

1. AZURE DATA FACTORY

FUNDAMENTALS



Follow us on :



Cloud And Data Universe



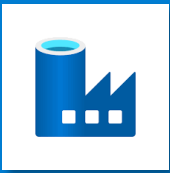
Azure Data Factory Overview

- Azure Data Factory a.k.a ADF is an ETL/ELT service provided by Microsoft in Azure cloud computing platform.
- ADF was offered publicly in 2015.

Follow us on :



Cloud And Data Universe



Azure Data Factory Concepts

Pipeline

Activity

**Linked
Service**

Dataset

**Integration
Runtime**

Follow us on :



Cloud And Data Universe



Pipeline

- A Pipeline controls the flow of tasks.
- Tasks are performed using activities.
- A pipeline can have one to many activities.
- These activities can be chained in serial or parallel depending on requirements.



Follow us on :



CDU

Cloud And Data Universe



Activity

- Activity executes a task.
- Activities are broadly classified into 3 types:

**Data
movement
activities**

**Data
Transformation
Activities**

**Control
Activities**

Follow us on :

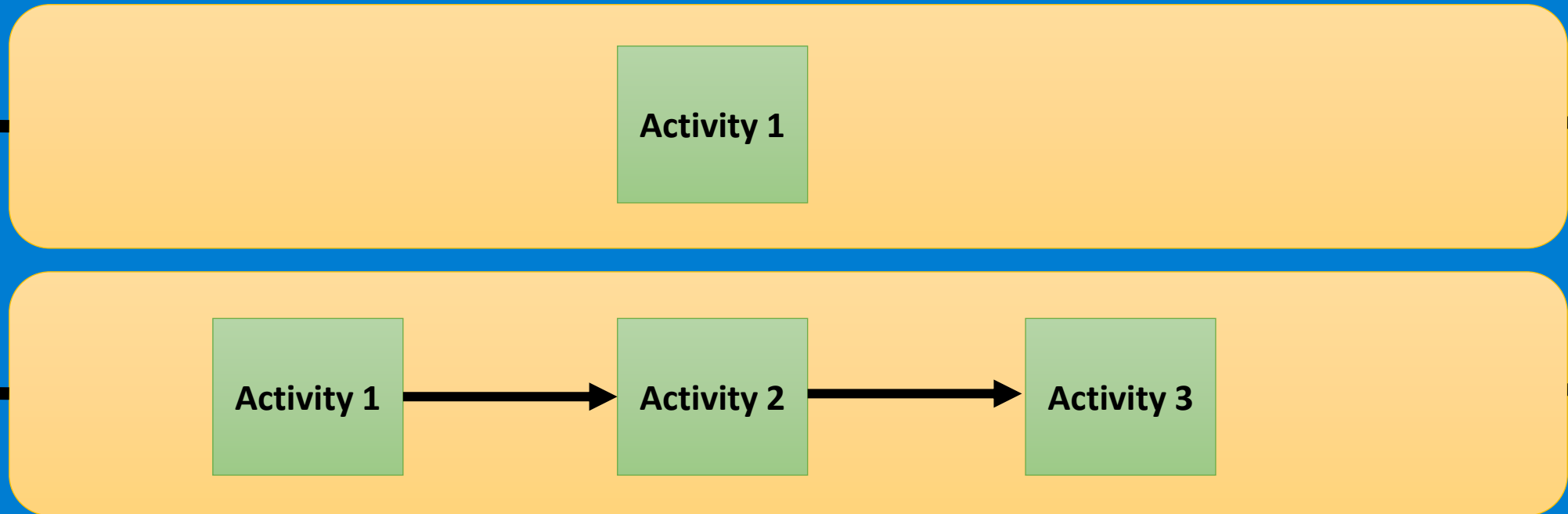


Cloud And Data Universe



Pipeline & Activity

Serial



Follow us on :



CDU

Cloud And Data Universe



Pipeline & Activity

Parallel

Activity 1

Activity 2

Activity 3

Follow us on :

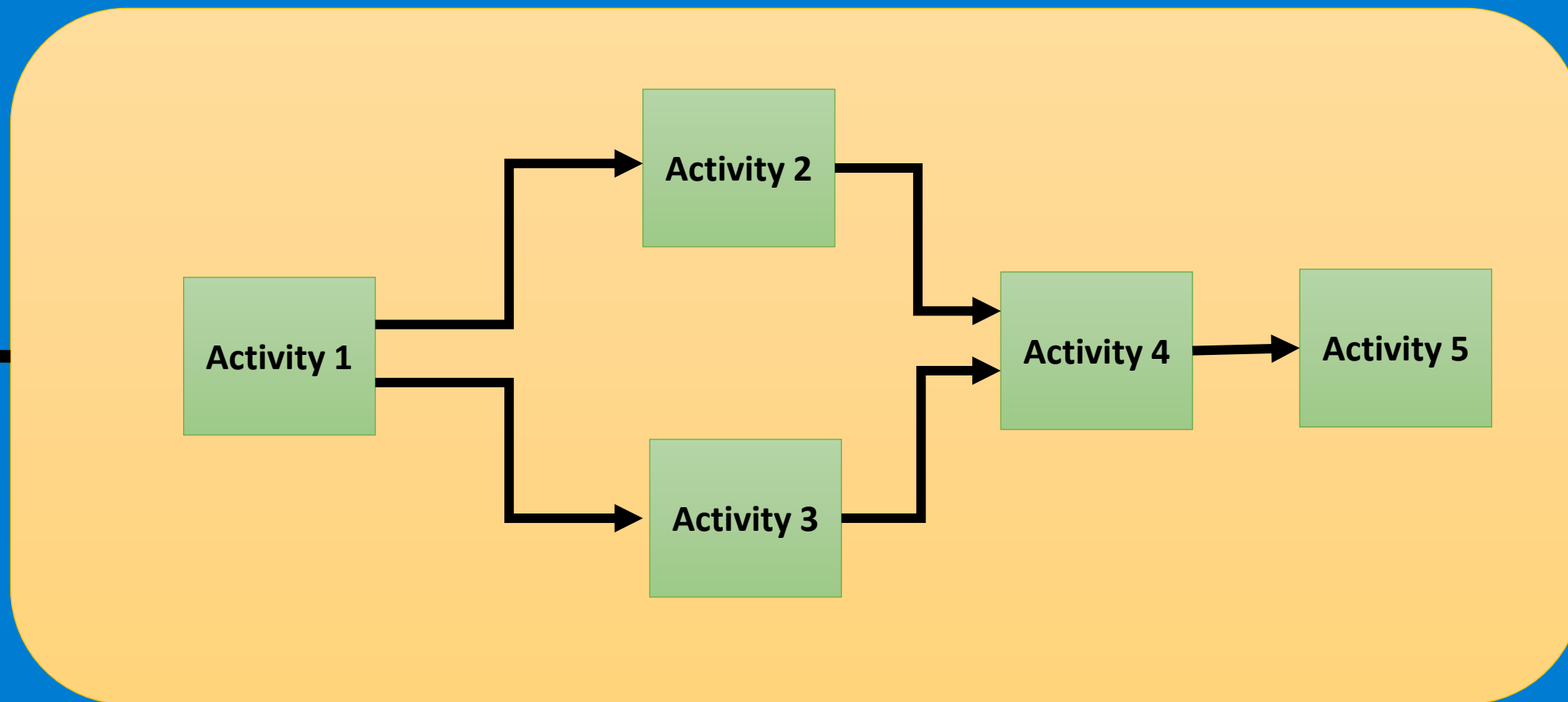


CDU

Cloud And Data Universe



Pipeline & Activity Combination



Follow us on :



CDU

Cloud And Data Universe



Activity

- There is an activity for each task.
- Different activities include copy data, if, for each(loop), set variable, etc.

Follow us on :



Cloud And Data Universe



Dataset

- Dataset is a pointer or reference to the application from where you need to extract or push the data.
- Depending on the service/application is has configurations which need to be set.
- When you create a dataset , ADF provides 90+ connectors which seamlessly help to connect to top services or applications.



Follow us on :



CDU

Cloud And Data Universe



Linked Service

- Linked service establishes connection to a service or application from where you need to extract or push the data.
- This is an authentication entity which performs the handshake between the data factory and source/target service or application



Follow us on :



CDU

Cloud And Data Universe



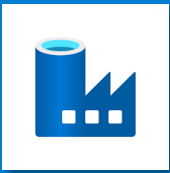
Linked Service

- Linked services are consumed in datasets

Follow us on :

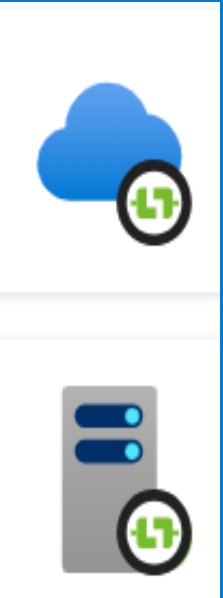


Cloud And Data Universe



Integration Runtime

- Integration Runtime a.k.a IR is the compute infrastructure needed to execute the pipelines in ADF.



Follow us on :

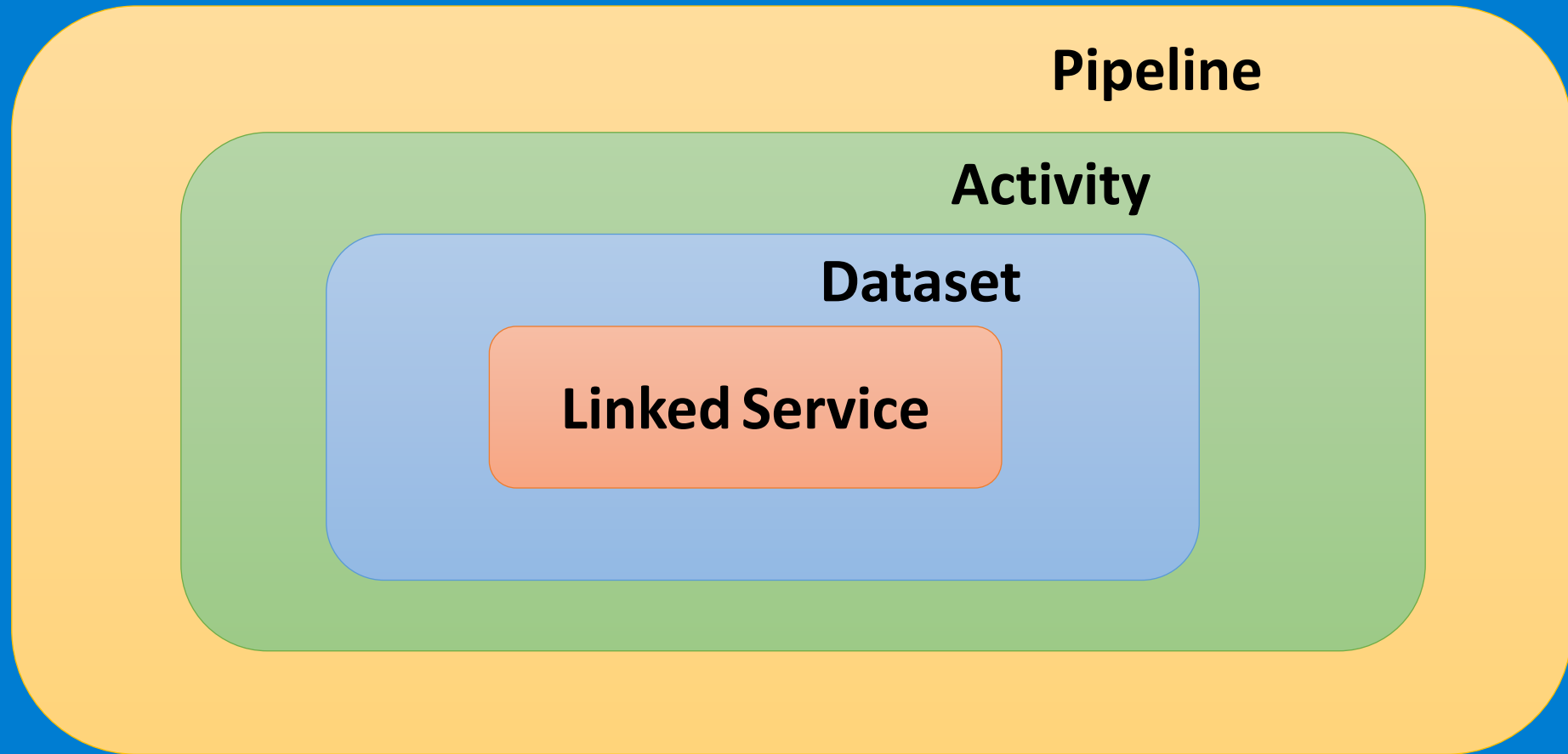


CDU

Cloud And Data Universe



Summary



Follow us on :



CDU

Cloud And Data Universe



Follow us on :



CDU

Cloud And Data Universe



2. AZURE DATA FACTORY

**COPY DATA
WITHIN DATALAKE**



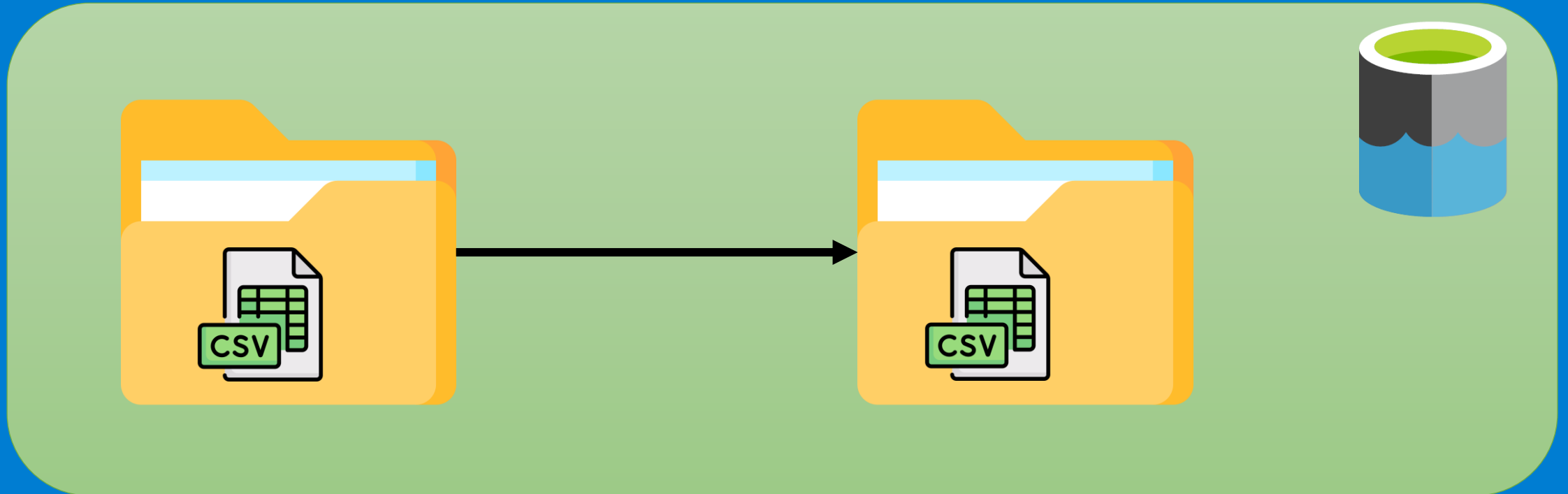
Follow us on :



Cloud And Data Universe



Copy data within ADLS gen2



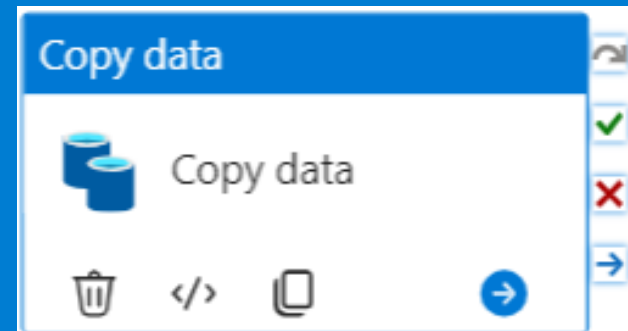
Follow us on :





Copy data within ADLS gen2

- The task of copying from one folder to another can be performed in ADF using a copy data activity.



Follow us on :



Cloud And Data Universe



2. AZURE DATA FACTORY

**COPY DATA
WITHIN DATALAKE**



Follow us on :



Cloud And Data Universe



3. AZURE DATA FACTORY

**COPY ENTIRE
FOLDER/SPECIFIC FILE
TYPES IN DATALAKE**



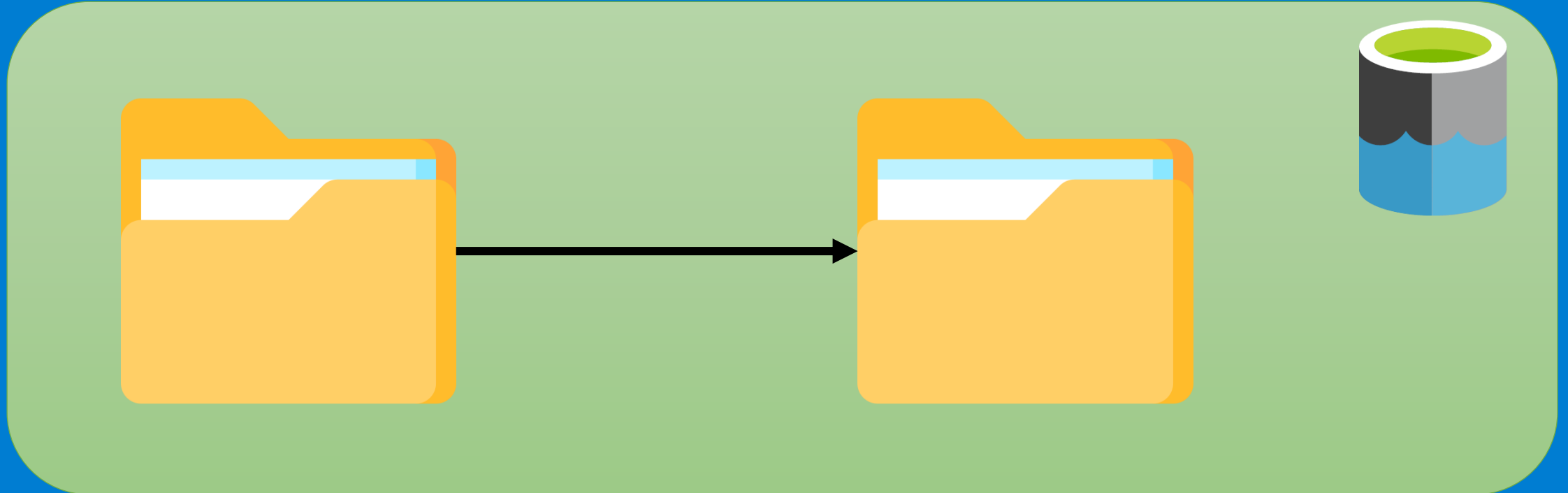
Follow us on :



Cloud And Data Universe



Copy folder within ADLS gen2



Follow us on :

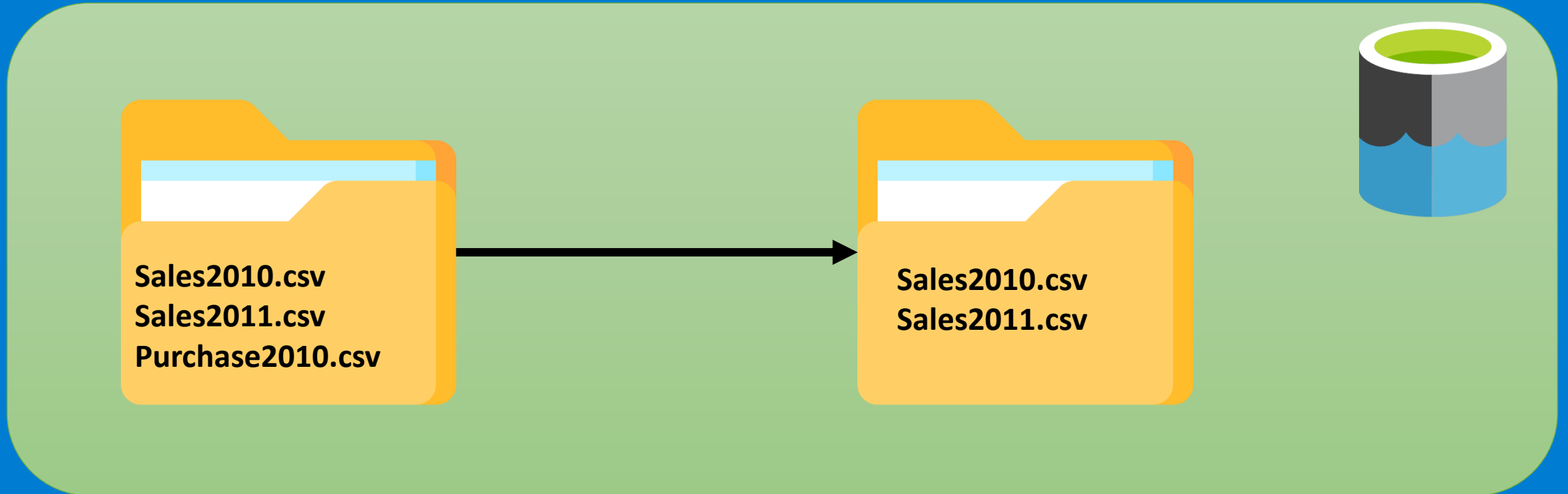


CDU

Cloud And Data Universe



Copy specific files within ADLS gen2



Follow us on :

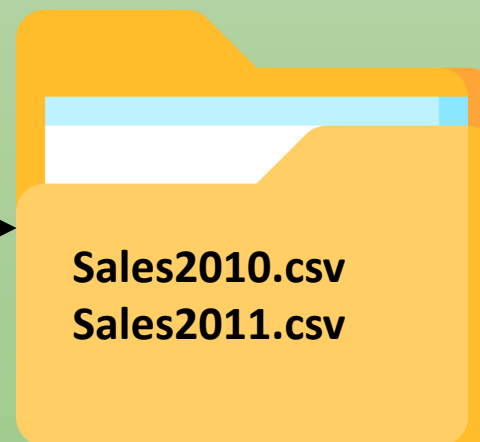
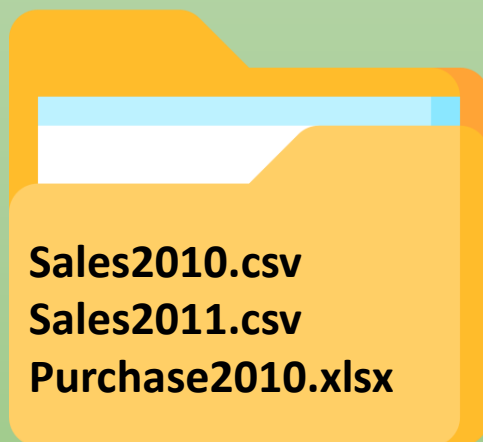


CDU

Cloud And Data Universe



Copy specific file types within ADLS gen2



Follow us on :



CDU

Cloud And Data Universe



3. AZURE DATA FACTORY

**COPY ENTIRE
FOLDER/SPECIFIC FILE
TYPES IN DATALAKE**



Follow us on :



Cloud And Data Universe



4.AZURE DATA FACTORY

**COPY MULTIPLE FILES
IN DATALAKE**



Follow us on :



CDU

Cloud And Data Universe