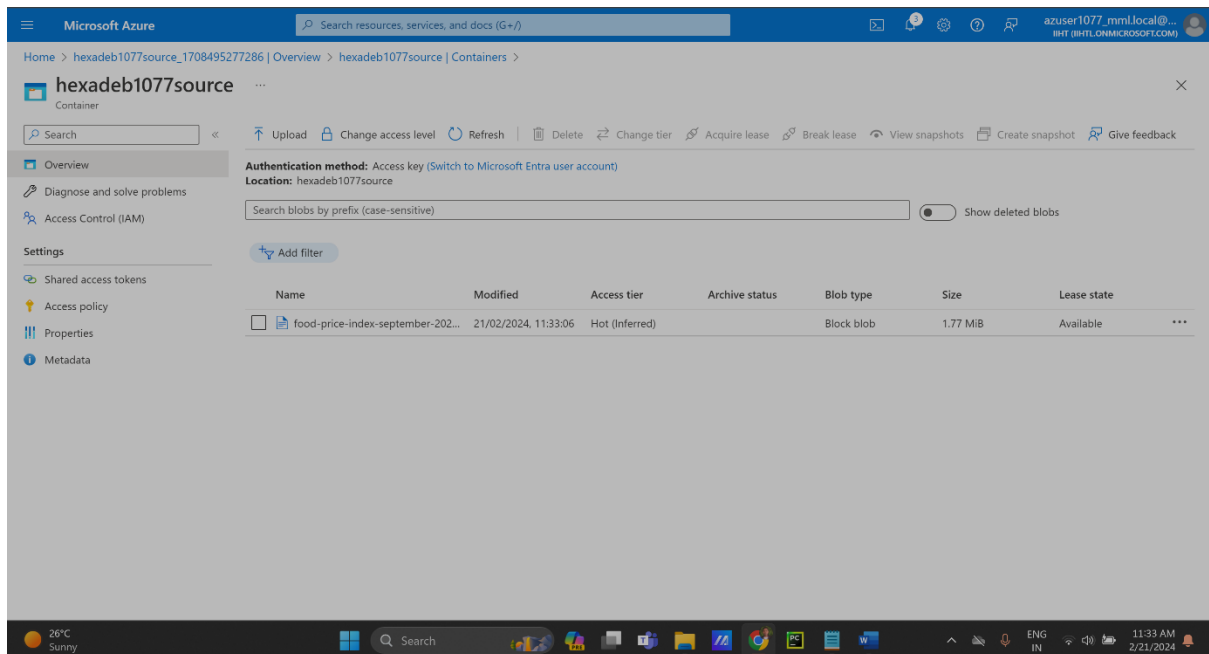
Work with an expert

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 >](#)



Adding data in source container



- Destination container

Microsoft Azure

Search resources, services, and docs (G+)

Home >

Create a storage account

Basics Advanced Networking Data protection Encryption Tags Review

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription * Azure subscription 1

Resource group * rg-azuser1077_mml.local-IWFLY [Create new](#)

Instance details

Storage account name ⓘ * hexadeb1077dest

Region ⓘ * (Asia Pacific) Central India [Deploy to an edge zone](#)

Performance ⓘ * ☒ Standard: Recommended for most scenarios (general-purpose v2 account)

[Review](#) < Previous Next: Advanced > [Give feedback](#)

Microsoft Azure

Search resources, services, and docs (G+)

Home >

hexadeb1077dest_1708495471863 | Overview

Deployment

Search < Delete Cancel Redeploy Download Refresh

Overview

Inputs

Outputs

Template

Your deployment is complete

Deployment name: hexadeb1077dest_1708495471863
Subscription: Azure subscription 1
Resource group: rg-azuser1077_mml.local-IWFLY

Start time: 21/02/2024, 11:34:34
Correlation ID: cf6d7134-7f1a-4ccc-872c-fb1a0687eee

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

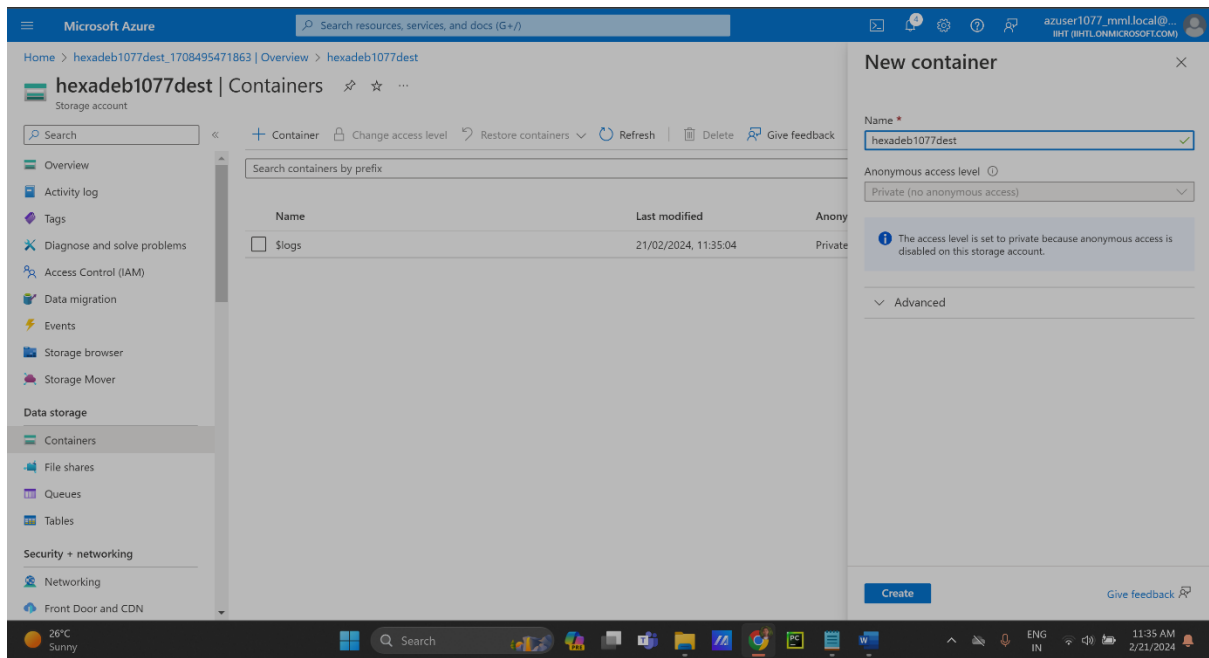
Secure your apps and infrastructure [Go to Microsoft Defender for Cloud >](#)

Free Microsoft tutorials

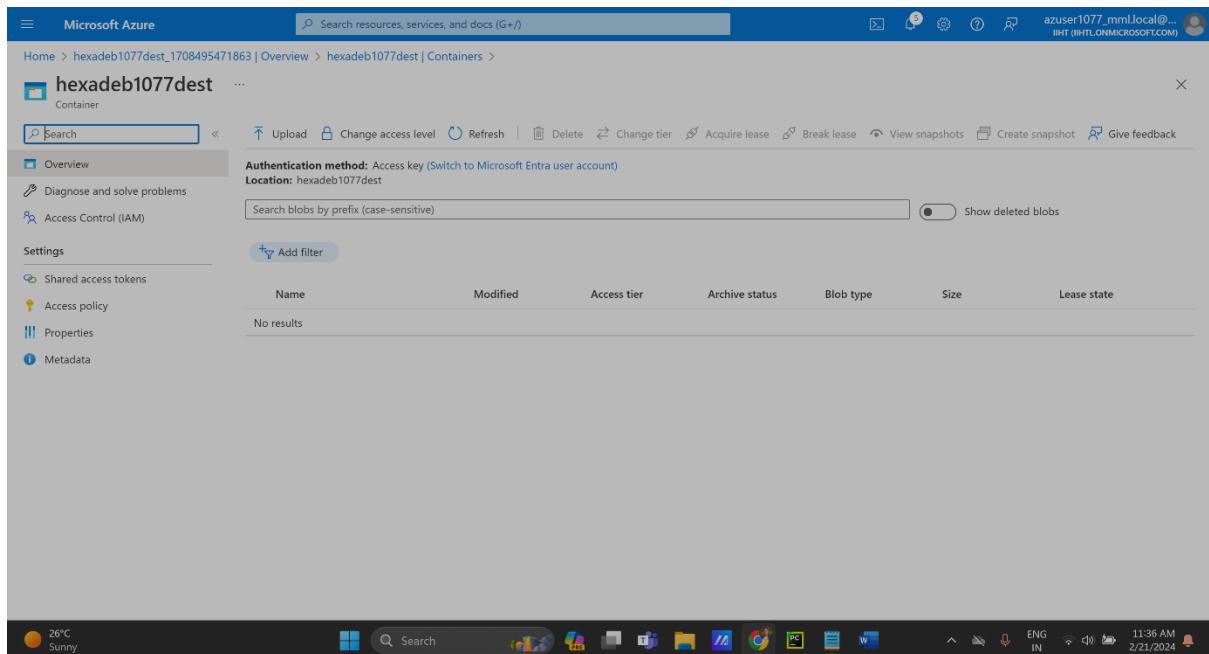
[Start learning today >](#)

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 >](#)



Keeping destination container empty.



➤ Creating Azure Data Factory

Microsoft Azure

Search resources, services, and docs (G+)

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-azuser1077_mml.local-iWFLY
[Create new](#)

Instance details

Name * ⓘ adfdeb1077 ✓

Region * ⓘ East US

Version * ⓘ V2

Previous | Next | Review + create

[Give feedback](#)

Microsoft Azure

Search resources, services, and docs (G+)

Home >

Microsoft.DataFactory-20240221113628 | Overview

Deployment

Search | Delete | Cancel | Redeploy | Download | Refresh

Overview | Inputs | Outputs | Template

✓ Your deployment is complete

Deployment name : Microsoft.DataFactory-20240221113628 Start time : 21/02/2024, 11:38:51

Subscription : Azure subscription 1 Correlation ID : 1756f12e-91c6-4512-9c25-2b62e6be7cfd

Resource group : rg-azuser1077_mml.local-iWFLY

> 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 >](#)

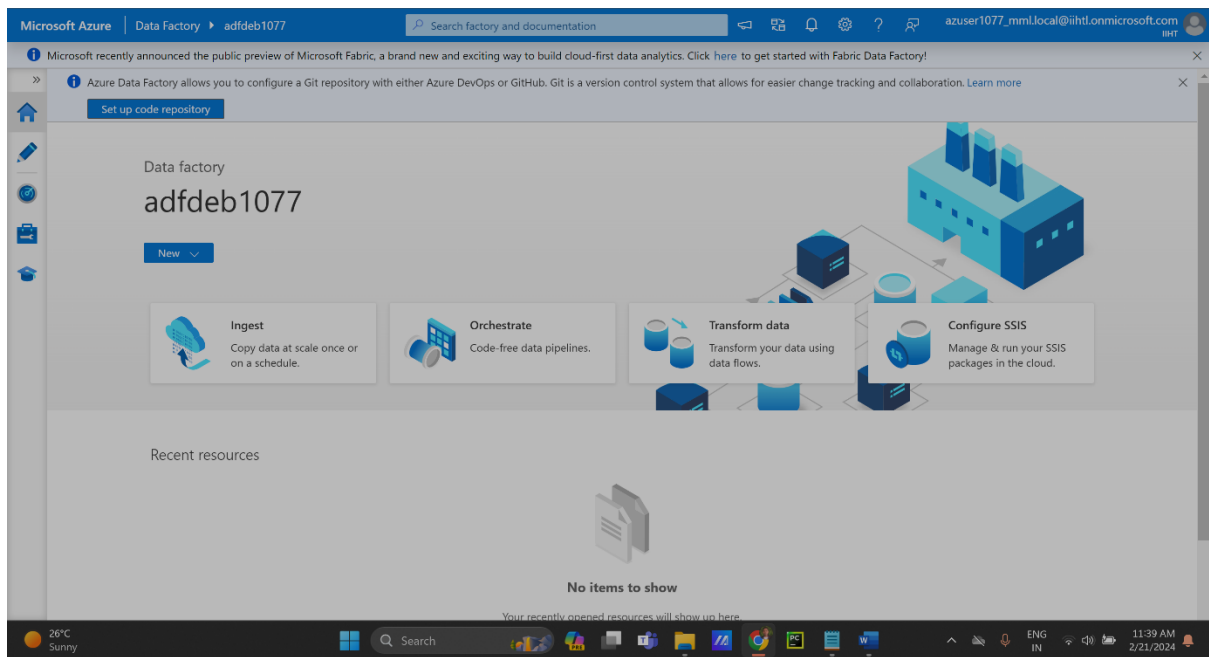
Free Microsoft tutorials

[Start learning today >](#)

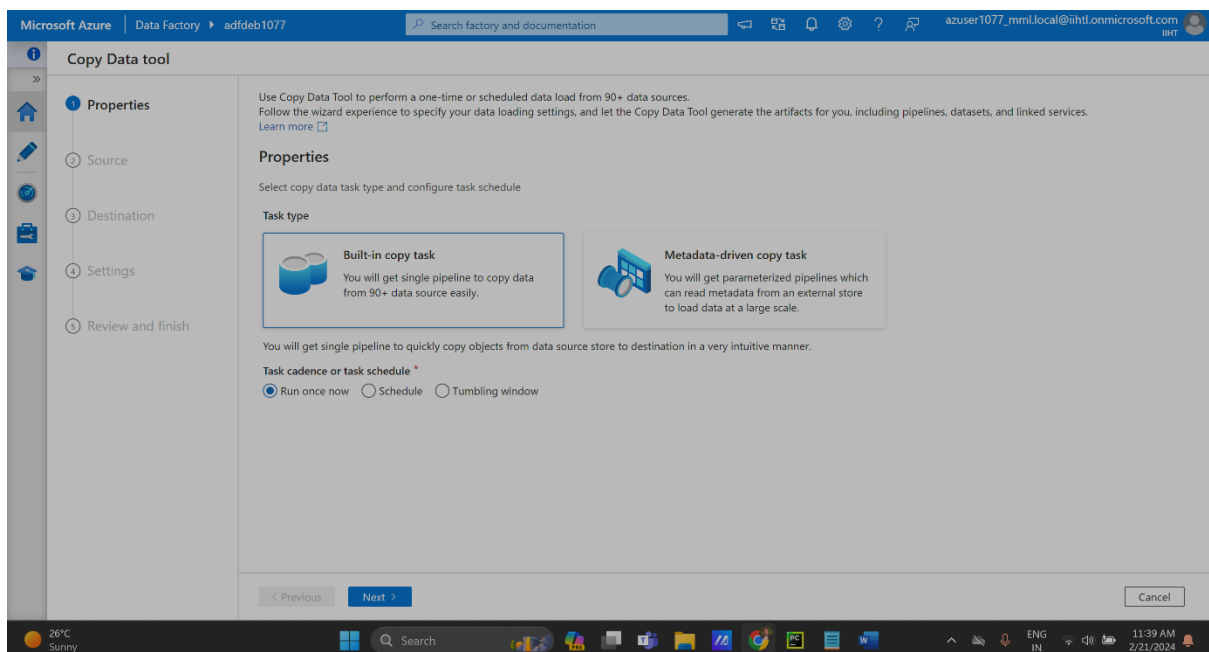
Work with an expert

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 >](#)



Starting Data Pipeline



Creating Connection with source data store

The screenshot shows the 'New connection' dialog in the Microsoft Azure Data Factory portal. The left sidebar indicates the 'Copy Data tool' is active, with the 'Source' step selected. The main panel is titled 'Source data store' and contains the following fields:

- Name:** AzureBlobStorage1
- Description:** (empty)
- Connect via integration runtime:** AutoResolveIntegrationRuntime
- Authentication type:** Account key
- Account selection method:** From Azure subscription (selected)
- Azure subscription:** Azure subscription 1 (984f097c-963c-4eb6-a20d-839457ae9f08)
- Storage account name:** hexadeb1077source

At the bottom right, a green checkmark indicates 'Connection successful', and a 'Test connection' button is available. The 'Create' button is highlighted.

The screenshot shows the 'Copy Data tool' configuration page in the Microsoft Azure Data Factory portal. The left sidebar indicates the 'Copy Data tool' is active, with the 'Source' step selected. The main panel is titled 'Source data store' and contains the following fields:

- Source type:** Azure Blob Storage
- Connection:** AzureBlobStorage1
- File or folder:** hexadeb1077source/
- Options:**
 - ☐ Binary copy
 - ☒ Recursively
 - ☐ Enable partitions discovery
- Max concurrent connections:** (empty)
- Filter by last modified:**
 - Start time (UTC):** (empty)
 - End time (UTC):** (empty)

A red error message at the bottom states: 'Please input or choose a folder or file'. The 'Next' button is highlighted.

Creating Connection with destination data store

The screenshot shows the 'New connection' dialog in the Microsoft Azure Data Factory interface. The dialog is titled 'New connection' and has a sub-header 'Azure Blob Storage'. The 'Name' field is set to 'AzureBlobStorage2'. The 'Description' field is empty. The 'Connect via integration runtime' dropdown is set to 'AutoResolveIntegrationRuntime'. The 'Authentication type' dropdown is set to 'Account key'. The 'Account selection method' is set to 'From Azure subscription'. The 'Azure subscription' dropdown is set to 'Azure subscription 1 (984f097c-963c-4eb6-a20d-839457ae9f08)'. The 'Storage account name' dropdown is set to 'hexadeb1077dest'. The 'Additional connection properties' section is empty. At the bottom, there are 'Create' and 'Cancel' buttons. A green checkmark and the text 'Connection successful' are visible in the bottom right corner.

Microsoft Azure | Data Factory | adfdeb1077

Search factory and documentation

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 Blob Storage

Connection: Select...

Name: AzureBlobStorage2

Description:

Connect via integration runtime: AutoResolveIntegrationRuntime

Authentication type: Account key

Account selection method: From Azure subscription

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

Storage account name: hexadeb1077dest

Additional connection properties:

Create Cancel

Connection successful

Test connection

The screenshot shows the 'Destination data store' configuration page in the Microsoft Azure Data Factory interface. The page is titled 'Destination data store'. The 'Destination type' dropdown is set to 'Azure Blob Storage'. The 'Connection' dropdown is set to 'AzureBlobStorage2'. The 'Folder path' field is set to 'hexadeb1077dest/'. The 'File name' field is empty. The 'Copy behavior' dropdown is set to 'Select...'. The 'Max concurrent connections' field is empty. The 'Block size (MB)' field is empty. The 'Metadata' section is empty. At the bottom, there are 'Previous', 'Next', and 'Cancel' buttons.

Microsoft Azure | Data Factory | adfdeb1077

Search factory and documentation

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 Blob Storage

Connection: AzureBlobStorage2

Folder path: hexadeb1077dest/

File name:

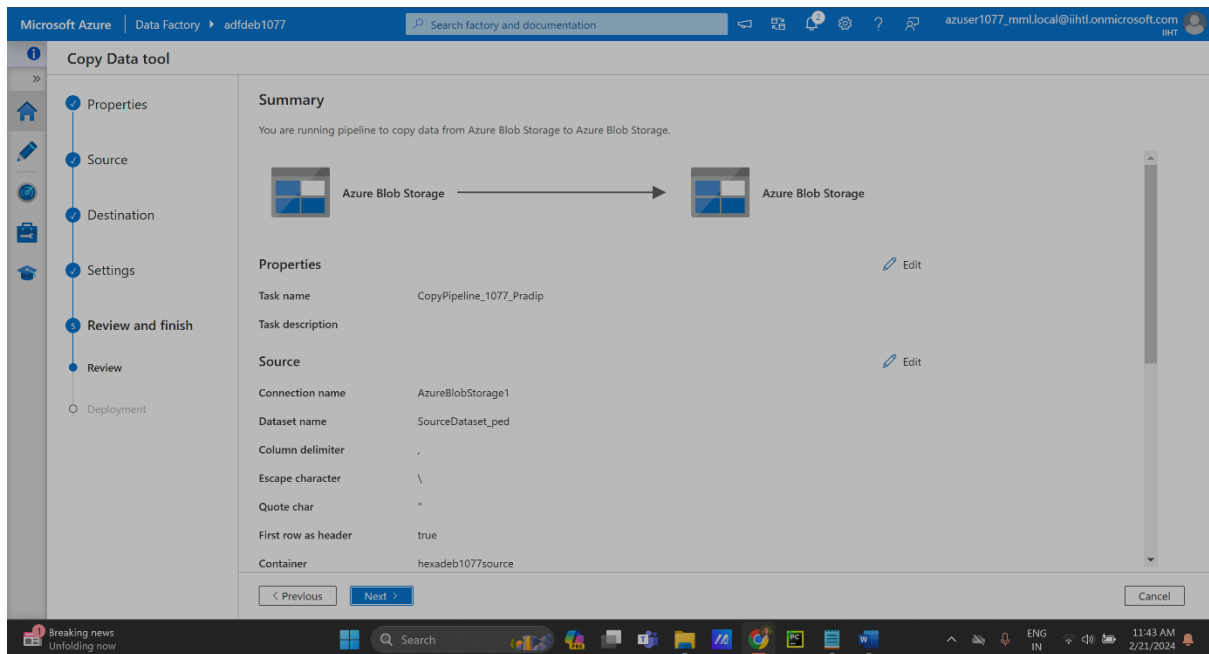
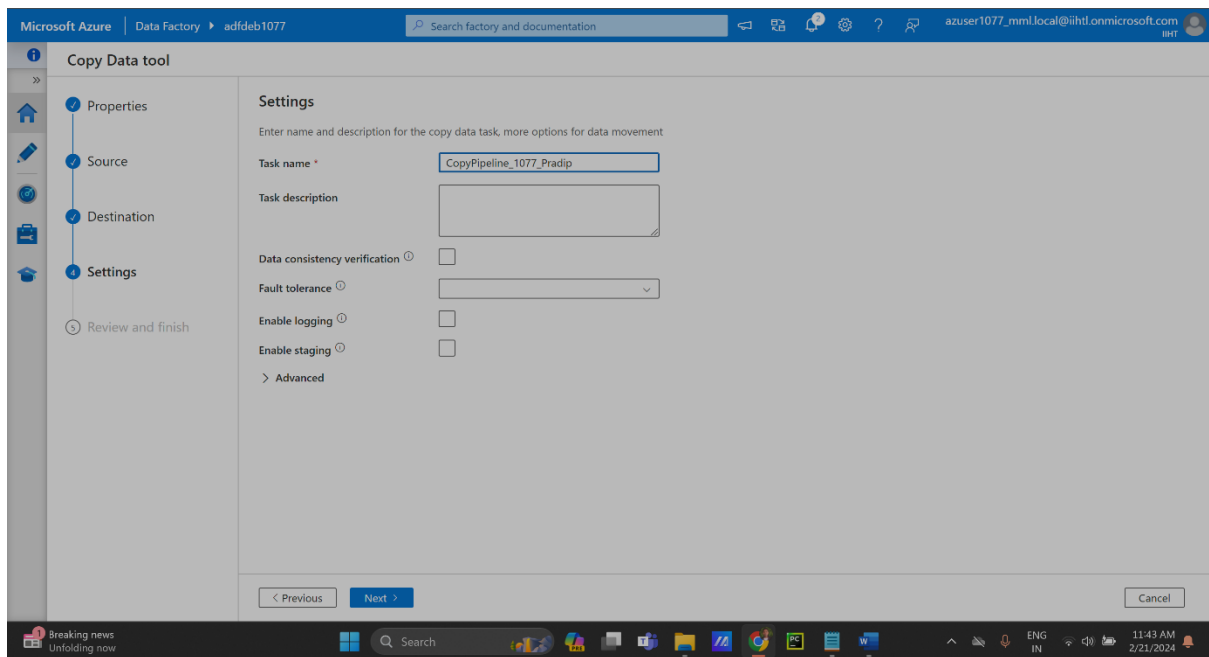
Copy behavior: Select...

Max concurrent connections:

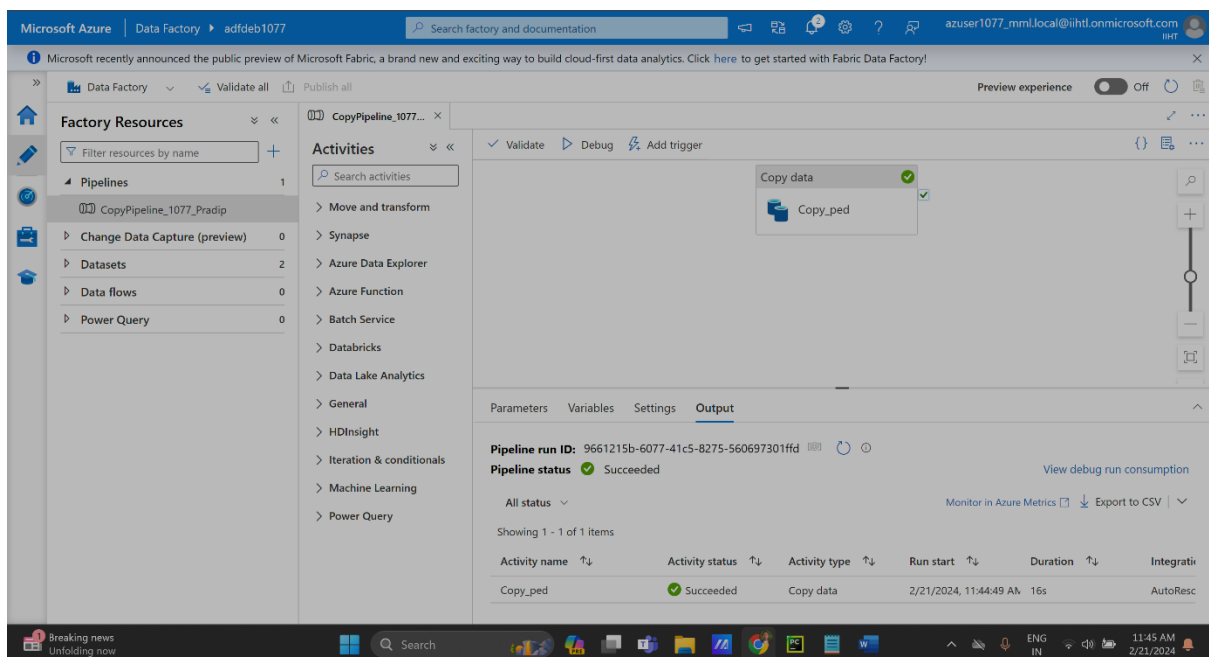
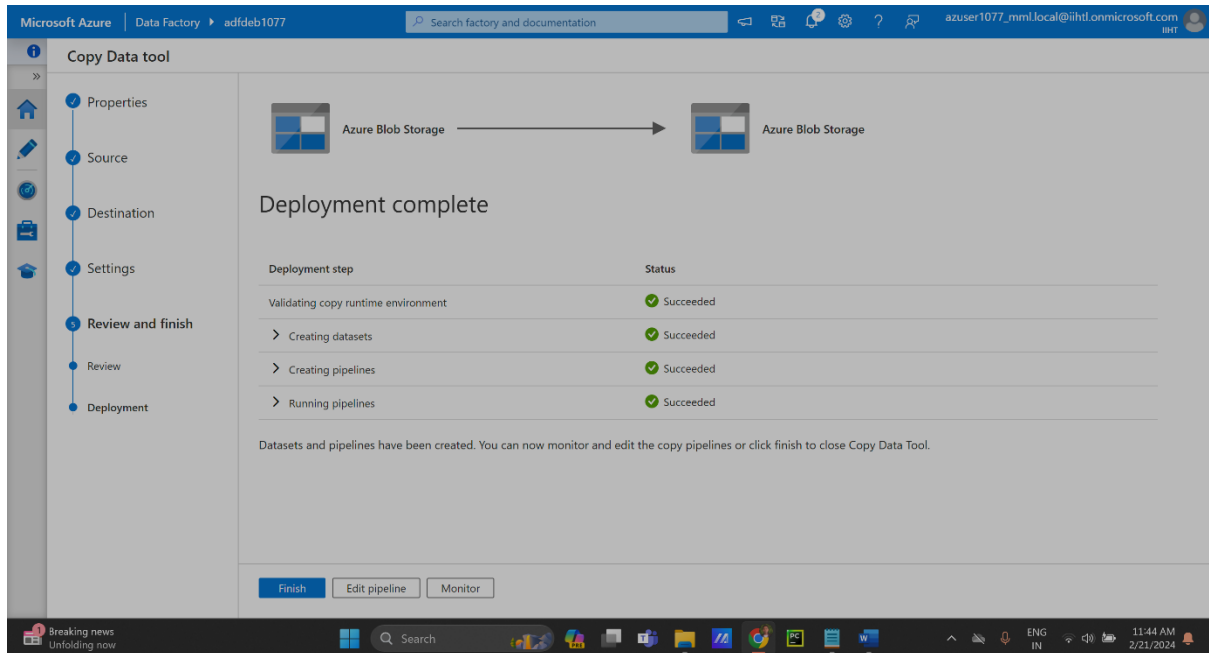
Block size (MB):

Metadata:

Previous Next Cancel



Deployment



Data in source container is copied in destination container by Copy Activity in ADF.

