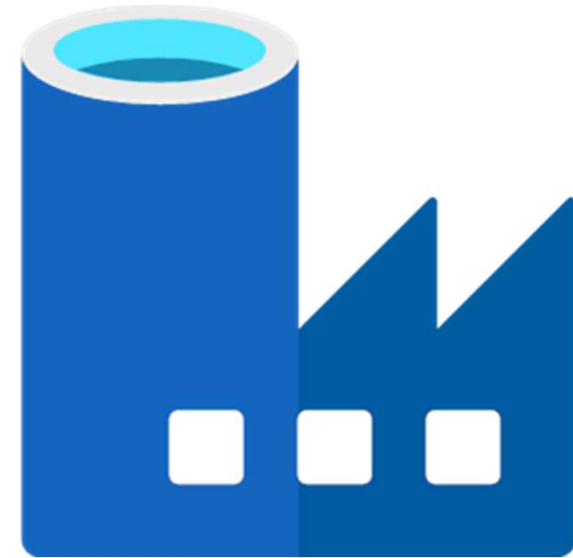


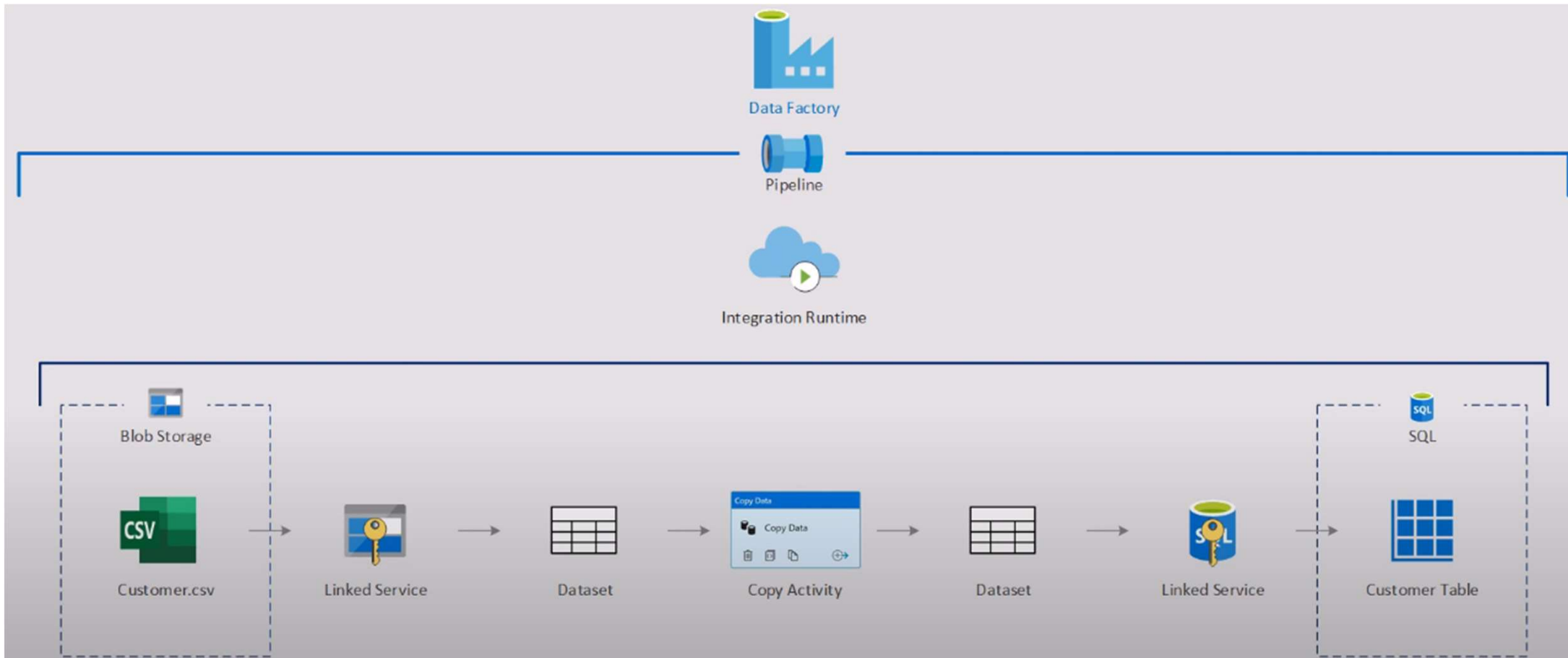
Azure Data Factory

Azure Data Factory (ADF) V2

- A powerful Cloud ETL tool
- ETL/ELT tool.
- Allow developers to integrate disparate data sources
- Provides access to
 - On-premises
 - Cloud data



Azure Data Factory



ADF allows you to..

Move data

- From on-premises and cloud sources to a centralized data store

Transform and integrate

- Big data processing and machine learning

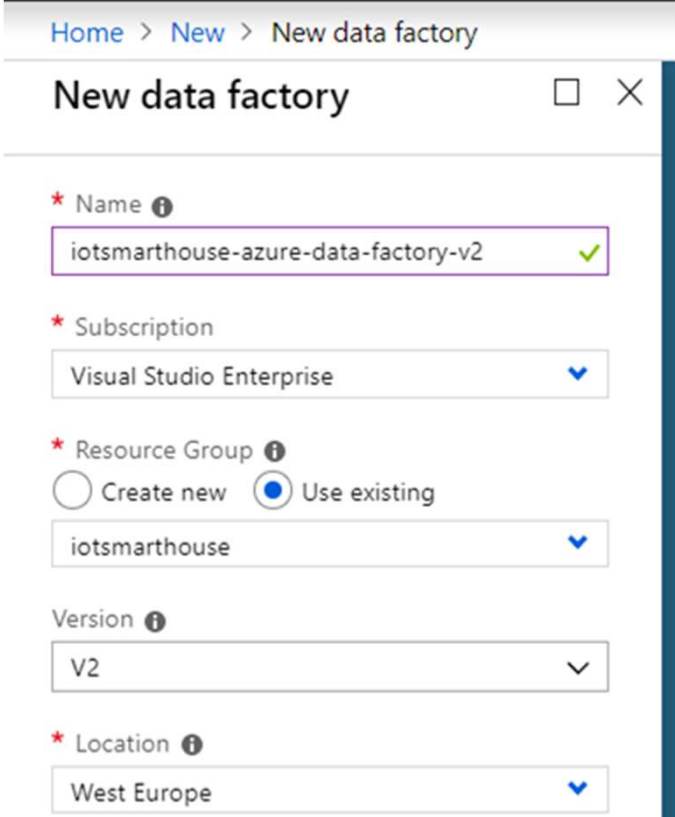
Has visual interface.

Invoke pipelines with

- Manual
- Event
- Scheduling

Hands-On: Create Azure Data Factory

- First, you do not install it, you create a service in Azure by:
 - New -> Analytics -> Data Factory
- Then, you need to set the
 - name,
 - select your subscription,
 - resource group,
 - version (1 or 2) and
 - location.



The screenshot shows the 'New data factory' form in the Azure portal. The breadcrumb navigation at the top reads 'Home > New > New data factory'. The form title is 'New data factory' with a close button (X) in the top right corner. The form contains several fields with red asterisks indicating required fields:

- Name**: A text input field containing 'iotsmarthouse-azure-data-factory-v2' with a green checkmark on the right.
- Subscription**: A dropdown menu showing 'Visual Studio Enterprise'.
- Resource Group**: A section with two radio buttons: 'Create new' (unselected) and 'Use existing' (selected). Below the radio buttons is a dropdown menu showing 'iotsmarthouse'.
- Version**: A dropdown menu showing 'V2'.
- Location**: A dropdown menu showing 'West Europe'.

What ADF can do

Data Pipelines

- Create
- Schedule
- Monitor

Accelerate

- Data integration with multiple native data connectors.

Modernize

- Data warehouse with big data integration.

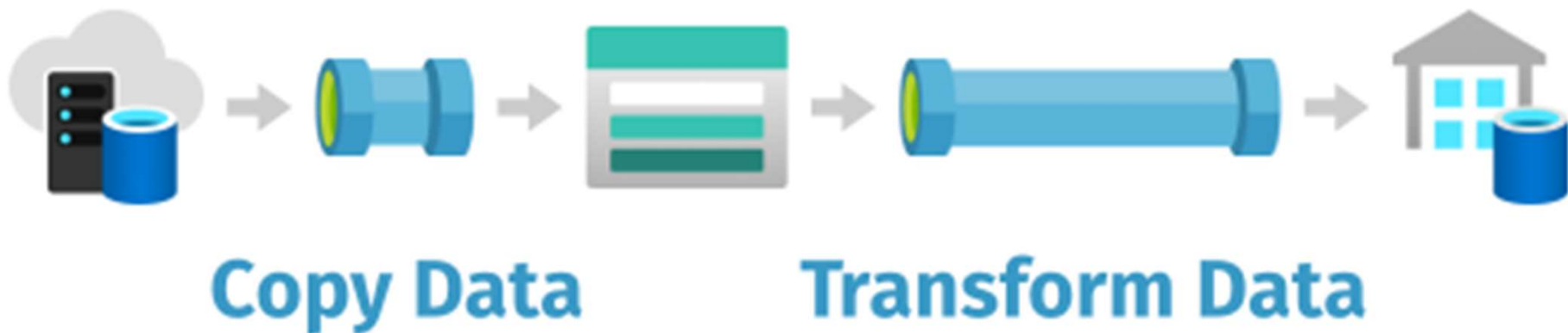
Orchestrate

- Data integration workflows wherever your data lives.

Hands-On: Explore Azure Data Factory

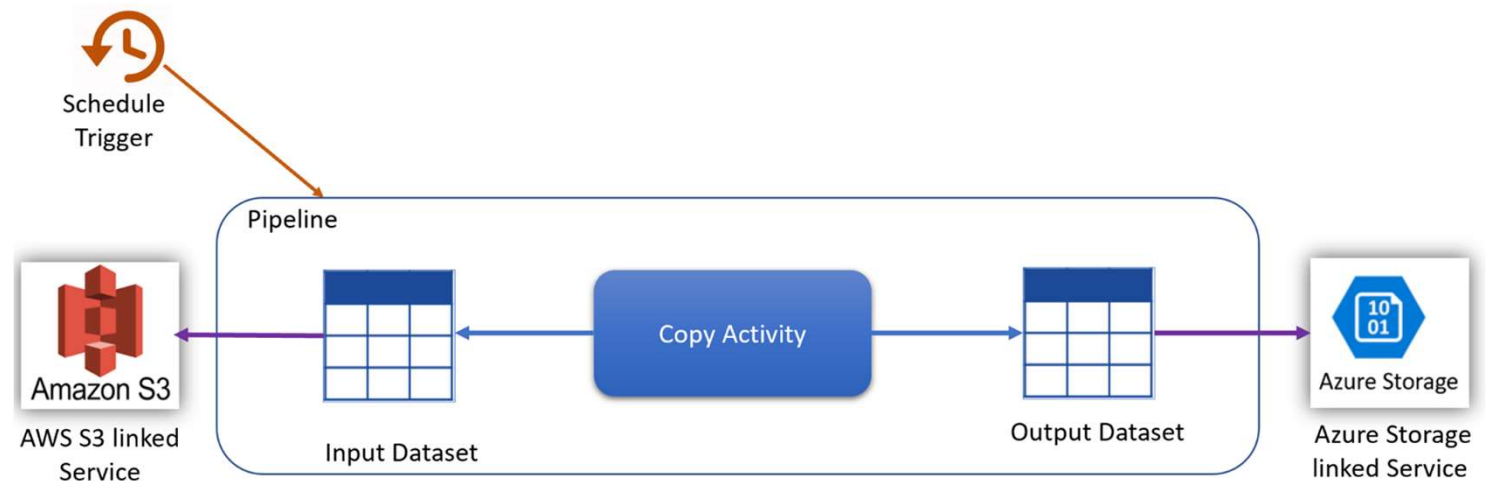
Pipeline

- Logical grouping of activities that performs a unit of work
- Together, the activities in a pipeline perform a task
- For example
 - A pipeline can contain a group of activities that ingests data from an Azure blob, and
 - Then runs a Hive query on an HDInsight cluster to partition the data.



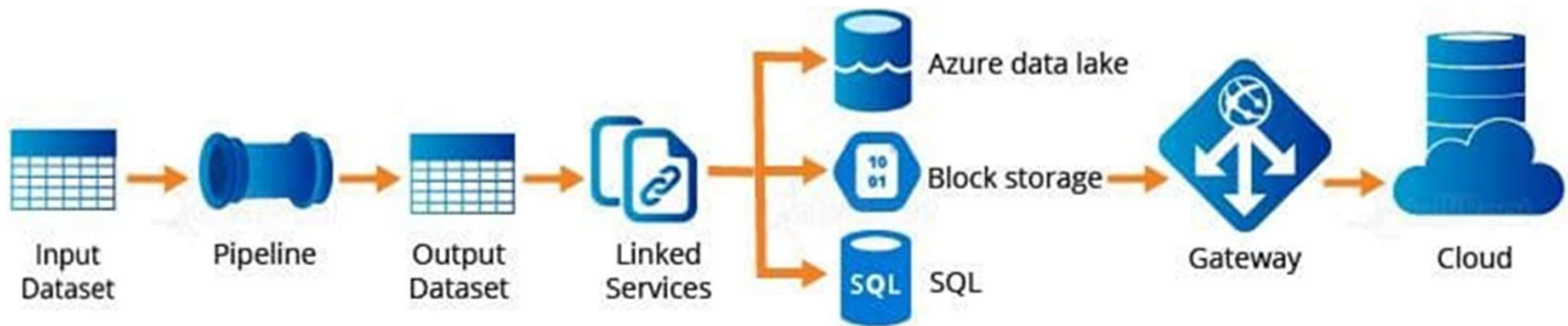
Activity

- Represent a processing step in a pipeline
- Example: Use a copy activity to copy data from one data store to another data store
- Data Factory supports three types of activities
 - Data movement
 - Data transformation
 - Control activities



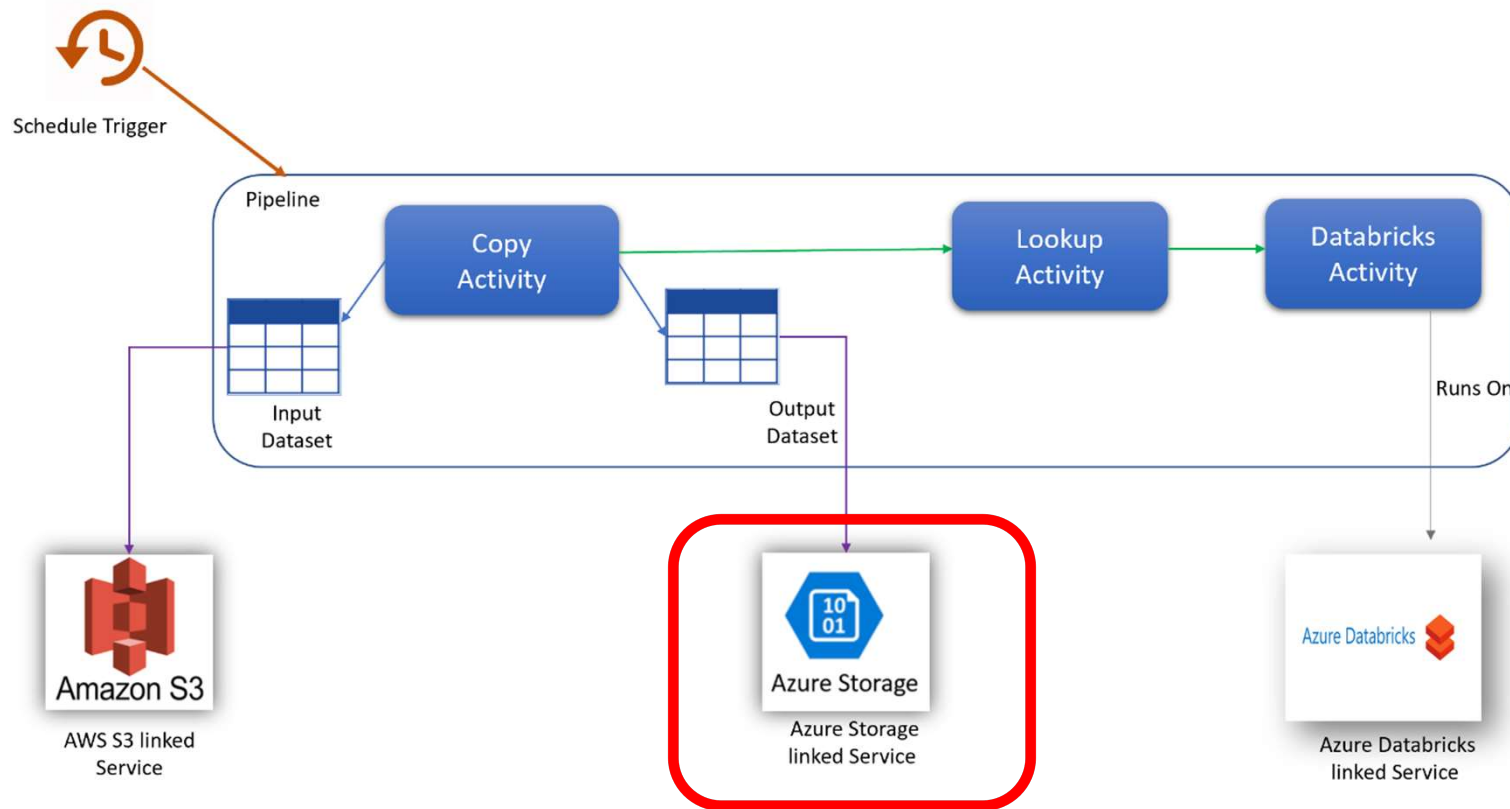
Datasets

- Represent data structures within the data stores
- It simply point to or reference the data you want to use in your activities as inputs or outputs.



Linked services

- Are much like connection strings, which define the connection information
- It's needed for Data Factory to connect to external resources

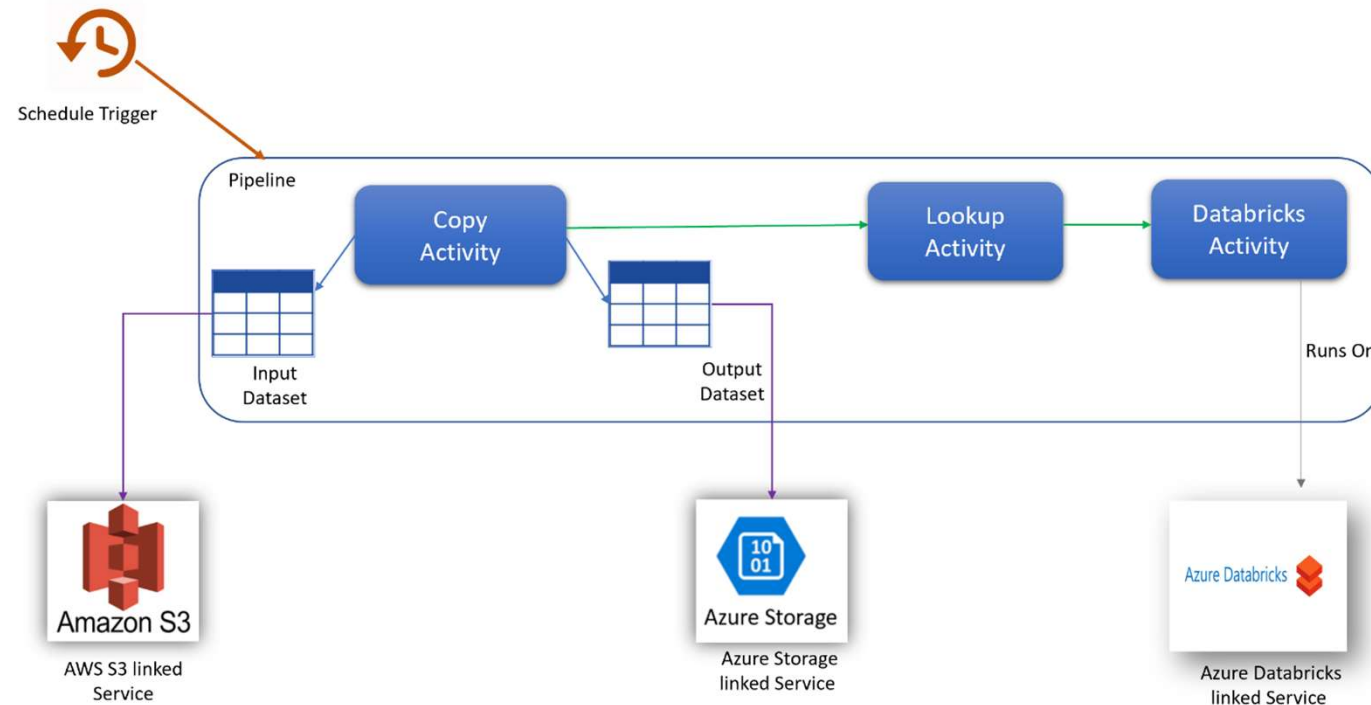


Hands-On: Explore Azure Data Factory

- Explore Azure Data Factory

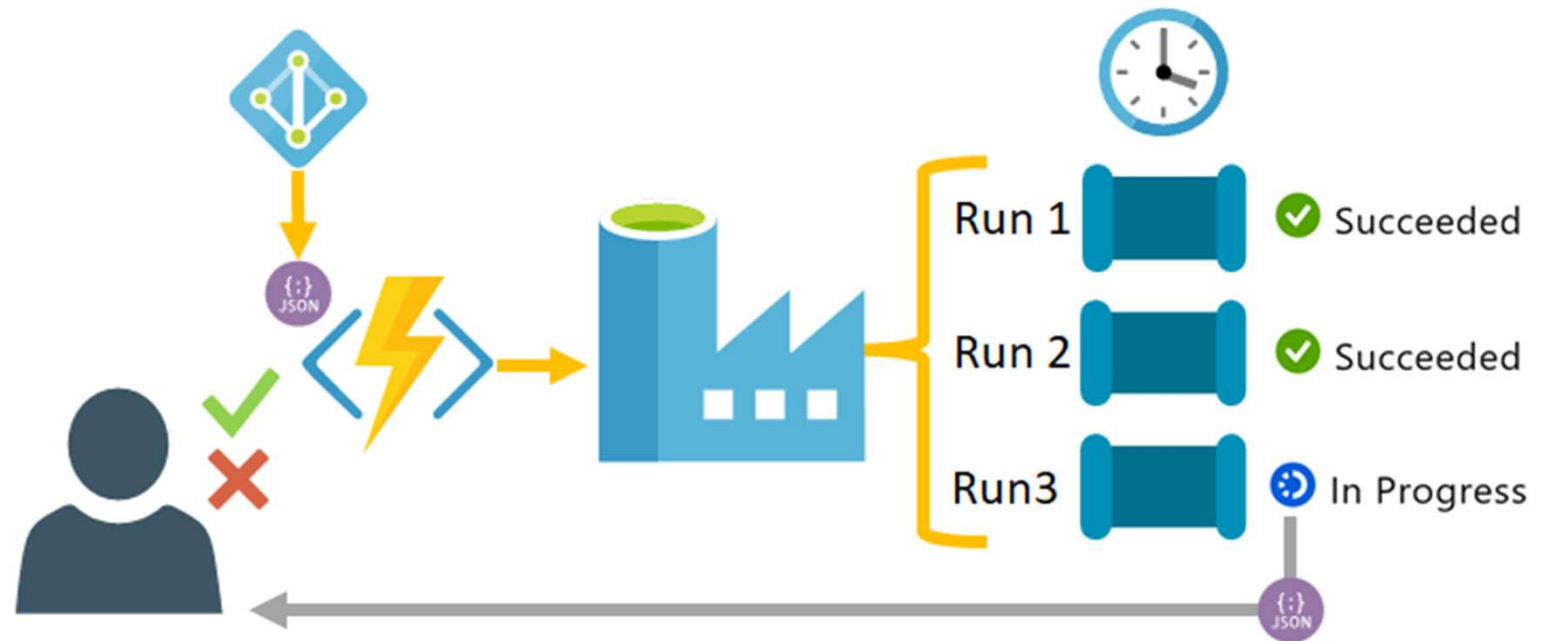
Triggers

- Triggers represent the unit of processing that determines when a pipeline execution needs to be kicked off
- There are different types of triggers for different types of events.



Pipeline runs

- An instance of the pipeline execution
- Pipeline runs are typically instantiated by passing the arguments to the parameters that are defined in pipelines
- The arguments can be passed manually or within the trigger definition.



Hands-On: Create a Pipeline and Run it

- Create a Pipeline and Run it

Thanks