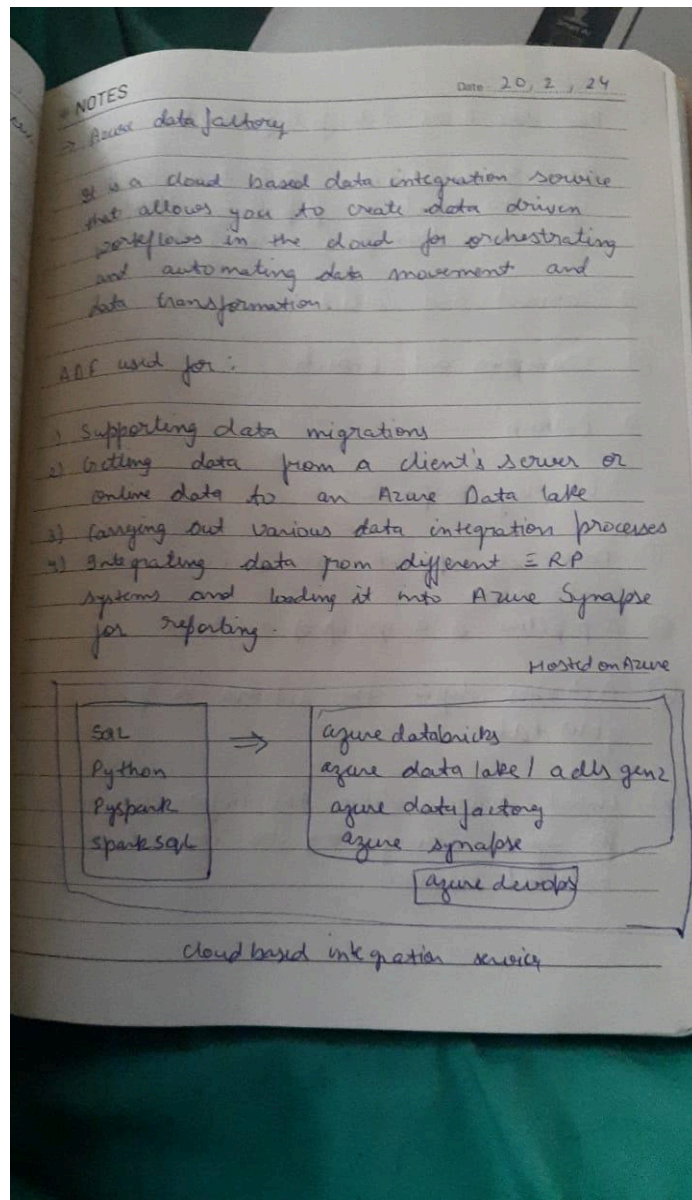


Asmita Porwal  
Batch-1  
Day-21  
20/02/2024  
Data engineering

## Assignment-22

### Handwritten Notes:



## ◆ NOTES

Date: \_\_\_\_\_

• How does Azure data factory work?

ADF service allows. ADF service allows you to create a pipeline that move, transform, run, data in

① Connect and collect

② Transform and Enrich

③ Publish

- ADF key components

① Data sets represents data structure within the data stores.

② A pipeline is a group of activities

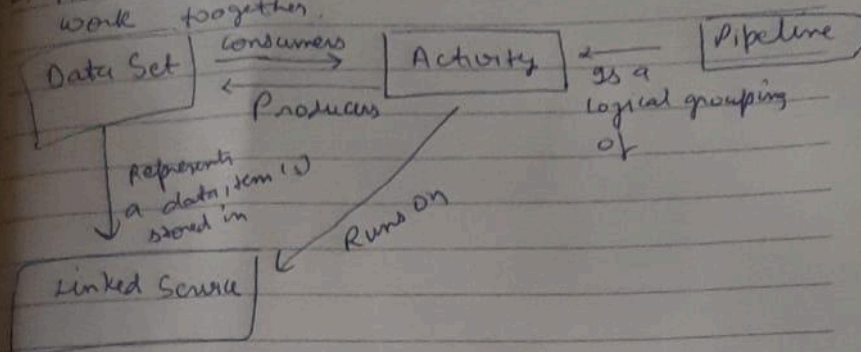
③ Activities define the actions to perform on your data

④ Linked services define the information needed for azure data factory to connect to external resources.

## NOTES

Date: \_\_\_\_\_

→ How the azure data factory components work together.



We can use the following tools or API's to create data pipeline in azure data factory:

- 1) Azure portal
- 2) Visual Studio
- 3) Powershell
- 4) .Net API
- 5) Rest API
- 6) Azure Resource
- 7) Manager template

# Azure datafactory copy activity.

Home > storage1060

storage1060 | Containers

Storage account

Search

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Data storage

Containers

File shares

Queues

Tables

Security + networking

Networking

+ Container

Change access level

Restore containers

Refresh

Delete

Give feedback

Search containers by prefix

Show deleted containers

Name	Last modified	Anonymous access level	Lease state	
<input type="checkbox"/> \$logs	16/02/2024, 14:09:54	Private	Available	...
<input type="checkbox"/> containerofstorage1060	21/02/2024, 09:57:08	Private	Available	...

Home > storage1060 | Containers >

containerofstorage1060

Container

Search

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

Upload

Add Directory

Refresh

Rename

Delete

Change tier

Acquire lease

Break lease

Give feedback

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

Location: containerofstorage1060

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size
<input type="checkbox"/> output					
<input type="checkbox"/> Marks_data.csv	21/02/2024, 09:57:35	Hot (Inferred)		Block blob	88 B

Home > storage1060 | Containers >

containerofstorage1060

Container

Search

UploadAdd DirectoryRefreshRenameDeleteChange tierAcquire leaseBreak leaseGive feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

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

Location: containerofstorage1060 / output

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size
[.]					
Marks_data.csv	21/02/2024, 10:01:44	Hot (Inferred)		Block blob	88 B

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!

Data factory

datafactory1060

New

Ingest

Copy data at scale once or on a schedule.

Orchestrate

Code-free data pipelines.

Transform data

Transform your data using data flows.

Configure SSIS

Manage & run your SSIS packages in the cloud.

Recent resources

No items to show

Your recently opened resources will show up here.

1

Copy Data tool

»

1 Properties

2 Source

3 Destination

4 Settings


5 Review and finish


Use Copy Data Tool to perform a one-time or scheduled data load from 90+ data sources. Follow the wizard experience to specify your data loading settings, and let the Copy Data Tool generate the artifacts for you, including pipelines, datasets, and linked services. [Learn more](#)

### Properties

Select copy data task type and configure task schedule

Task type

**Built-in copy task**  
You will get single pipeline to copy data from 90+ data source easily.

**Metadata-driven copy task**  
You will get parameterized pipelines which can read metadata from an external store to load data at a large scale.

You will get single pipeline to quickly copy objects from data source store to destination in a very intuitive manner.

**Task cadence or task schedule \***

☒ Run once now ☐ Schedule ☐ Tumbling window

< Previous

Next >

Cancel

1

Copy Data tool

»

1 Properties

2 Source

3 Dataset

4 Configuration

5 Destination

6 Settings

7 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**

**Connection \***  [Edit](#) [+ New connection](#)

**File or folder \***

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.

[Browse](#)

**Options**

☒ Binary copy [?](#)

**Compression type**

☒ Recursively [?](#)

☐ Delete files after completion [?](#)

**Max concurrent connections** [?](#)

< Previous

Next >

Cancel



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

Destination typeAll

Connection \*Select...

< PreviousNext >

New connection

Search

AllAzureFile

Azure Blob Storage

Azure Data Lake Storage Gen1

Azure Data Lake Storage Gen2

Azure File Storage

File system

Microsoft Fabric Lakehouse

ContinueCancel

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 typeAll

Connection \*AzureBlobStorage1Edit+ New connection

Folder path \*  
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.  
containerofstorage1060/outputBrowse

File name

Compression typeNone

Copy behavior ①Select...

Max concurrent connections ①

< PreviousNext >

Cancel

Microsoft Azure

Data Factory > datafactory1060

Search factory and documentation

azuser1060\_mml.local@iiht.onmicrosoft.com

Copy Data tool

»

Properties

Source

Destination

Settings

Review and finish

Settings

Enter name and description for the copy data task, more options for data movement

Task name \*ADFHHelloDemow

Task description

Data consistency verification

Enable logging

Enable staging

Advanced

< Previous

Next >

Cancel

Microsoft Azure

Data Factory > datafactory1060

Search factory and documentation

azuser1060\_mml.local@iiht.onmicrosoft.com

Copy Data tool

»

Properties

Source

Destination

Settings

Review and finish

Review

Deployment

Summary

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

Azure Blob Storage

Azure Blob Storage

Properties

Task nameADFHHelloDemow

Task description

Source

Connection nameAzureBlobStorage1

Dataset nameSourceDataset\_sn6

File nameMarks\_data.csv

Containercontainerofstorage1060

< Previous

Next >

Cancel



Copy Data tool

»

Properties

Source

Destination

Settings

Review and finish

Review

Deployment

Azure Blob Storage

Azure Blob Storage

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

Home >

datafactory1060

Data factory (V2)

Search

Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Networking

Managed identities

Properties

Locks

Getting started

Quick start

Monitoring

Alerts

Metrics

PipelineRuns

2

1.5

1

0.5

0

12

18

21 Feb

06

UTC+05:30

Succeeded pipeline runs ... datafactory1060

2

Failed pipeline runs metrics... datafactory1060

0

ActivityRuns

2

1.5

1

0.5

0

12

18

21 Feb

06

UTC+05:30

Succeeded activity runs m... datafactory1060

2

Failed activity runs metrics... datafactory1060

0

TriggerRuns

100

80

60

Integration Runtime CPU

100%

80%

60%