Data Movement and Transformation using Azure Data Factory

DataGrillen 2022

03rd JUN 2022 1:30 am (Europe/London)   https://evals.datagrillen.com/

Who I am

Alpa Buddhabhatti, Cluster Reply, UK

Microsoft Certified: Azure Data Engineer Associate
Microsoft

Microsoft Certified Trainer 2021-2022
Microsoft

Microsoft Certified: Azure Developer Associate
Microsoft

Microsoft Certified: Azure Data Scientist Associate
Microsoft

https://twitter.com/AlpaB7

alpaBuddhabhatti/DG2022-ADF (github.com)

Alpa Buddhabhatti | LinkedIn

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?
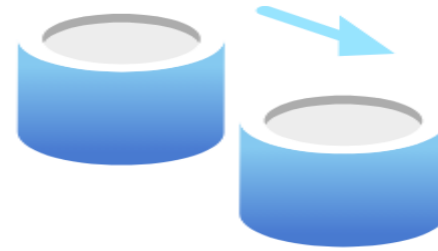
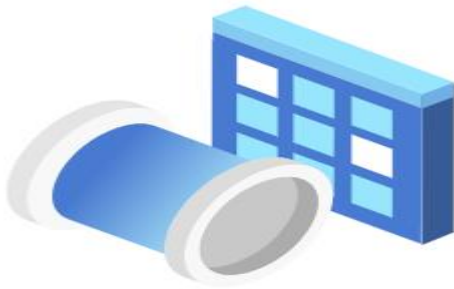2. DEMO –
    2.1 ADF Quick Tour !!!
    2.2 Case Study - Data Movement and Transformation using Azure Data Factory

What is Azure Data Factory?

A cloud-based data integration service that <u>orchestrates</u> and <u>automates</u> the <u>movement</u> and <u>transformation</u> of data
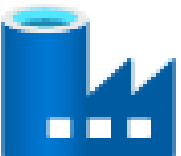
What is Azure Data Factory?

✓Azure Platform as a Service (PaaS)

✓It is Serverless

✓Only pay for what you have use

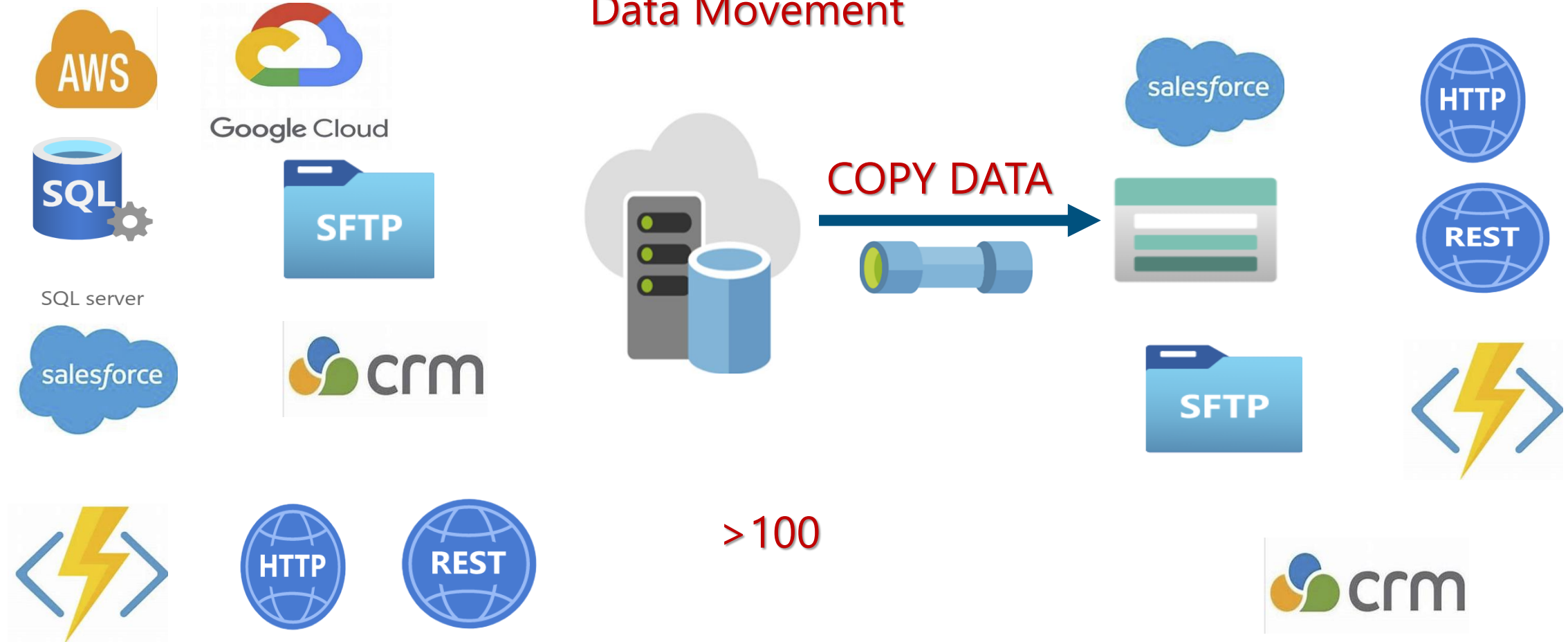✓Low-code and no-code Solutions

What are Key Components of ADF?

1   Linked services

2   Datasets

3   Activities

4   Pipelines

5   Triggers

What can you do in ADF?

Data Movement

COPY DATA

>100

Copy activity – Azure Data Factory & Azure Synapse | Microsoft Docs

What can you do in ADF?

Data Transformation

TRANSFORMATION

Snowflake

SQL server

Azure SQL Database
Managed Instance

REST

~18

REST

Snowflake

SQL server

Azure SQL Database
Managed Instance
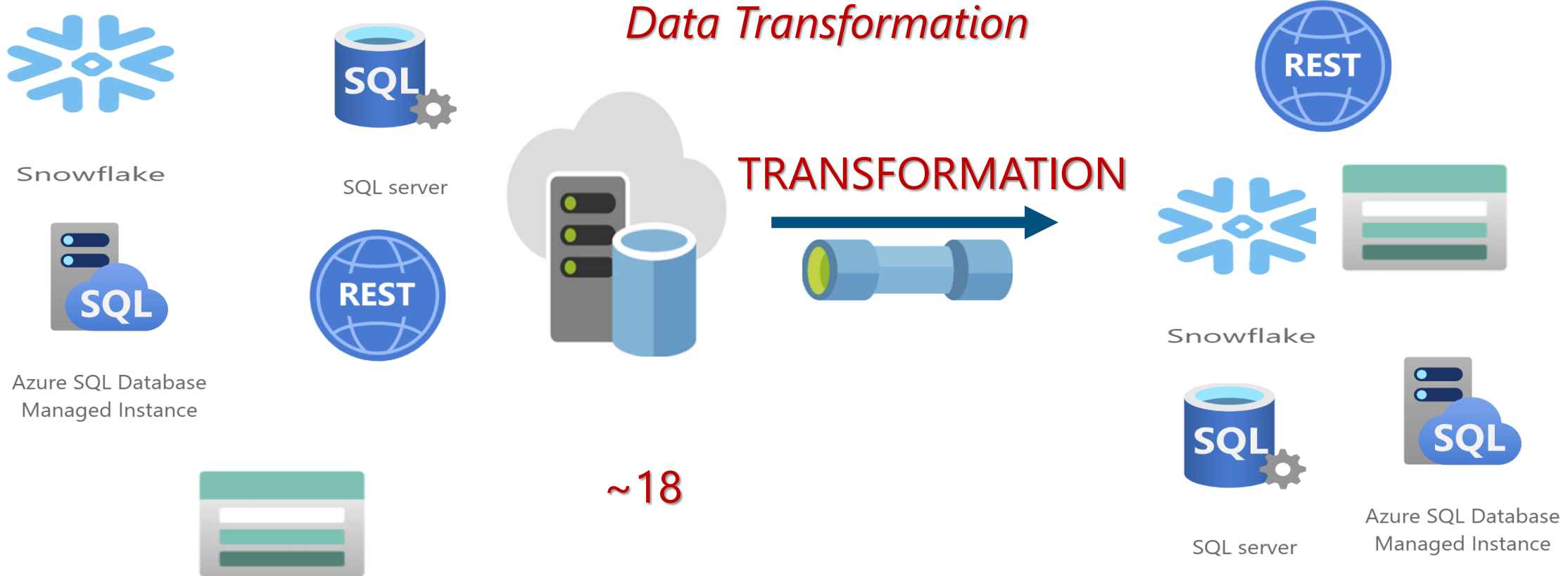
Source transformation in mapping data flow - Azure Data Factory & Azure Synapse | Microsoft Docs

What can you do in ADF?

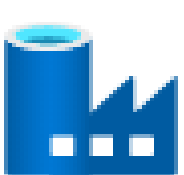- ➢ Pipelines are JSON, ready for Source Control
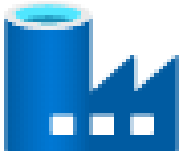
- ➢ Git Integration built in

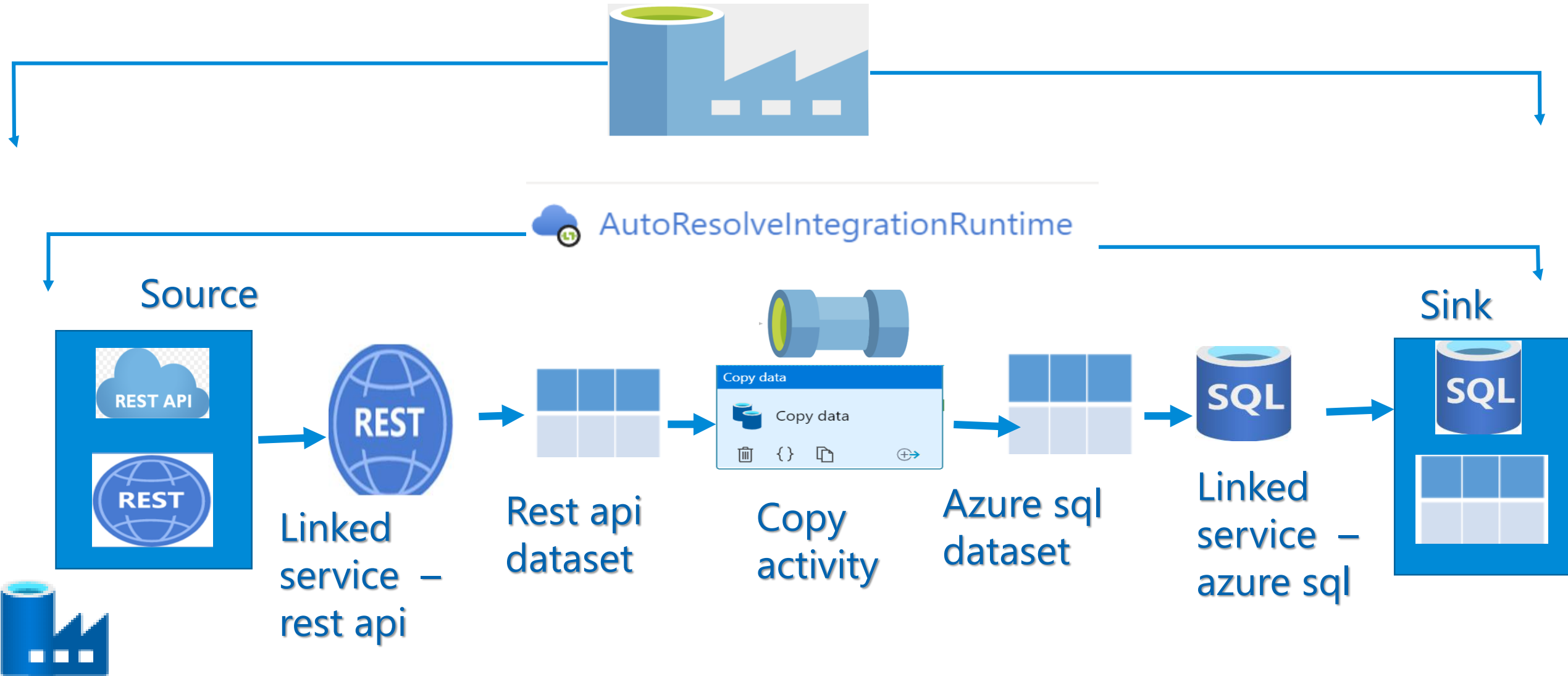- ➢ Monitor

- ➢ 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

AutoResolveIntegrationRuntime

Source

Sink

Linked service – rest api

Rest api dataset

Copy data

Copy activity

Azure sql dataset

Linked service – azure sql

DEMO

Lets have a Quick ADF Tour!!!

DEMO

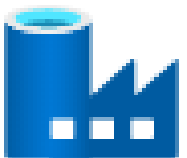A Case Study - Data Movement & Data Transformation

Azure Resources

Scenarios :

Moving data from Azure Blob storage to Azure Data Lake Gen2.
Apply Transforming on data using Azure data flow and store to Azure SQL Server.
Following Azure resources needed per environment

1. Azure Resource group
2. Azure Data factory
3. Azure Data Lake
4. Azure Blob Storage
5. Azure SQL Server
6. Azure Key Vault

Data Movement - Pipeline 1
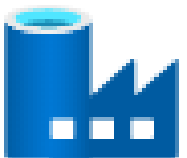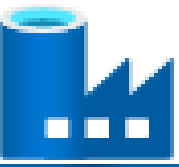
Azure Blob Storage

Azure Data Lake Gen2 /Blob Storage

1

2

Data Movement

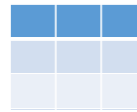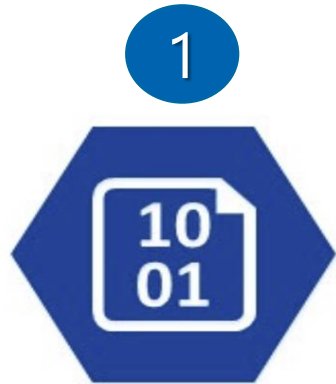Container →movie-in

Container → movie-staging

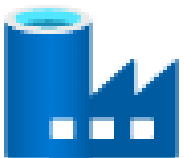Data Transformation

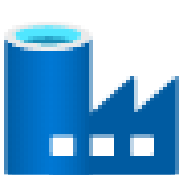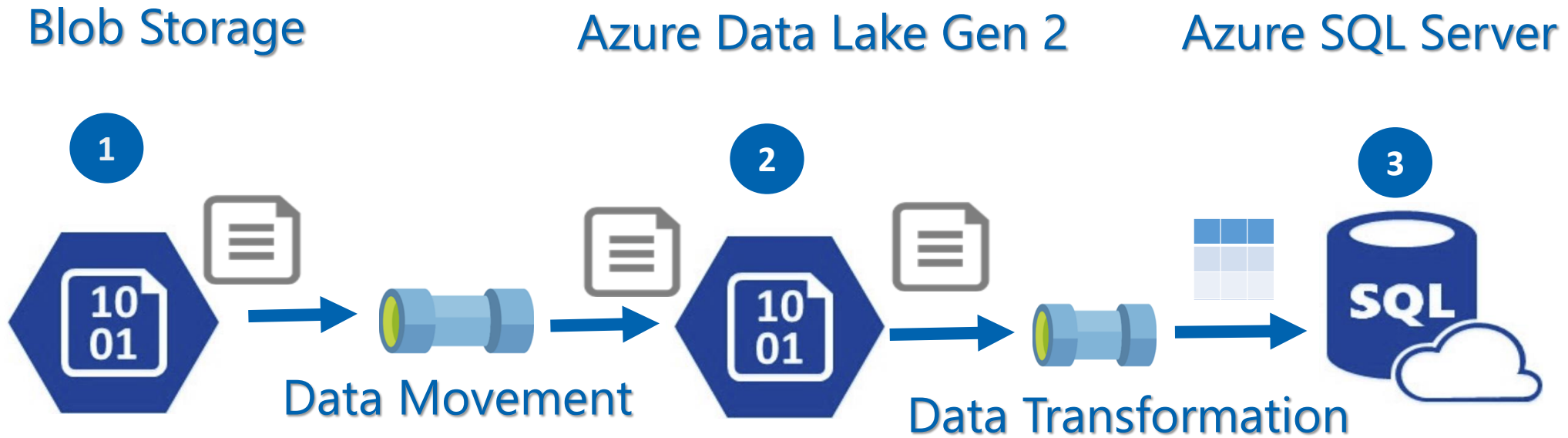# Data Transformation – Pipeline 2

Azure Data Lake Gen2

Azure SQL Server

1

2

Container → movie-staging

Mapping Data Flow

Table → dbo.MovieDB

Extract, Load and Transform (ELT)

Blob Storage

Azure Data Lake Gen 2

Azure SQL Server

1

2

3

Data Movement

Data Transformation

Conclusion

Quick Overview of Azure Data

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

Key Component of Azure Data Factory

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

ELT or ETL

1. Move data from

   Blob storage ➔ Data Lake Gen 2
2. Data transformation

   Azure Data Lake Gen 2 ➔ Azure SQL

.

Thank you!!!

Alpa Buddhabhatti, Cluster Reply, UK

Microsoft Certified: Azure Data Engineer Associate
Microsoft

Microsoft Certified Trainer 2021-2022
Microsoft

Microsoft Certified: Azure Developer Associate
Microsoft

Microsoft Certified: Azure Data Scientist Associate
Microsoft

https://twitter.com/AlpaB7

alpaBuddhabhatti (github.com)

Alpa Buddhabhatti | LinkedIn

Q&A

Alpa Buddhabhatti, Cluster Reply, UK

Microsoft Certified:
Azure Data Engineer
Associate
Microsoft

Microsoft Certified
Trainer 2021-2022
Microsoft

Microsoft Certified:
Azure Developer
Associate
Microsoft

Microsoft Certified:
Azure Data Scientist
Associate
Microsoft

https://twitter.com/AlpaB7

alpaBuddhabhatti/DG2022-ADF (github.com)

Alpa Buddhabhatti | LinkedIn

https://evals.datagrillen.com/