Asmita Porwal
Data Engineering
Batch-1
21/02/2024

Coding Challenge -4 Question-3

Execute & explain, Azure datafactory and its copy activity.

Azure Data Factory (ADF):

Azure Data Factory is a cloud-based data integration service provided by Microsoft Azure.

It allows you to create, schedule, and manage data pipelines that can move data between supported on-premises and cloud-based data stores.

ADF enables you to compose data storage, processing, and movement services into streamlined, scalable, and reliable data production pipelines.

Key components of Azure Data Factory include:

Pipelines: A collection of data processing activities, including data movement and transformation, that are executed sequentially.
Datasets: Represent the data structures within data stores, such as tables, files, or folders. Datasets define the schema and location of the data.
Linked Services: Define the connection information to connect to external data stores or compute services.
Activities: The processing steps in a pipeline, such as copying data or transforming data.

Copy Activity:

Copy Activity is one of the key activities within Azure Data Factory.

It allows you to copy data from one data store to another.

This could involve moving data between on-premises and cloud storage, copying data between cloud data stores, or performing transformations during the copy process.

Key features of Copy Activity:

Source and Destination: You define the source and destination linked services for the data copy operation. These linked services contain the connection information for the data stores.
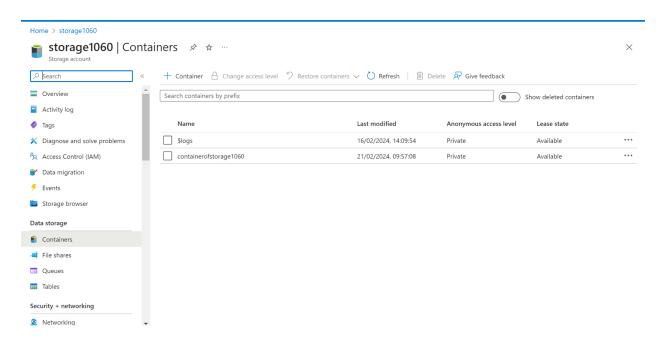
Mappings: You can specify mappings to transform data during the copy process. This is useful when the schema of the source and destination data is different.

Data Movement Modes: Copy Activity supports different data movement modes, including bulk insert, table-to-table insert, and stored procedure execution.
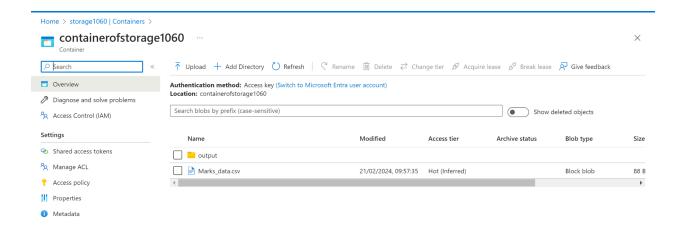
Data Integration Runtime: The copy operation is executed by a Data Integration Runtime (DIR), which is an infrastructure responsible for moving data securely and efficiently across network boundaries.

Monitoring and Logging: Azure Data Factory provides monitoring and logging capabilities, allowing you to track the progress and status of your data copy operations.

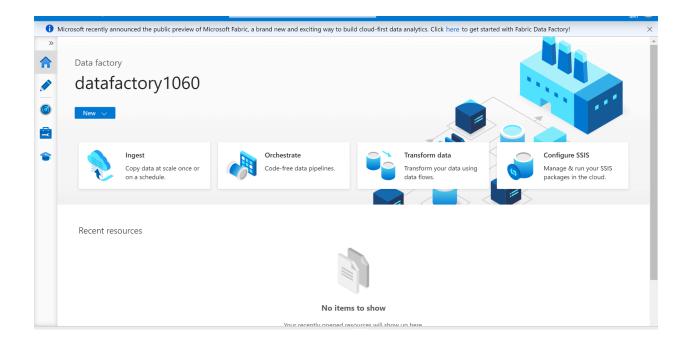## Created the container inside the storage
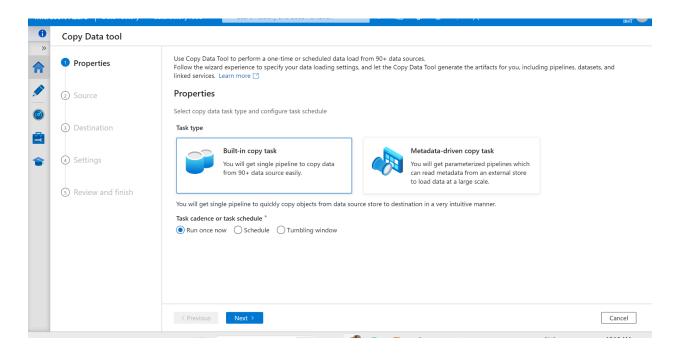
Uploaded the files Marks_data.csv

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.
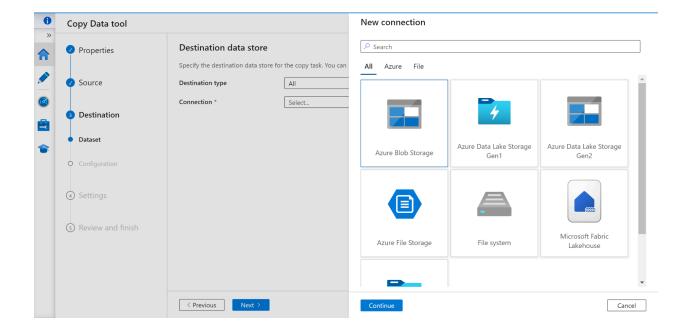
2.On the Properties page of the Copy Data tool, choose **Built-in** copy task under Task type, then select Next.
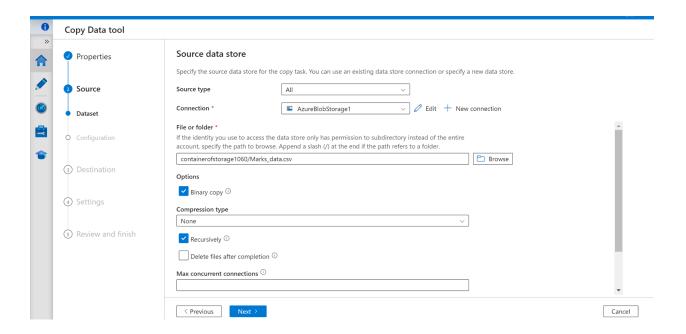

Step 2: Complete source configuration

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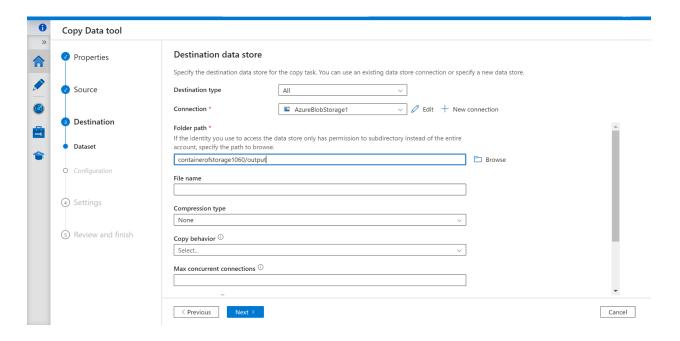
3.On the New connection (Azure Blob Storage) page, specify a name for your connection. Select your Azure subscription from the Azure subscription list and your storage account from the Storage account name list, test connection, and then select Create.

4.Select the newly created connection in the Connection block.

5.In the File or folder section, select Browse to navigate to the **containerofstorage1060/Marks_data.csv** folder, select the **Marks_data.csv** file, and then click OK.
6.Select the Binary copy checkbox to copy file as-is, and then select Next.

Step 3: Complete destination configuration
1. Select the AzureBlobStorage connection that you created in the Connection block.
2. In the Folder path section, enter **containerofstorage1060/output** for the folder path.

3.Leave other settings as default and then select Next.

Step 4: Review all settings and deployment
1. On the Settings page, specify a name for the pipeline and its description, then select Next to use other default configurations.

2.On the Summary page, review all settings, and select Next.
3.On the Deployment complete page, select Monitor to monitor the pipeline that you created.

Step 5: Monitor the running results

Output folder created and that marks_data.csv file is copied

containerofstorage1060 ...
Container

Search

- ▦ Overview
- 🔧 Diagnose and solve problems
- 👥 Access Control (IAM)

Settings

- 🔗 Shared access tokens
- 👥 Manage ACL
- 🔑 Access policy
- ⫴ Properties
- ⓘ Metadata

Show deleted objects

---

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 / output

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