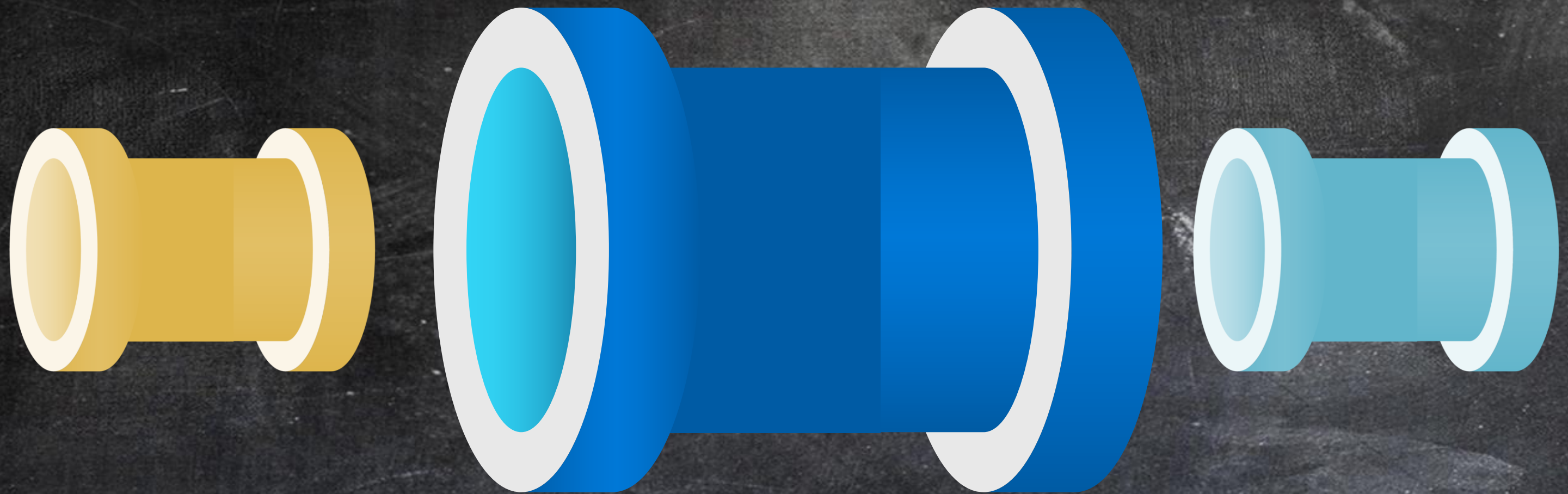


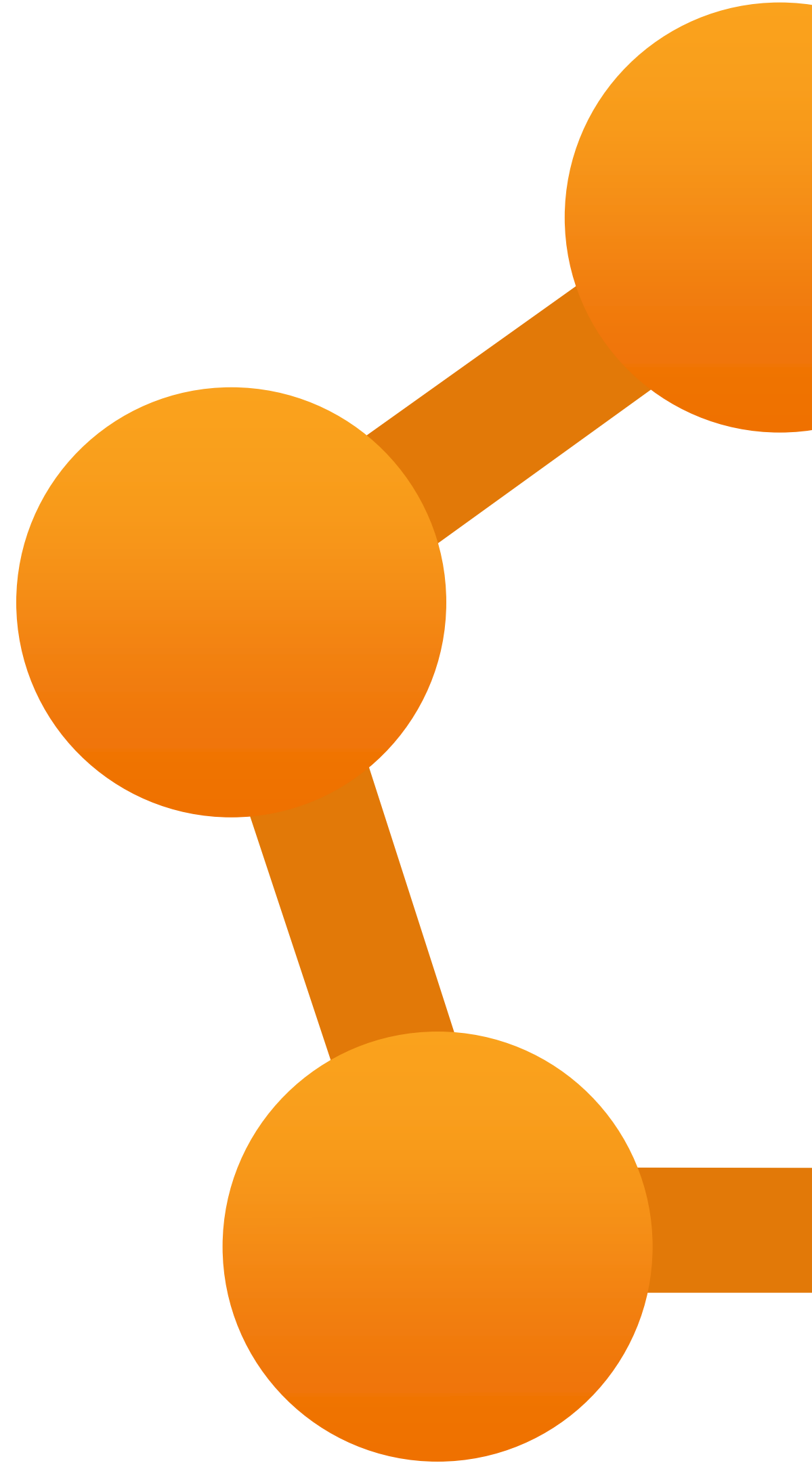
Integration Pipelines



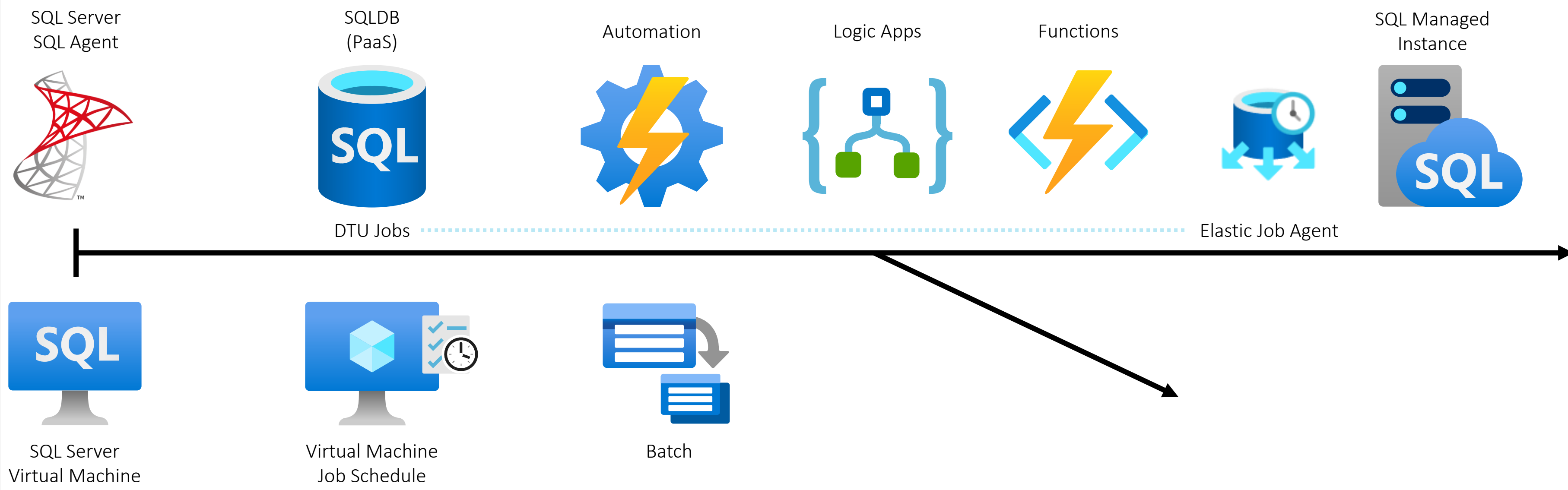
Module 1 – Pipeline Fundamentals

An Evolution of Orchestration
Services

Cloud Formations



A Quick History Lesson



A Quick History Lesson



What is Azure Data Factory (ADF)?



[Home](#) / [Products](#) / [Data Factory](#)

Data Factory

Hybrid data integration service that simplifies ETL at scale

[Start for free >](#)

Already an Azure customer? [Getting started >](#)

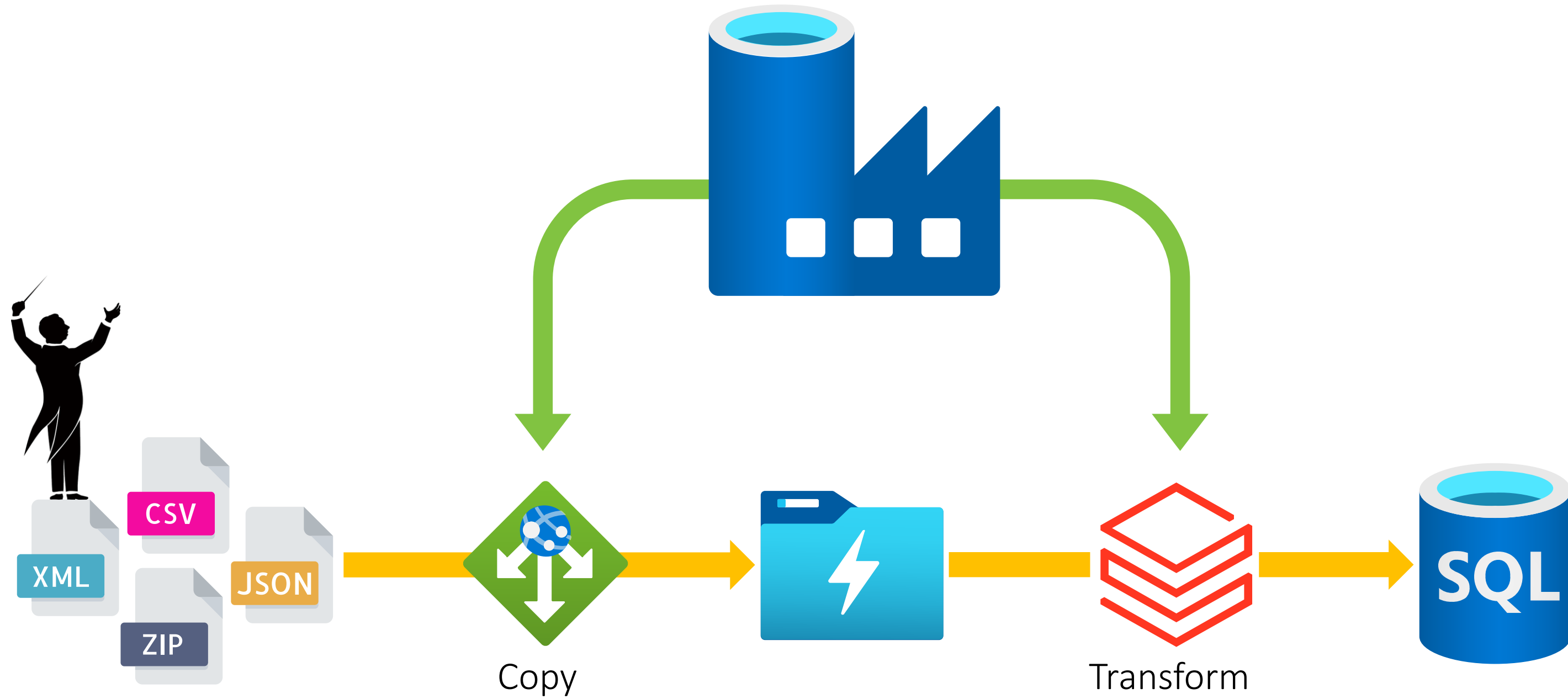
[Product overview](#) [Features](#) [Security](#) [Pricing](#) [Customer stories](#) [Getting started](#) [Documentation](#) [FAQs](#)

Accelerate data integration

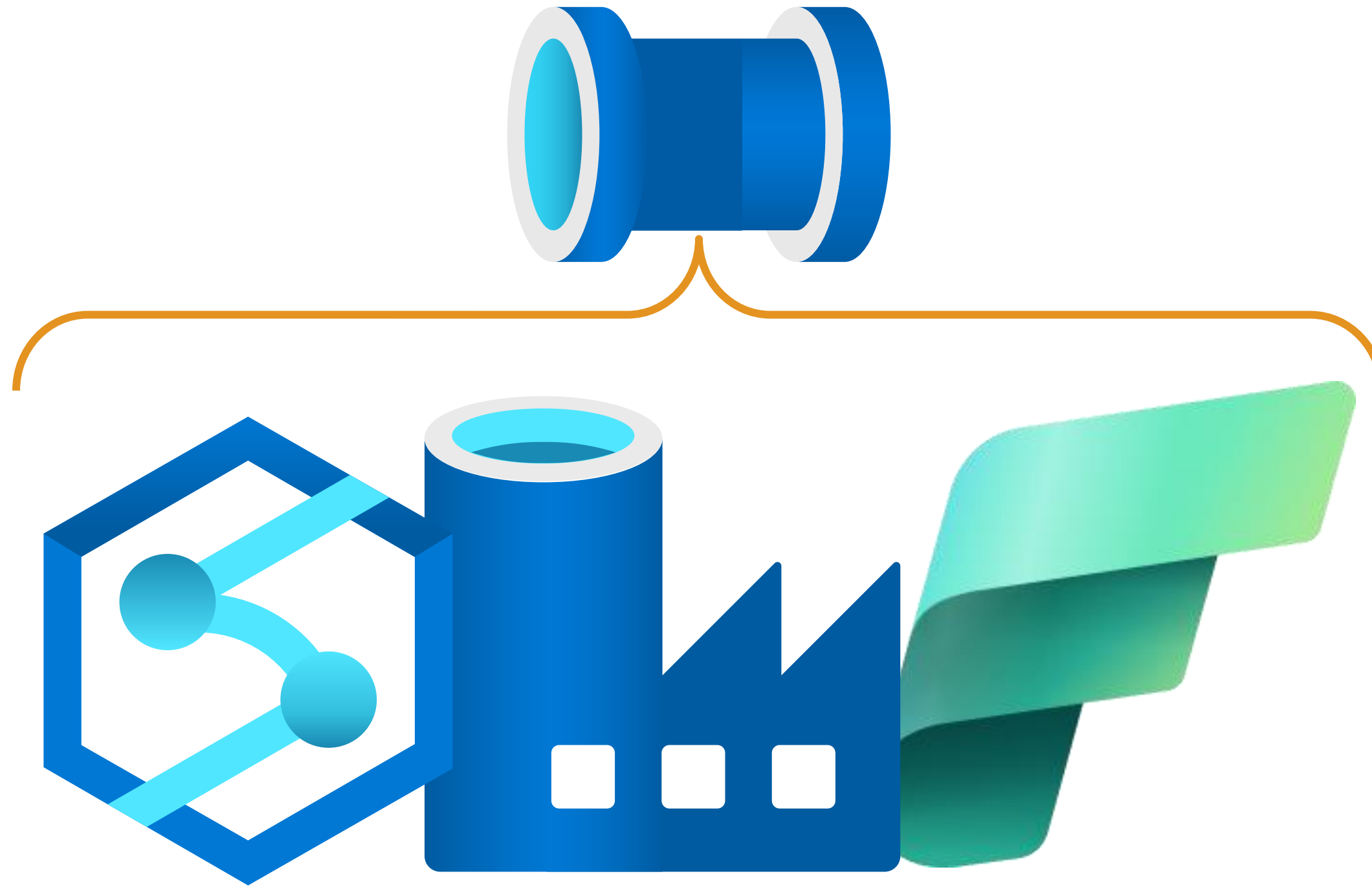
Integrate data silos with Azure Data Factory, a service built for all data integration needs and skill levels. Easily construct ETL and ELT processes code-free within the intuitive visual environment, or write your own code. Visually integrate data sources using more than 90+ natively built and maintenance-free connectors at no added cost. Focus on your data – the serverless integration service does the rest.



What is Azure Data Factory (ADF)?



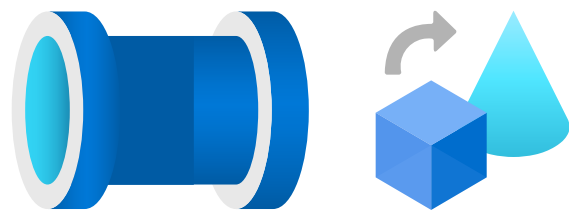
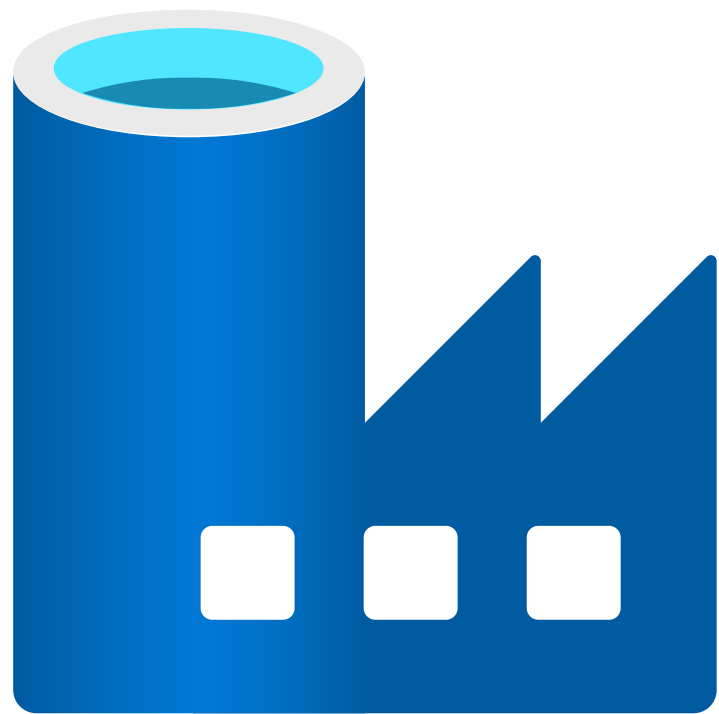
What are Integration Pipelines?



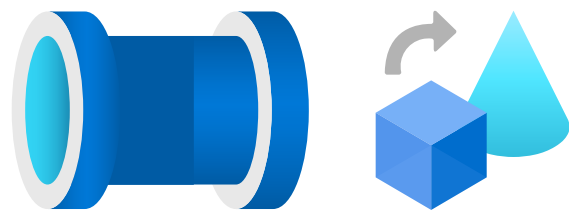
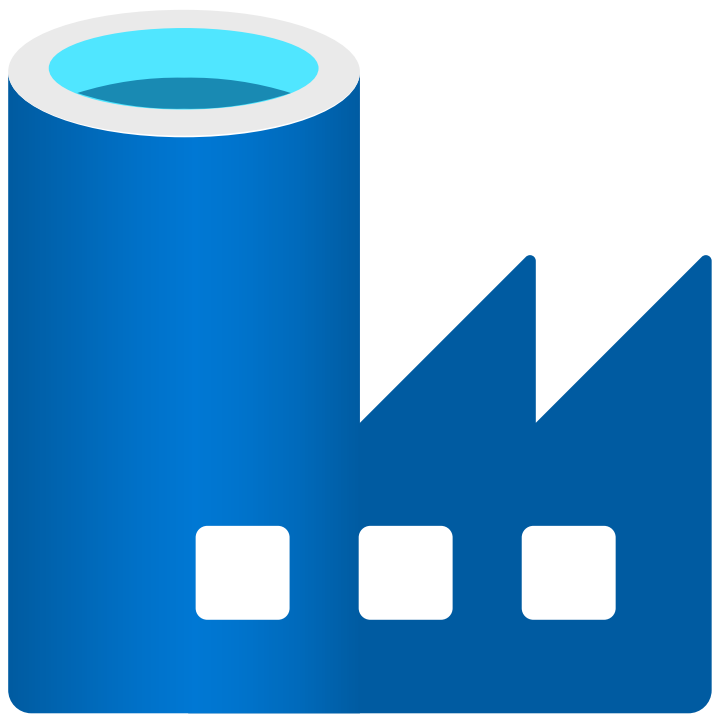
Synapse Analytics vs Data Factory



<https://docs.microsoft.com/en-us/azure/synapse-analytics/data-integration/concepts-data-factory-differences>



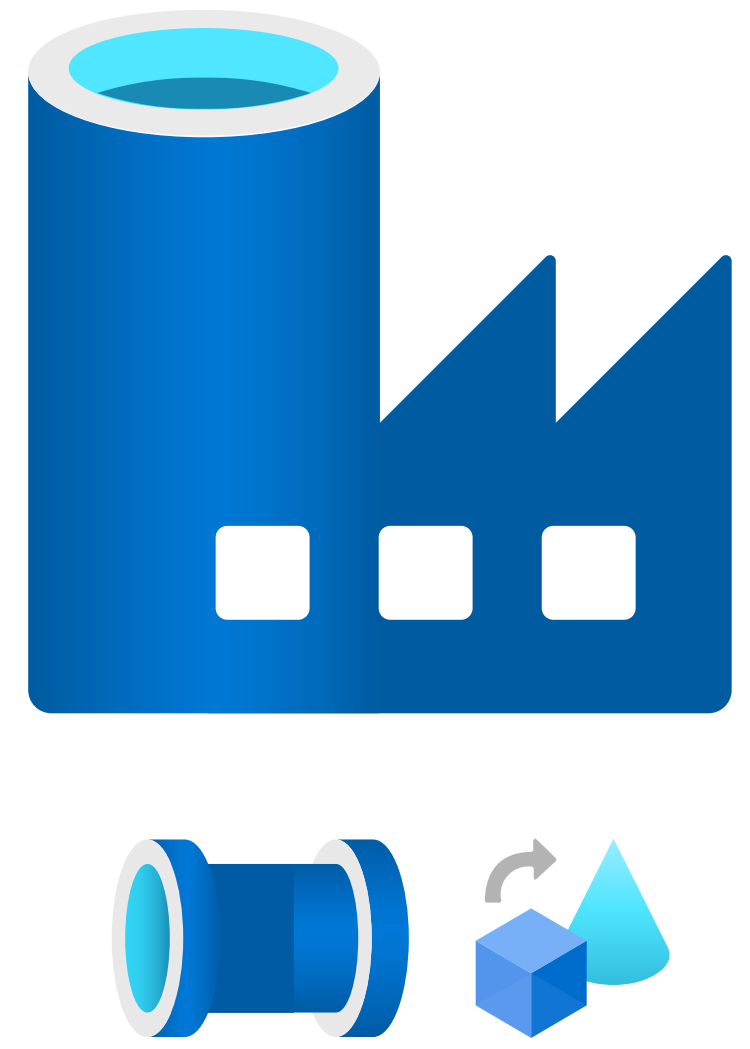
Synapse Analytics vs Data Factory



Microsoft Fabric vs Data Factory



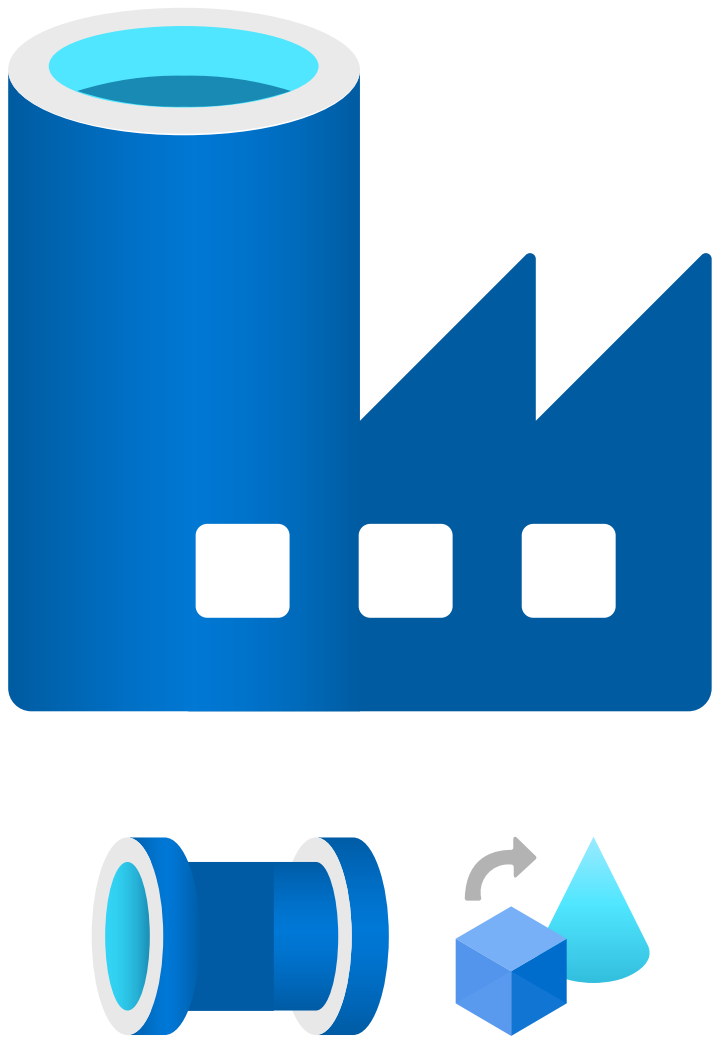
<https://mrpaulandrew.com/2023/05/31/what-is-microsoft-fabric-my-point-of-view/>



Microsoft Fabric vs Data Factory



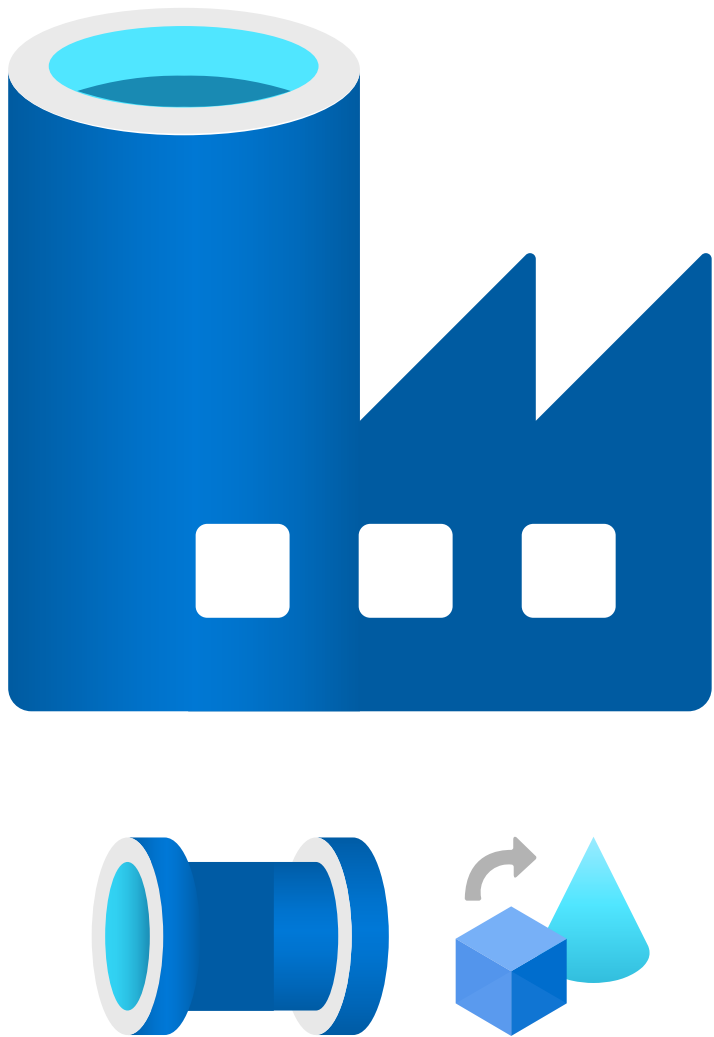
<https://mrpaulandrew.com/2023/05/31/what-is-microsoft-fabric-my-point-of-view/>



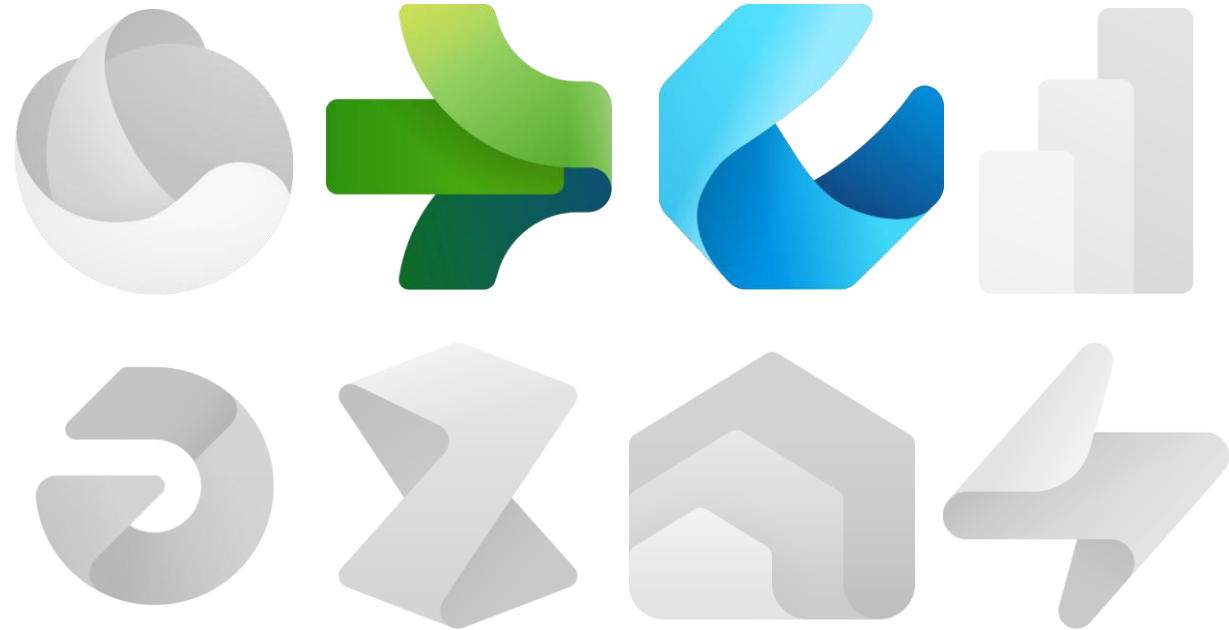
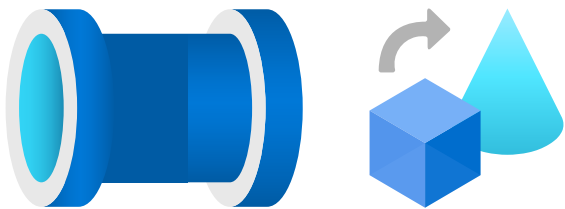
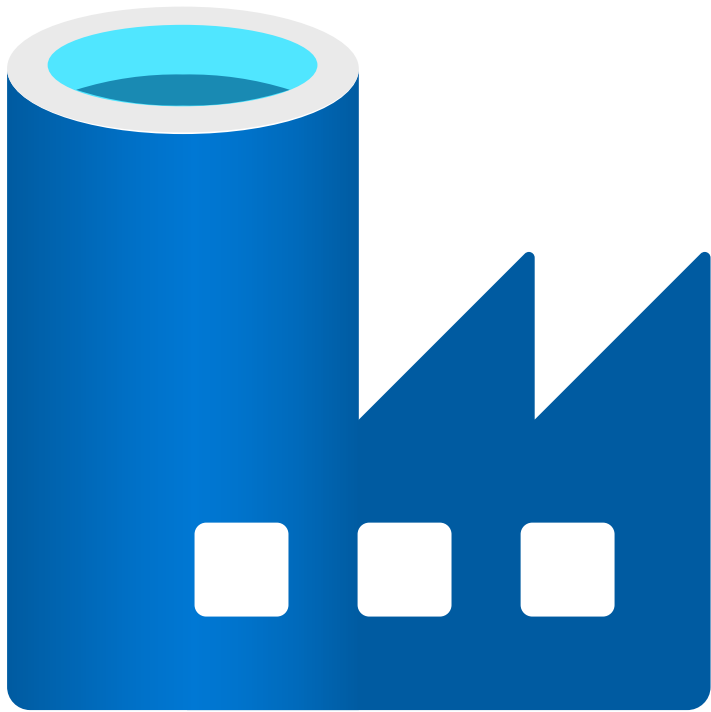
Microsoft Fabric vs Data Factory



<https://mrpaulandrew.com/2023/05/31/what-is-microsoft-fabric-my-point-of-view/>



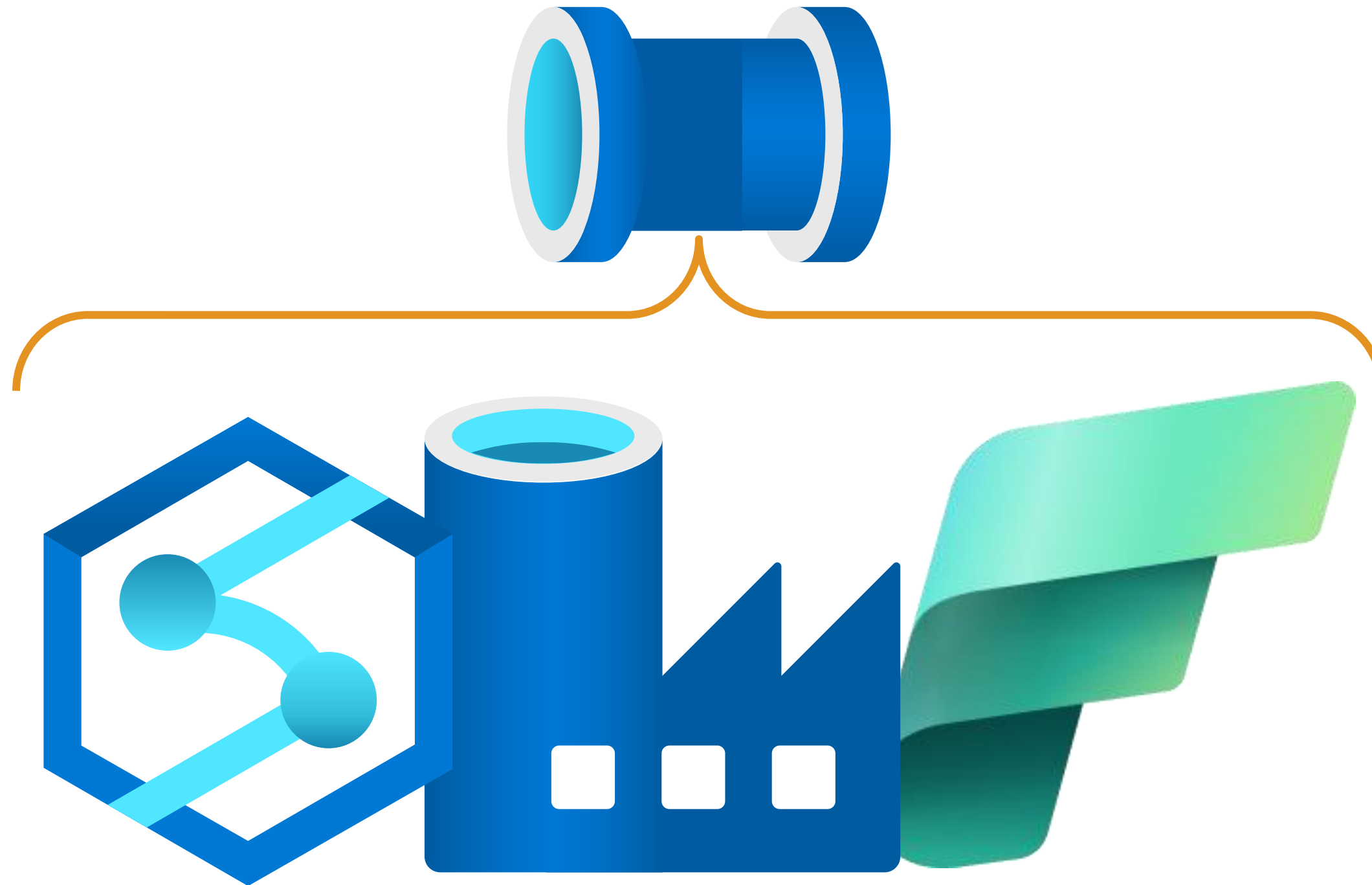
Data Factory vs Synapse Analytics vs Microsoft Fabric



What are Integration Pipelines?



Answer: A data integration pipeline is a logical package of work that can be used move, transform and orchestrate data tasks from and to various compute and storage resources.

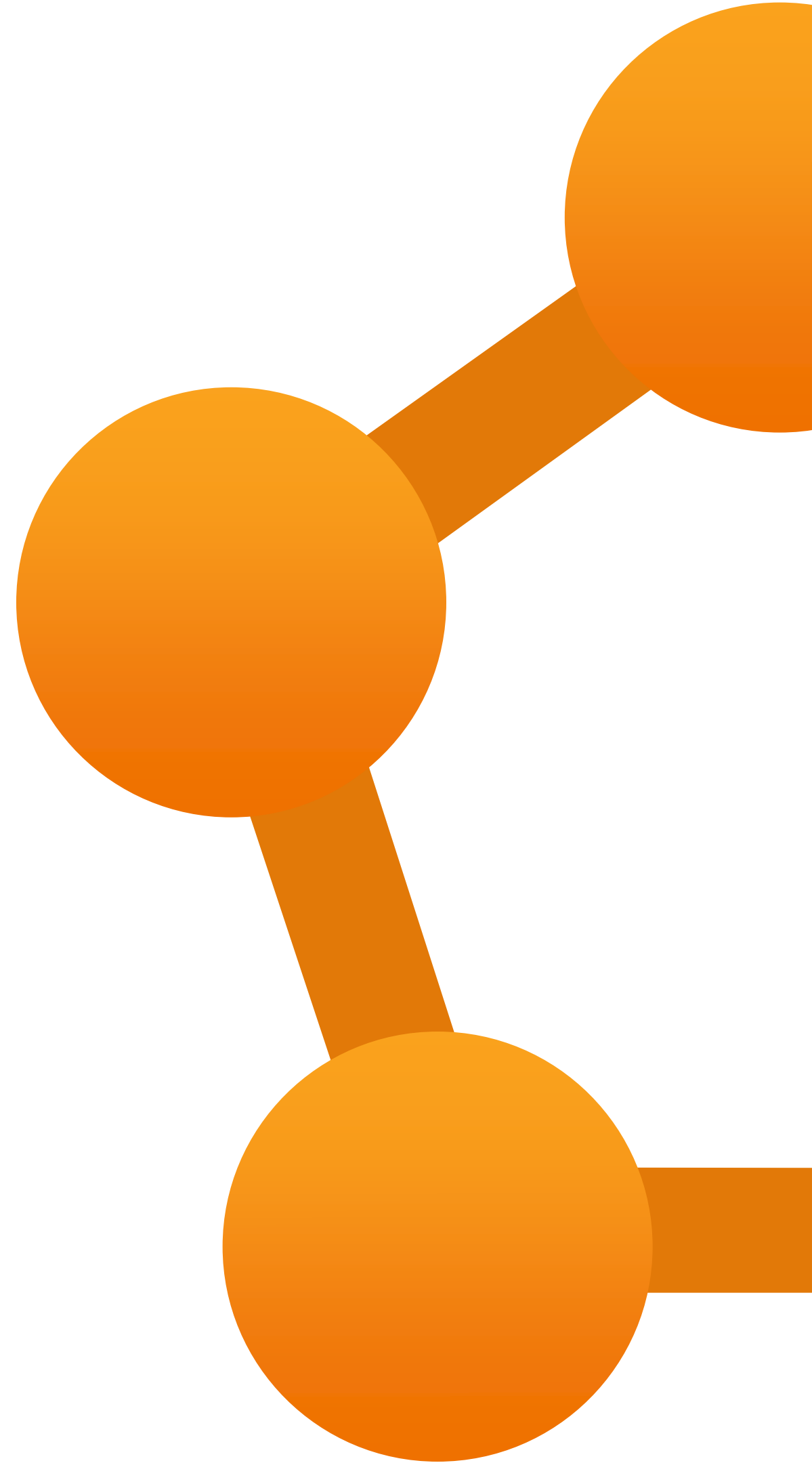


In the context of Azure Data Factory, Azure Synapse Analytics, and Microsoft Fabric, integration pipelines represent a common component or tool across products for engineering teams to utilise in the delivery of a data platform solution.

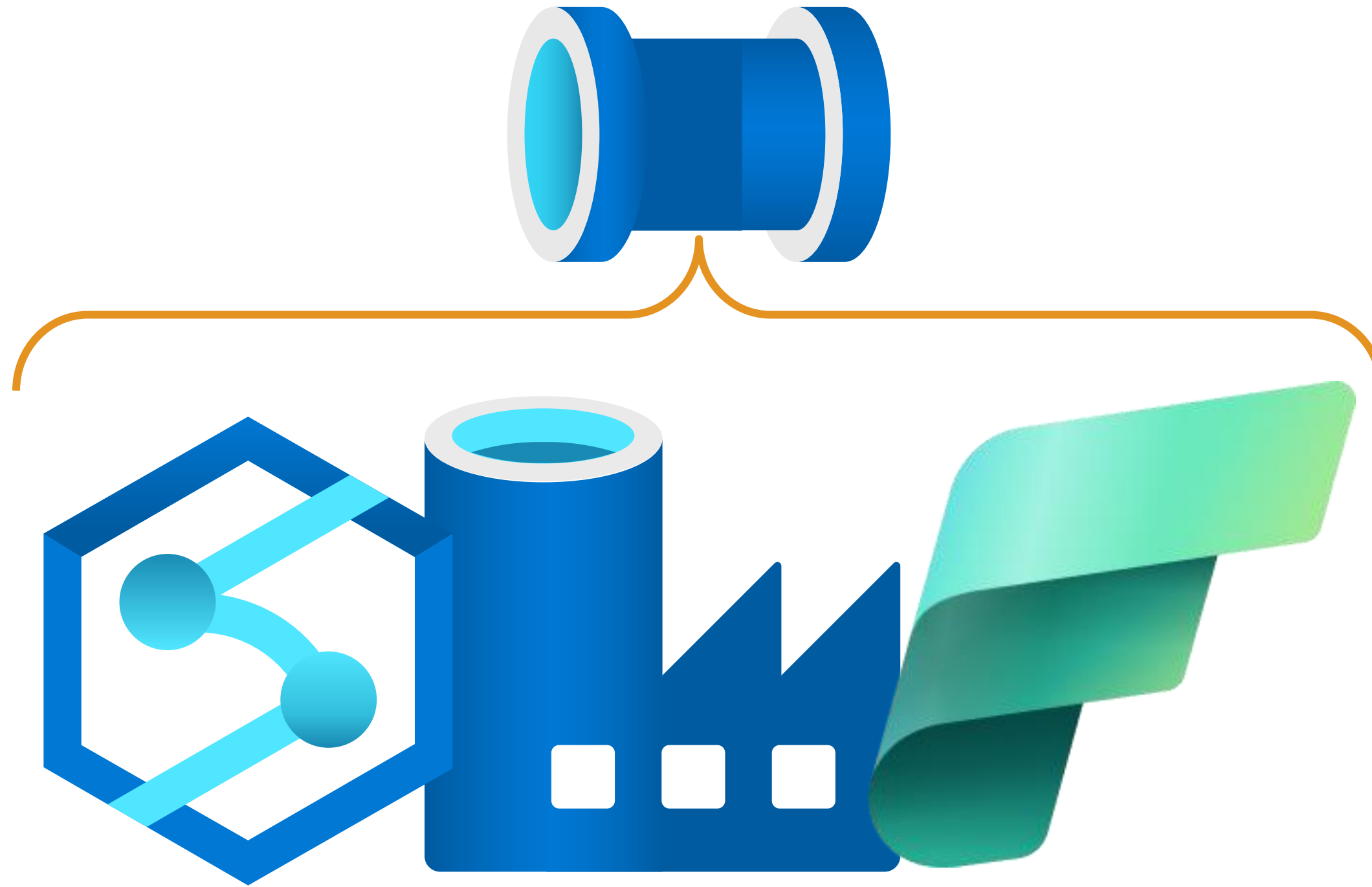
Module 1 – Pipeline Fundamentals

Core Components

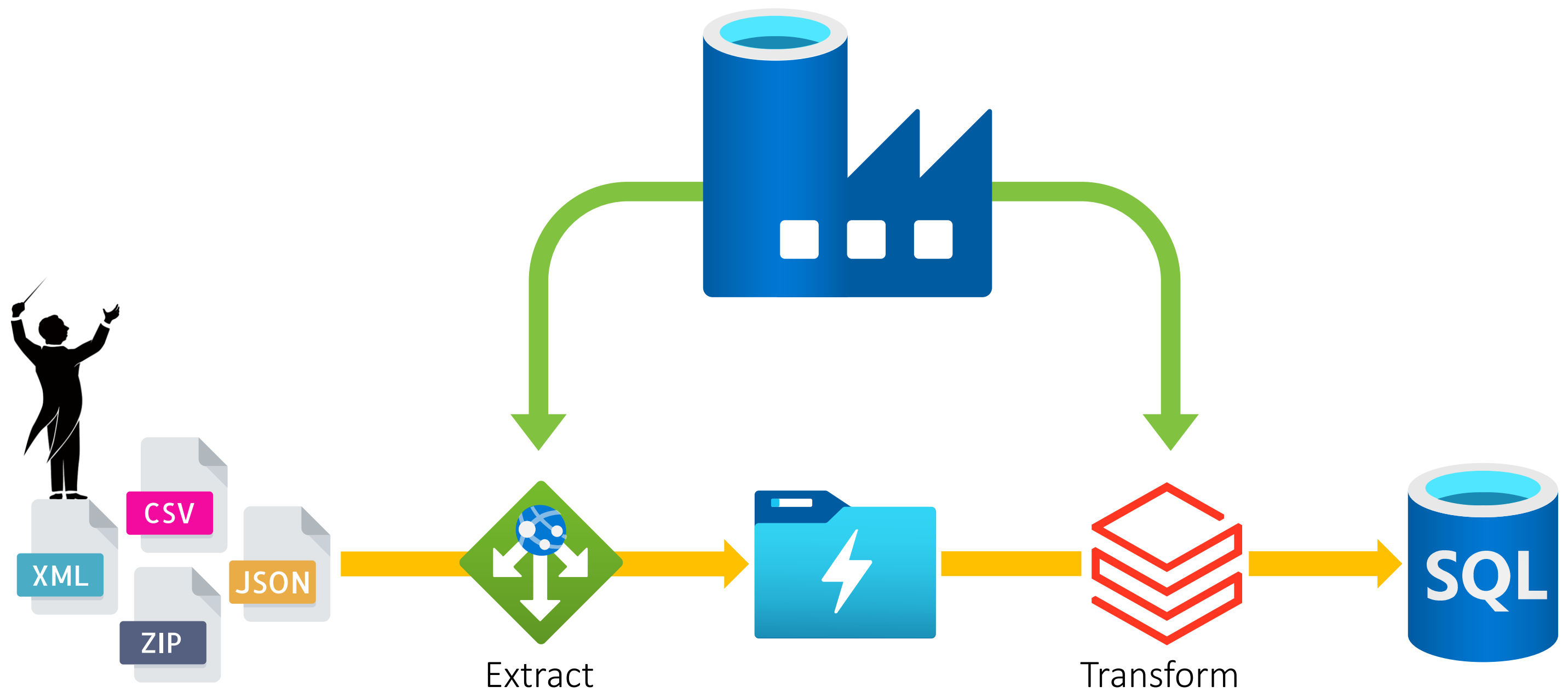
Cloud Formations



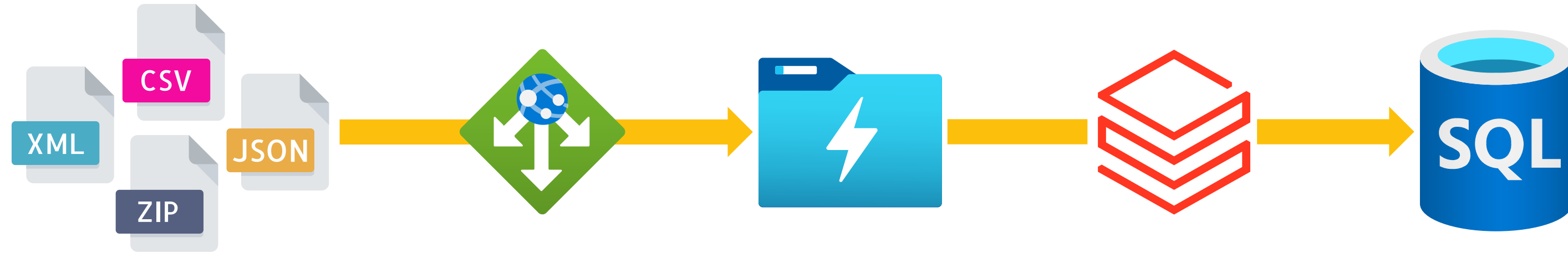
Data Factory Core Components



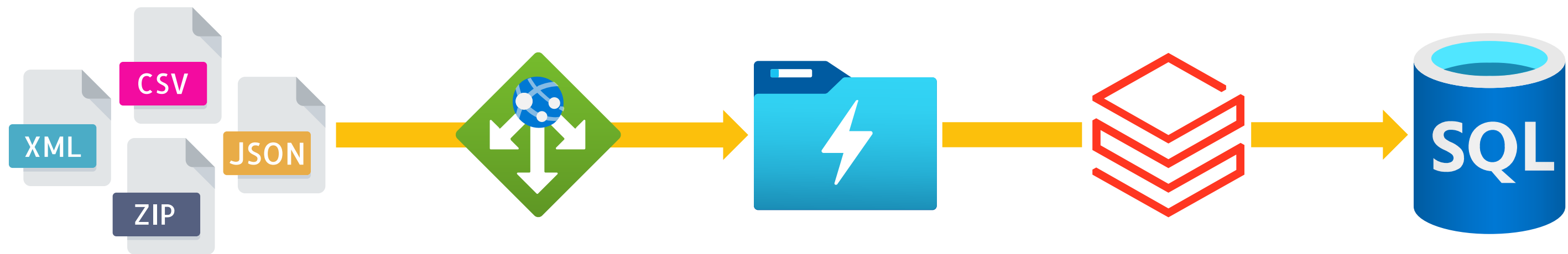
Data Factory Core Components



Data Factory Core Components

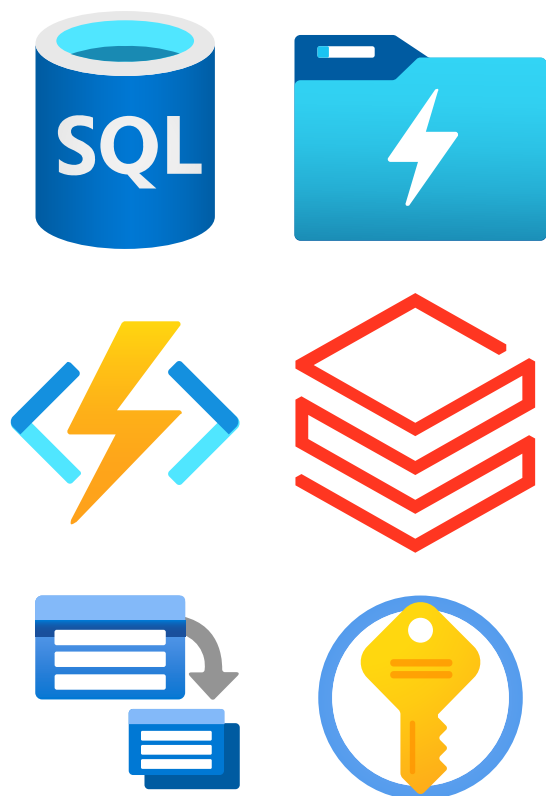


Data Factory Core Components



1

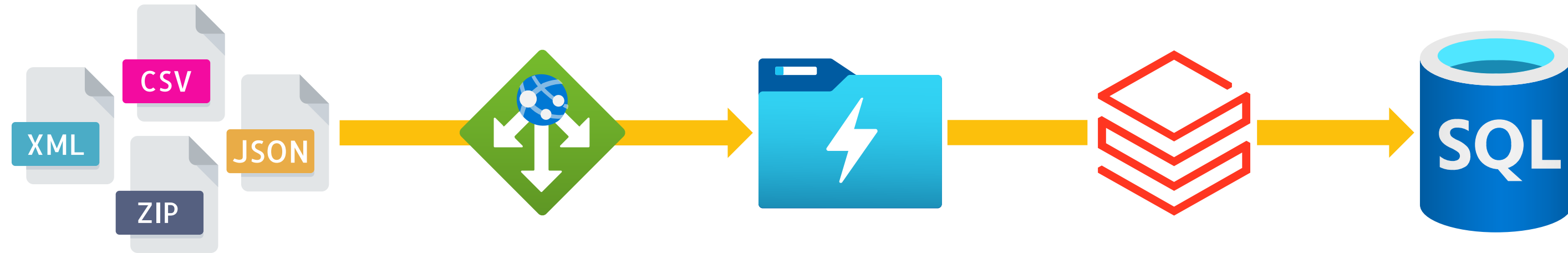
Linked Services – What to interact with and how?



SQLDBLinkedService

ConnectionString: *Server=MyServer;Database=myDataBase*
UserName: *"MrPaulAndrew"*
Password: *******

Data Factory Core Components

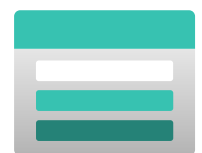


1 Linked Services

2 Datasets – Where is my data? What format? What file path/table do I need?

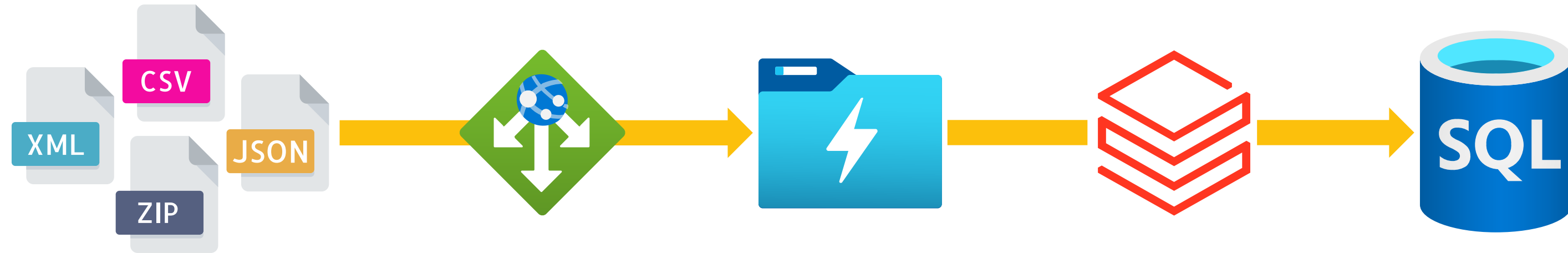


[dbo].[SalesOrders]



/RAW/Orders/2018/01/01/SalesOrders.csv

Data Factory Core Components



1 Linked Services

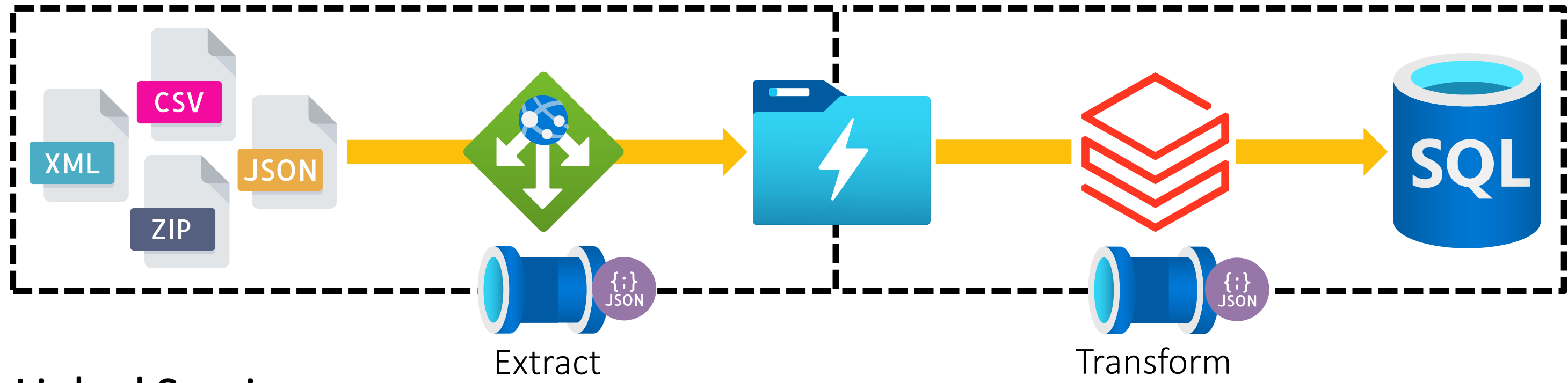
2 Datasets

3 **Activities** – What do we want to happen when we invoke a Linked Service?
With what conditions?

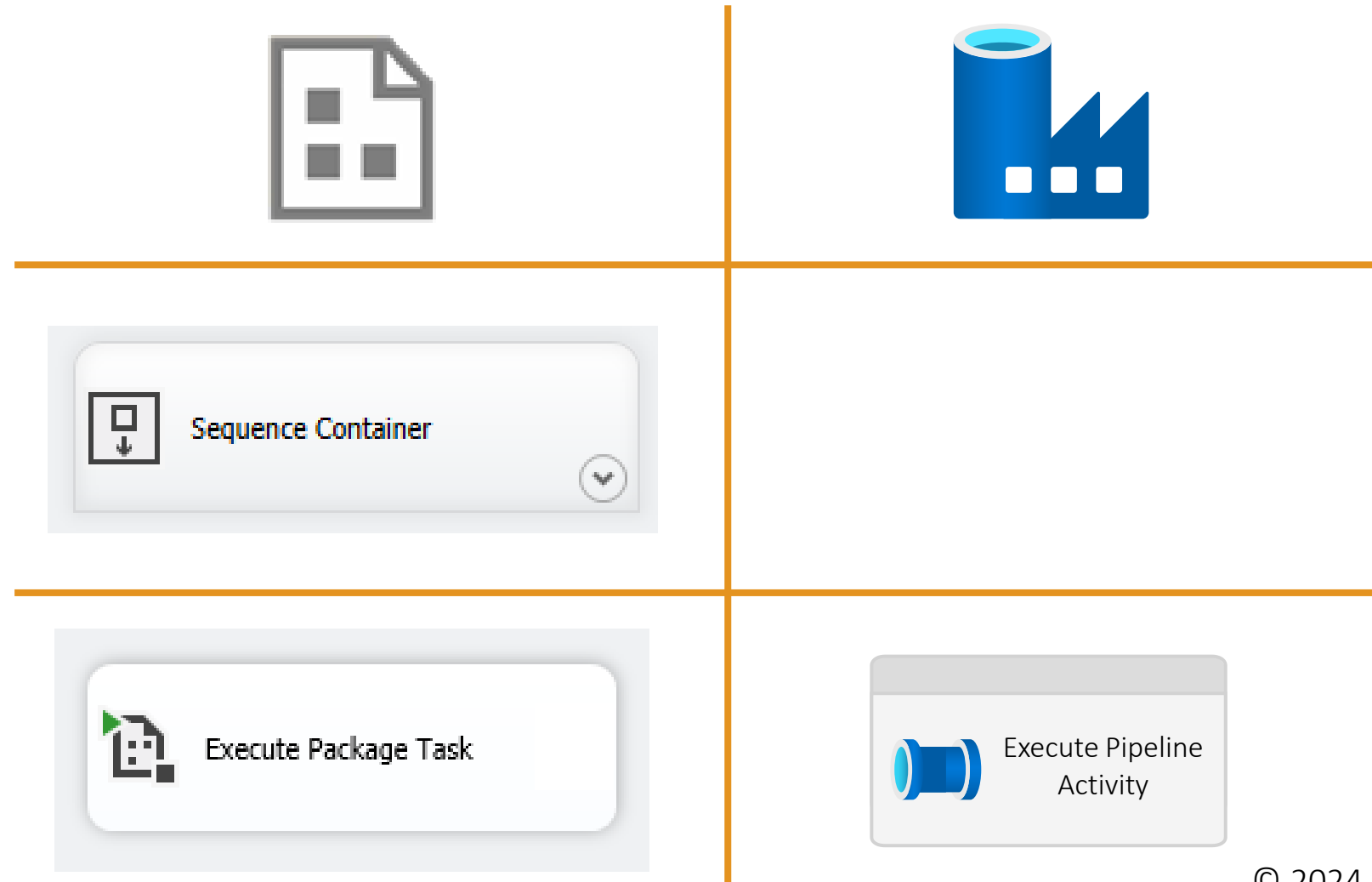
Databricks Notebook Activity

```
notebookPath: /Playground/Playing
baseParameters: Testing
libraries[jar]: dbfs:/lib1.jar
linkedServiceName: BricksOfData01
```

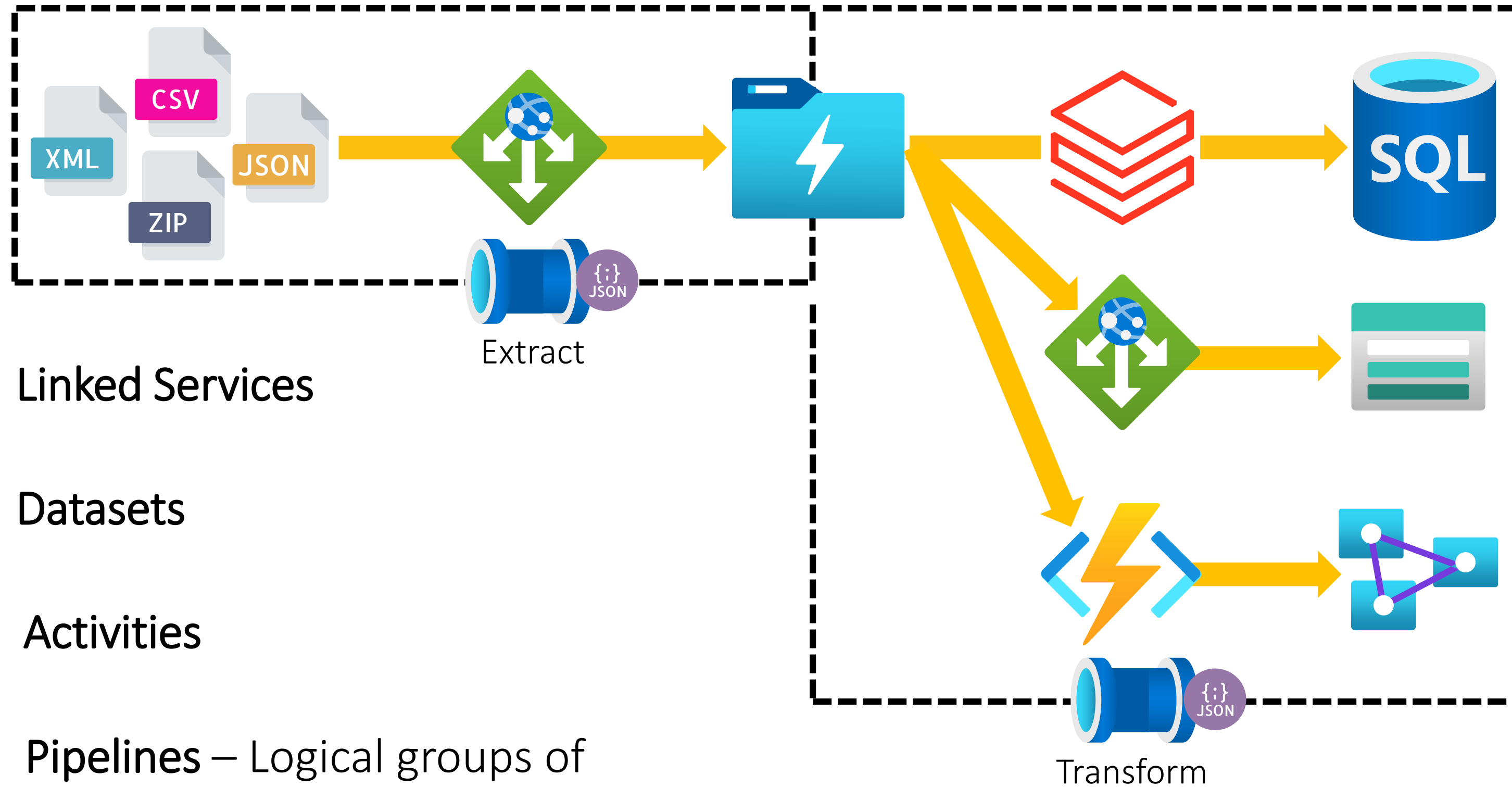
Data Factory Core Components



- 1 Linked Services
- 2 Datasets
- 3 Activities
- 4 Pipelines – Logical groups of work that can be executed.



Data Factory Core Components



1

Linked Services

2

Datasets

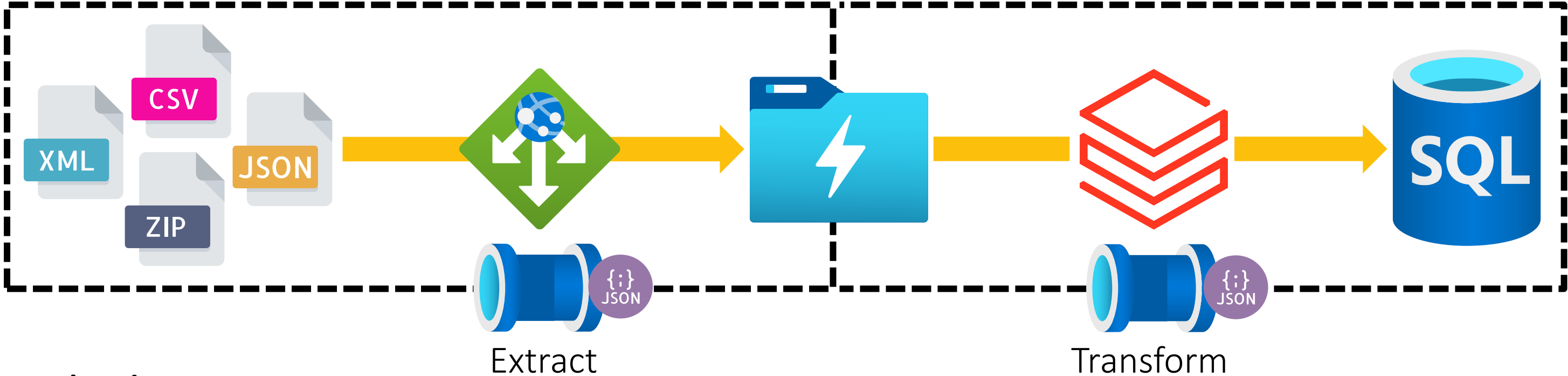
3

Activities

4

Pipelines – Logical groups of work that can be executed.

Data Factory Core Components



1

Linked Services

2

Datasets

3

Activities

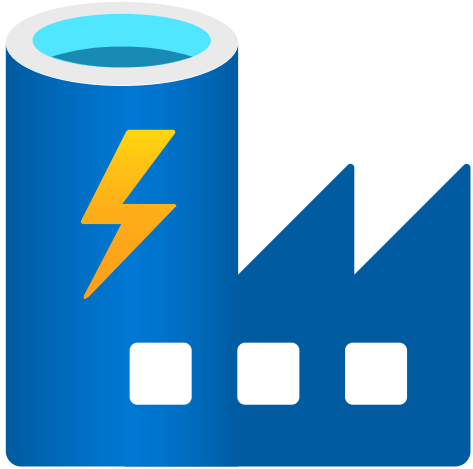
4

Pipelines

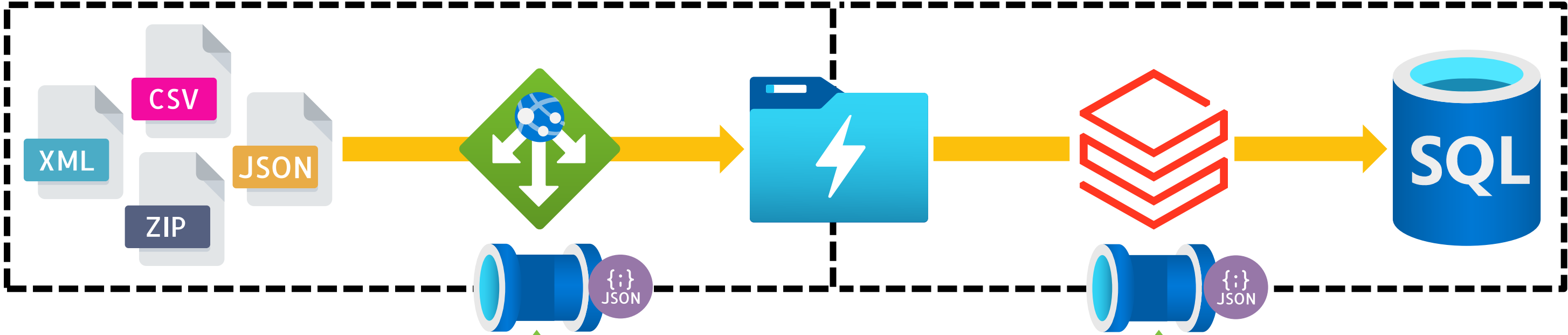
5

Triggers – Telling our when pipelines to run.

- Manually
- Programmatically
- Schedule
- Tumbling Windows
- Storage Events
- Custom Events



Data Factory Core Components

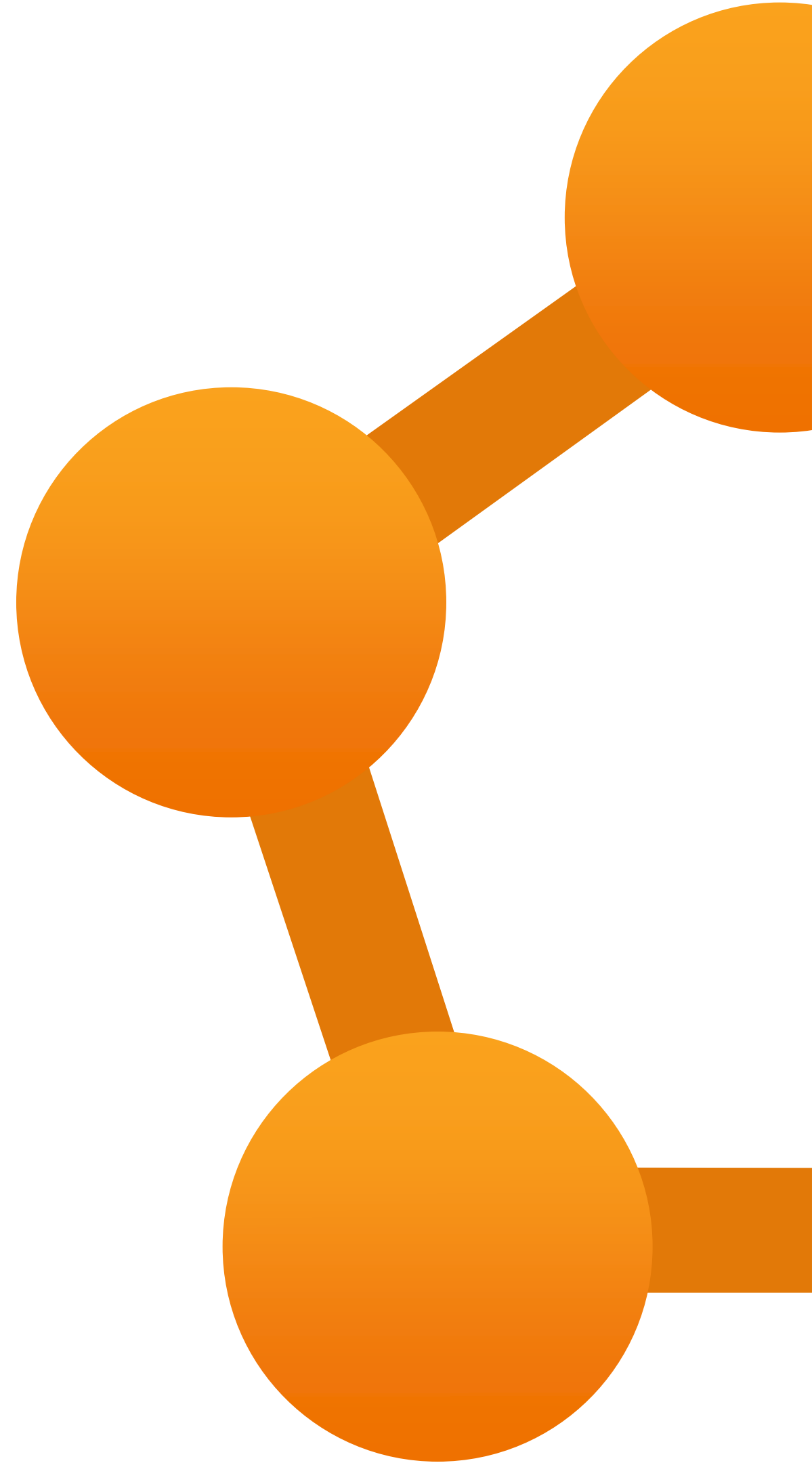


- 1 Linked Services
- 2 Datasets
- 3 Activities
- 4 Pipelines
- 5 Triggers

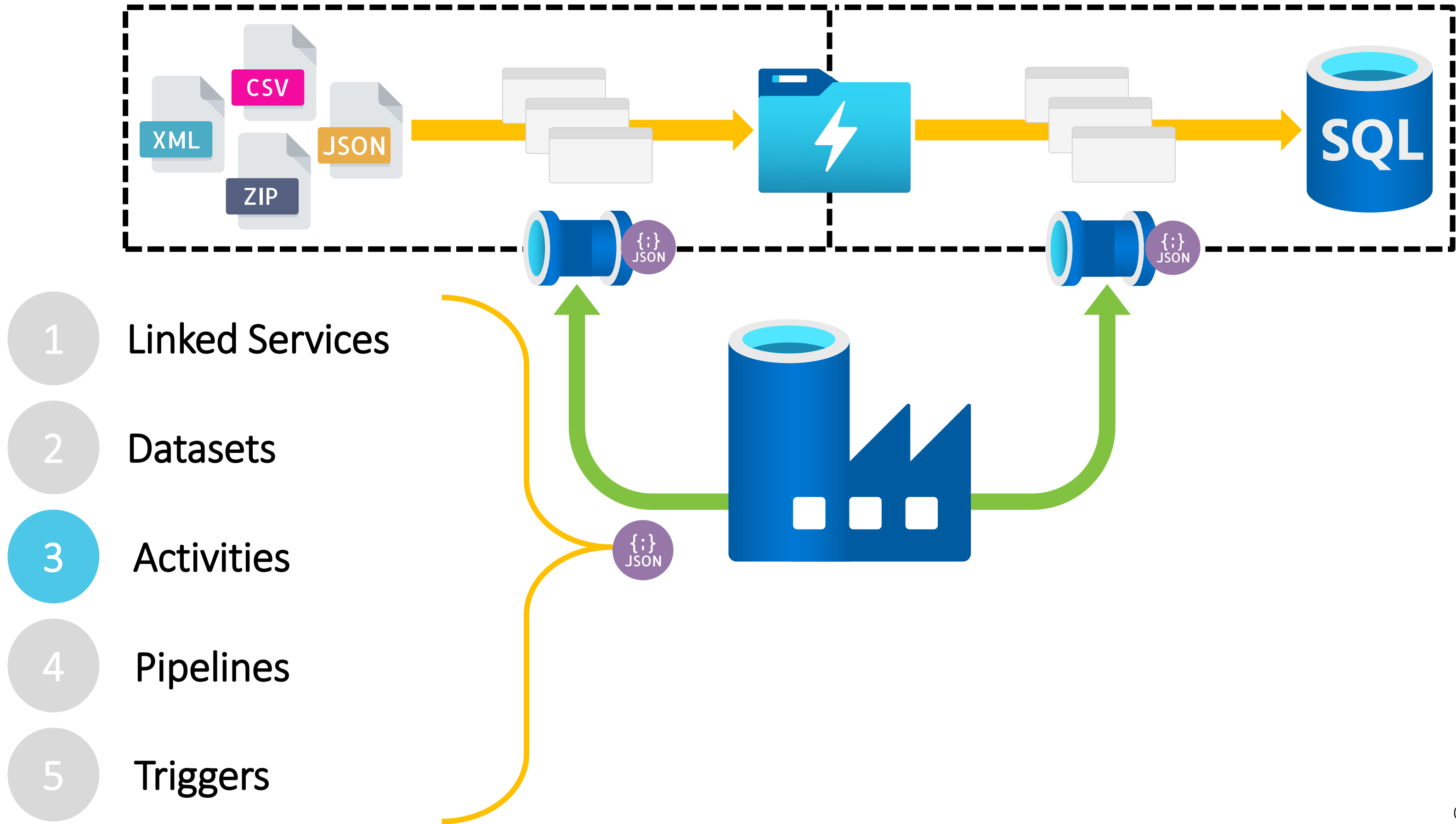
Module 1 – Pipeline Fundamentals

Common Activities

Cloud Formations



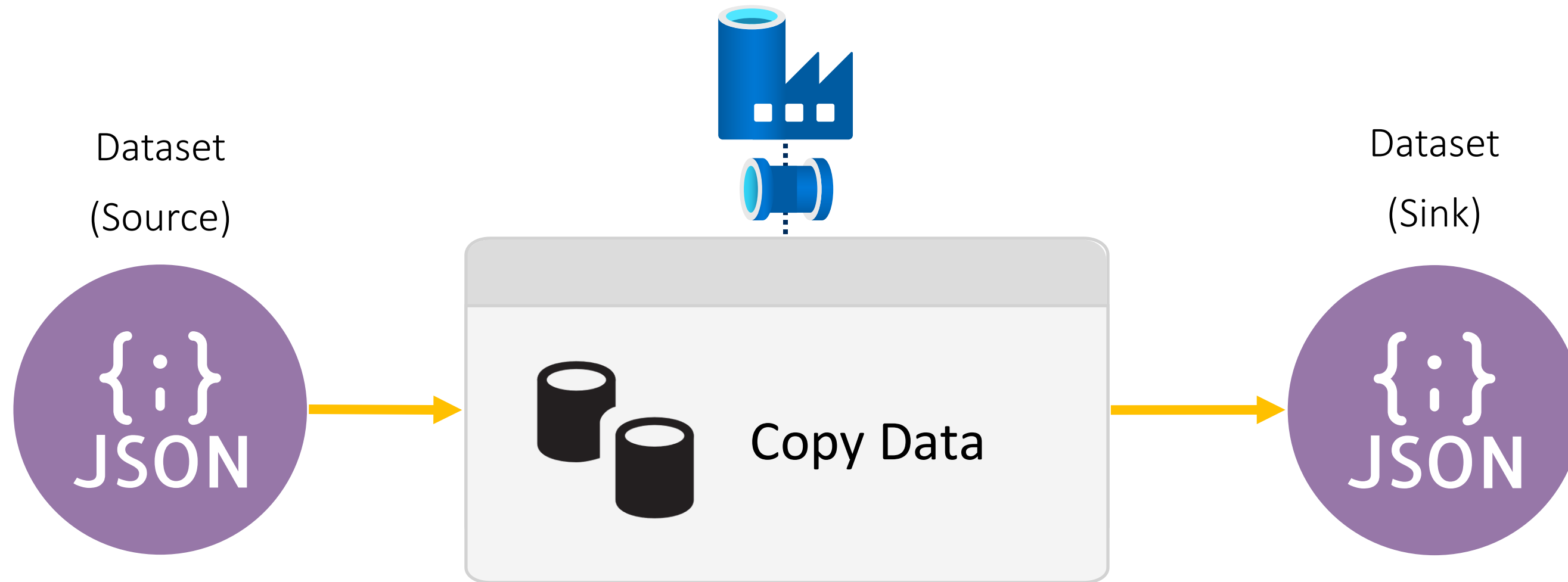
Data Factory Common Activities





Copy

Getting your data from A to B (not a Move operation)

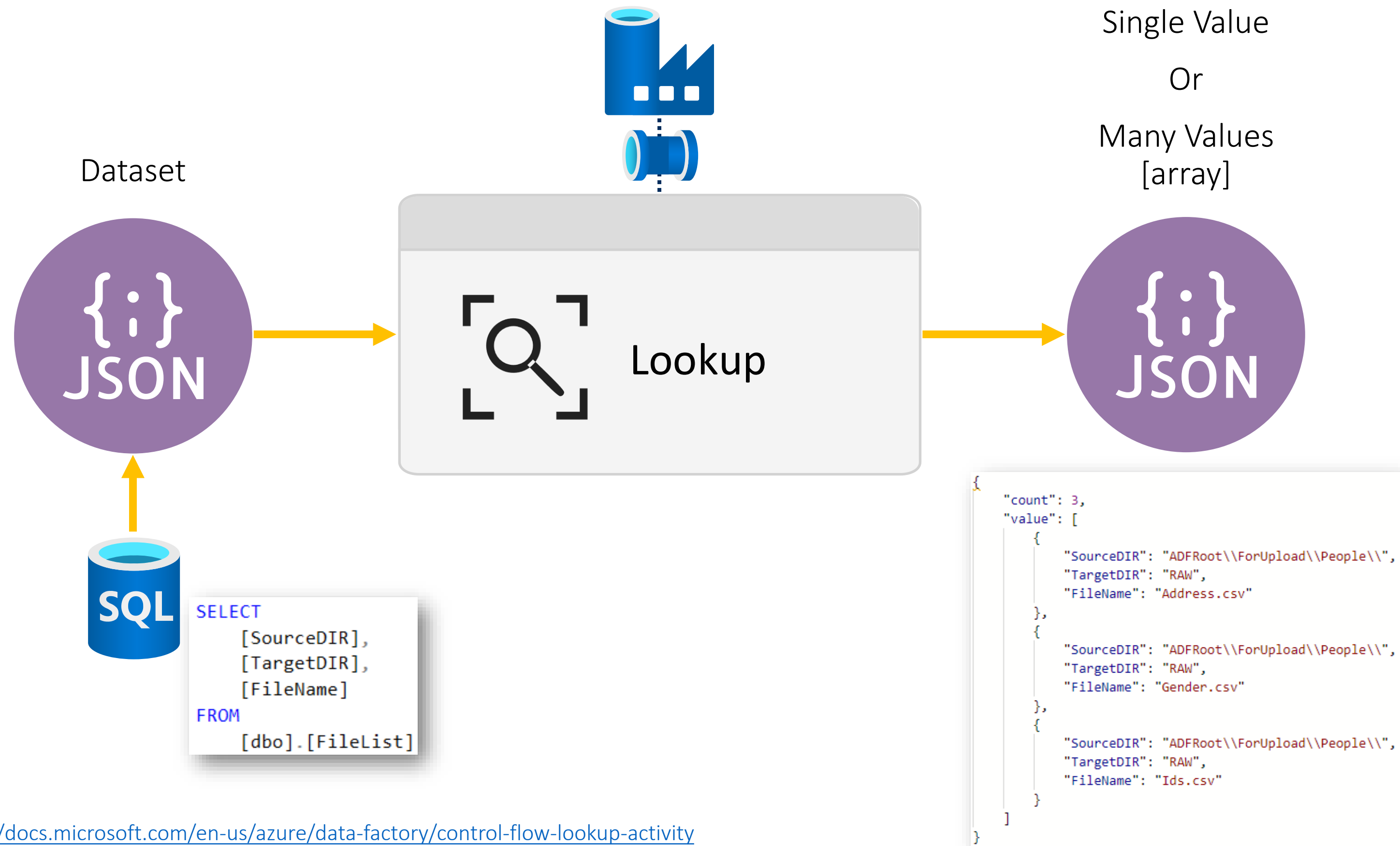


- ☐ Auto Scaling
- ☐ Transactional Restarts
- ☐ Handle Zip Compression
- ☐ Attribute Mapping and Schema Drift
- ☐ Handle Failed Rows
- ☐ Add Custom Attributes
- ☐ Parse Excel & JSON Files

Lookup



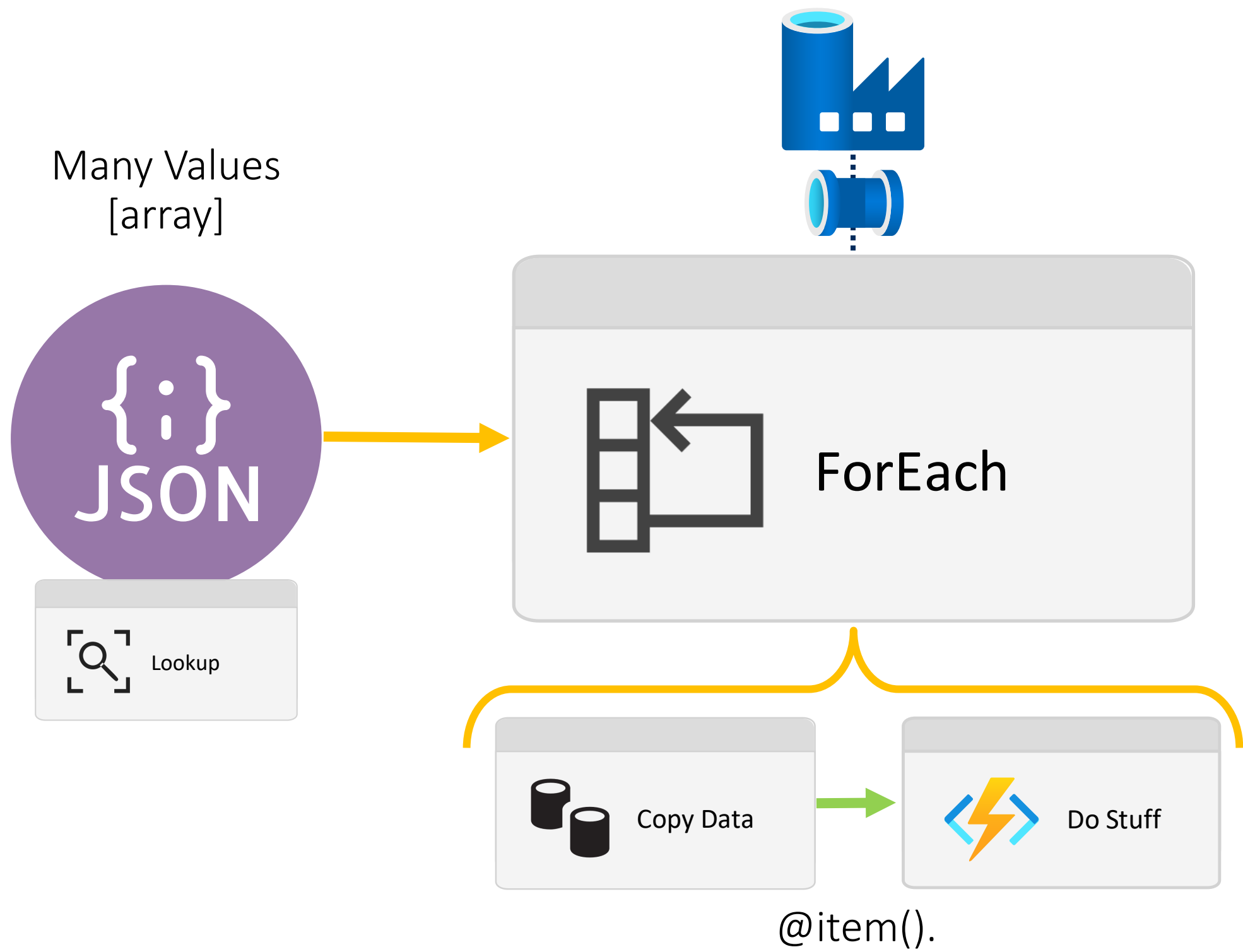
Get value(s) to support other control flow activities



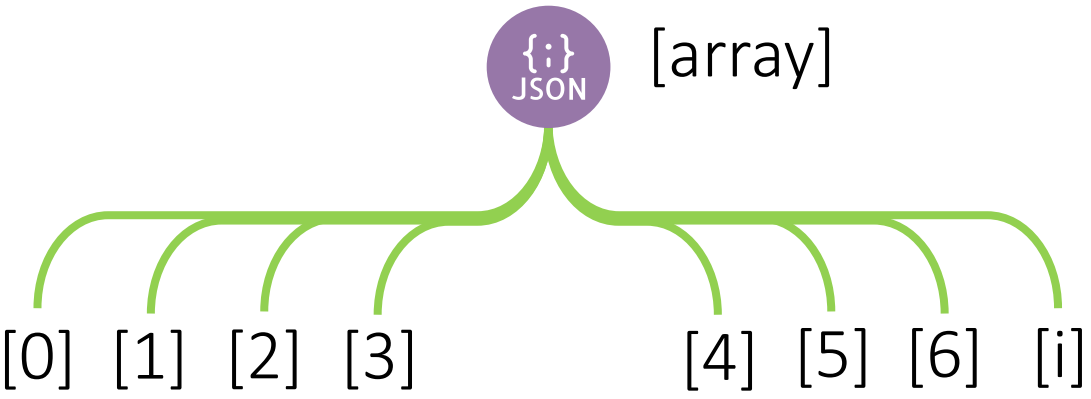
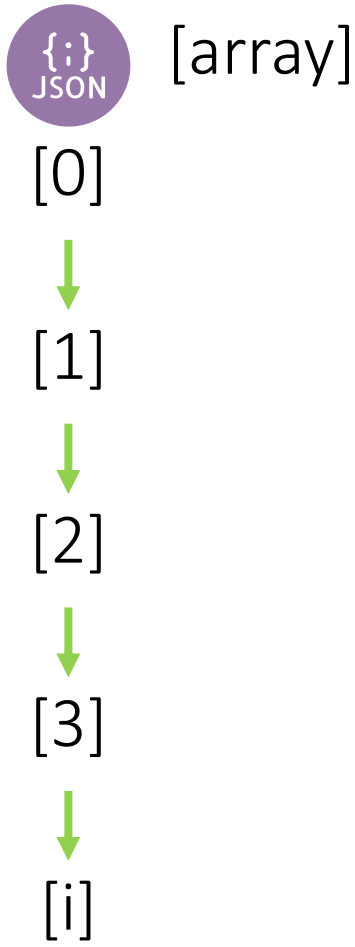
For Each



Iterating over other control flow activities



IsSequential:
true



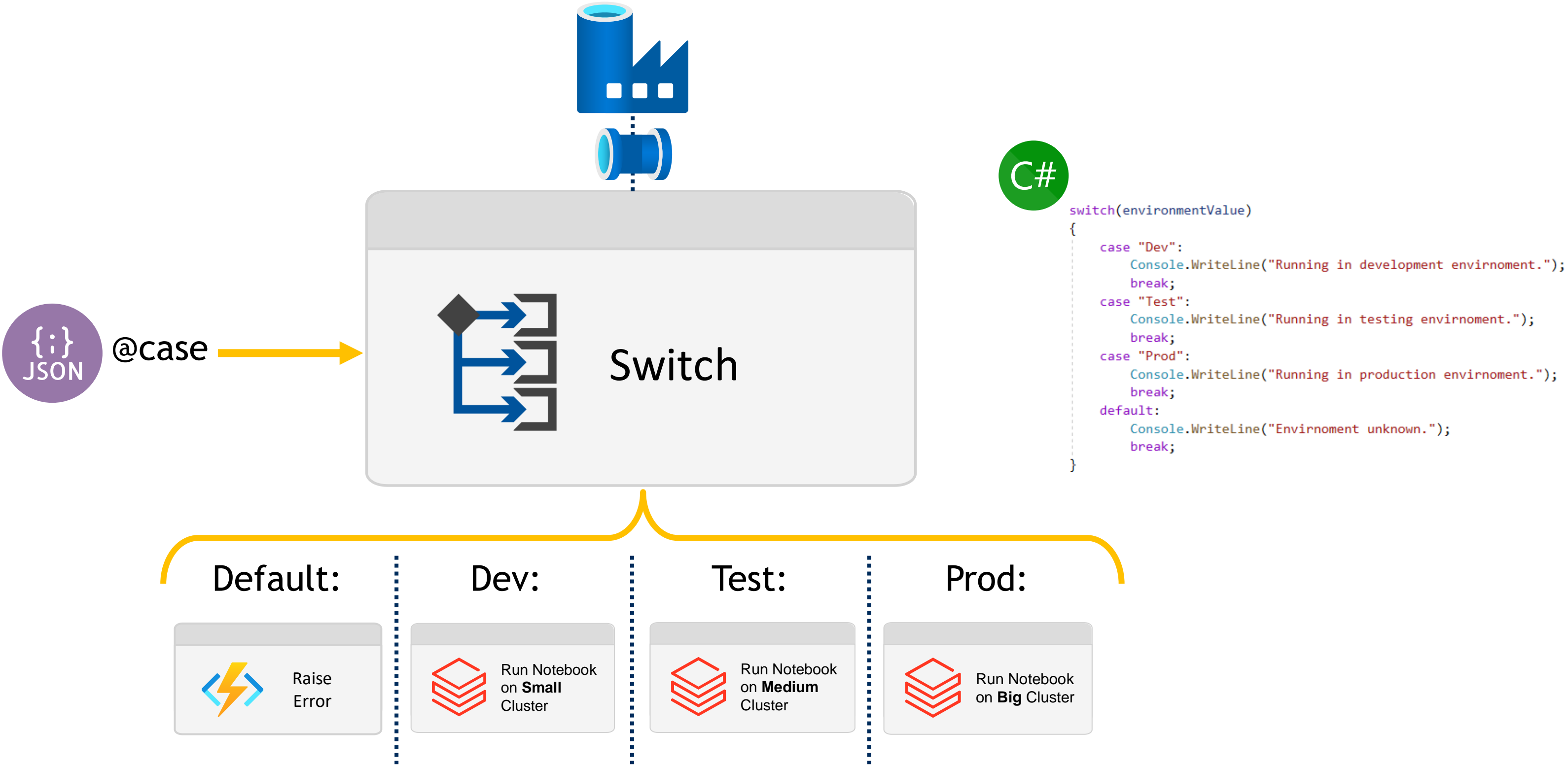
Batch Count Default: 20

Batch Count Max: 50

Switch



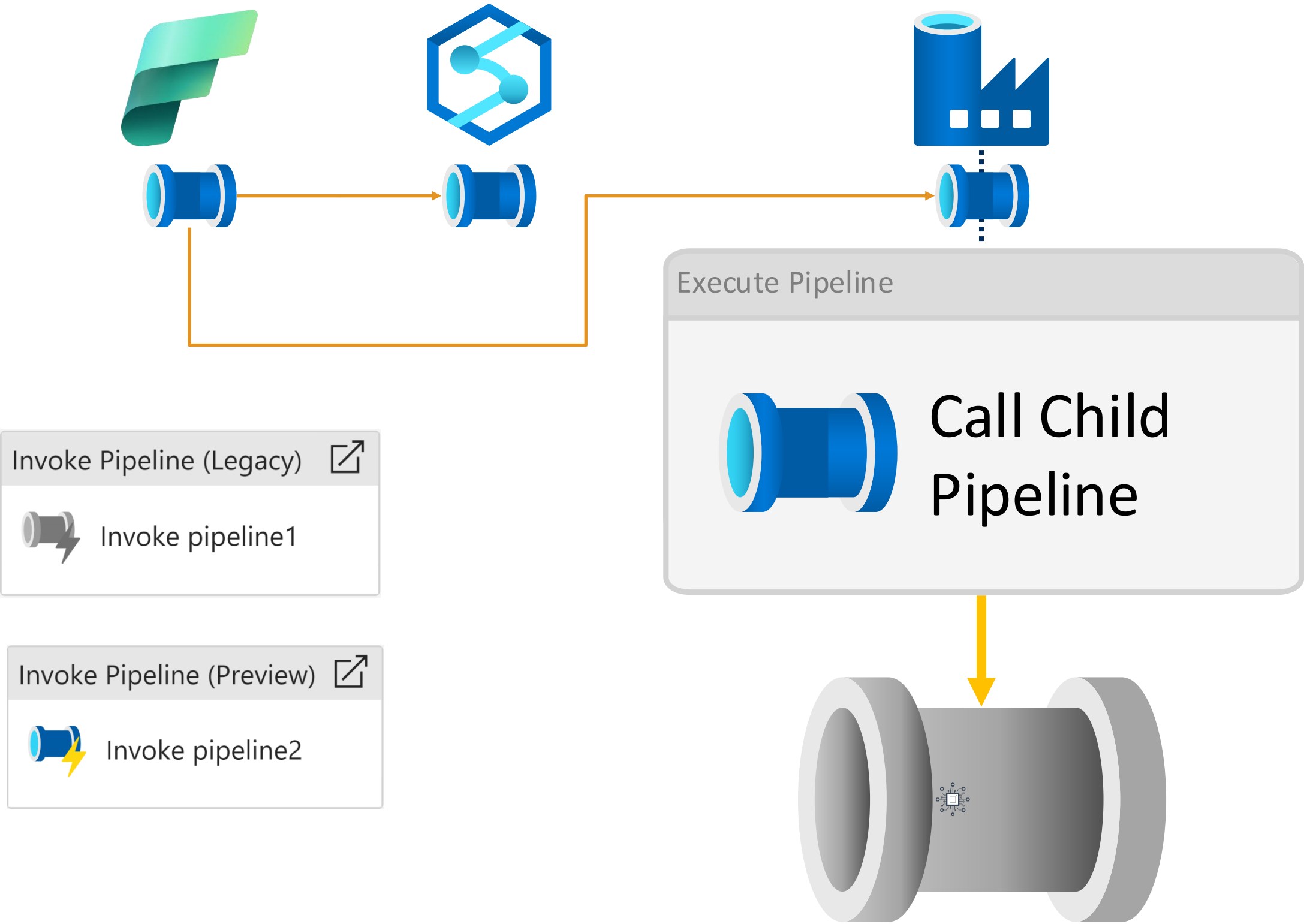
Execute other control flow components based on a provided condition



Execute Pipeline



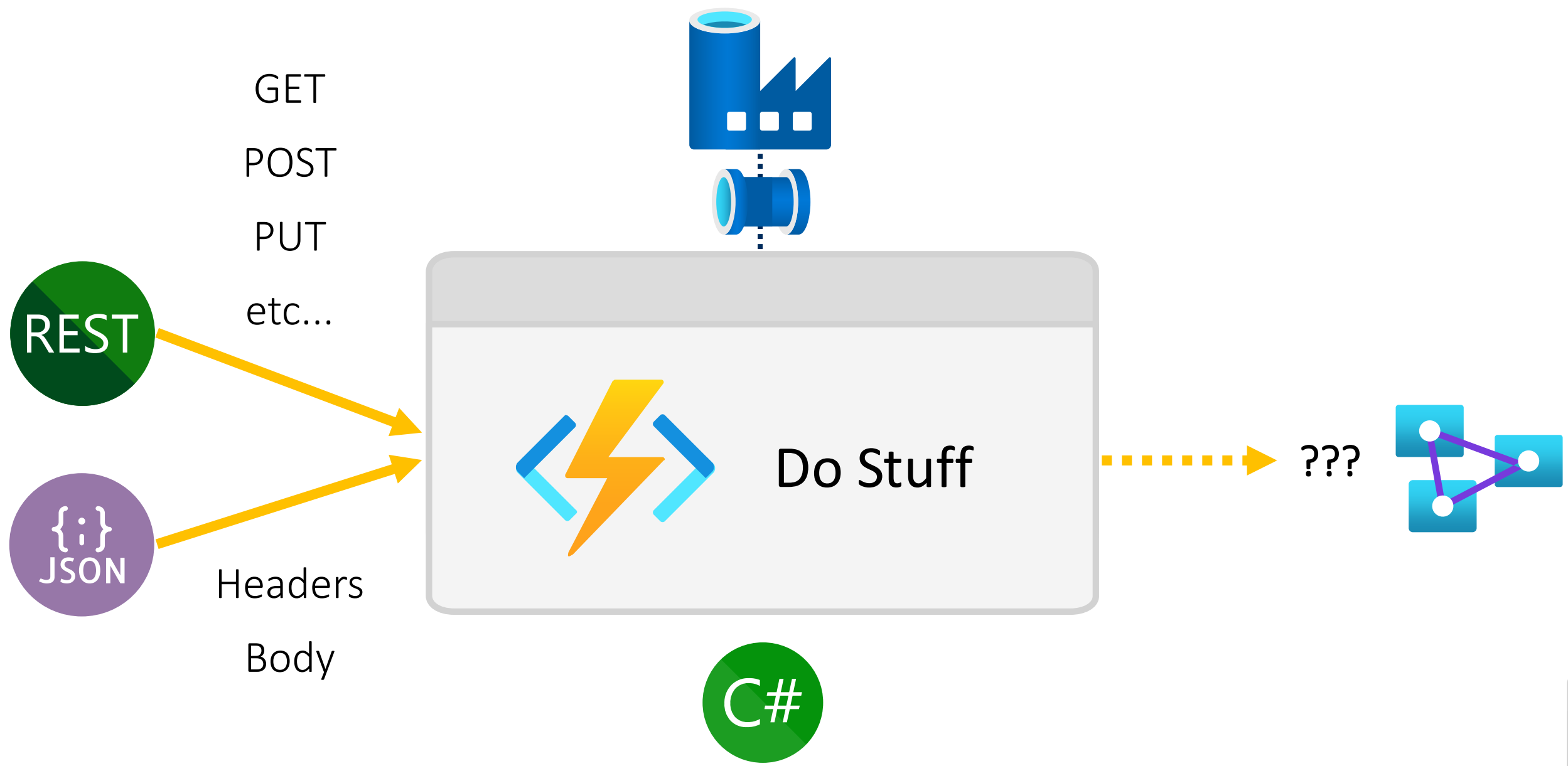
Chaining pipeline executions via an activity




Azure Function



Extend Data Factory with custom serverless code executions via REST calls

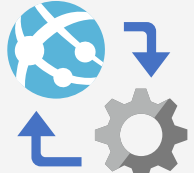


Web



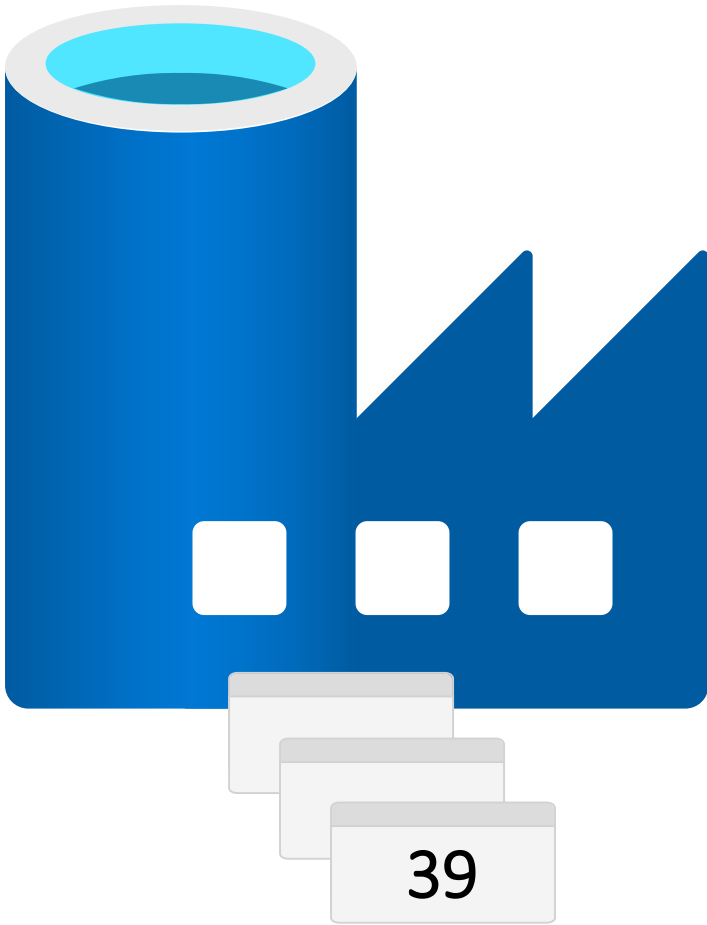
Do Stuff

Web Hook



Do Stuff

Activities Comparison Between Tools

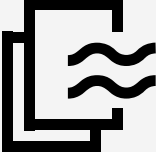


Fabric Activity Comparison

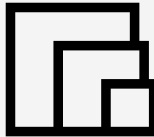


Not available Microsoft Fabric


Streaming

 Streaming


Map Reduce

 Map Reduce


Machine Learning Execute Pipeline

 ML Pipeline


Pig

 Pig


Spark

 Native Spark Job


Machine Learning Update Resource

 ML Update


U-SQL

 U-SQL


Execute SSIS Package

 SSIS


Power Query

 Power Query


Hive

 Hive


Databricks

 Jar

Databricks


 Python

Spark Job Definition


 Spark Job Definition

Only available Microsoft Fabric

Teams

 Post to Chat

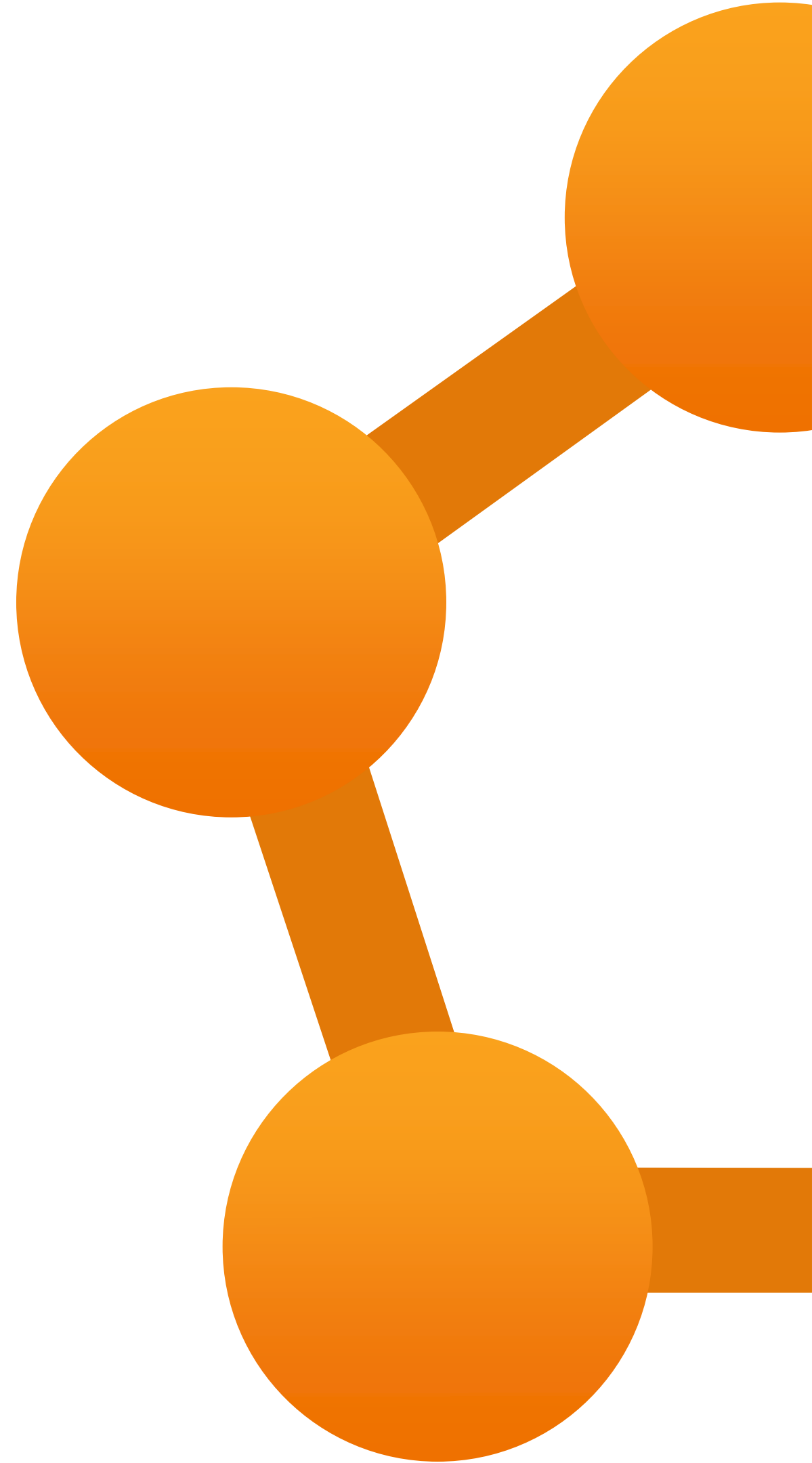
Outlook

 Send an Email

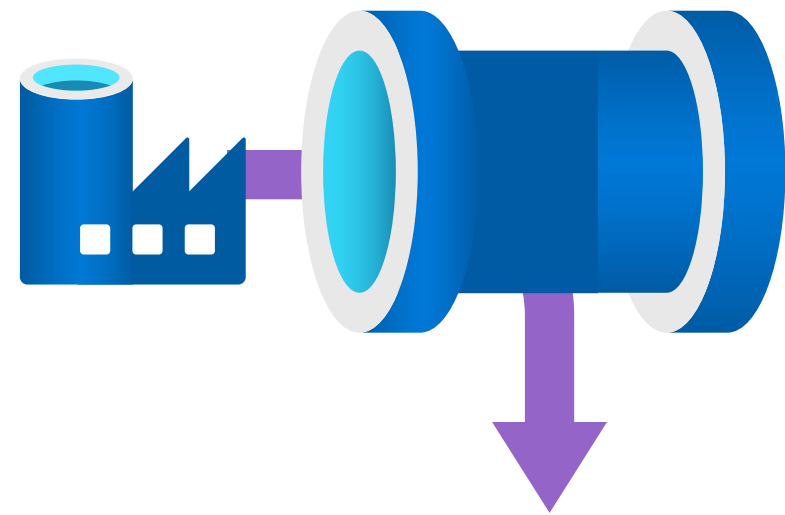
Module 1 – Pipeline Fundamentals

Execution Dependencies

Cloud Formations



Execution Dependency Options



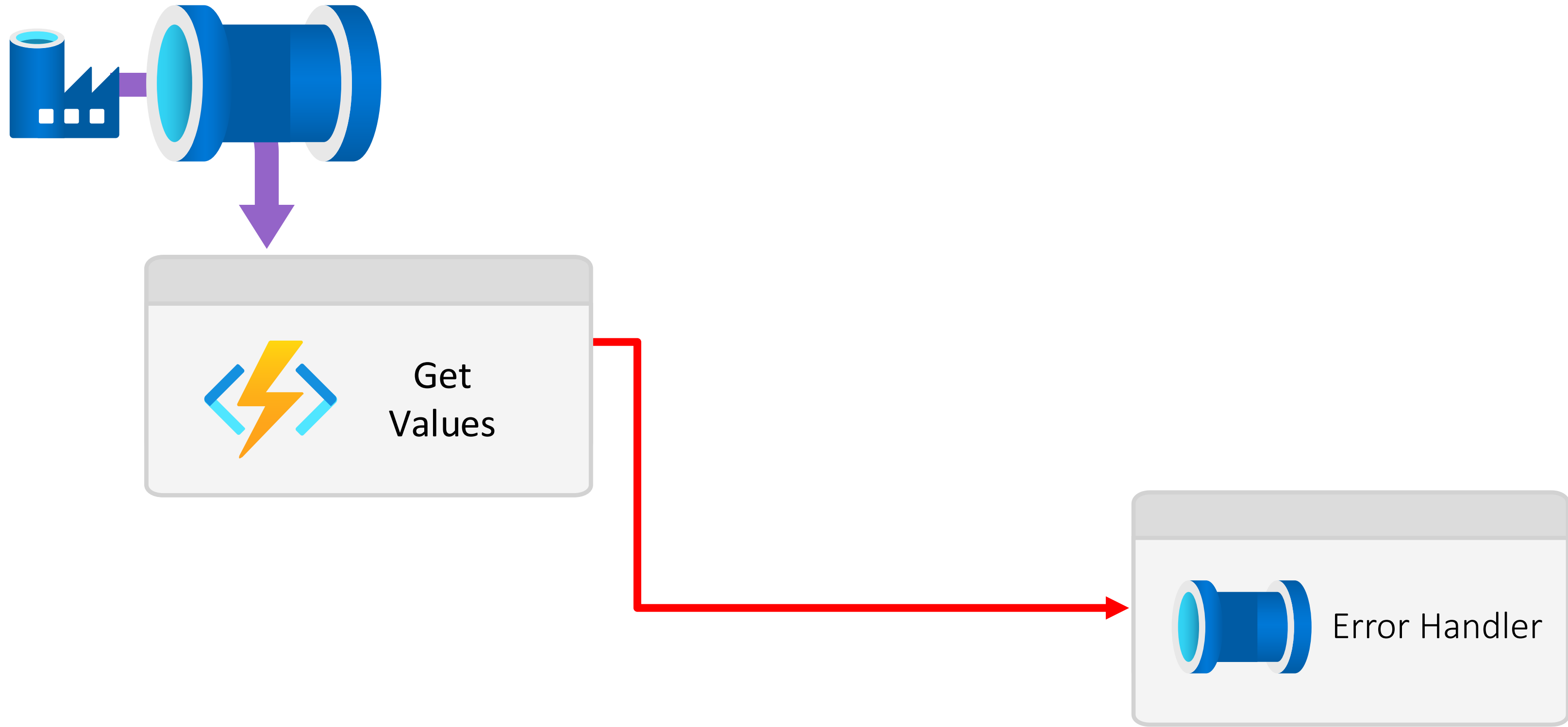
Success, when the activity completes successfully, any downstream activities that are dependent on this activity will proceed.

Fail, when the activity fails downstream activities that are set to execute upon failure will be triggered. This allows for error handling and alternative processing paths.

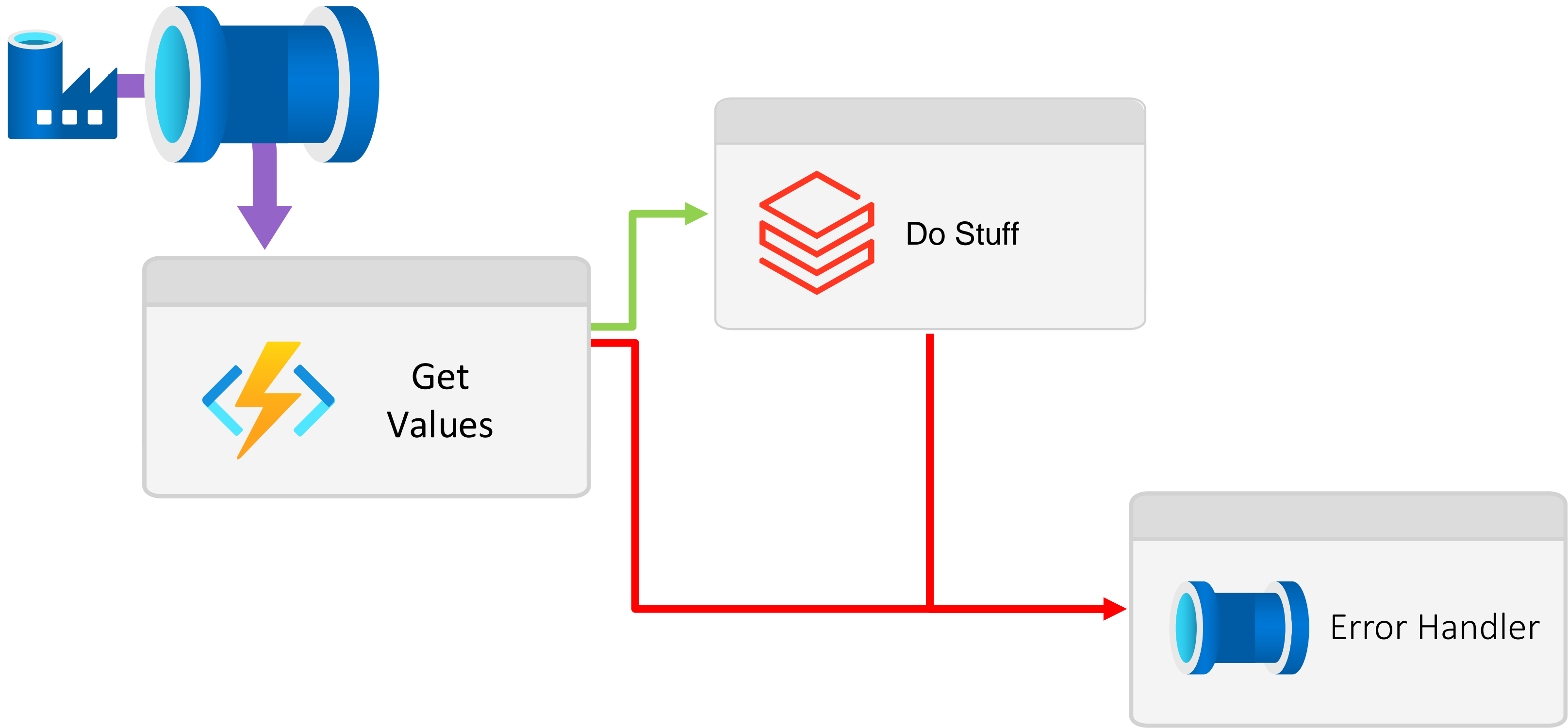
Complete, triggers downstream activities regardless of whether the preceding activity succeeded or failed. Useful for cleanup operations and logging. *Try > Catch > Finally*.

Skip, when the activity is skipped it means the activity was not executed due to a condition not being met. Downstream activities that are dependent on the skipped activity will also be skipped unless they are configured to run upon skip.

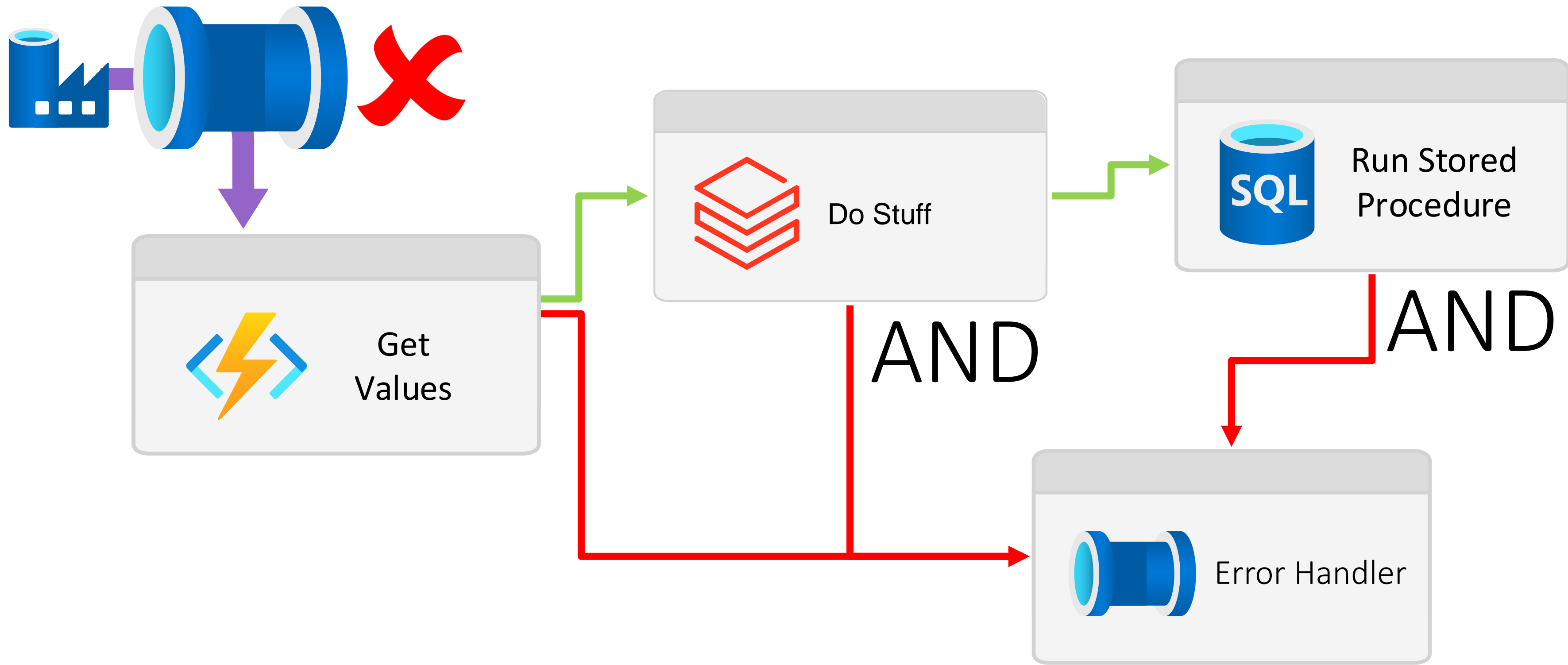
Execution On Failure



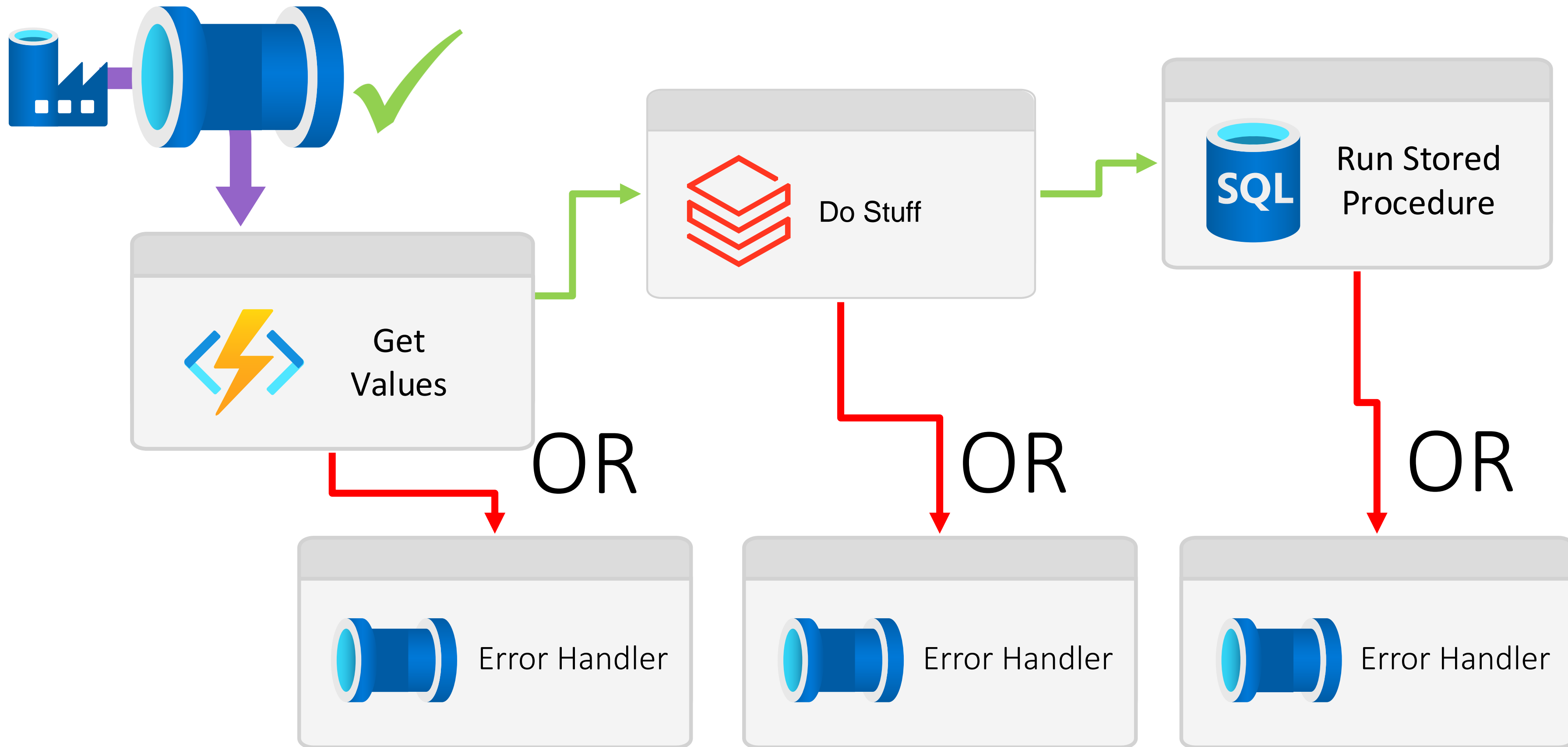
Execution On Failure or On Success



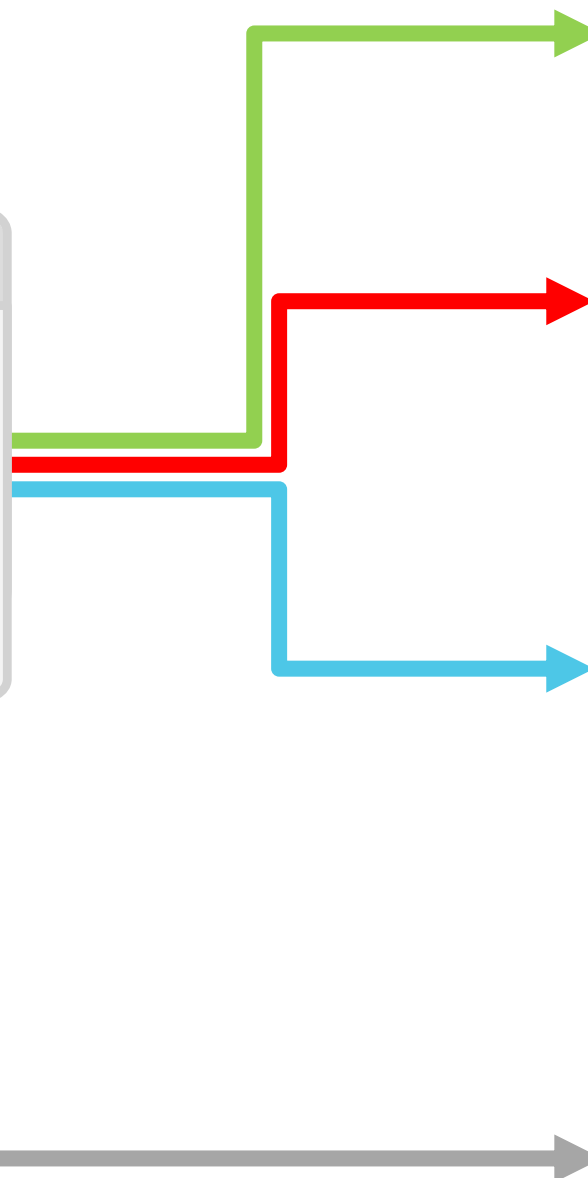
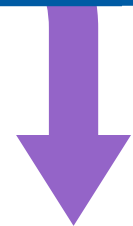
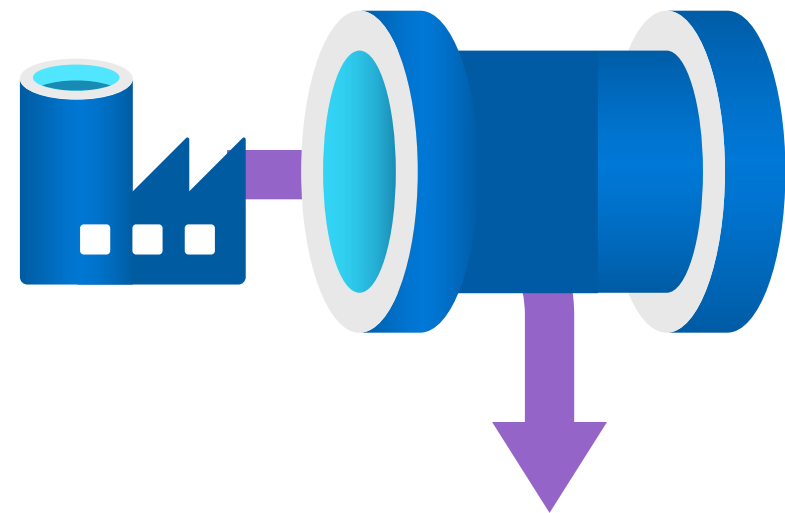
Execution On ???



Execution On Failure or On Success



Execution Dependency Options



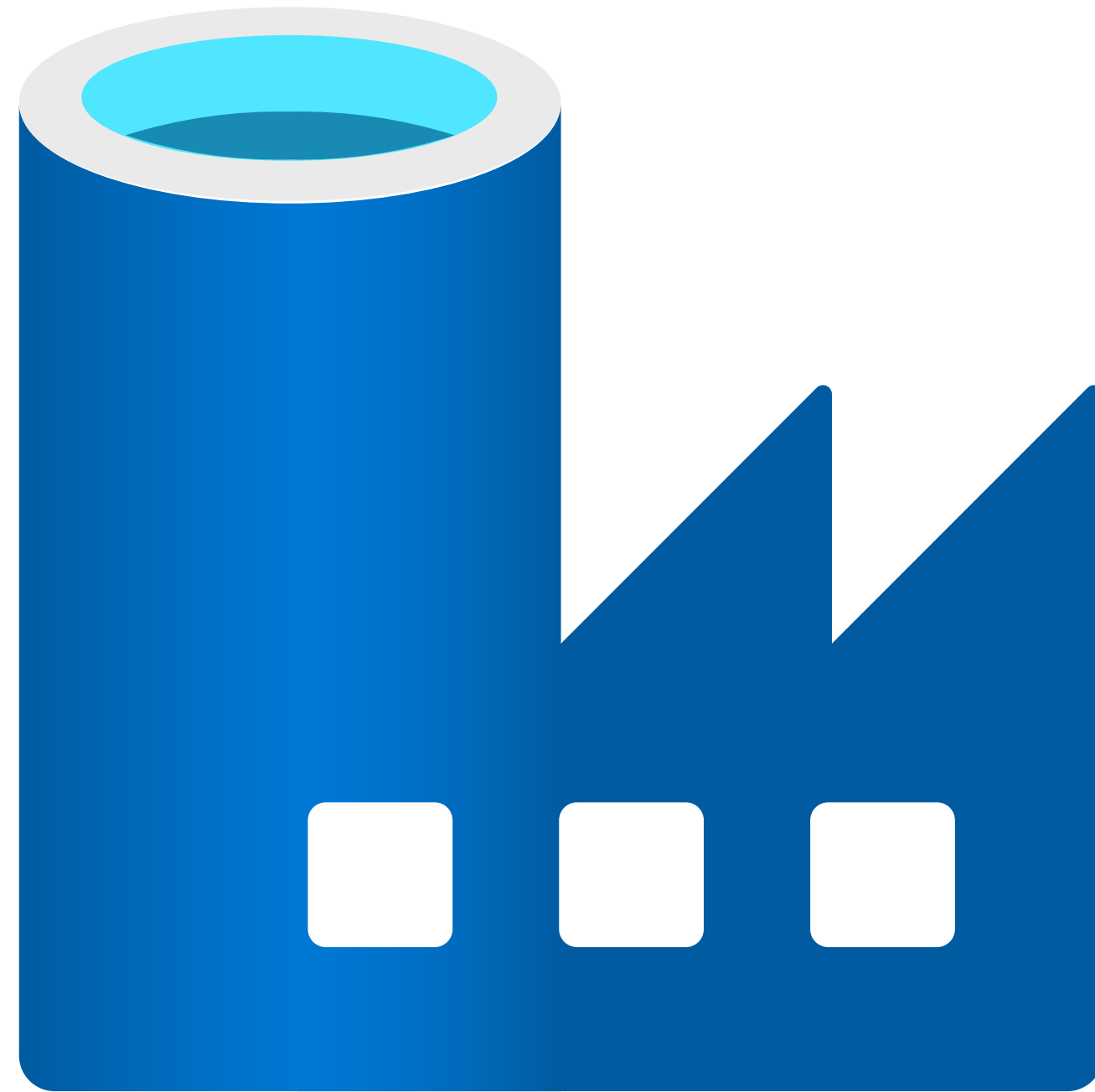
Success, when the activity completes successfully, any downstream activities that are dependent on this activity will proceed.

Fail, when the activity fails downstream activities that are set to execute upon failure will be triggered. This allows for error handling and alternative processing paths.

Complete, triggers downstream activities regardless of whether the preceding activity succeeded or failed. Useful for cleanup operations and logging. *Try > Catch > Finally.*

Skip, when the activity is skipped it means the activity was not executed due to a condition not being met. Downstream activities that are dependent on the skipped activity will also be skipped unless they are configured to run upon skip.

Handy for debugging and testing.



Module 1

Pipeline Fundamentals



Any questions?

Cloud Formations

