Vinodh Kumar TD

Azure Data Architect

Microsoft Certified Trainer

C# Corner MVP

Blogger

21st Jan 2023

Data transformation using Databricks & ADF

Azure Data Factory

ADF Layers

Pipeline: The graphical user interface where you place widgets and draw data paths

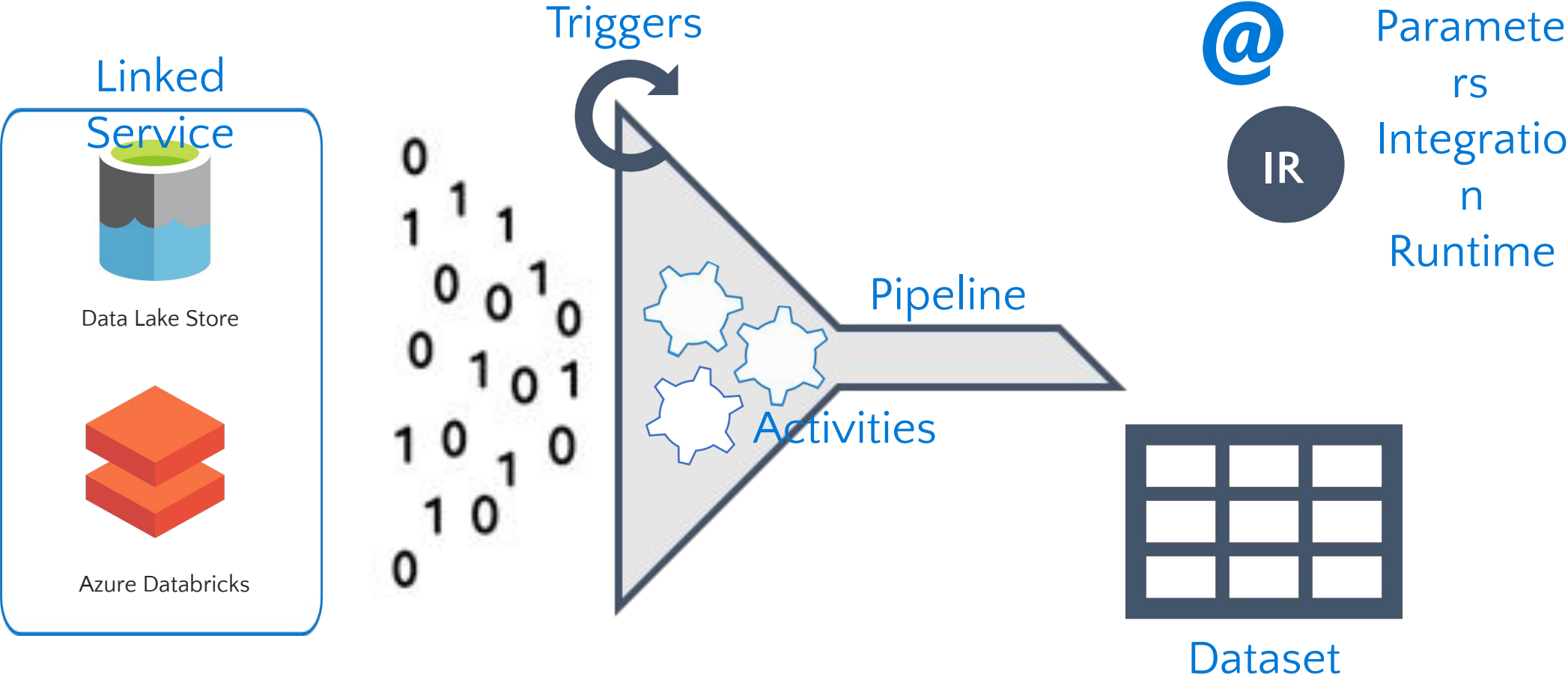Activity: A graphical widget that does something to your data

Source and Sink: The parts of an activity that specify where data is coming from and going to

Data Set: An explicitly defined set of data that ADF can operate on

Linked Service: The connection information that allows ADF to access a specific outside data resource

Integration Runtime: A glue/gateway layer that lets ADF talk to software outside of itself

Data integration with Azure Data Factory

Linked Service

Data Lake Store

Azure Databricks

Triggers

Pipeline

Activities

@ Parameters

IR Integration Runtime

Dataset

Linked Services

Configure and manage secrets in Azure Key Vault

Azure Key Vault protects

1. Secrets

2. Keys

3. Certificates

Data Engineers are typically concerned with accessing the data contained in Key Vault to apply to linked services

Datasets

Activities and pipelines
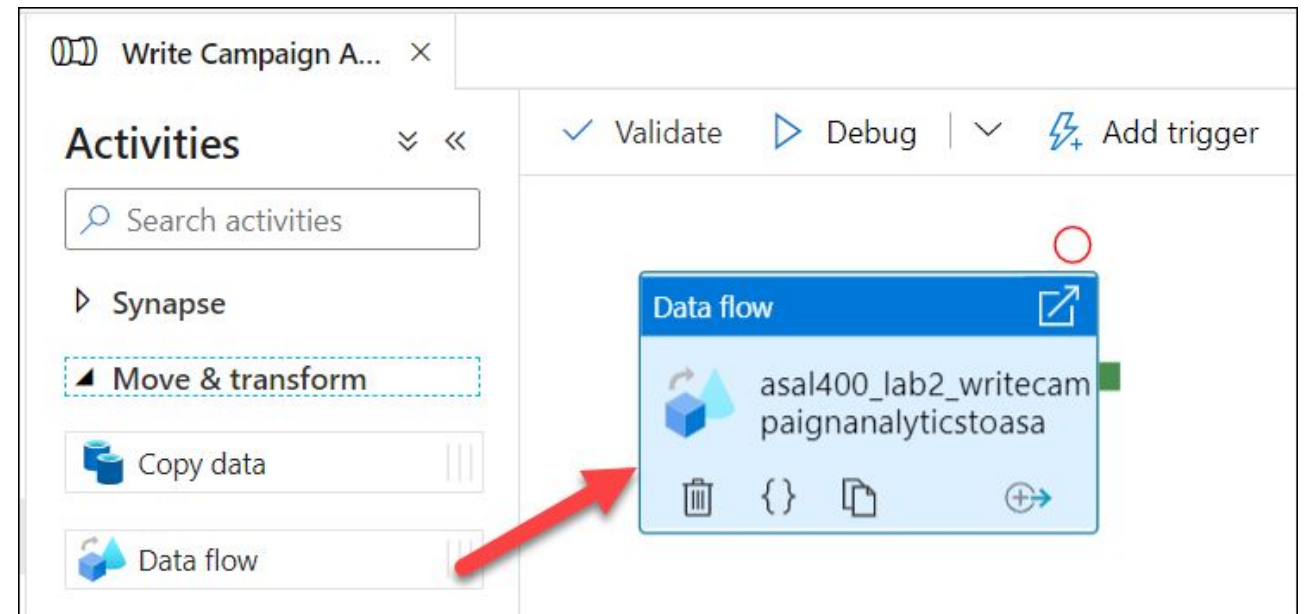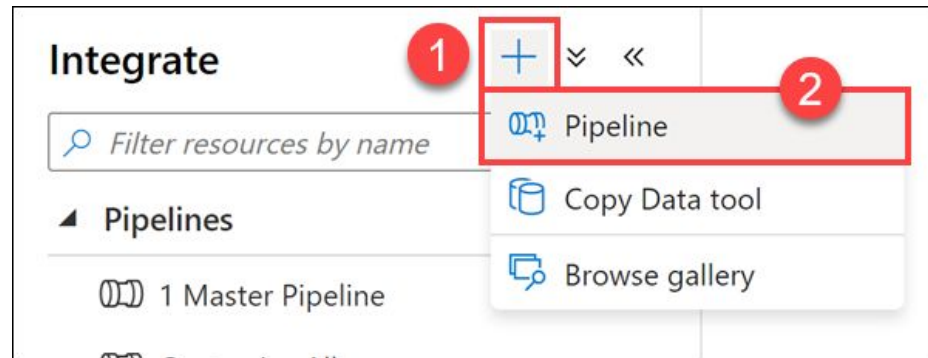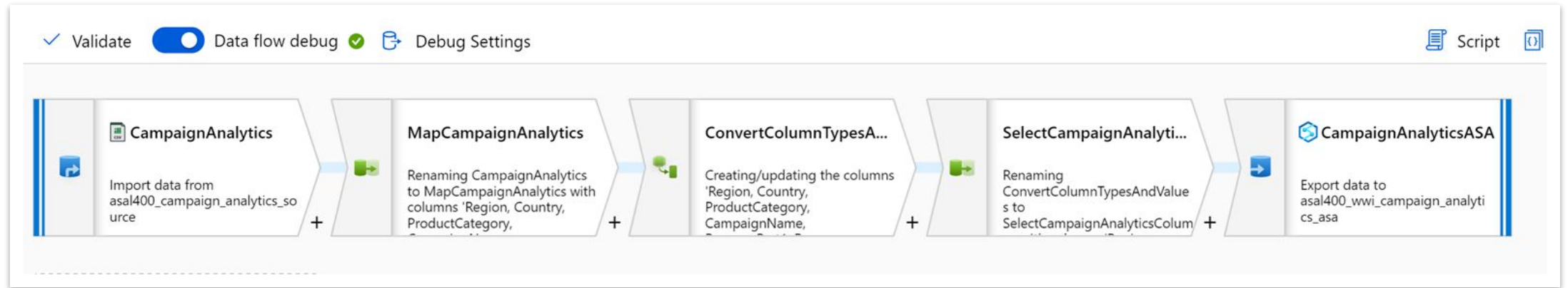
Processing Big Data with Azure – Typical usage
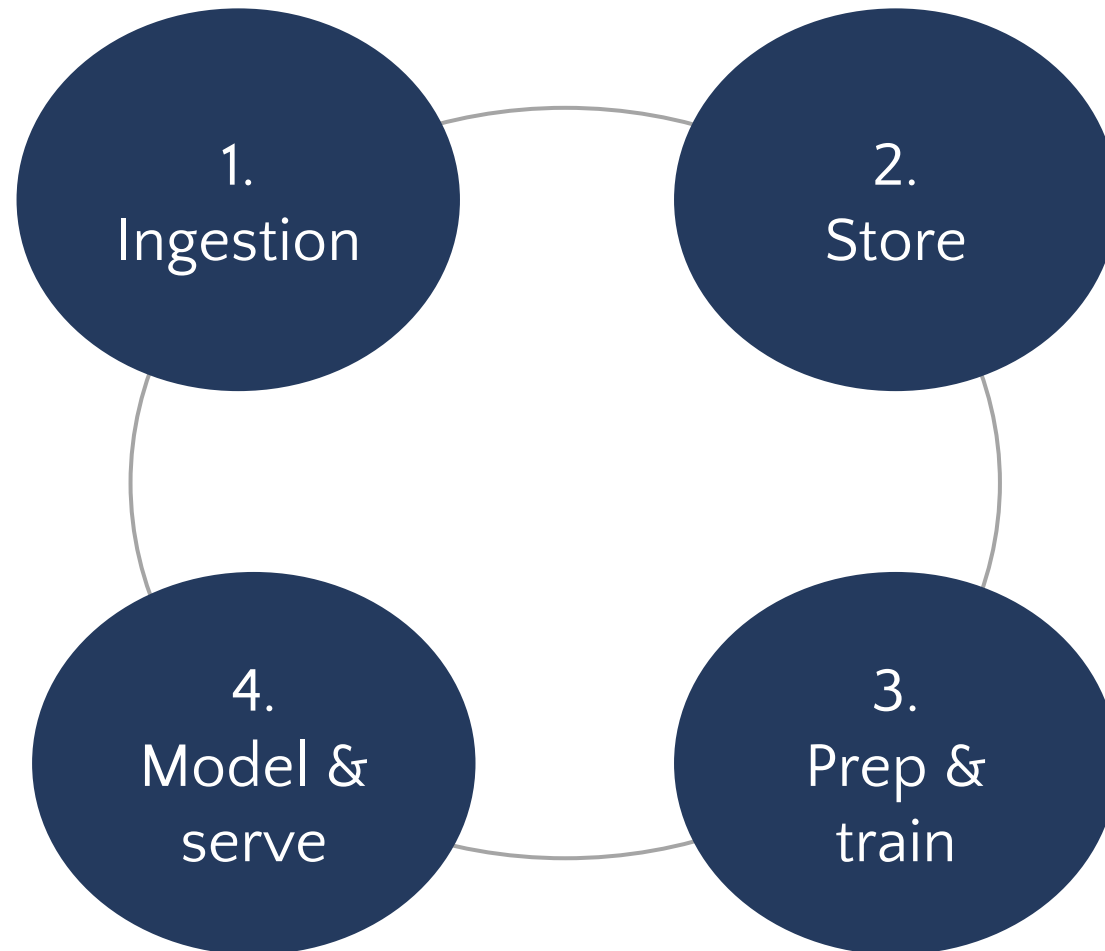
Integrating data from Notebooks with Azure Data Factory

1.
Use Storage Account

2.
Use Azure Data Factory

3.
Create data workflow pipeline

4.
Add a Notebook Activity to the pipeline

5.
Set parameters & dependency conditions

Programmatically creating Azure Data Factory

Controlling notebook execution in a pipeline

Viewing the output or results of pipeline activities

Viewing the output or results of pipeline activities