Azure Data Factory Masterclass

About Me

Alpa Buddhabhatti

/alpaBuddhabhatti/

/alpabuddhabhatti/

/AlpaB7

/@alpabuddhabhatti

/@meetalpa

Microsoft Certified: Azure Data Engineer Associate
Microsoft

Microsoft Certified: Azure Data Scientist Associate
Microsoft

Microsoft Certified: Azure Developer Associate
Microsoft

Microsoft Certified Trainer 2021-2022
Microsoft

Agenda

1. Azure Data Factory (ADF) Overview
    - 1.1 What is Azure Data Factory?
    - 1.2 What are Key Components of Azure Data Factory?
    - 1.3 What can you do in ADF?
    - 1.4 How Azure Data Factory Works?
2. Labs –
    - 2.1 ADF Quick Tour !!!!
    - 2.2 Azure Resources Creations
    - 2.3 Scenarios
        - Creating Linked Services & Datasets
        - Creating Pipeline – Data Movements
        - Debugging Pipeline
        - Triggering Pipeline

1.Azure Data Factory (ADF)

What is Azure Data Factory(ADF)?

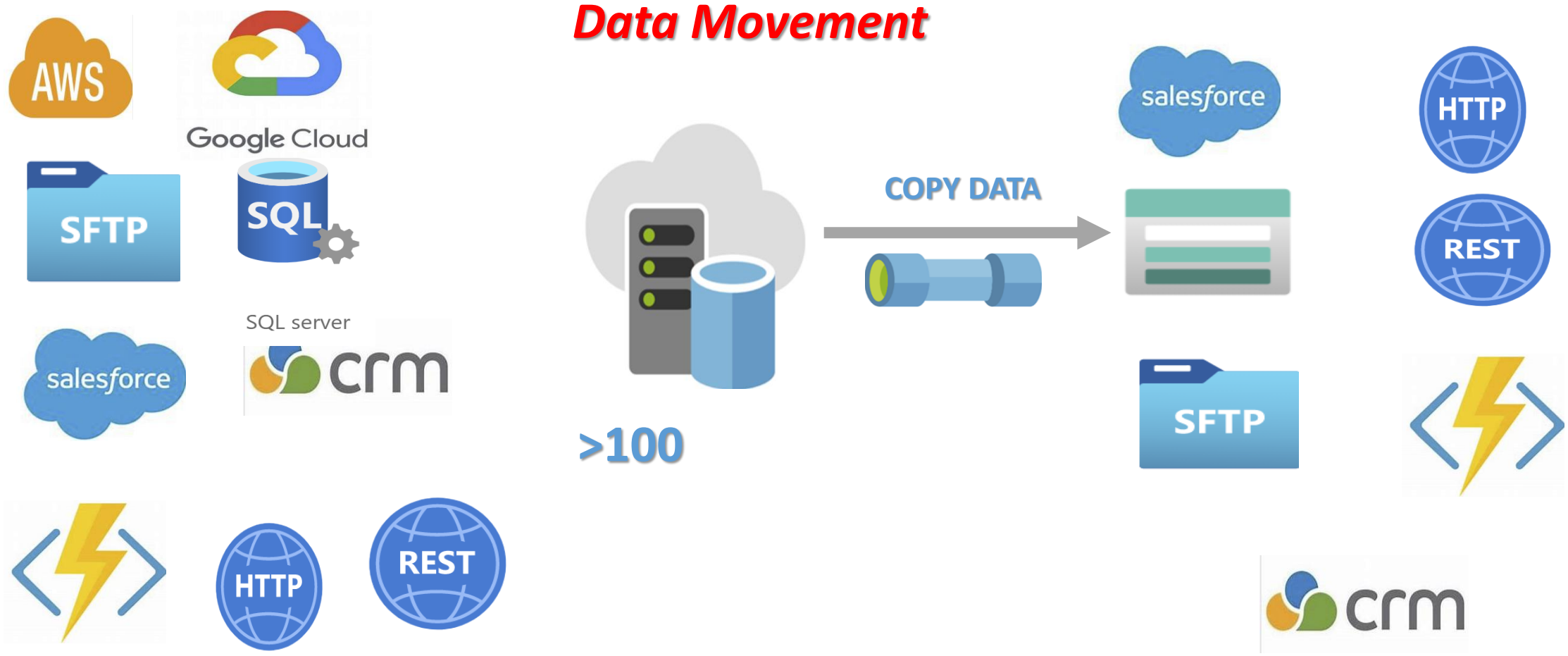
A cloud-based data integration service that ***orchestrates*** and ***automates*** the ***movement*** and ***transformation*** of data.

- ✓Azure Platform as a Service (PaaS)
- ✓It is Serverless
- ✓Only pay for what you have use
- ✓Low-code and no-code Solutions

What are Key Components of ADF?

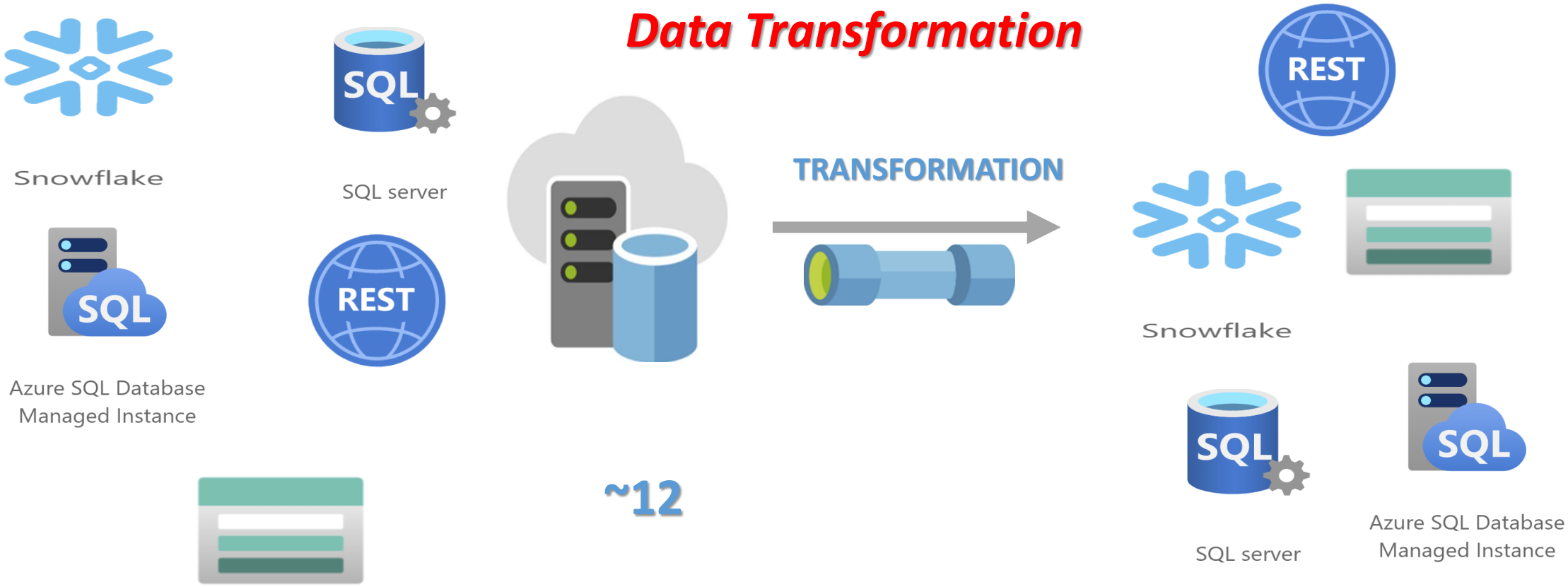
1. Linked services
2. Datasets
3. Activities
4. Pipelines
5. Triggers

What *can you do in ADF?*



[Copy activity - Azure Data Factory & Azure Synapse | Microsoft Docs]

What can you do in ADF?



[Source transformation in mapping data flow - Azure Data Factory & Azure Synapse | Microsoft Docs]

What can you do in ADF?

✓Pipelines are JSON, ready for Source Control

✓Git Integration built in

✓Monitor

✓Manage

What can you do in ADF?

✓Trigger pipeline

A Specific Time

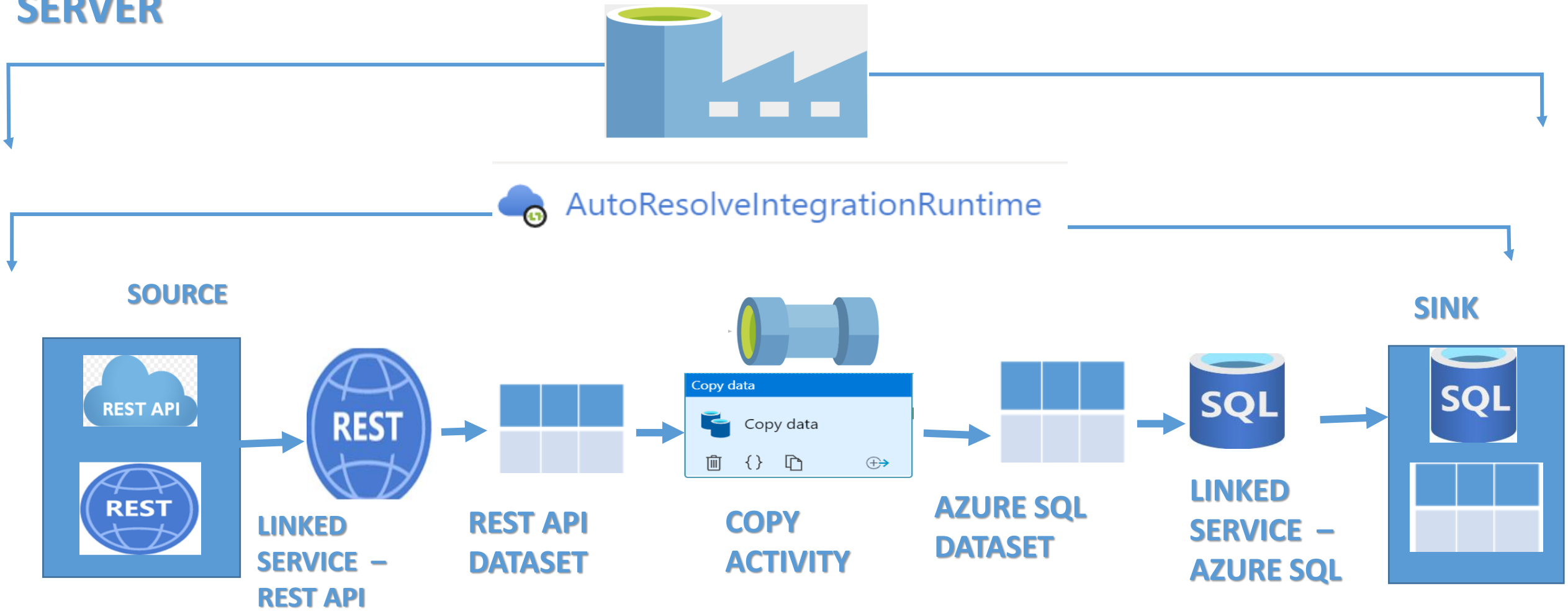
File created or deleted from Blob Storage

Customer Activity such as




How Data Factory Works?

Let's Imagine we have scenarios to move Data from REST API to Azure SQL SERVER



2. Labs

2.1 Lab - A Quick ADF Tour!!!

2.2 Lab - Azure Resources creation

2.3 Lab - Pipeline Designs

Lab 2.3

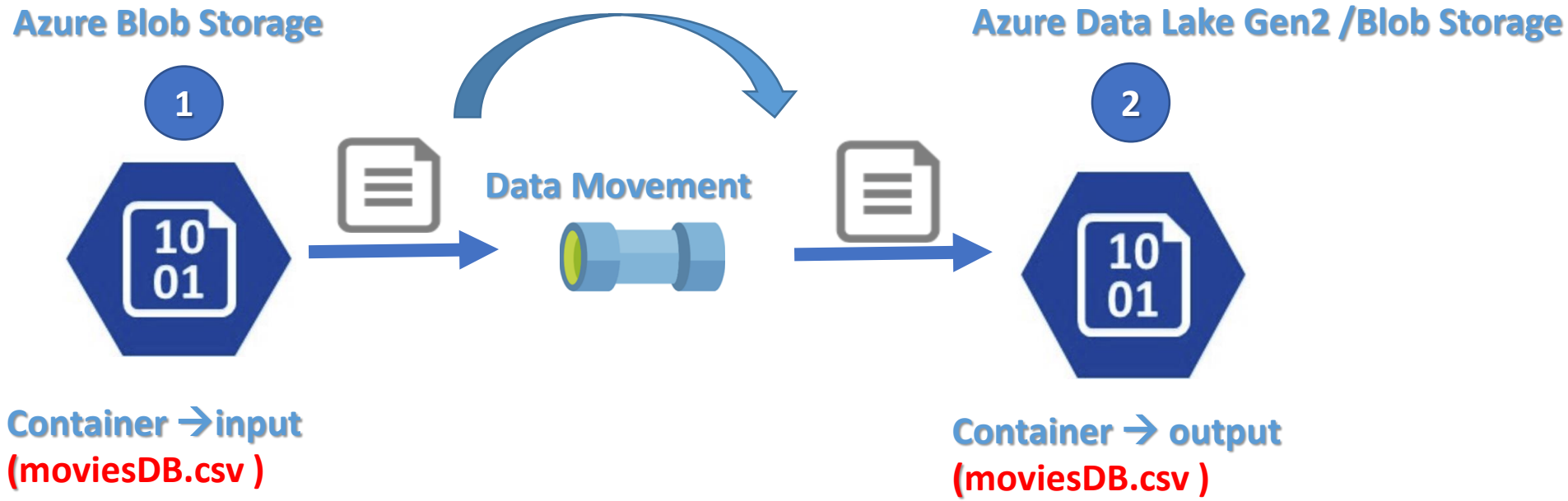
Scenarios 1:

Moving file from Azure Blob storage to Azure Data Lake Gen2.

Following Azure resources needed per environment

- ✓ *Azure Resource group*
- ✓ *Azure Data factory*
- ✓ *Azure Data Lake*
- ✓ *Azure Blob Storage*

Lab 2.3 – Scenarios 1



Lab 2.3

Scenarios 2:

Moving file/s (csv, txt, .pdf,.jpg, .dump) from Azure Blob storage to Azure Data Lake Gen2.

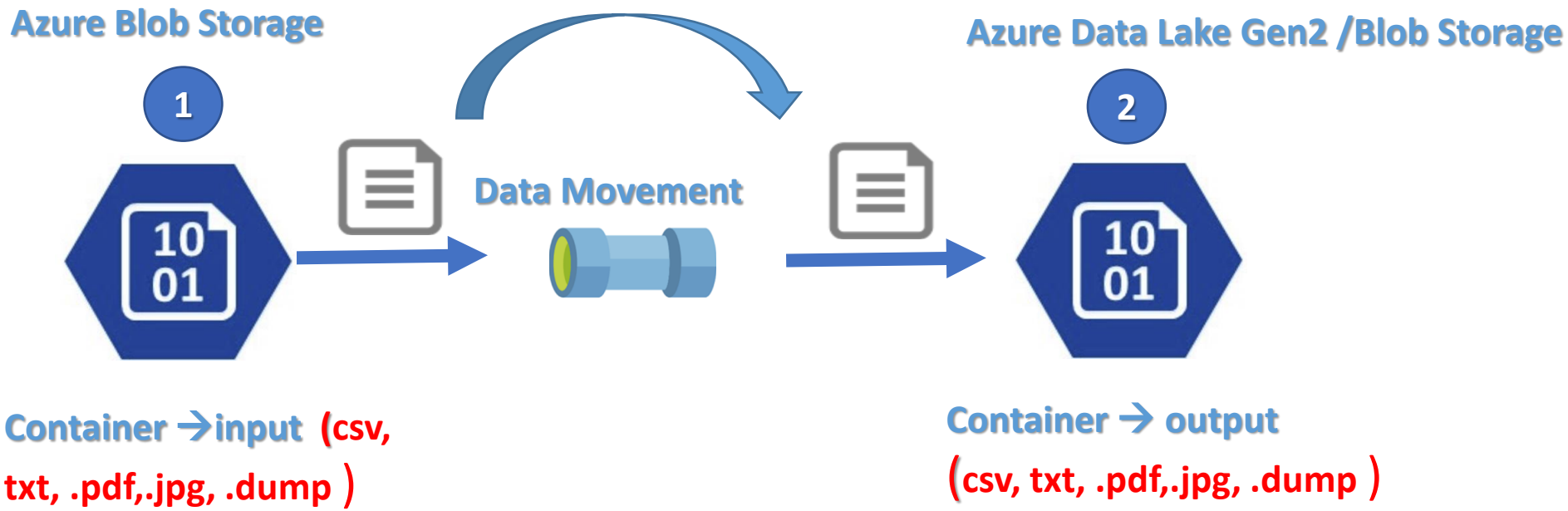
Following Azure resources needed per environment

- ✓ *Azure Resource group*
- ✓ *Azure Data factory*
- ✓ *Azure Data Lake*
- ✓ *Azure Blob Storage*

NOTE - Here two Datasets(ABLB, ADLG) and two Linked Services (ABLB, ADLG) Needed

- One Pipeline with Copy Activity

Lab 2.3 – Scenarios 2



Lab 2.3

Scenarios 3:

Moving file from Azure Blob Storage to Azure SQL Database

Following Azure resources needed per environment

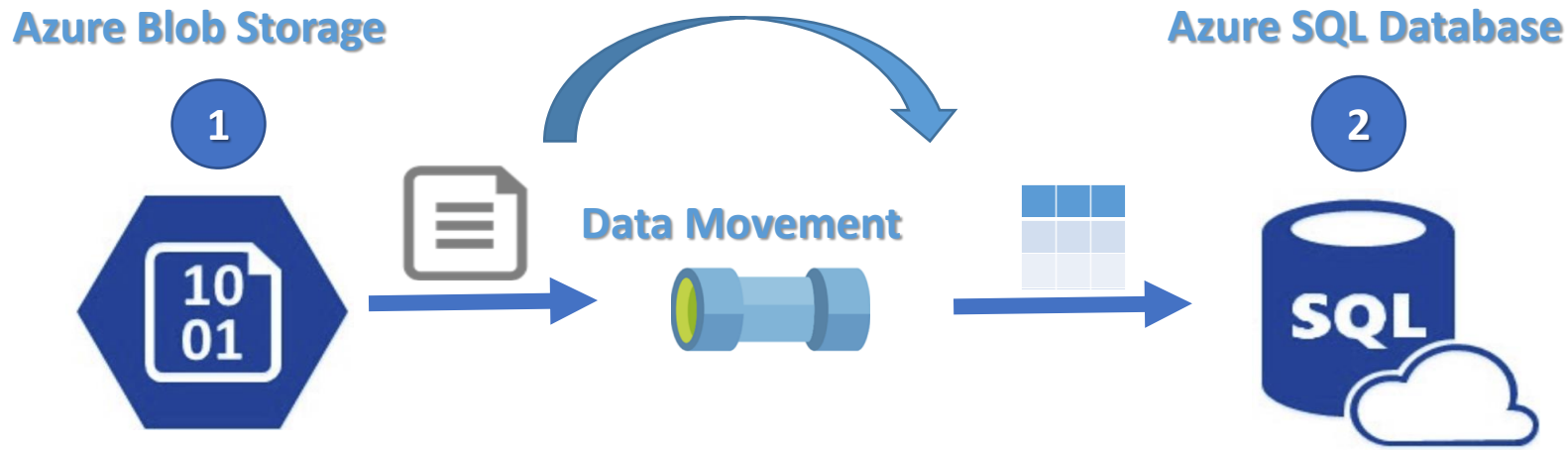
- ✓ *Azure Resource group*
- ✓ *Azure Data factory*
- ✓ *Azure Blob Storage*
- ✓ *Azure SQL Database*

Note –

- Here two Datasets(ABLB, ASQL) and two Linked Services (ABLB, ASQL) Needed

- One Pipeline with Copy Activity

Lab 2.3 – Scenarios 3



Lab 2.3

Scenarios 4:

Moving date from Http Server to Azure Data Lake Gen2.

Following Azure resources needed per environment

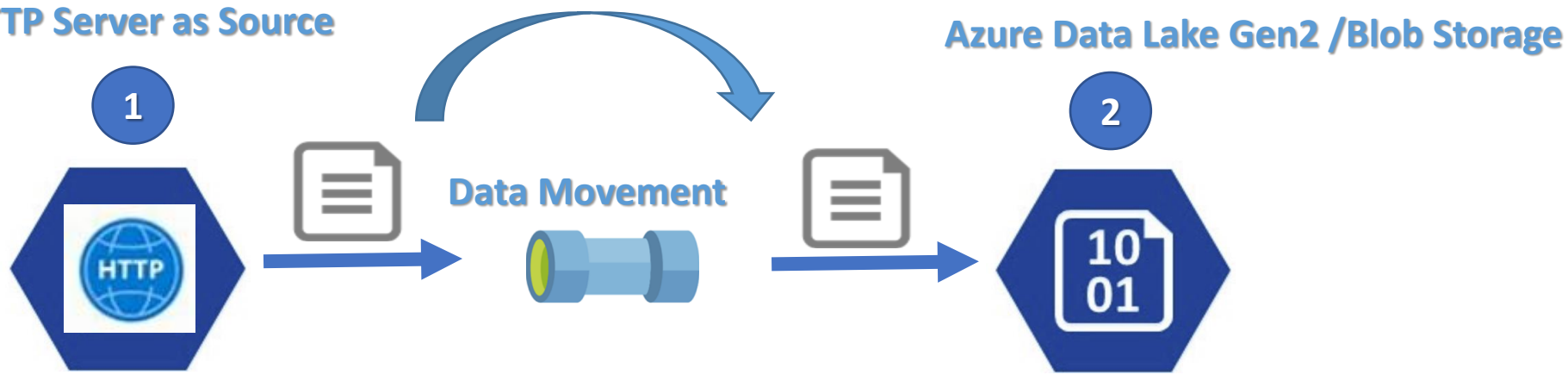
- ✓ *Azure Resource group*
- ✓ *Azure Data factory*
- ✓ *Azure Data Lake*

Note –

- Here two Datasets(HTTP, ADLG) and two Linked Services (HTTP, ADLG) Needed

- One Pipeline with Copy Activity

Lab 2.3 – Scenarios 4



Conclusions

Quick Overview of Azure Data

1. Data Movements
2. Orchestration
3. Automation
4. Easy to Build Data Integration solution
5. Modern Data Warehouse (ELT or ETL)
6. Serverless and Azure PaaS

Key Component of Azure Data Factory

1. Linked Services
2. Datasets
3. Pipelines
4. Activities
5. Triggers
6. Integration Runtimes

Extract & Load (Data Movements)

1. Move data from

Blob storage(Single file) ➔ Data Lake Gen 2

Blob storage (Multiple Files) ➔ Data Lake Gen 2

Blob storage(Single File) ➔ Azure SQL

Http Server ➔ Azure Blob Storage