

Name: Pradip Bochare

## ✚ Notes

Page No.	
Date	

\* Azure Data Factory

- Azure data factory is a cloud based data integration service that allows you to create data-driven workflows in cloud for orchestrating and automating data movement and data transformation.
- ADF does not store any data itself. It allows to create data-driven workflows
- It also allows you to monitor and manage workflows using both programmatic and UI mechanisms.

\* ADF Use cases

- Supporting data migrations
- Getting data from 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

- ADF service allows you to create data pipelines that move, transform data and then pipelines on specified schedule.
- ADF pipelines typically perform three steps.

### Step 1: Connect and collect

- Connect to all required sources of data and processing such as SaaS services, file shares, FTP & web services.
- Then move data as needed to centralized location for subsequent processing by using copy activity in data pipeline.

### Step 2: Transform and Enrich

- Once data is present in a centralized data store in cloud it is transformed using compute services such as HD Insight Hadoop, Spark, Azure Data Lake Analytics and machine learning.

### Step 3: Publish

- Delivers transformed data from 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.



## \* Azure Data Factory key components

- ADF has four key components that work together to define & output data, processing events & schedule & resources required to execute desired data flow.

① Datasets represent data structure within data stores.

- input dataset represents input for activity in pipeline  
output dataset represents output for activity.

② Pipeline is a group of activities.

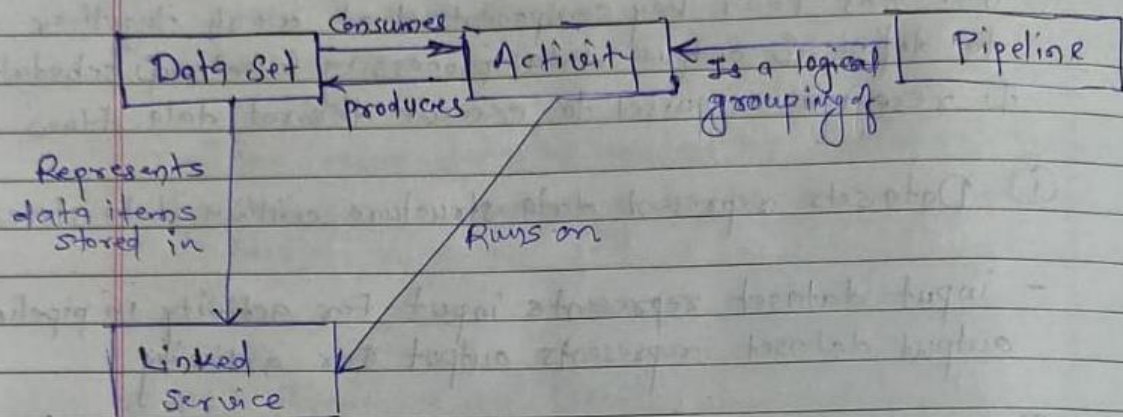
- They are used to group activities into a unit that together performs tasks. A data factory may have one or more pipelines.

③ Activities define actions to perform on your data.

- currently ADF supports two types of activities data movement and data transformation.

④ Linked Services define information needed for Azure data factory to connect to external resources

## \* How ADF components work together \*

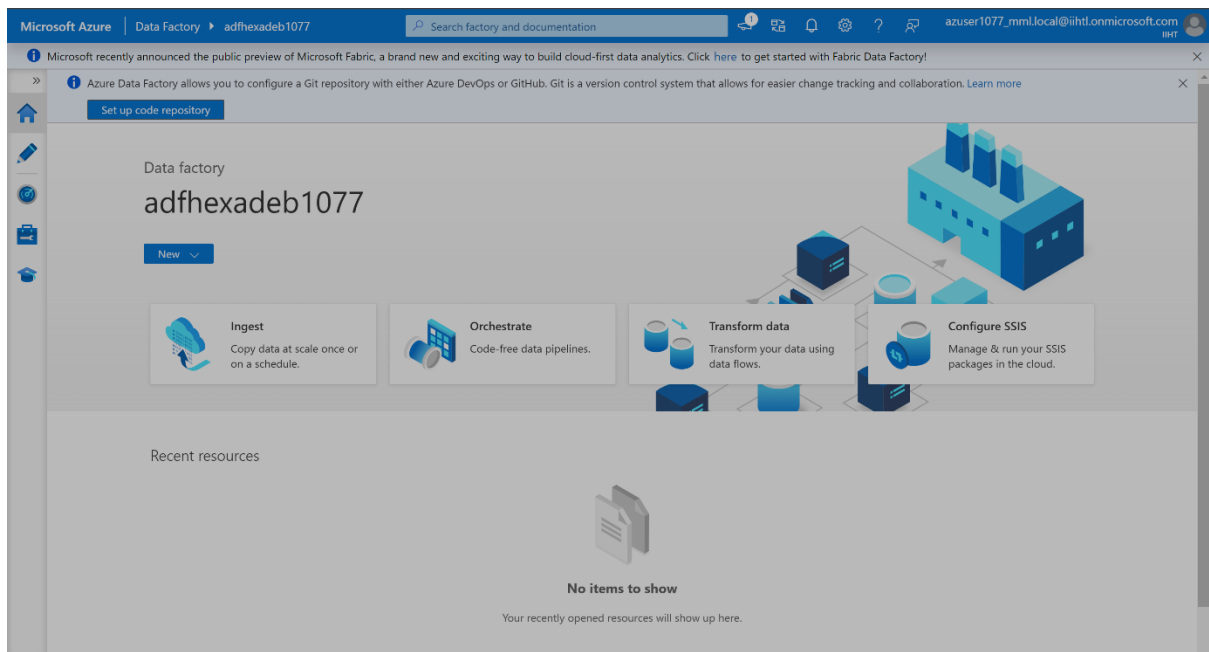
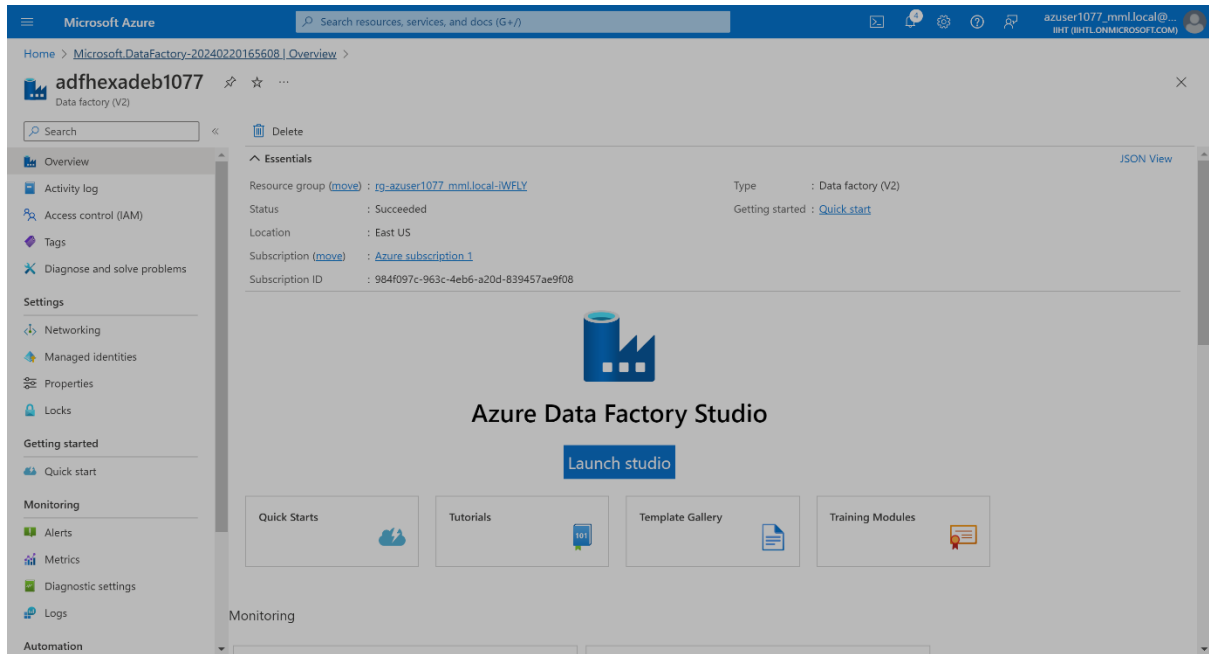


\* You can also use one of following tools or APIs to create data pipelines in ADF.

- Azure portal
- Visual Studio
- Powershell
- .NET API
- REST API

# Azure Data Factory

- Making Azure Data Factory



- Making Data Pipeline

The screenshot shows the 'Copy Data tool' wizard in the Microsoft Azure Data Factory portal. The left sidebar contains a navigation pane with steps: 1 Properties, 2 Source, 3 Destination, 4 Settings, and 5 Review and finish. The main content area is titled 'Copy Data tool' and includes a description: '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 [link]'. Below this is the 'Properties' section with the instruction 'Select copy data task type and configure task schedule'. There are two task type options: 'Built-in copy task' (selected) and 'Metadata-driven copy task'. The 'Built-in copy task' description says 'You will get single pipeline to copy data from 90+ data source easily.' The 'Metadata-driven copy task' description says 'You will get parameterized pipelines which can read metadata from an external store to load data at a large scale.' Below these is the 'Task cadence or task schedule' section with three radio buttons: 'Run once now' (selected), 'Schedule', and 'Tumbling window'. At the bottom are buttons for '< Previous', 'Next >', and 'Cancel'.

- Source Data

The screenshot shows the 'Copy Data tool' wizard in the Microsoft Azure Data Factory portal, Step 2: Source. The left sidebar shows the navigation pane with steps: 1 Properties, 2 Source (selected), 3 Destination, 4 Settings, and 5 Review and finish. The main content area is titled 'Source data store' and includes the instruction 'Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.' There are two dropdown menus: 'Source type' (set to 'Azure Blob Storage') and 'Connection \*' (set to 'AzureBlobStorage1'). Below these is a 'File or folder \*' section with a text input field and a 'Browse' button. The 'Options' section has three checkboxes: 'Binary copy' (unchecked), 'Recursively' (checked), and 'Enable partitions discovery' (unchecked). Below this is a 'Max concurrent connections' section with a text input field. At the bottom is a 'Filter by last modified' section with two text input fields: 'Start time (UTC)' and 'End time (UTC)'. At the bottom are buttons for '< Previous', 'Next >', and 'Cancel'.

Microsoft Azure | Data Factory | adfhexadeb1077 | Search factory and documentation | azuser1077\_mm1.local@iiht.onmicrosoft.com | IHT

### Copy Data tool

- Properties
- Source
- Dataset
- Configuration
- Destination
- Settings
- Review and finish

#### File format settings

File format: DelimitedText Detect text format Preview data

Column delimiter: Comma (,) Edit

Row delimiter: Default (\r\n, or \n) Edit

☒ First row as header ?

> Advanced

Compression type: Select... ?

Additional columns ?

+ New

< Previous Next > Cancel

- Connecting with Source Container

Microsoft Azure | Data Factory | adfhexadeb1077 | Search factory and documentation | azuser1077\_mm1.local@iiht.onmicrosoft.com | IHT

### 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 or create a new one.

Destination type: Azure Blob Storage

Connection: Select... +

#### New connection

Azure Blob Storage Learn more

Name: AzureBlobStorage2

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: hexadeb1077dest ?

Additional connection properties

+ New

< Previous Next > Create Cancel

✓ Connection successful Test connection

Microsoft Azure | Data Factory | adfhexadeb1077 | Search factory and documentation | azuser1077\_mm1.local@iihtl.onmicrosoft.com | IIHT

### 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 \*

Task description

Data consistency verification ☐

Fault tolerance

Enable logging ☐

Enable staging ☐

> Advanced

< Previous Next > Cancel

- Summary



Microsoft Azure | Data Factory | adfhexadeb1077 | Search factory and documentation | azuser1077\_mm1.local@iihtl.onmicrosoft.com | IIHT

### 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 name CopyPipeline\_zkn [Edit](#)

Task description [Edit](#)

##### Source

Connection name AzureBlobStorage1 [Edit](#)

Dataset name SourceDataset\_zkn

Column delimiter ,

Escape character \

Quote char "

First row as header true

Container myhexadeb1077

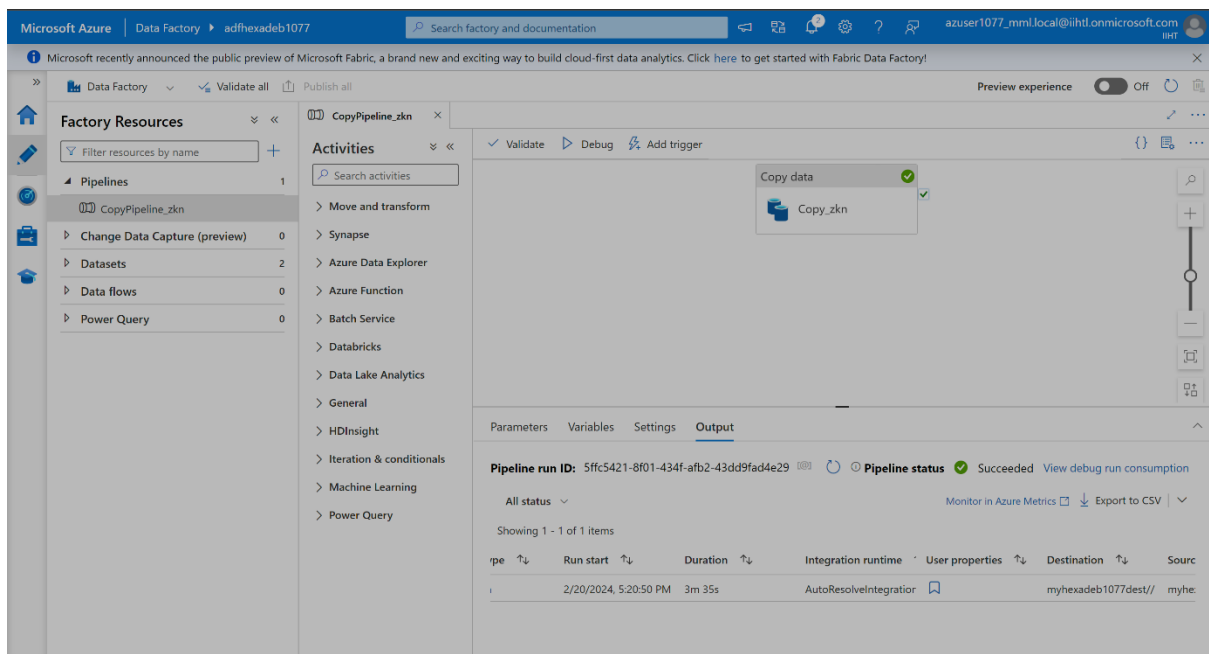
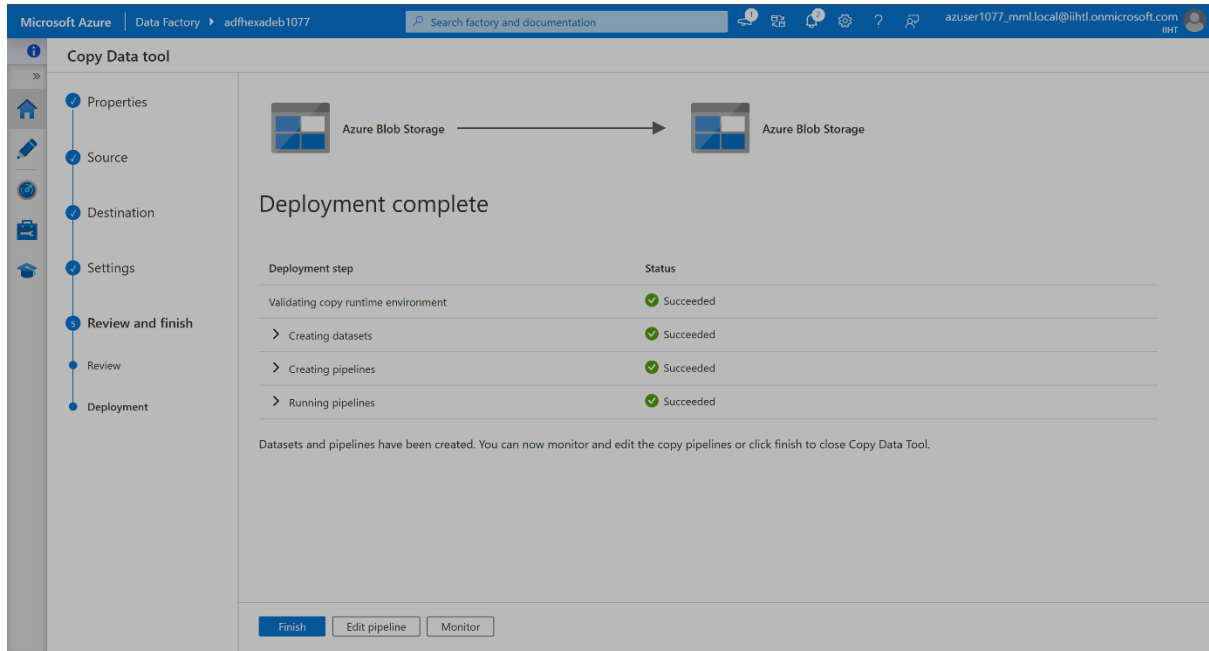
##### Destination

[Edit](#)

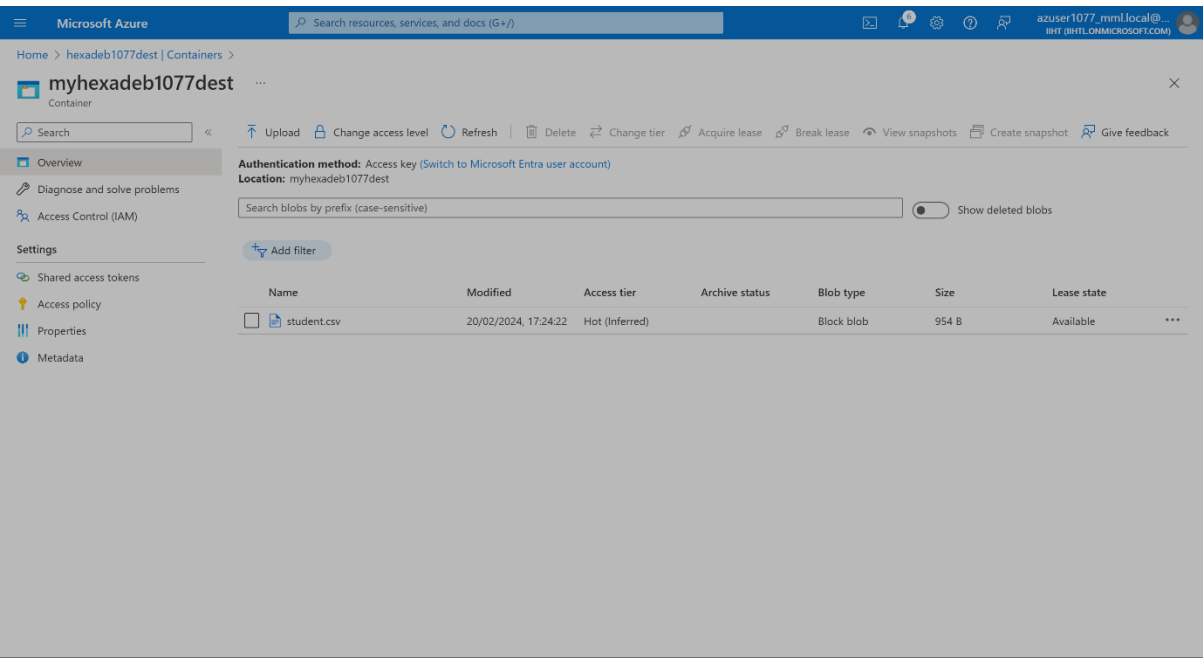
< Previous Next > Cancel



- Deployment



- Destination Container



- Source Container

