

Coding challenge

Azure Databricks

Name: Rohan Vinayak Chaudhari

Email:chaudharirohan24@gmail.com

Batch: Data Engineering 1

Question3: Execute & explain, Azure datafactory and its copy activity.

Source Container which contains data:

[Home](#) > [Source Data](#) > Containers > sourcedata

sourcedata

Container

Search

Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots Create snapshot Give feedback

Authentication method: Access key ([Switch to Microsoft Entra user account](#))
Location: sourcedata

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"/> output_file.csv | 21/02/2024, 11:01:32 | Hot (Inferred) | | Block blob | 61.42 KiB | Available |

Destination Container which is empty:

Microsoft Azure

Home > fordestination | Containers >

destinationdata

Container

[Upload](#)
[Change access level](#)
[Refresh](#)
[Delete](#)
[Change tier](#)
[Acquire lease](#)
[Break lease](#)
[View snapshots](#)
[Create snapshot](#)
[Give feedback](#)

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Access policy

Properties

Metadata

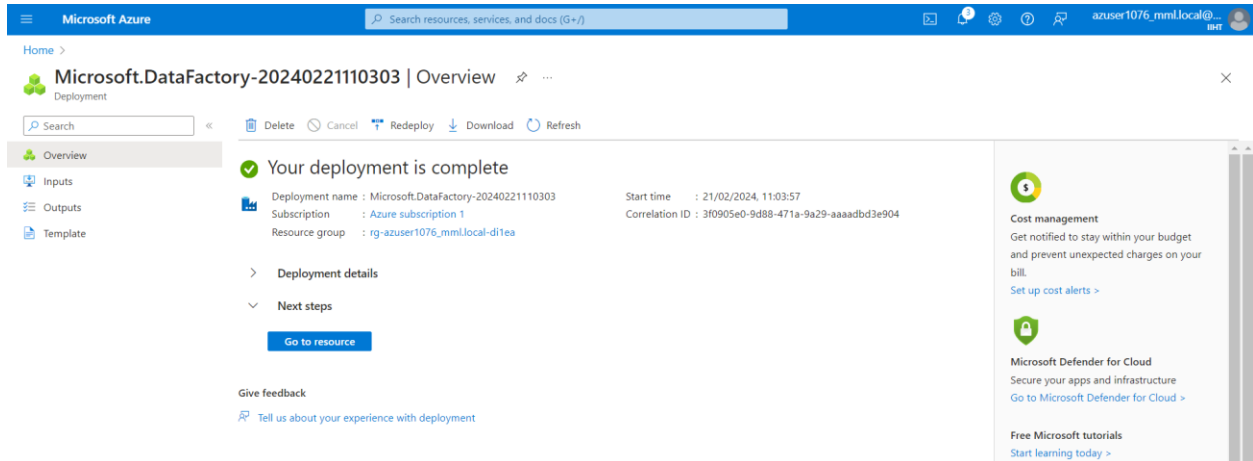
Authentication method: Access key (Switch to Microsoft Entra user account)

Location: destinationdata

☐ Show deleted blobs

| Name | Modified | Access tier | Archive status | Blob type | Size | Lease state |
|------------|----------|-------------|----------------|-----------|------|-------------|
| No results | | | | | | |

Created DataFactory:



Microsoft Azure

Search resources, services, and docs (G+)

Home >

Microsoft.DataFactory-20240221110303 | Overview

Deployment

Search

Delete Cancel Redeploy Download Refresh

Overview

Inputs

Outputs

Template

✓ Your deployment is complete

Deployment name : Microsoft.DataFactory-20240221110303

Subscription : Azure subscription 1

Resource group : rg-azuser1076_mml.local-di1ea

Start time : 21/02/2024, 11:03:57

Correlation ID : 3f0905e0-9d88-471a-9a29-aaaadb3e904

> 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

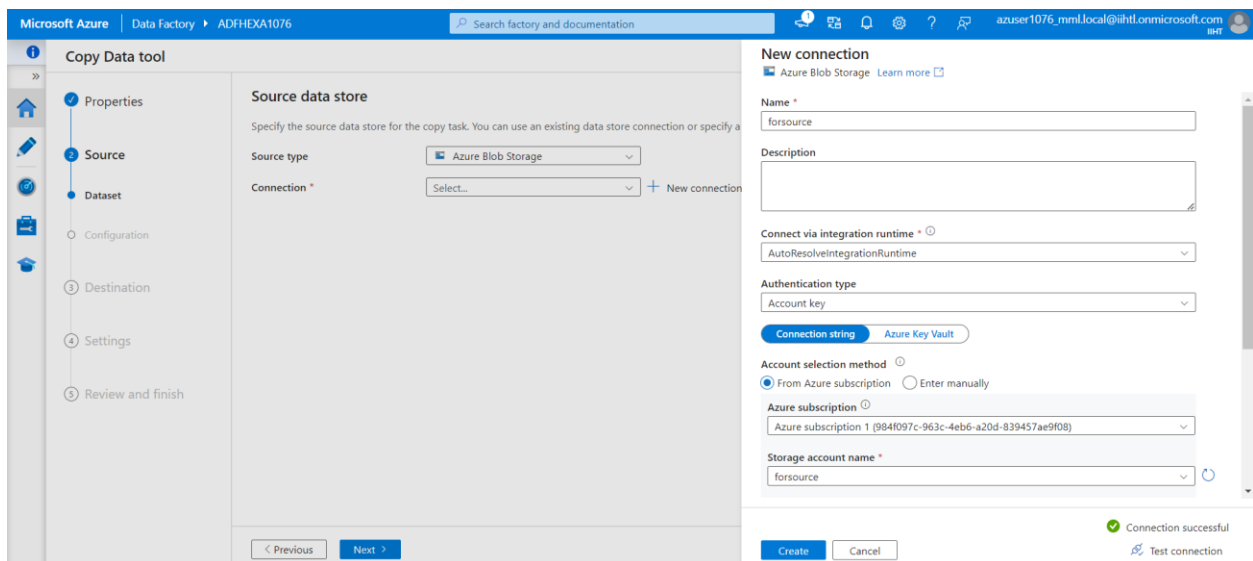
Secure your apps and infrastructure

Go to Microsoft Defender for Cloud >

Free Microsoft tutorials

Start learning today >

Created Source Dataset & Tested it successfully:



Microsoft Azure

Data Factory > ADFHDXA1076

Search factory and documentation

Copy Data tool

Properties

Source

Dataset

Configuration

Destination

Settings

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 one.

Source type: Azure Blob Storage

Connection: Select... + New connection

New connection

Azure Blob Storage Learn more

Name: forsource

Description:

Connect via integration runtime: AutoResolveIntegrationRuntime

Authentication type: Account key

Connection string: Azure Key Vault

Account selection method: From Azure subscription (selected) Enter manually

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

Storage account name: forsource

Create Cancel

✓ Connection successful

Test connection

Microsoft Azure | Data Factory | ADFHXA1076 | Search factory and documentation | azuser1076_mml.local@ihhtl.onmicrosoft.com

Copy Data tool

- Properties
- Source
- Dataset
- Configuration
- Destination
- Settings
- 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: Azure Blob Storage

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

File or folder: sourcedata/ [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. Append a slash (/) at the end if the path refers to a folder.

Options

- ☒ Binary copy
- Compression type: None
- ☒ Recursively
- ☐ Delete files after completion
- Max concurrent connections:

Filter by last modified

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

Created destination Dataset&Tested it:

Microsoft Azure | Data Factory | ADFHXA1076 | Search factory and documentation | azuser1076_mml.local@ihhtl.onmicrosoft.com

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... [+ New connection](#)

New connection

Azure Blob Storage [Learn more](#)

Name: fordestination

Description:

Connect via integration runtime: AutoResolveIntegrationRuntime

Authentication type: Account key

[Connection string](#) [Azure Key Vault](#)

Account selection method: ☒ From Azure subscription ☐ Enter manually

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

Storage account name: fordestination [Refresh](#)

[Create](#) [Cancel](#)

✓ Connection successful [Test connection](#)

Assigning Container path:

This screenshot shows the 'Destination data store' configuration step in the 'Copy Data tool' wizard. The left sidebar indicates the current step is 'Destination', with 'Source' and 'Dataset' also visible. The main panel is titled 'Destination data store' and includes instructions: 'Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.' The configuration fields are as follows:

- Destination type:** Azure Blob Storage
- Connection:** fordestination (with 'Edit' and 'New connection' links)
- Folder path:** destinationdata/ (with a 'Browse' button)
- File name:** Filenames are defined by source
- Compression type:** None
- Copy behavior:** Select...
- Max concurrent connections:** (empty field)
- Block size (MB):** (empty field)

Navigation buttons at the bottom include '< Previous', 'Next >', and 'Cancel'.

Created Task:

This screenshot shows the 'Settings' configuration step in the 'Copy Data tool' wizard. The left sidebar indicates the current step is 'Settings', with 'Properties', 'Source', 'Destination', and 'Dataset' also visible. The main panel is titled 'Settings' and includes instructions: 'Enter name and description for the copy data task, more options for data movement'. The configuration fields are as follows:

- Task name:** ADF_Test
- Task description:** (empty text area)
- Data consistency verification:** (unchecked checkbox)
- Fault tolerance:** Skip missing files
- Enable logging:** (unchecked checkbox)
- Enable staging:** (unchecked checkbox)
- Advanced:** (expandable section)

Navigation buttons at the bottom include '< Previous', 'Next >', and 'Cancel'.

Reviewing:

The screenshot shows the 'Copy Data tool' configuration in the 'Review and finish' step. The left sidebar contains a navigation pane with steps: Properties, Source, Destination, Settings, Review and finish (selected), Review, and Deployment. The main area displays a summary of the pipeline: 'You are running pipeline to copy data from Azure Blob Storage to Azure Blob Storage.' Below this is a diagram showing data flow from 'Azure Blob Storage' to 'Azure Blob Storage'. The 'Properties' section lists: Task name (ADF_Test), Task description, Source (Connection name: forsource, Dataset name: SourceDataset_v35, Container: sourcedata), and Destination (Connection name: fordestination). Each section has an 'Edit' link. At the bottom are buttons for '< Previous', 'Next >', and 'Cancel'.

Microsoft Azure | Data Factory | ADFHEXA1076 | Search factory and documentation | azuser1076_mml.local@ihtl.onmicrosoft.com | ihtl

Copy Data tool

Summary

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

Properties [Edit](#)

Task name: ADF_Test

Task description:

Source [Edit](#)

Connection name: forsource

Dataset name: SourceDataset_v35

Container: sourcedata

Destination [Edit](#)

Connection name: fordestination

< Previous | Next > | Cancel

Deployment:

The screenshot shows the 'Deployment complete' screen. The left sidebar navigation pane is the same as the previous screen, but 'Deployment' is now selected. The main area shows a diagram of the data flow from 'Azure Blob Storage' to 'Azure Blob Storage'. Below the diagram, the text 'Deployment complete' is displayed. A table lists the deployment steps and their status:

| Deployment step | Status |
|-------------------------------------|-------------|
| Validating copy runtime environment | ✓ Succeeded |
| > Creating datasets | ✓ Succeeded |
| > Creating pipelines | ✓ Succeeded |
| > Running pipelines | ✓ Succeeded |

Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.

Buttons at the bottom: Finish, Edit pipeline, Monitor.

Microsoft Azure | Data Factory | ADFHEXA1076 | Search factory and documentation | azuser1076_mml.local@ihtl.onmicrosoft.com | ihtl

Copy Data tool

Deployment complete

| Deployment step | Status |
|-------------------------------------|-------------|
| Validating copy runtime environment | ✓ Succeeded |
| > Creating datasets | ✓ Succeeded |
| > Creating pipelines | ✓ Succeeded |
| > Running pipelines | ✓ Succeeded |

Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.

Finish | Edit pipeline | Monitor

Here we can see the files are uploaded in destination container:

Microsoft Azure

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

Home > fordestination | Containers >

destinationdata

Container

Search

Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots Create snapshot Give 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: destinationdata

Search blobs by prefix (case-sensitive)

Show deleted blobs

| Name | Modified | Access tier | Archive status | Blob type | Size | Lease state | |
|-----------------|----------------------|----------------|----------------|------------|-----------|-------------|-----|
| output_file.csv | 21/02/2024, 11:13:25 | Hot (Inferred) | | Block blob | 61.42 KIB | Available | ... |

Now we will try to upload new files:

We have debugged the pipeline for copying new file and it is successfully copied:

Microsoft Azure

Data Factory > ADFHEXA1076

Search factory and documentation

azuser1076_mml.local@ihlt.onmicrosoft.com

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click here to get started with Fabric Data Factory!

Preview experience Off

Factory Resources

Filter resources by name

Pipelines

ADF_Test

Change Data Capture (preview) 0

Datasets 2

Data flows 0

Power Query 0

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

Copy_v35

Parameters Variables Settings Output

Pipeline run ID: d9c6d6fd-85c8-418a-9454-9019c98db61d

Pipeline status Succeeded

View debug run consumption

Monitor in Azure Metrics Export to CSV

Showing 1 - 1 of 1 items

| Activity name | Activity status | Activity type | Run start | Duration | Integration runtime |
|---------------|-----------------|---------------|------------------------|----------|-----------------------|
| Copy_v35 | Succeeded | Copy data | 2/21/2024, 11:16:19 AM | 17s | AutoResolveIntegrator |

In our destination folder New file has be copied:

Microsoft Azure

Search resources, services, and docs (G+7)

azuser1076_mm1.local@...

ENT

Home > fordestination | Containers >

destinationdata

Container

Search

«

Upload

Change access level

Refresh

Delete

Change tier

Acquire lease

Break lease

View snapshots

Create snapshot

Give 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: destinationdata

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"/> |  output_file.csv | 21/02/2024, 11:16:32 | Hot (Inferred) | | Block blob | 61.42 KiB | Available | *** |
| <input type="checkbox"/> |  output_file1.json | 21/02/2024, 11:16:32 | Hot (Inferred) | | Block blob | 161.8 KiB | Available | *** |