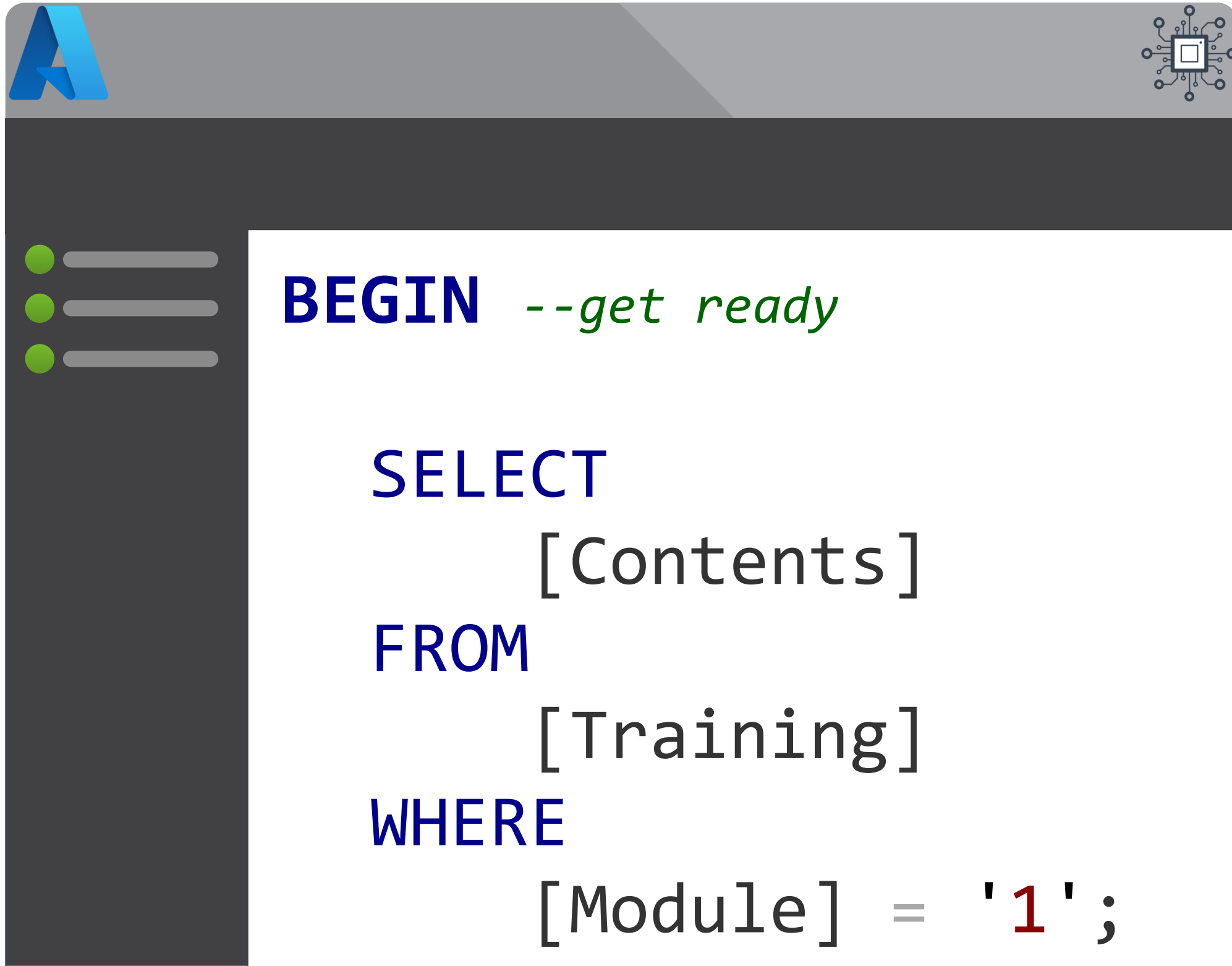


Module 1

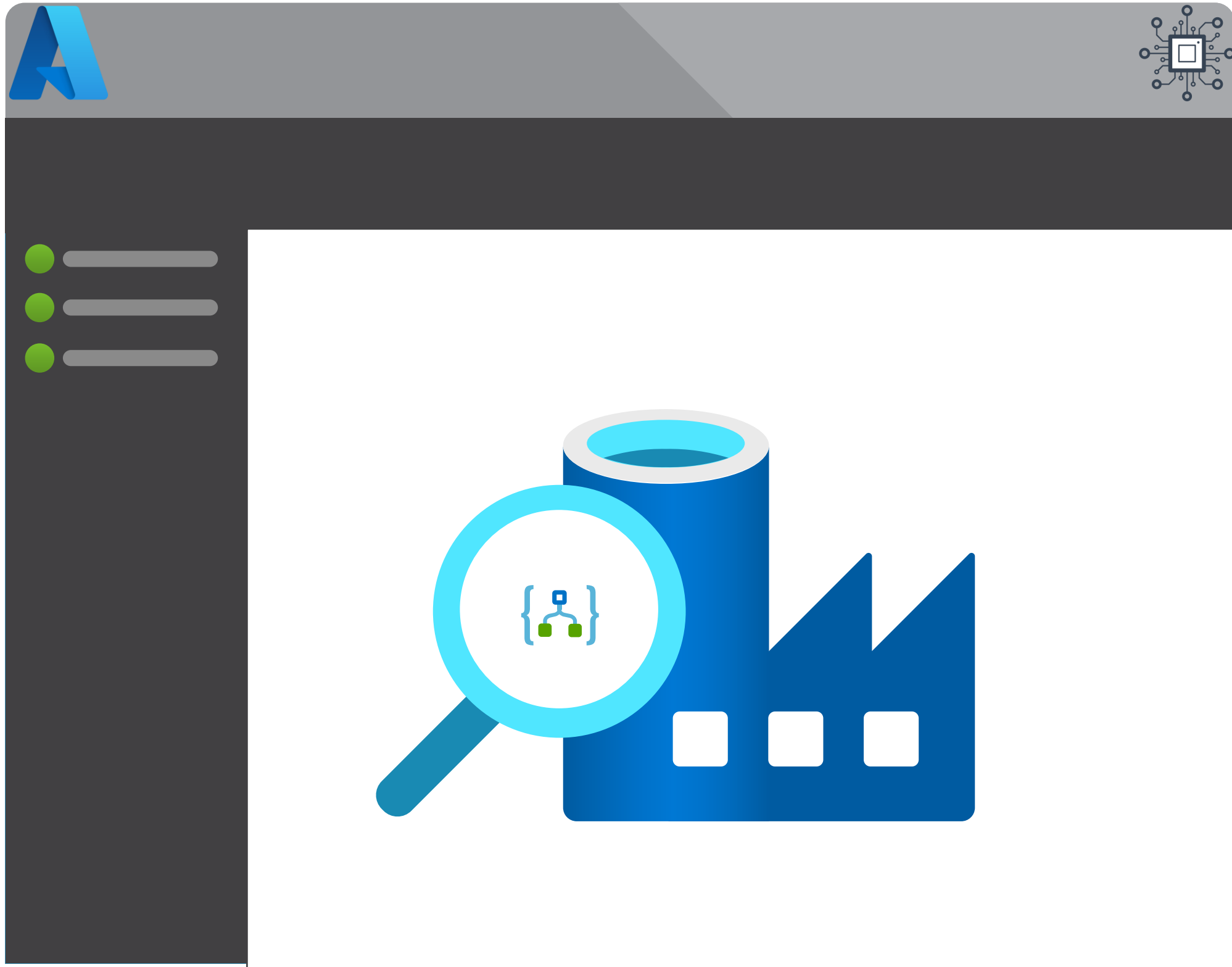
Pipeline Fundamentals



- The History of Azure Orchestration
- Synapse Analytics vs Data Factory
- Integration Components
- Common Activities
- Execution Dependencies

Module 1

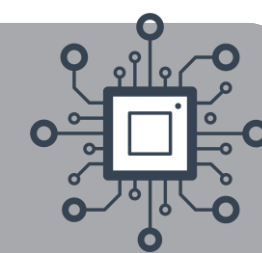
Pipeline Fundamentals



- The History of Azure Orchestration
- Synapse Analytics vs Data Factory
- Integration Components
- Common Activities
- Execution Dependencies



A Quick History Lesson



SQL Server
SQL Agent



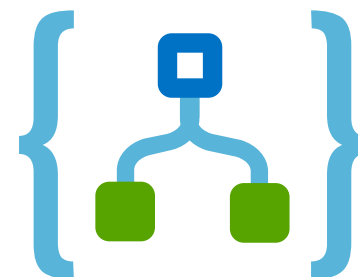
SQLDB
(PaaS)



Automation



Logic Apps



Functions



SQL Managed
Instance

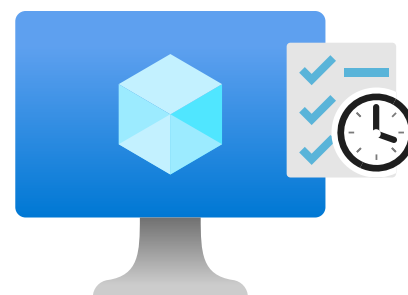


DTU Jobs

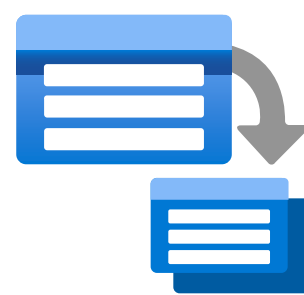
Elastic Job Agent



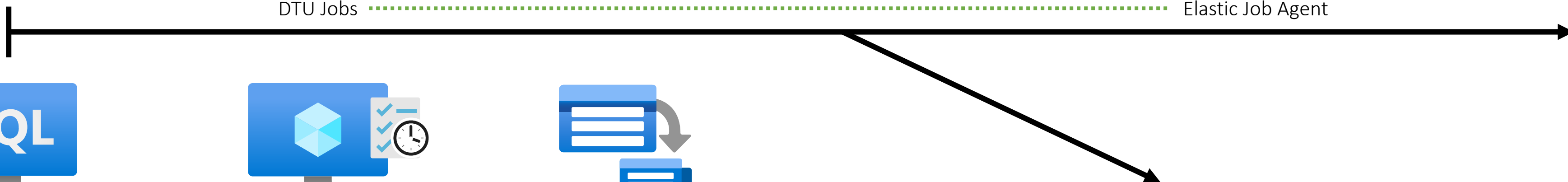
SQL Server
Virtual Machine



Virtual Machine
Job Schedule

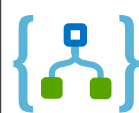
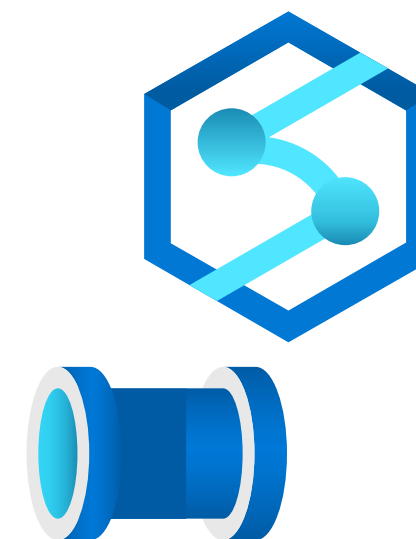
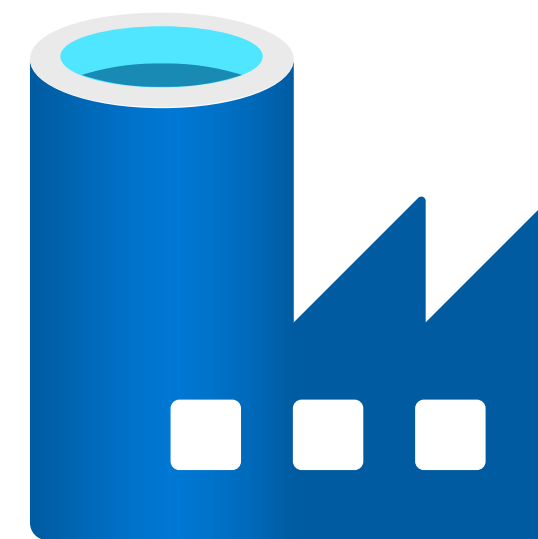
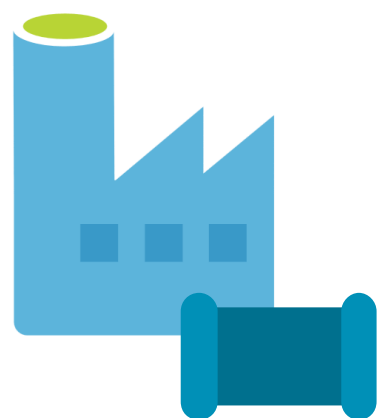
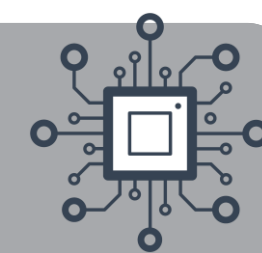


Batch





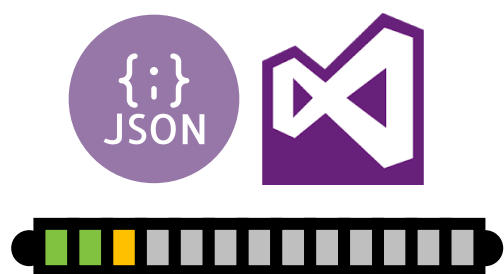
A Quick History Lesson



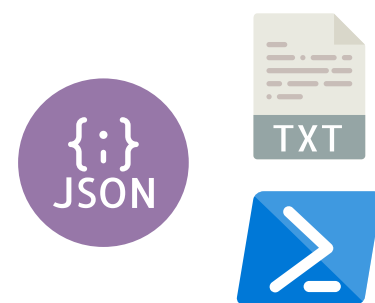
v1
July
2016

v2
Sept
2017

Oct
2020



Time Slices
VS2015 Projects



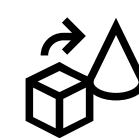
Notepad
Development



New UI



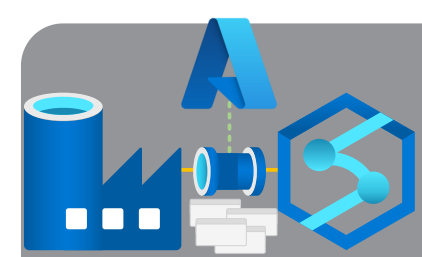
Source
Control



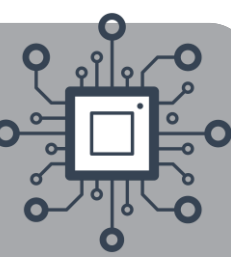
Data
Flows



Private
Endpoints



What is Azure Data Factory (ADF)?



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

Data Factory

Hybrid data integration service that simplifies ETL at scale

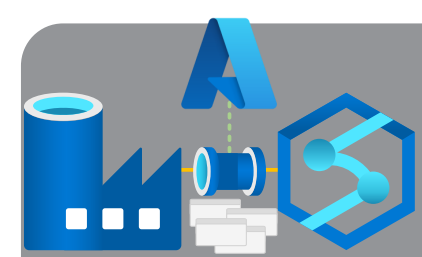
[Start for free >](#)

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

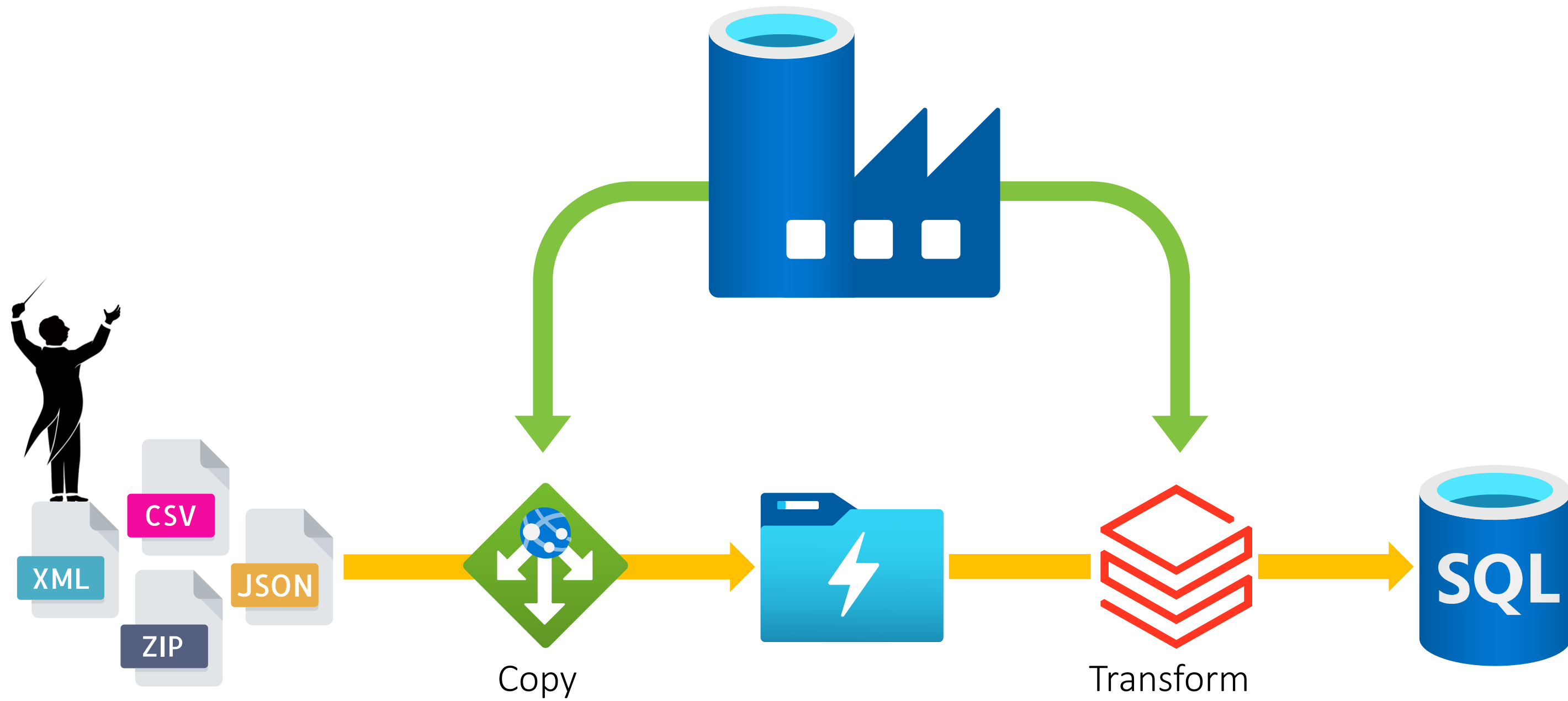
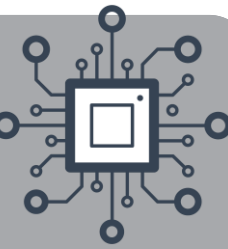
[Product overview](#) [Features](#) [Security](#) [Pricing](#) [Customer stories](#) [Getting started](#) [Documentation](#) [FAQs](#)

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.

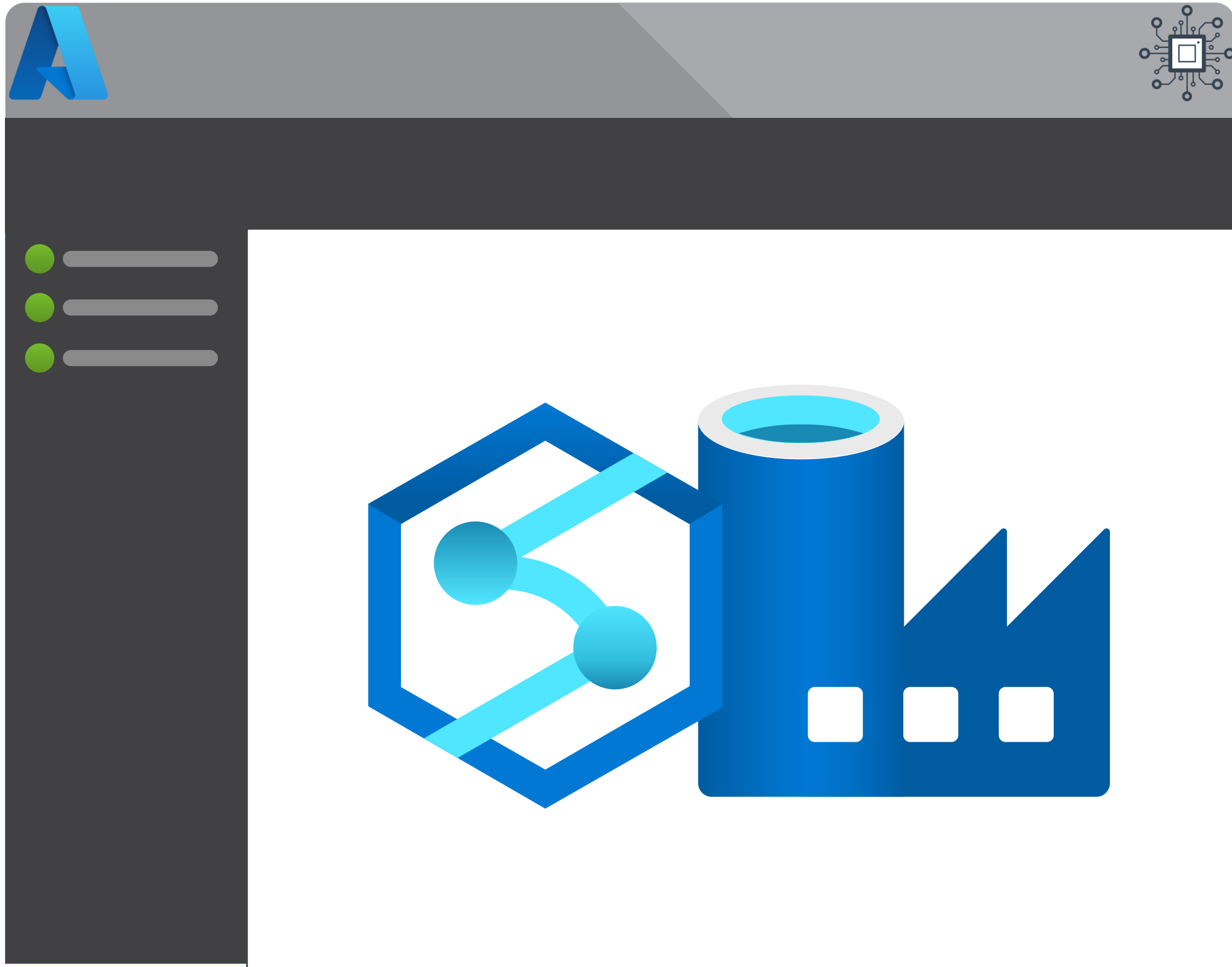


What is Azure Data Factory (ADF)?

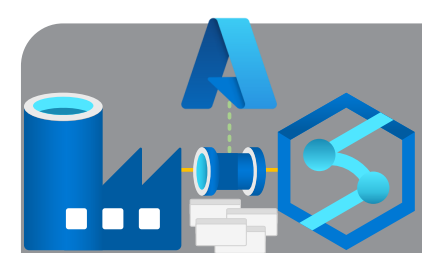


Module 1

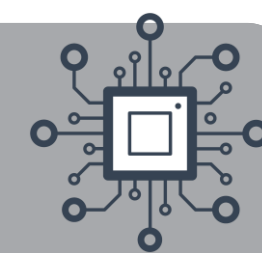
Pipeline Fundamentals



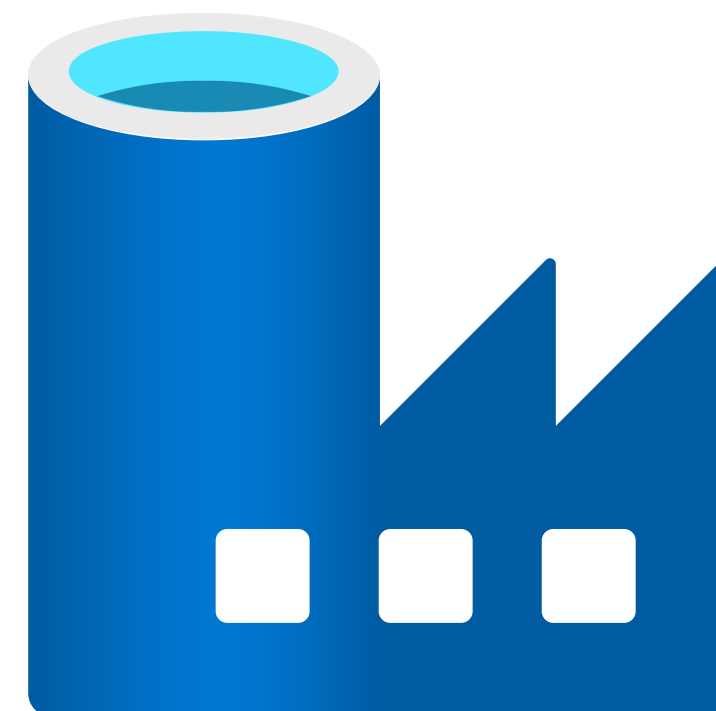
- The History of Azure Orchestration
- Synapse Analytics vs Data Factory
- Integration Components
- Common Activities
- Execution Dependencies



Synapse Analytics vs Data Factory

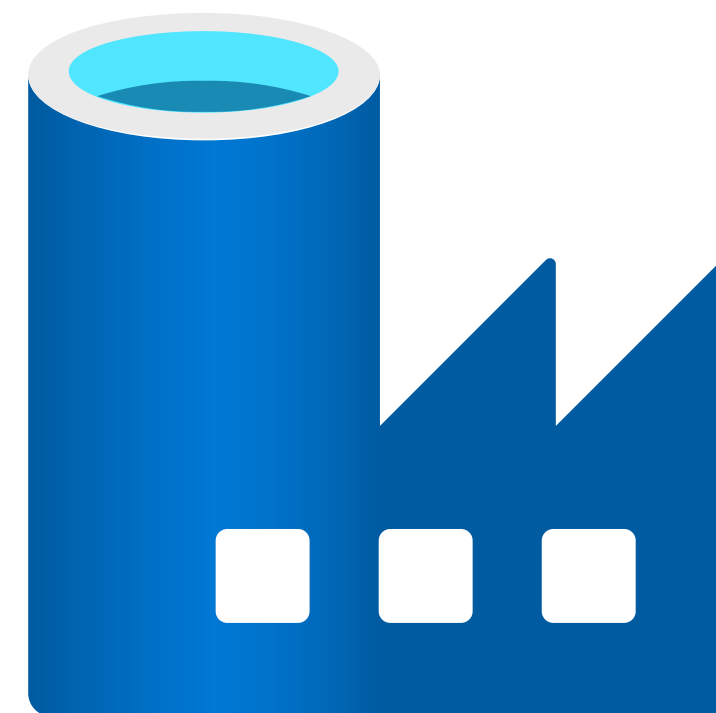
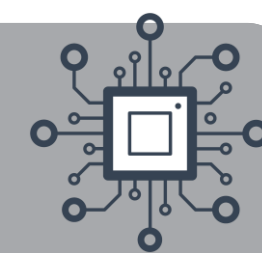


<https://docs.microsoft.com/en-us/azure/synapse-analytics/data-integration/concepts-data-factory-differences>



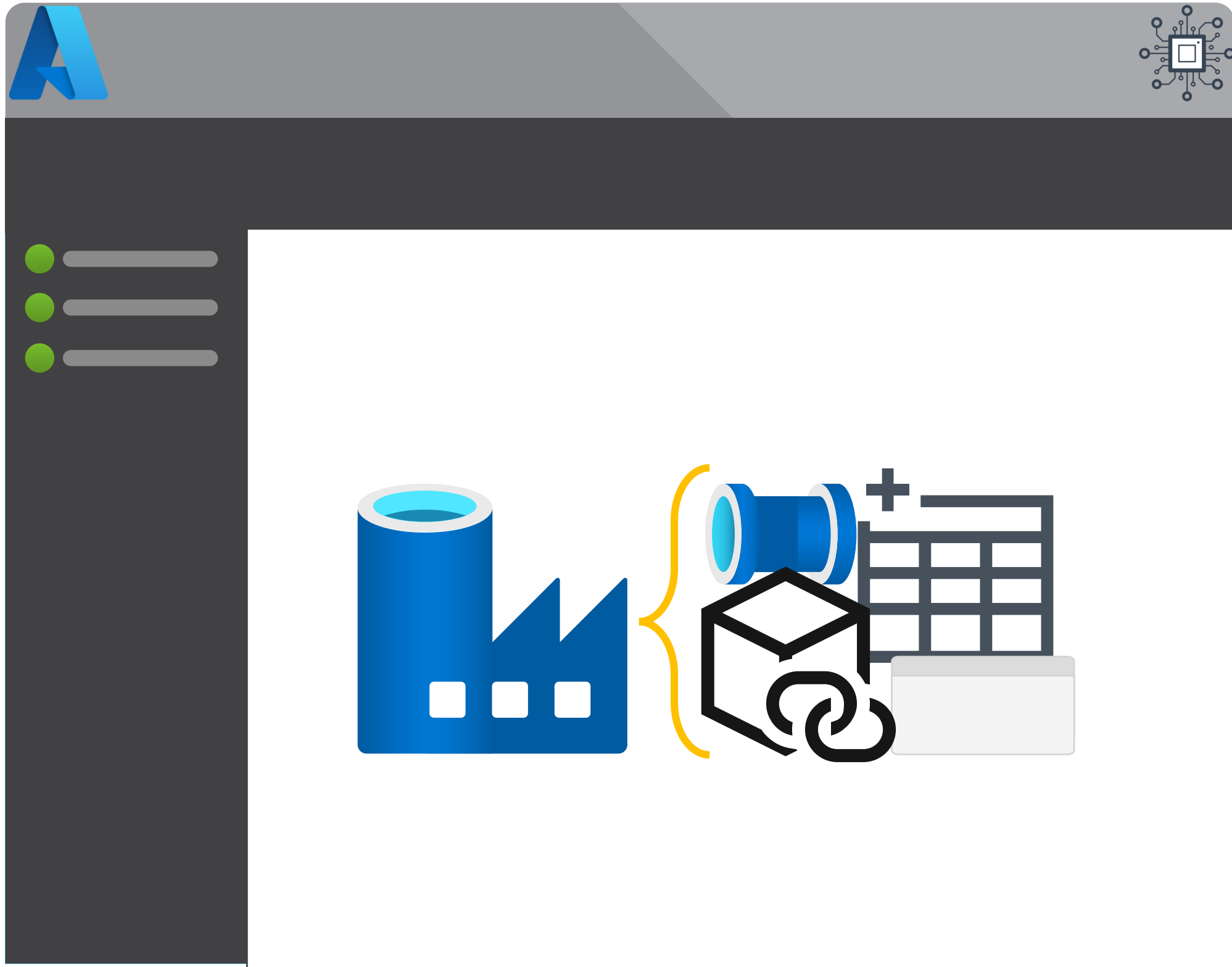


Synapse Analytics vs Data Factory

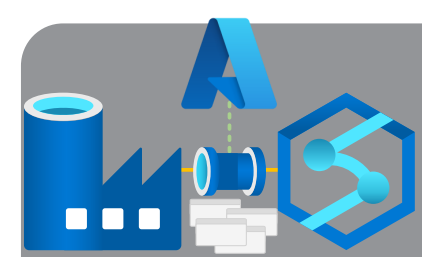


Module 1

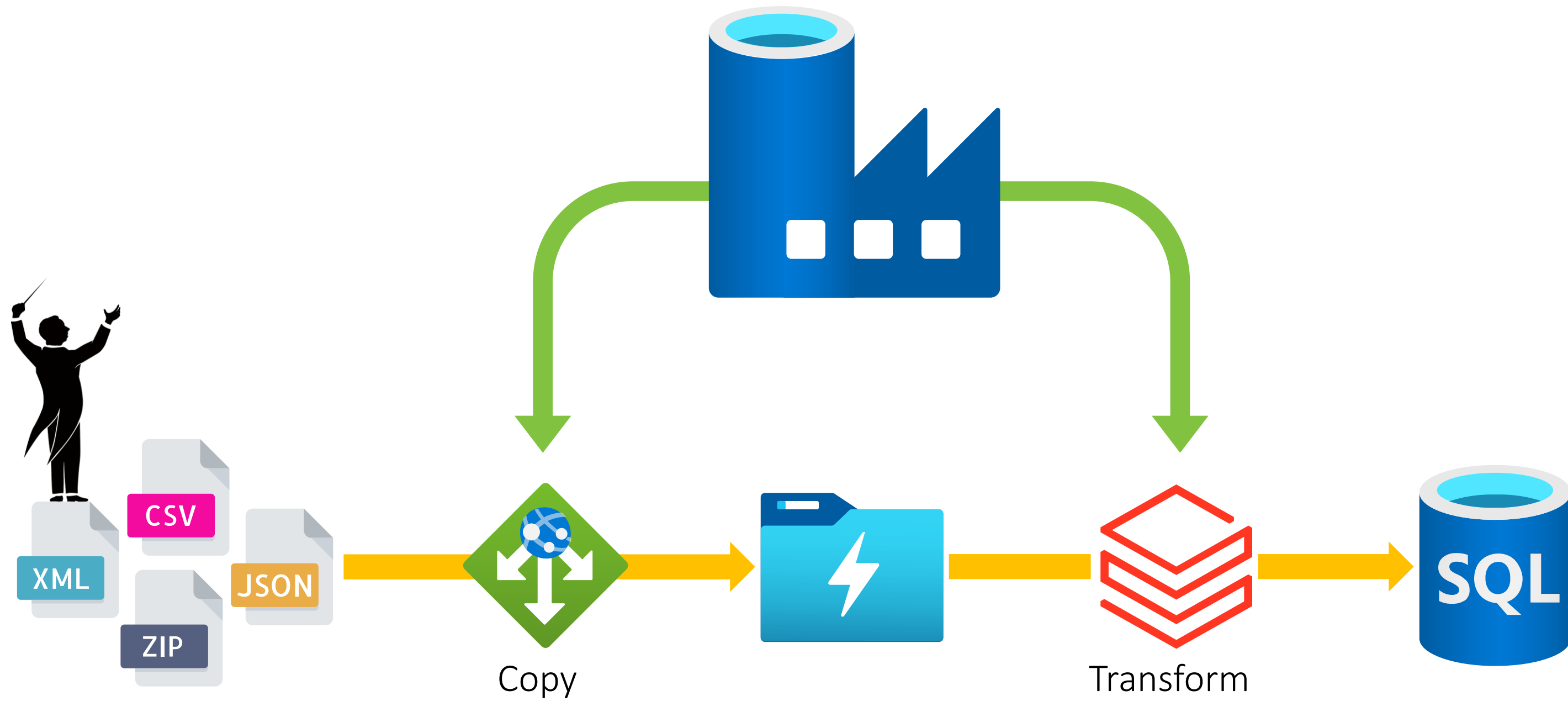
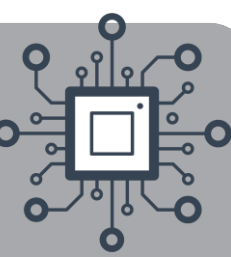
Pipeline Fundamentals



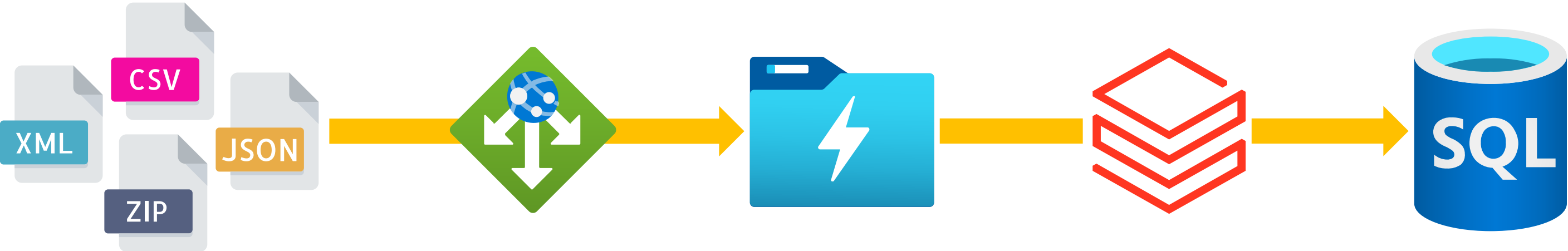
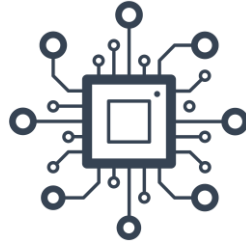
- The History of Azure Orchestration
- Synapse Analytics vs Data Factory
- **Integration Components**
- Common Activities
- Execution Dependencies



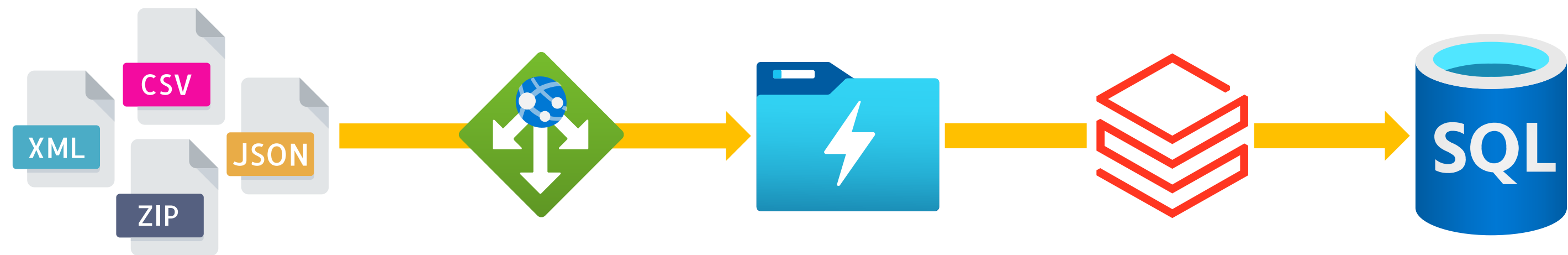
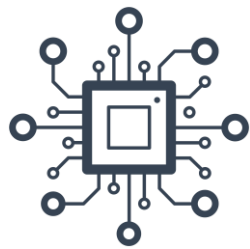
Data Factory Components



Data Factory Components

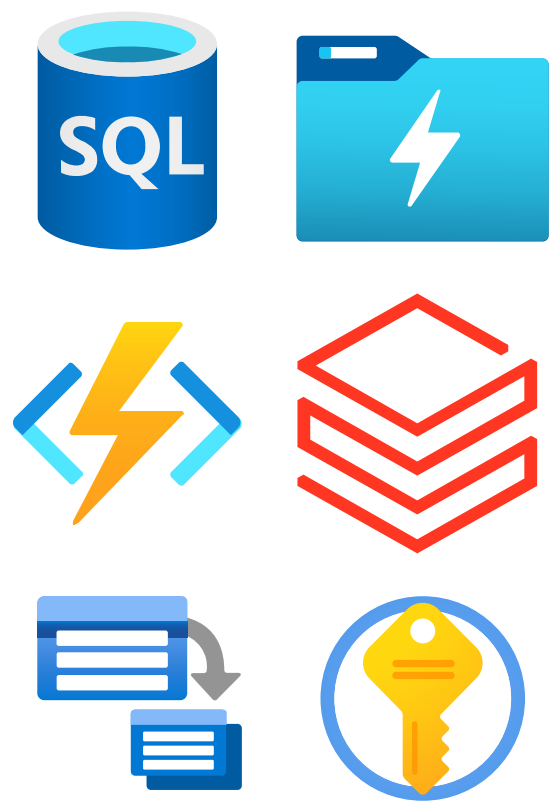


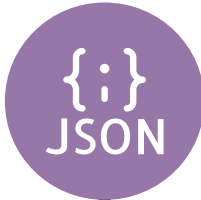
Data Factory Components



1

Linked Services – What to interact with and how?

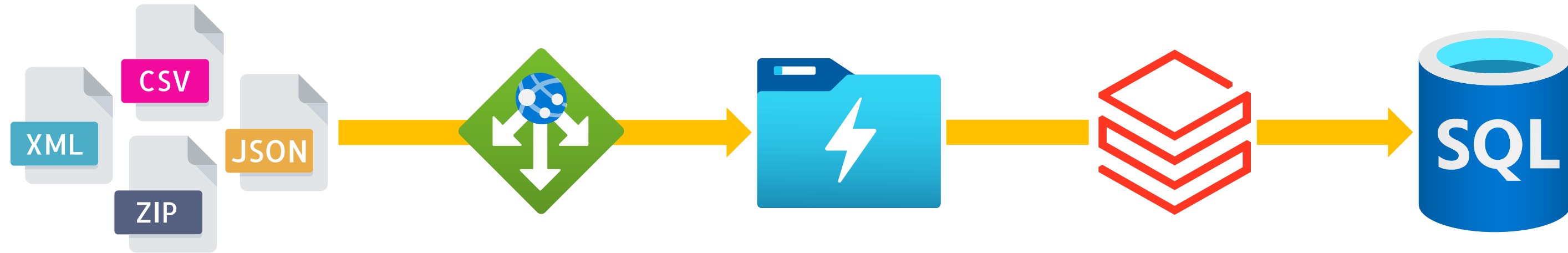
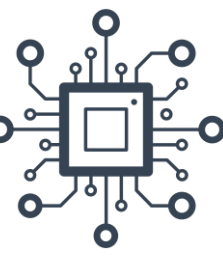




```
SQLDBLinkedService

ConnectionString: Server=MyServer;Database=myDataBase
UserName: "MrPaulAndrew"
Password: *****(12)*****
```

Data Factory Components



1 Linked Services

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

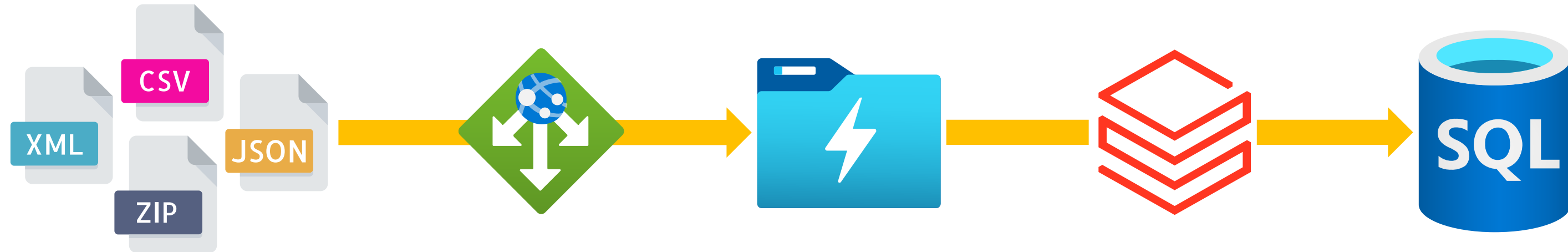
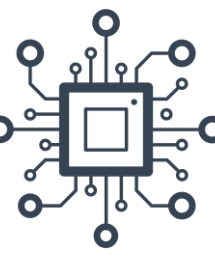


[dbo].[SalesOrders]



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

Data Factory Components



1 Linked Services

2 Datasets

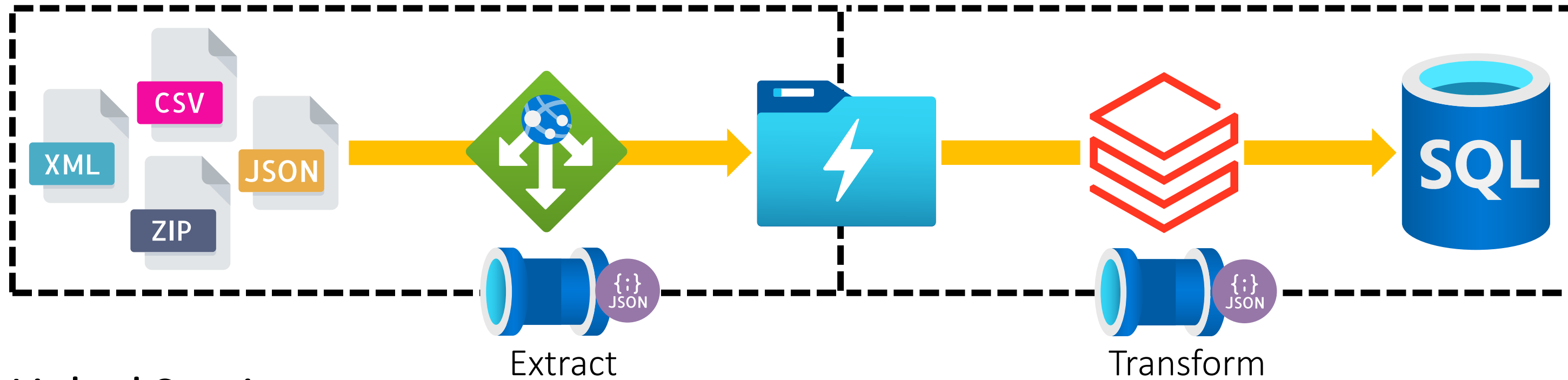
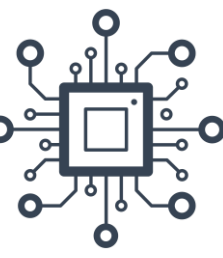
3 **Activities** – What do we want to happen when we invoke a Linked Service?
With what conditions?

{:}
JSON

Databricks Notebook Activity

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

Data Factory Components

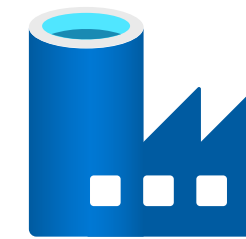
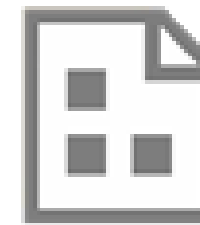


1 Linked Services

2 Datasets

3 Activities

4 **Pipelines** – Logical groups of work that can be executed.



Sequence Container

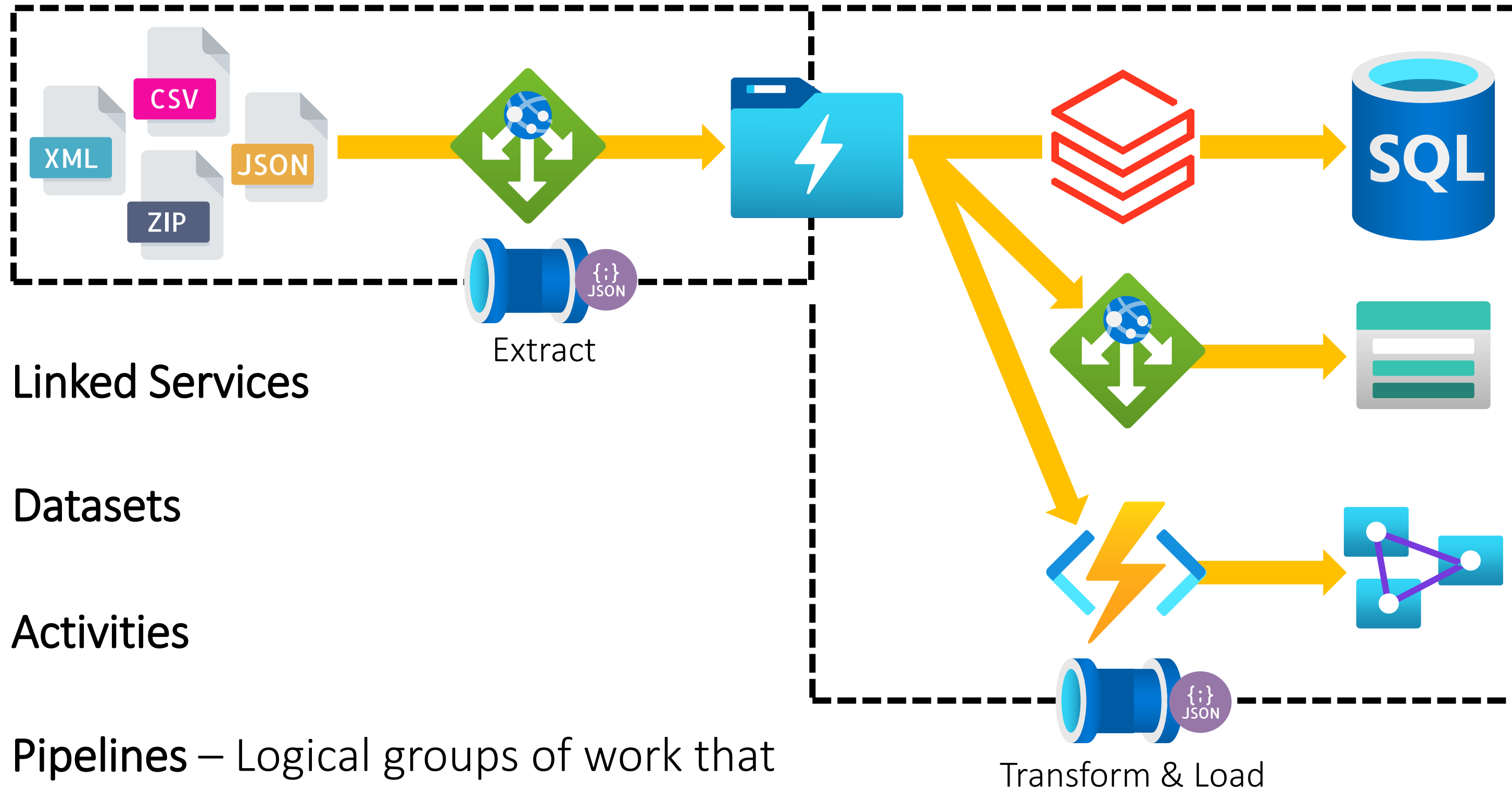
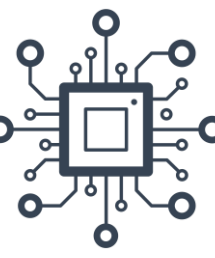


Execute Package Task

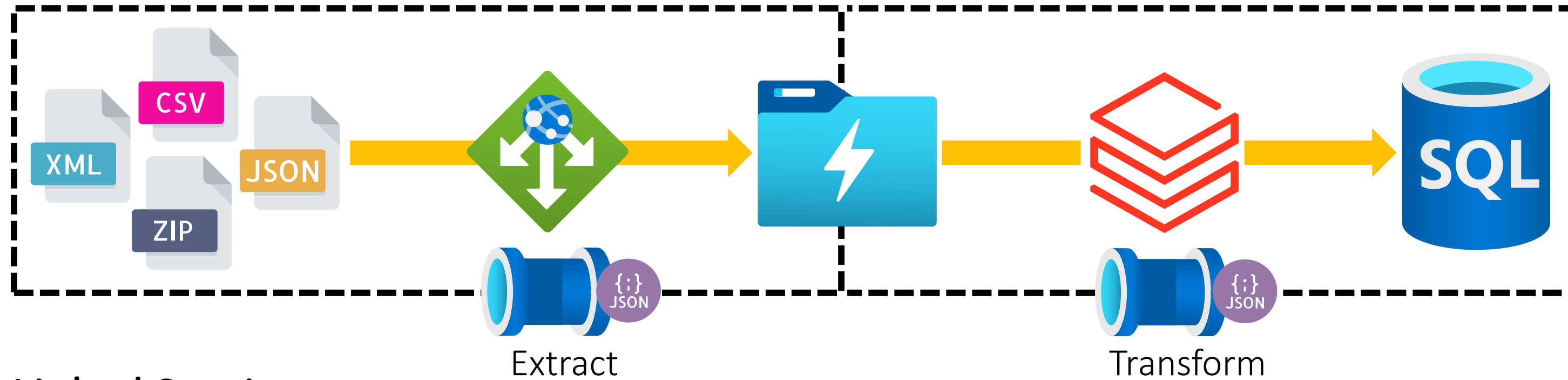
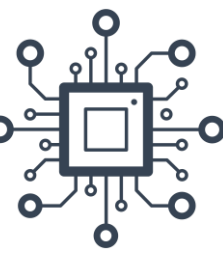


Execute Pipeline Activity

Data Factory Components



Data Factory Components



1 Linked Services

2 Datasets

3 Activities

4 Pipelines

5 Triggers – Telling our when pipelines to run.

Manually

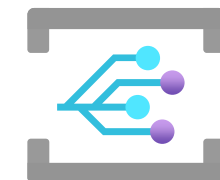
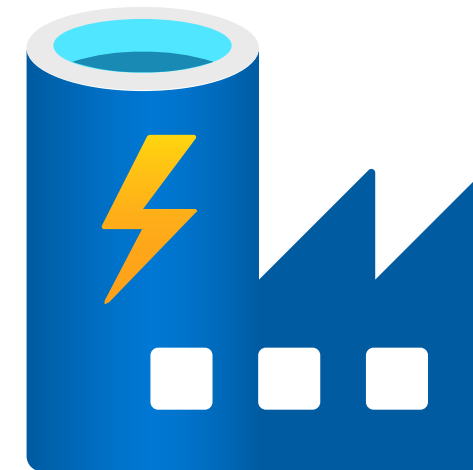
Programmatically

Schedule

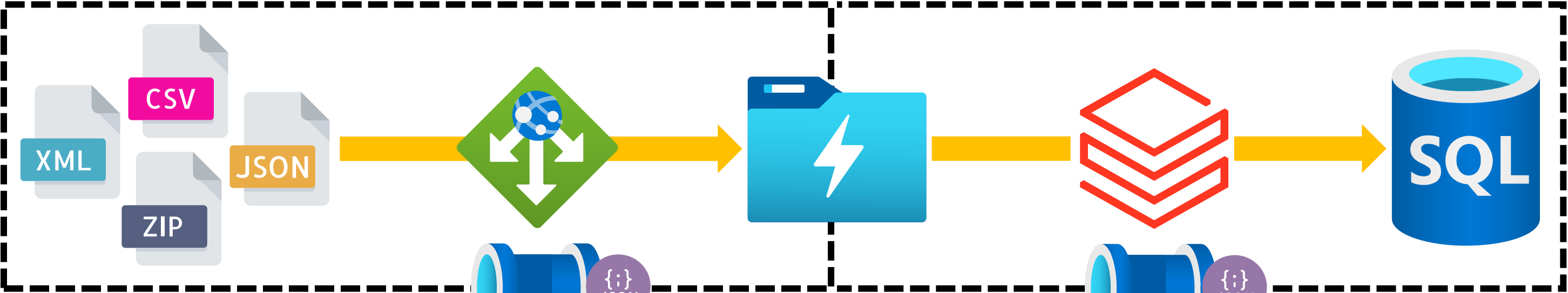
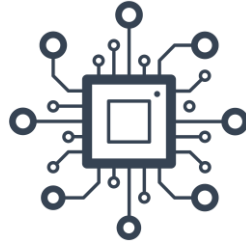
Tumbling Windows

Storage Events

Custom Events



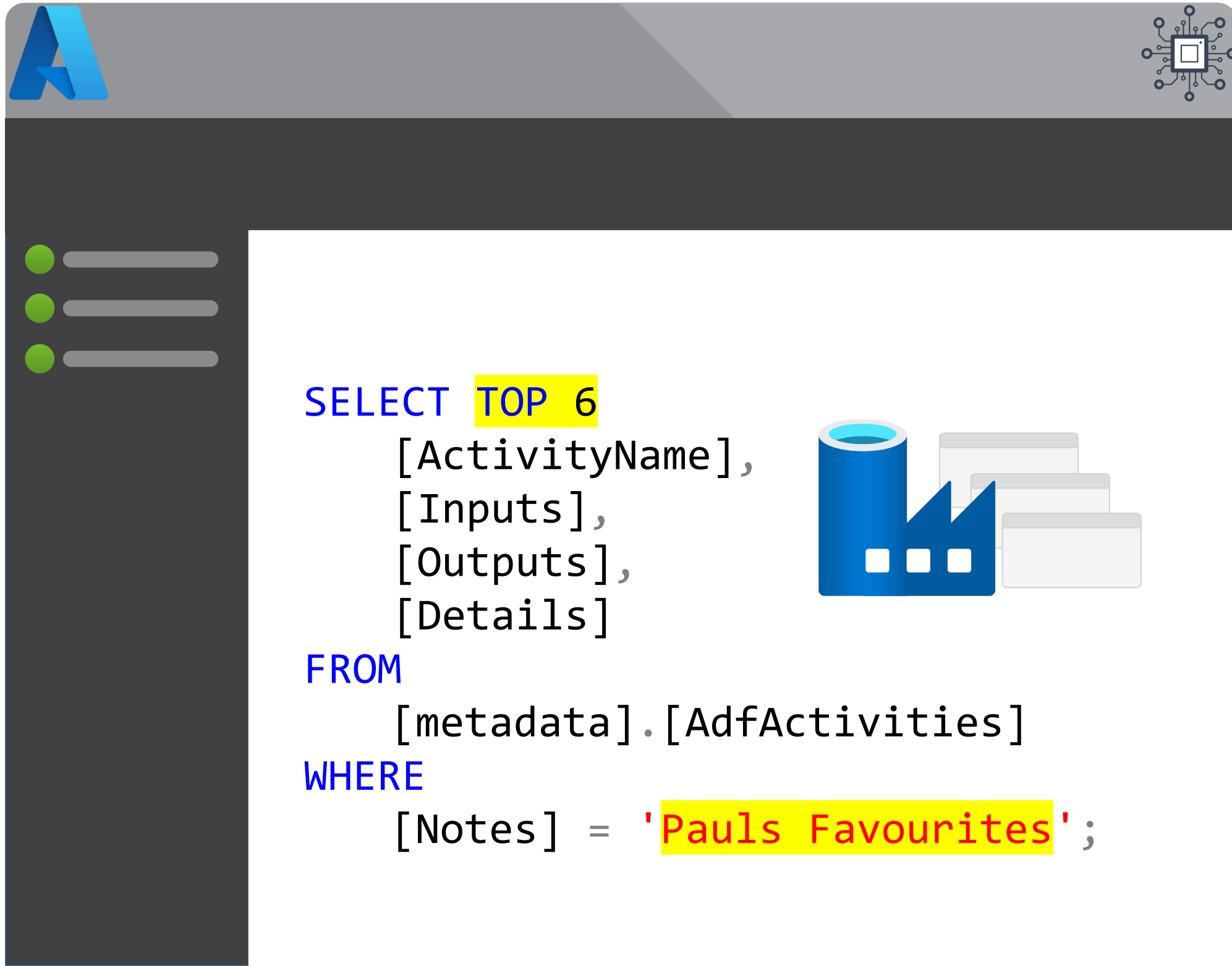
Data Factory Components



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

Module 1

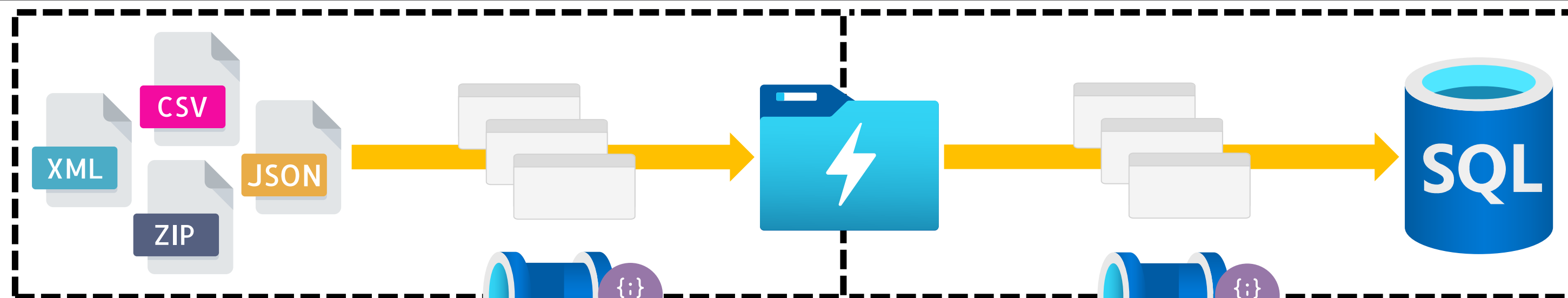
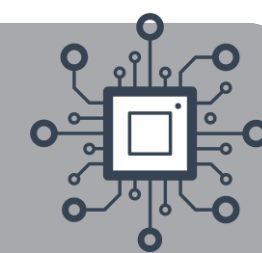
Pipeline Fundamentals



- The History of Azure Orchestration
- Synapse Analytics vs Data Factory
- Integration Components
- Common Activities
- Execution Dependencies



Data Factory Common Activities



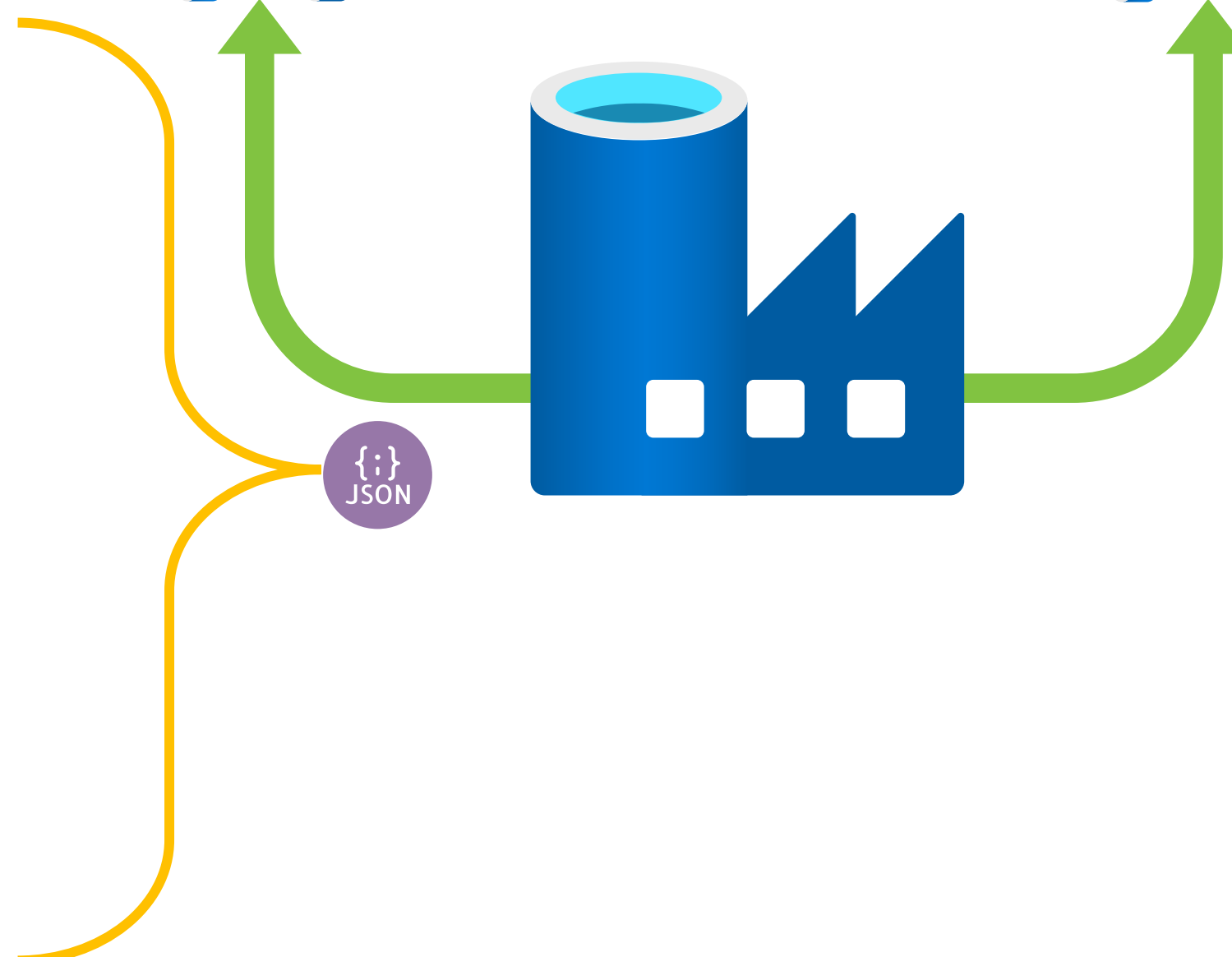
1 Linked Services

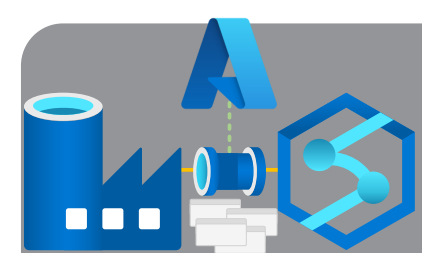
2 Datasets

3 Activities

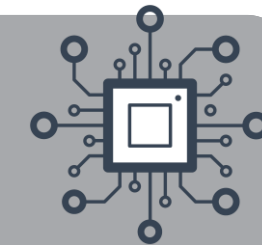
4 Pipelines

5 Triggers

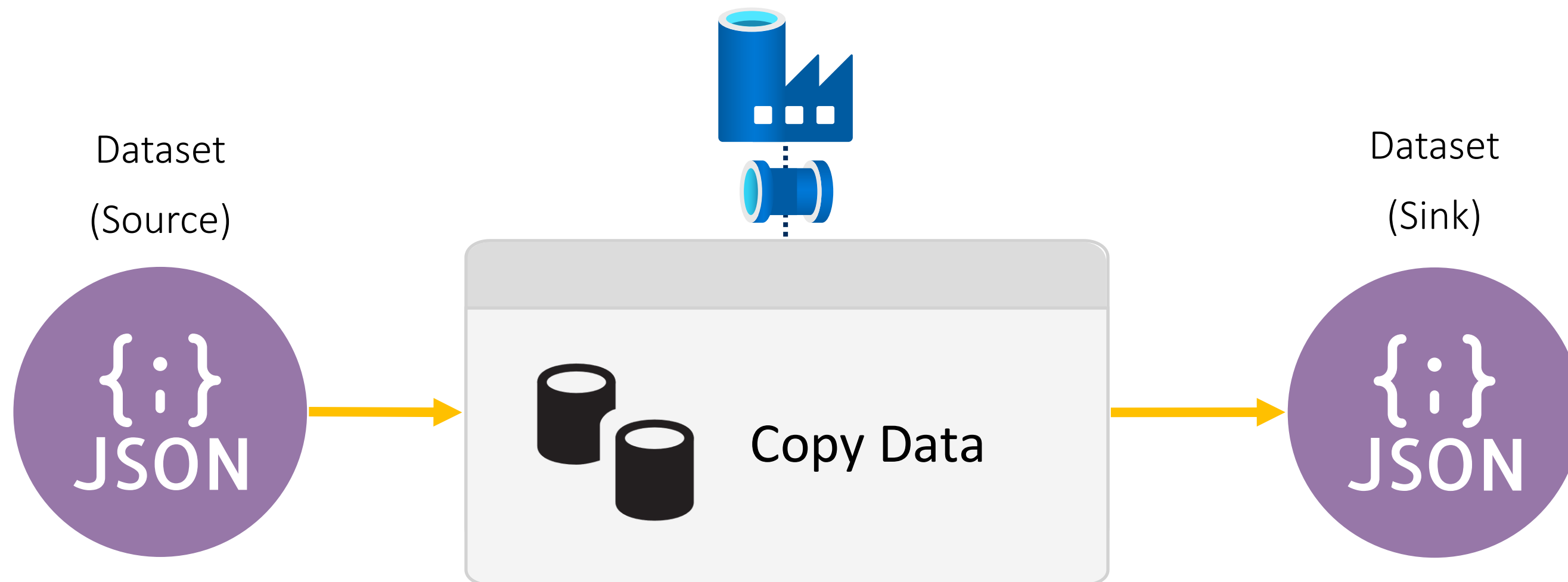




Copy



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



☐ Auto Scaling

☐ Transactional Restarts

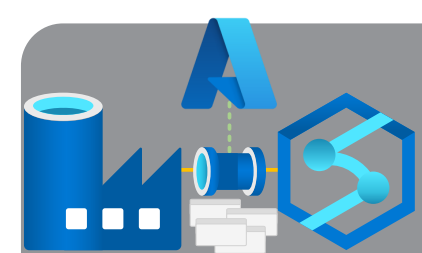
☐ Handle Zip Compression

☐ Attribute Mapping and Schema Drift

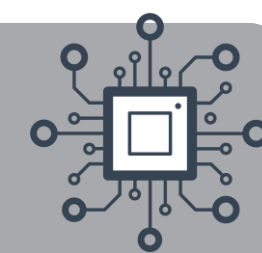
☐ Handle Failed Rows

☐ Add Custom Attributes

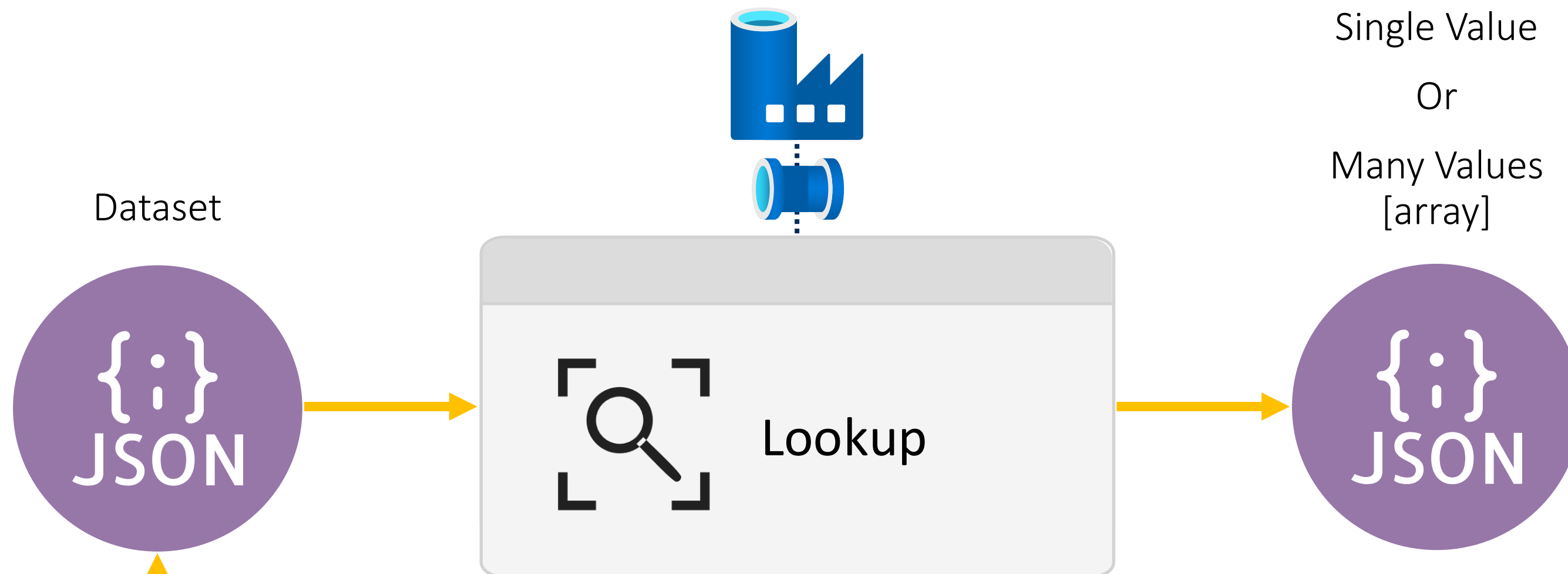
☐ Parse Excel & JSON Files



Lookup

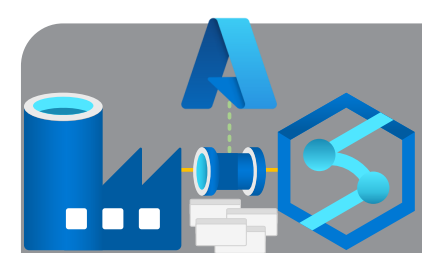


Get value(s) to support other control flow activities

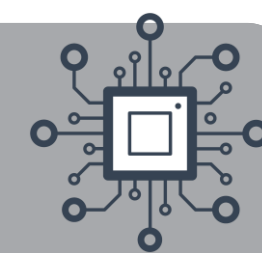


```
SELECT  
  [SourceDIR],  
  [TargetDIR],  
  [FileName]  
FROM  
  [dbo].[FileList]
```

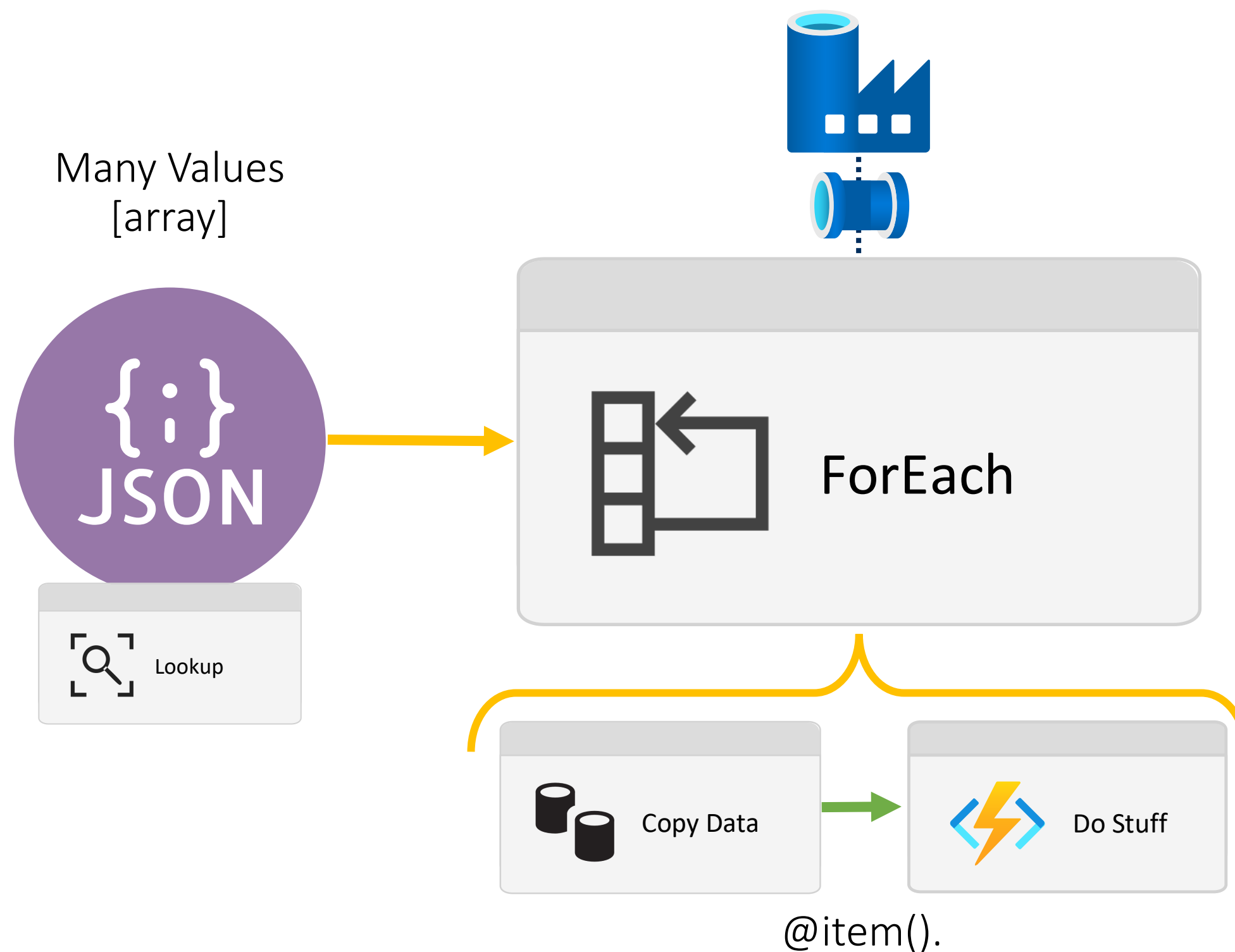
```
{  
  "count": 3,  
  "value": [  
    {  
      "SourceDIR": "ADFRoot\\ForUpload\\People\\",  
      "TargetDIR": "RAW",  
      "FileName": "Address.csv"  
    },  
    {  
      "SourceDIR": "ADFRoot\\ForUpload\\People\\",  
      "TargetDIR": "RAW",  
      "FileName": "Gender.csv"  
    },  
    {  
      "SourceDIR": "ADFRoot\\ForUpload\\People\\",  
      "TargetDIR": "RAW",  
      "FileName": "Ids.csv"  
    }  
  ]  
}
```



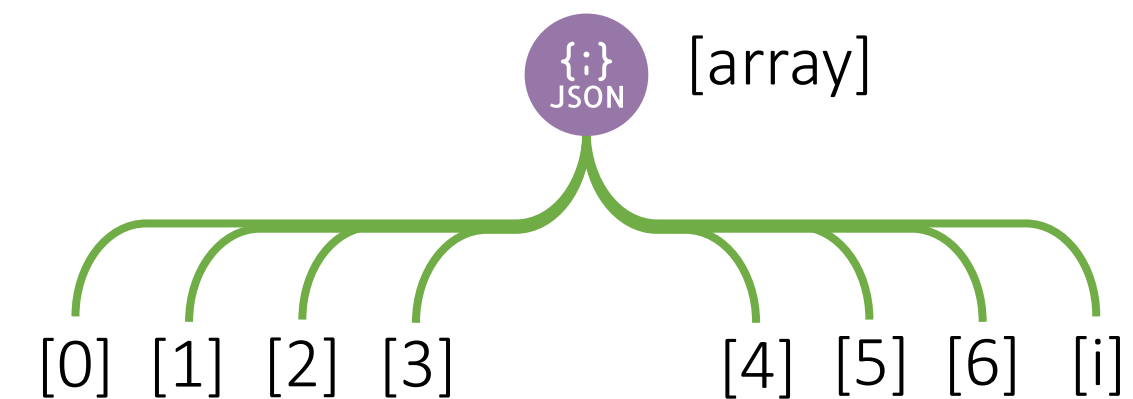
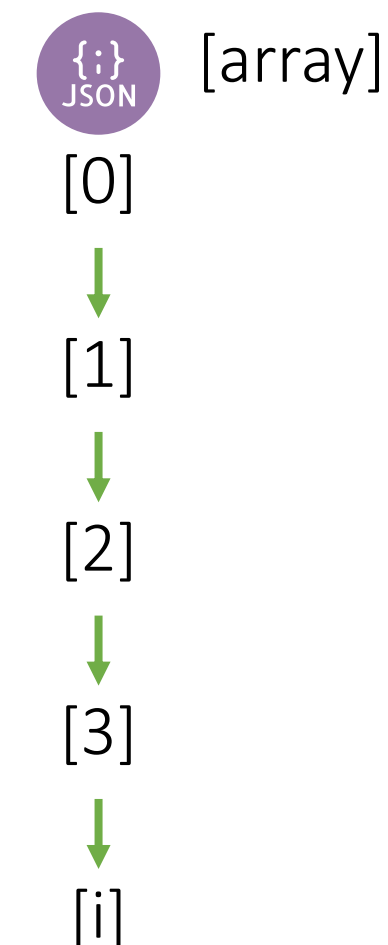
For Each



Iterating over other control flow activities



IsSequential:
true

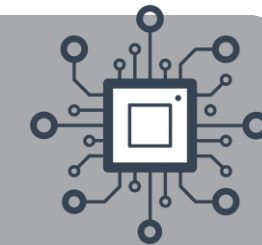


Batch Count Default: 20

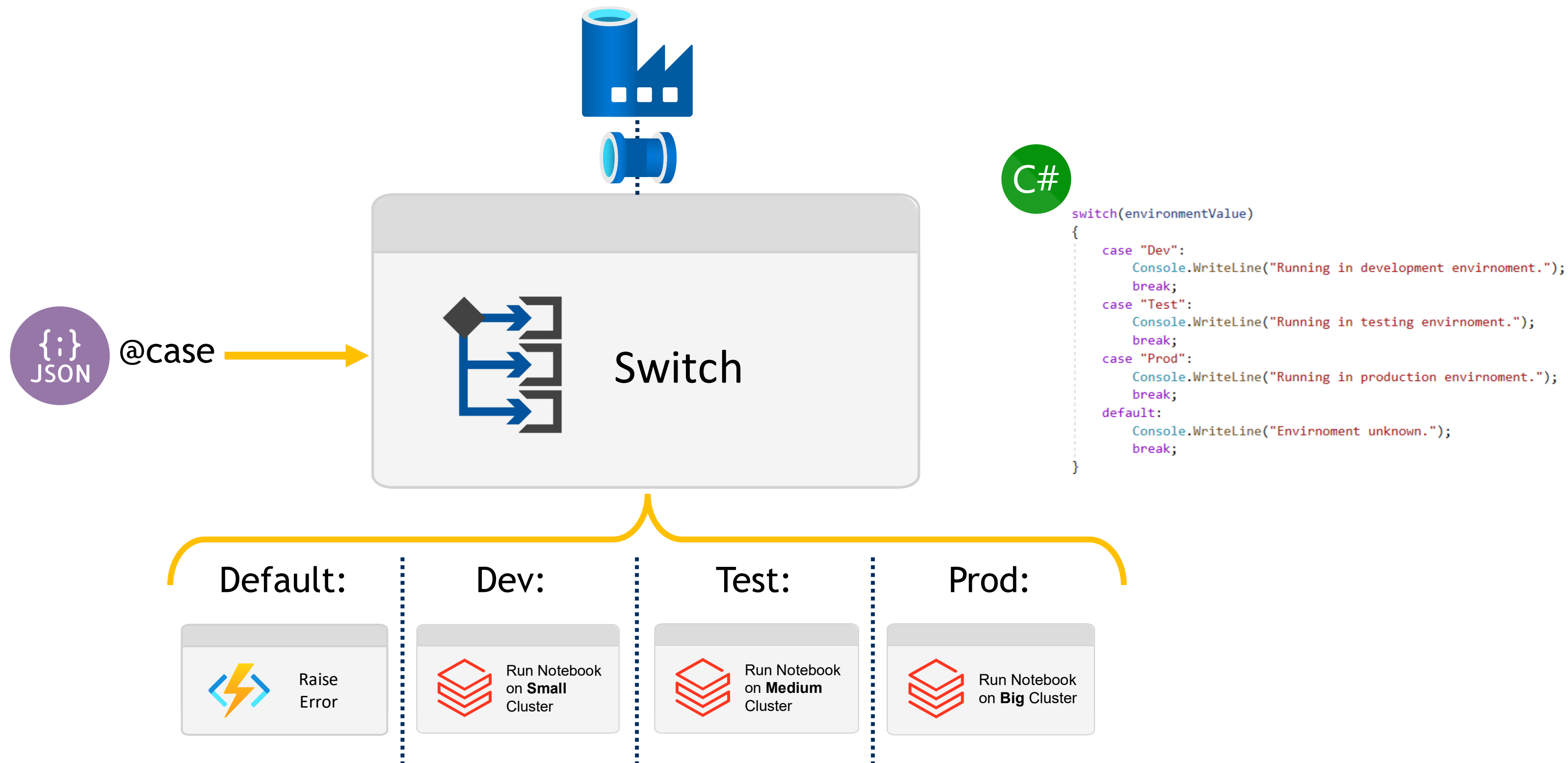
Batch Count Max: 50



Switch

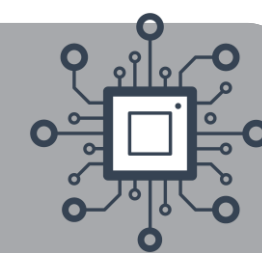


Execute other control flow components based on a provided condition

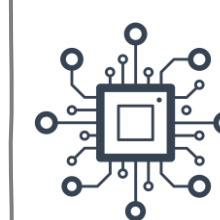
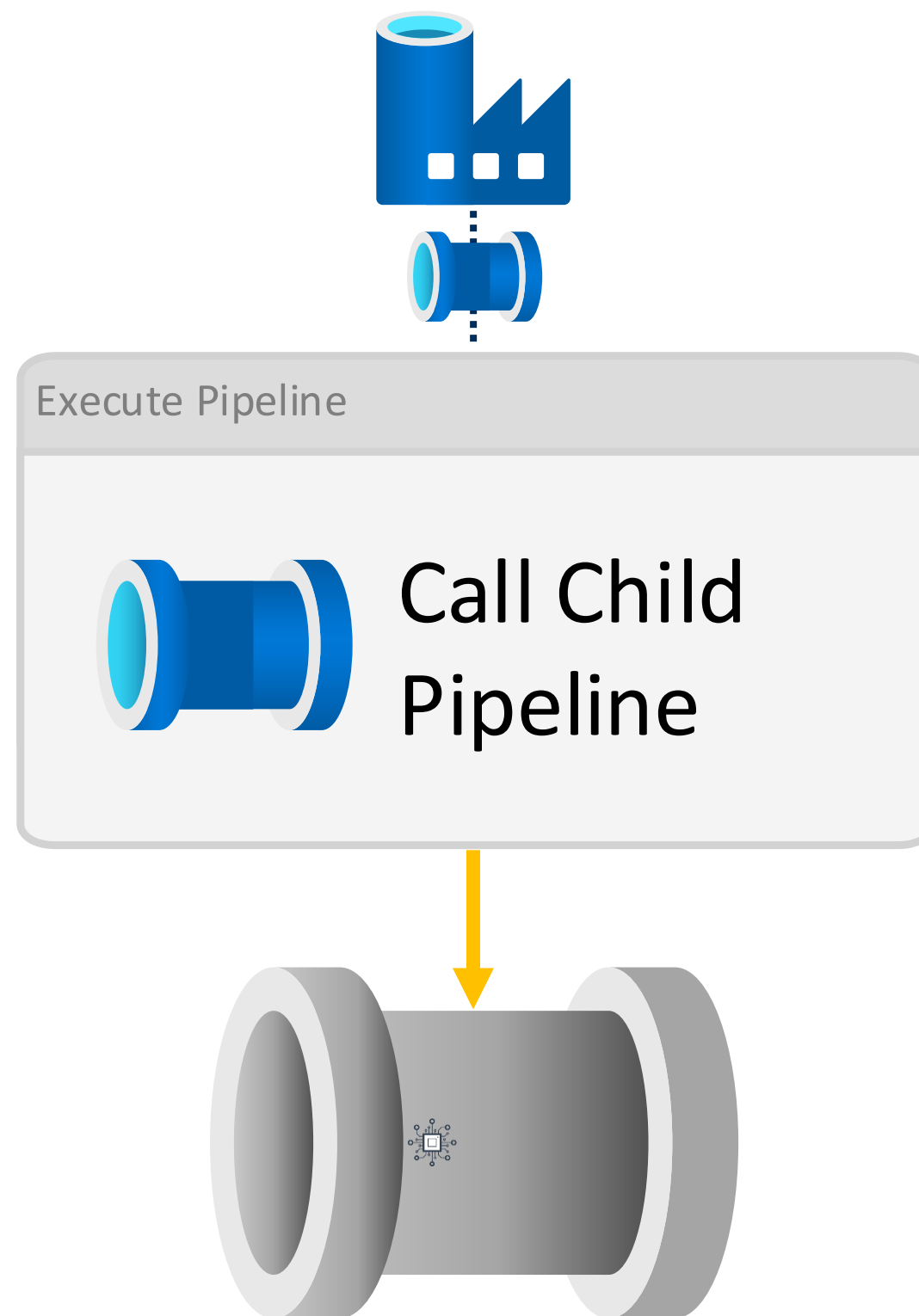




Execute Pipeline

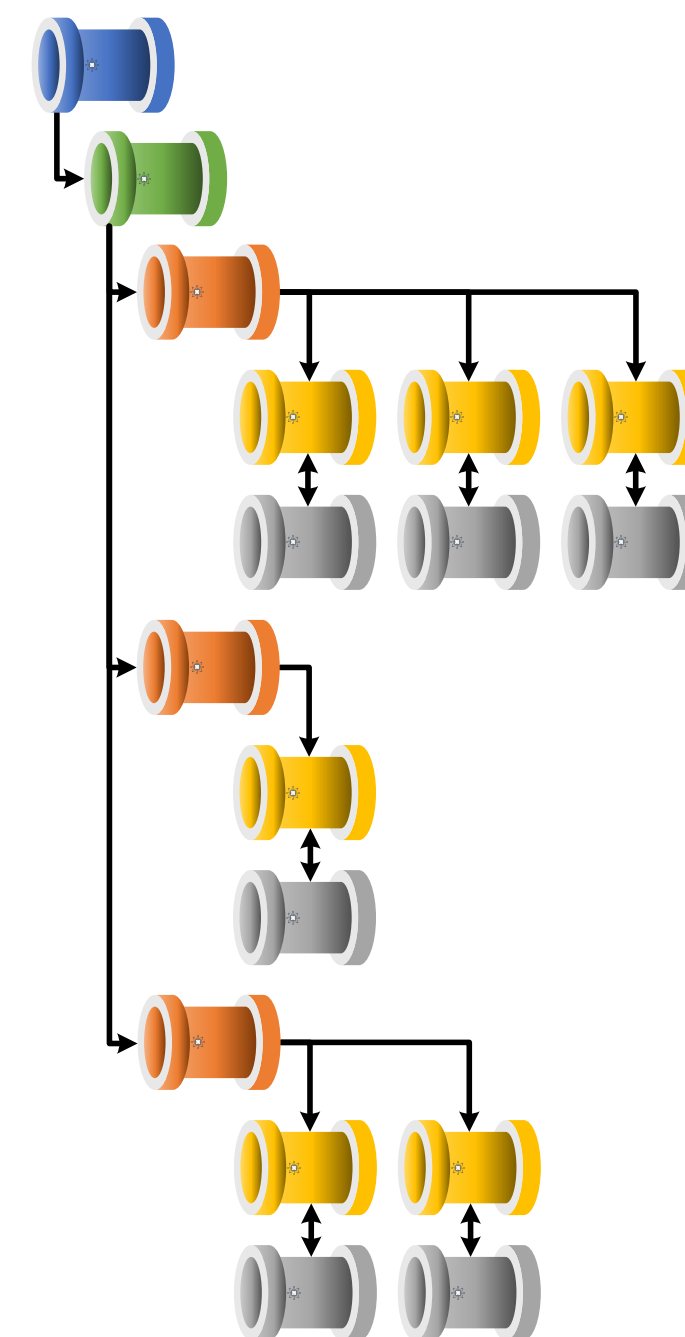


Chaining pipeline executions via an activity



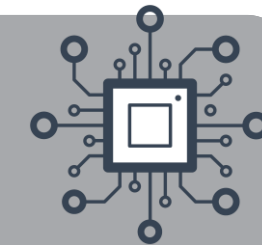
Pipeline Hierarchies Generation Control

<https://mrpaulandrew.com/2019/09/25/azure-data-factory-pipeline-hierarchies-generation-control>

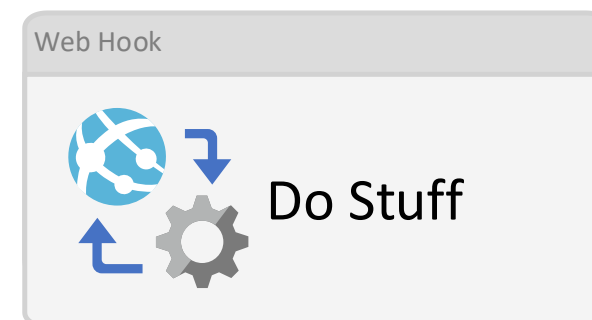
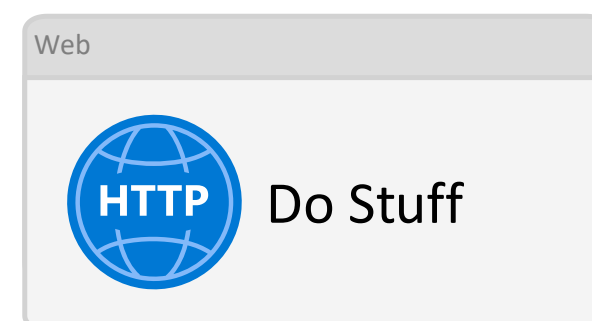
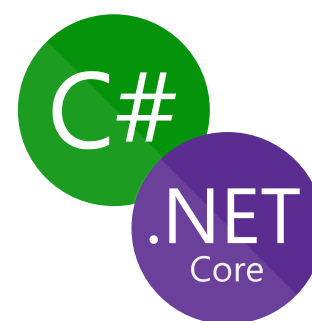
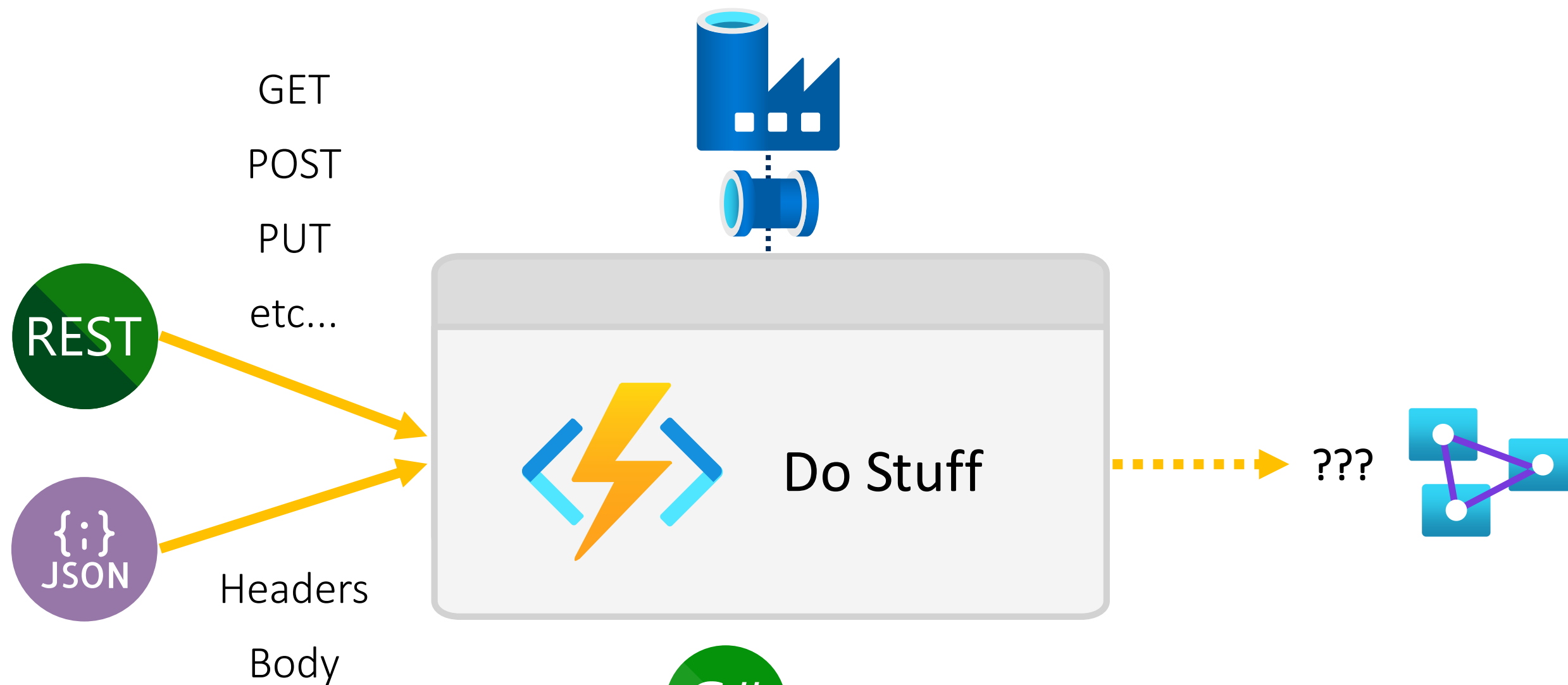




Azure Function

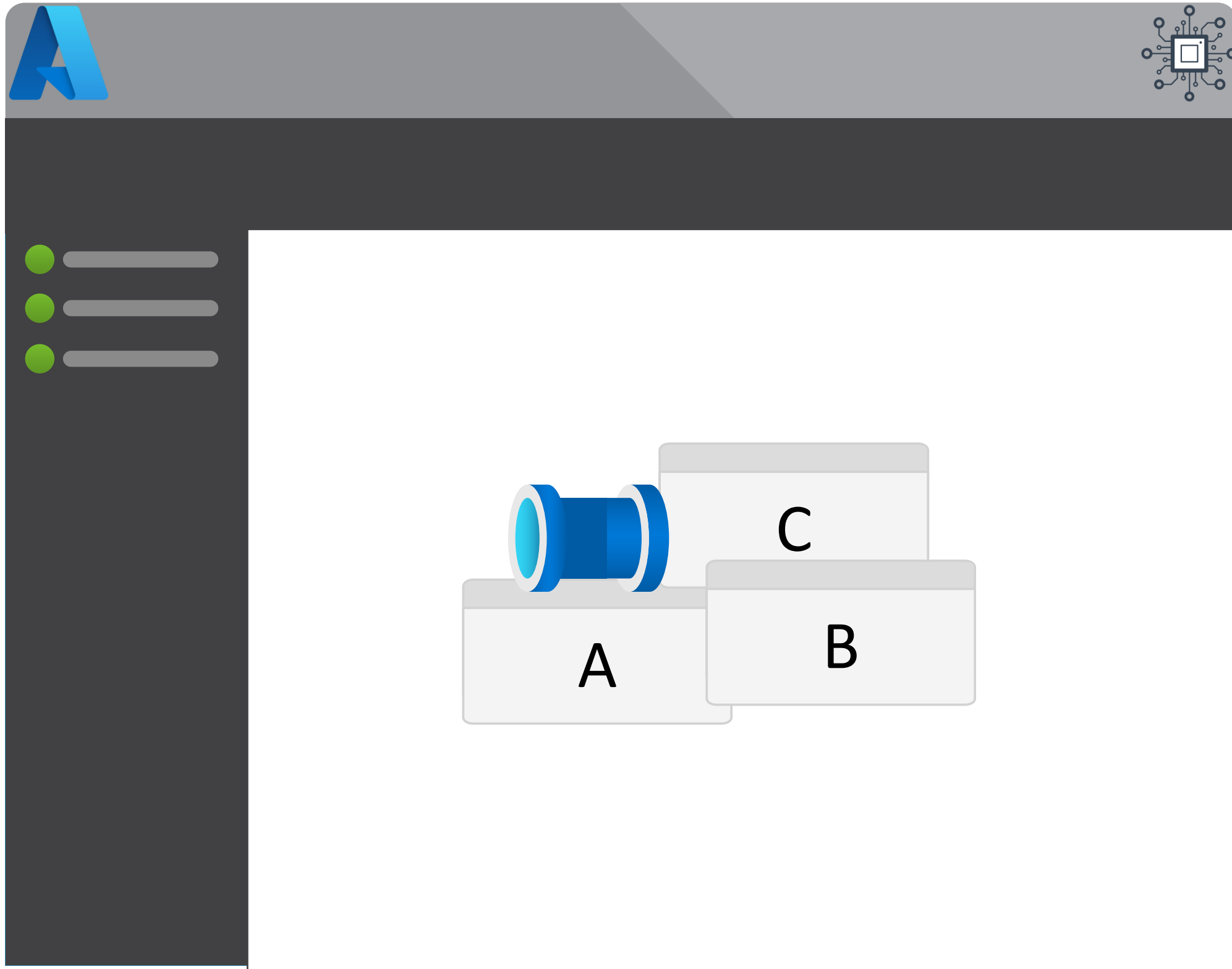


Extend Data Factory with custom serverless code executions via REST calls

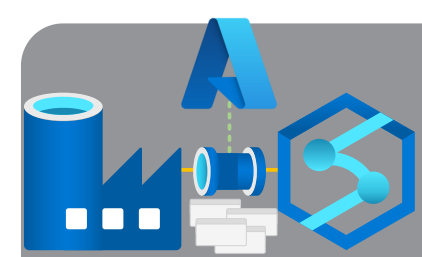


Module 1

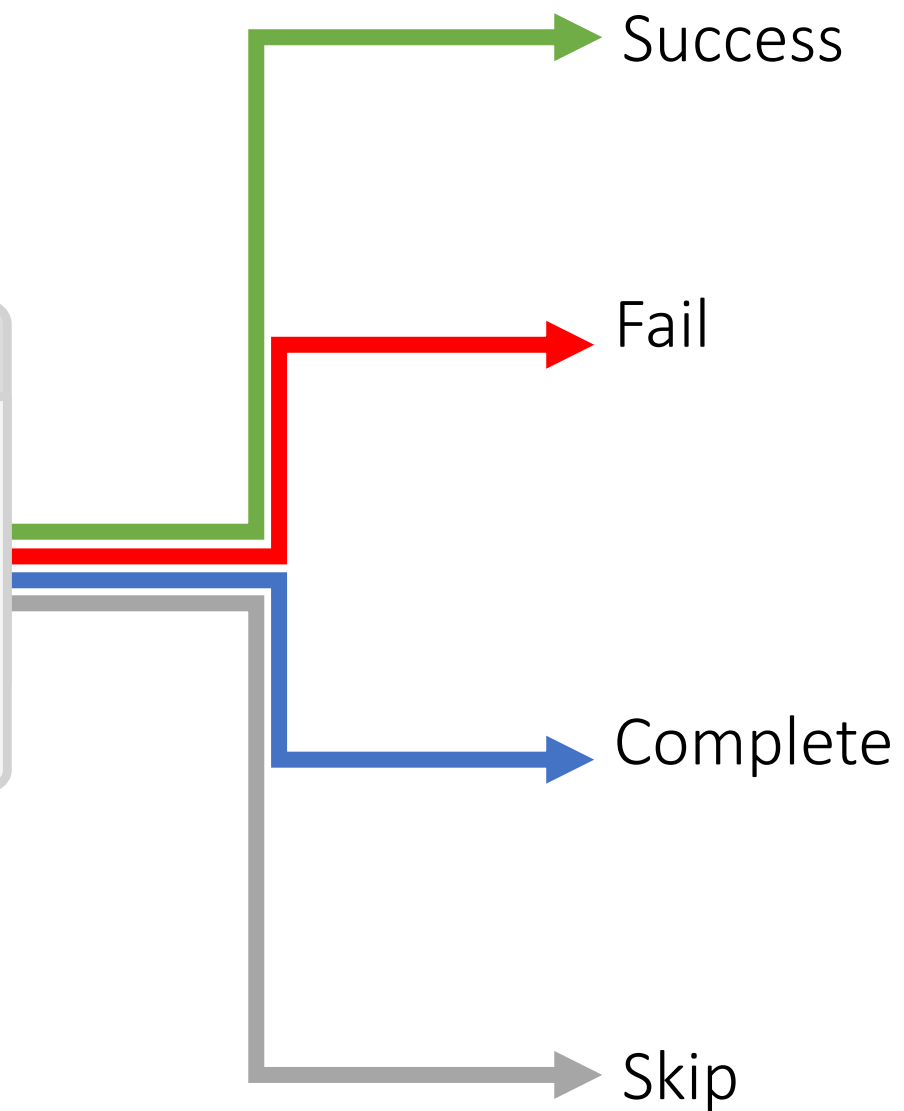
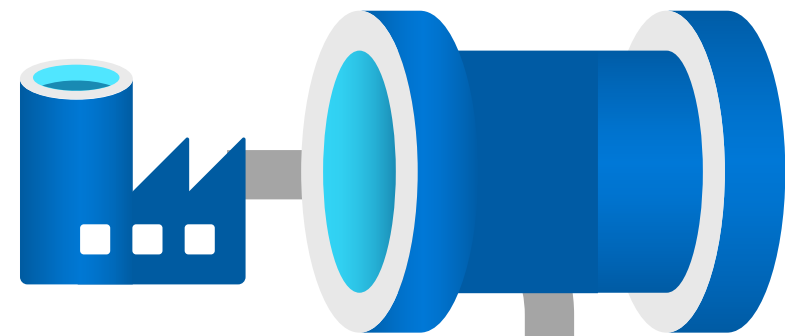
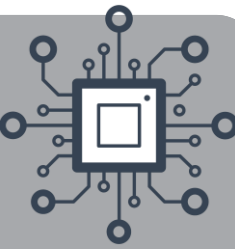
Pipeline Fundamentals

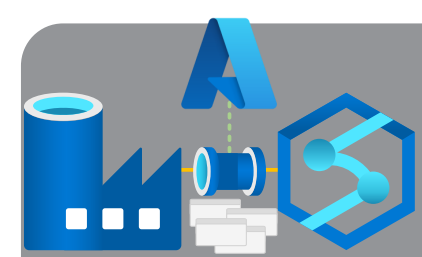


- The History of Azure Orchestration
- Synapse Analytics vs Data Factory
- Integration Components
- Common Activities
- Execution Dependencies

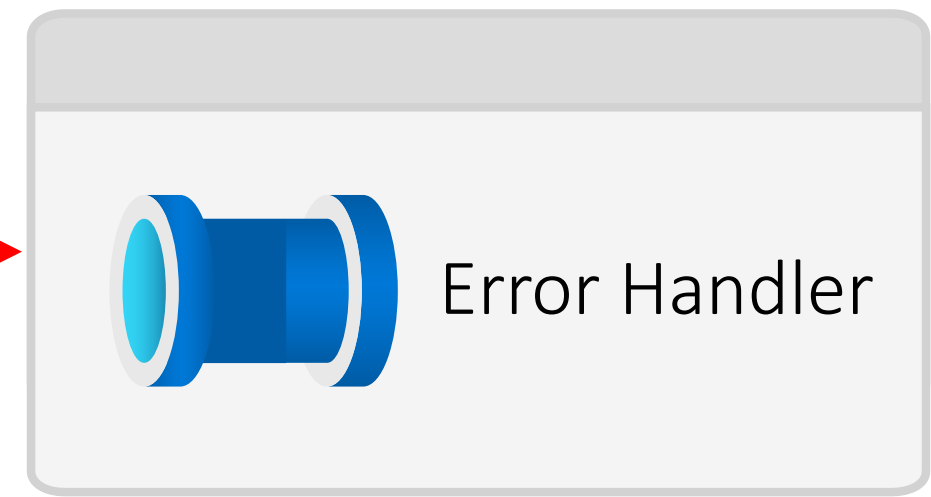
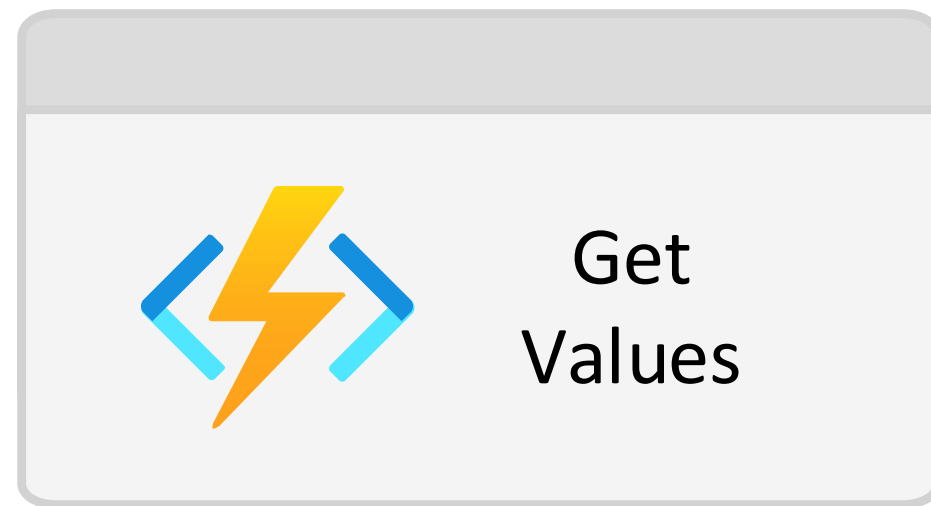
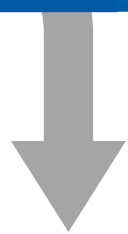
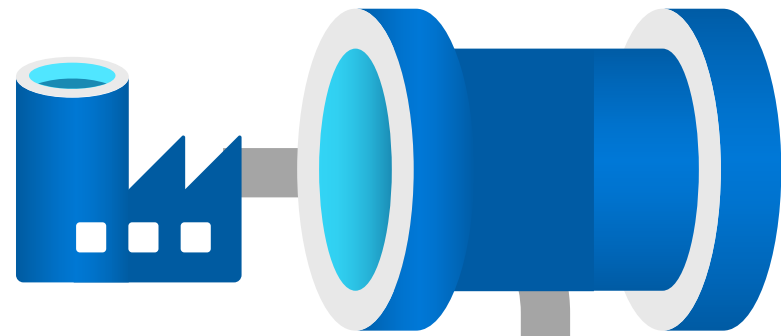
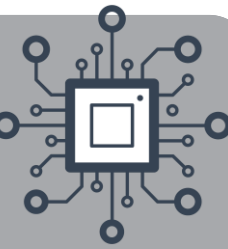


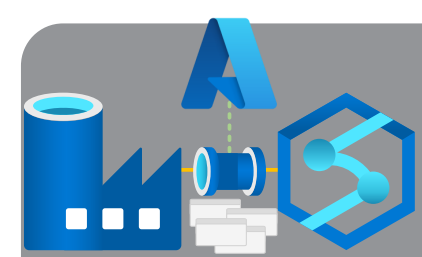
Execution Dependency Options



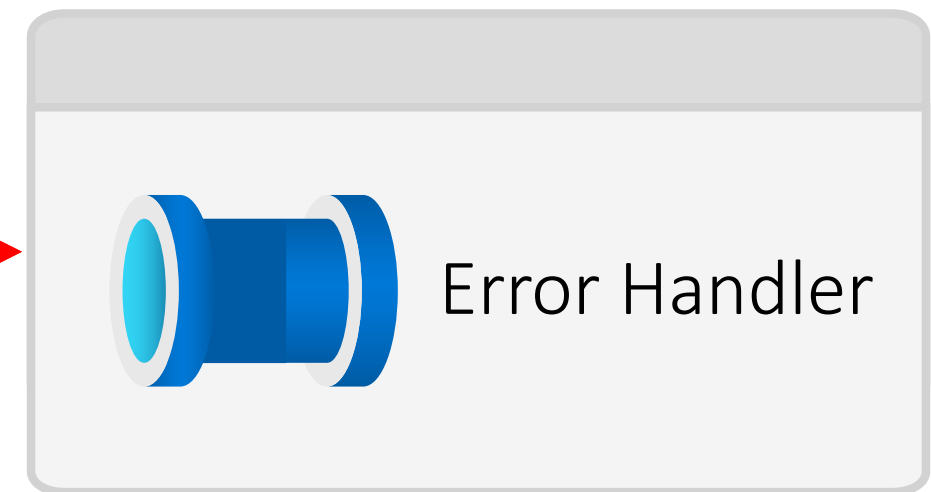
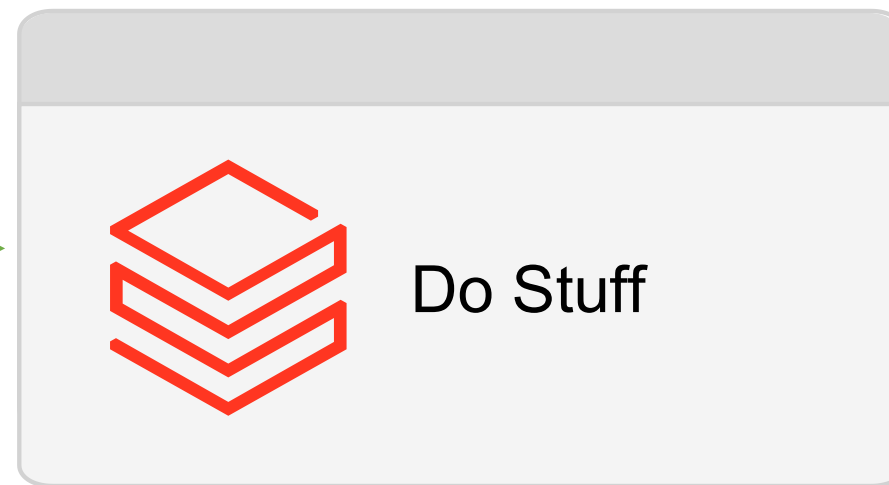
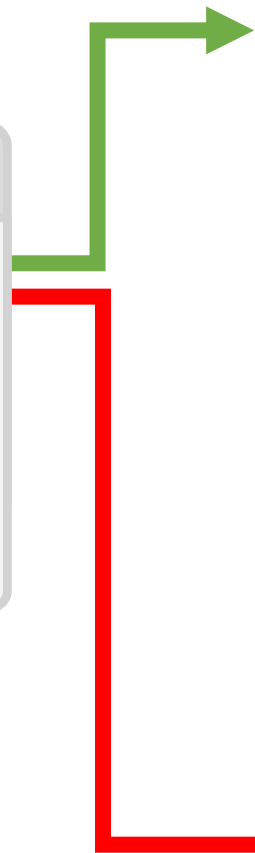
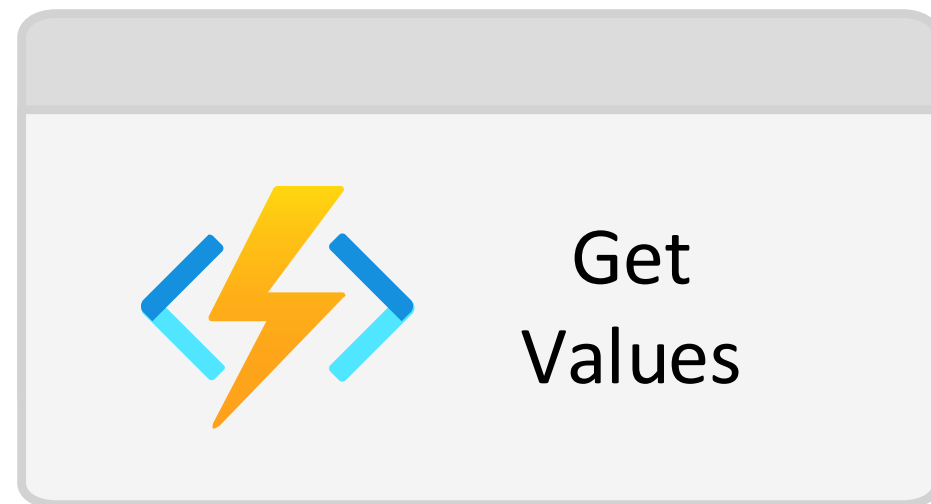
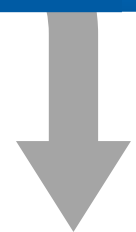
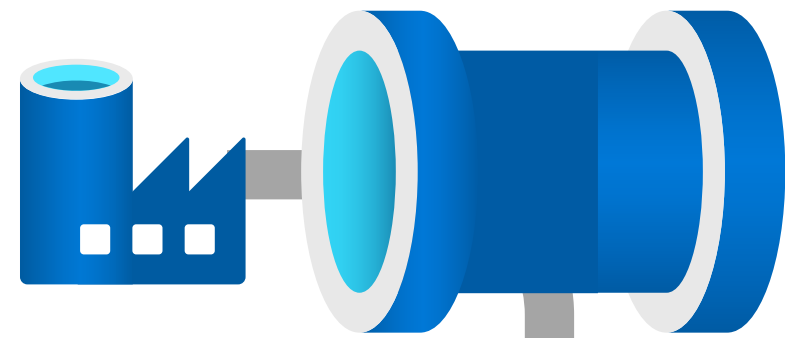
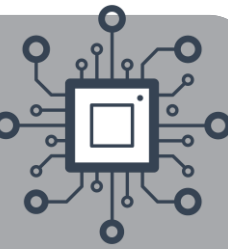


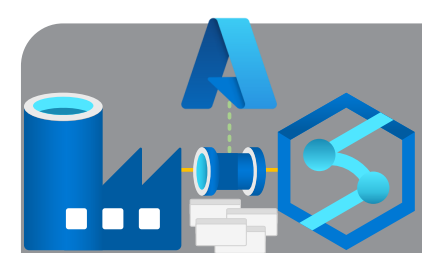
Execution On Failure



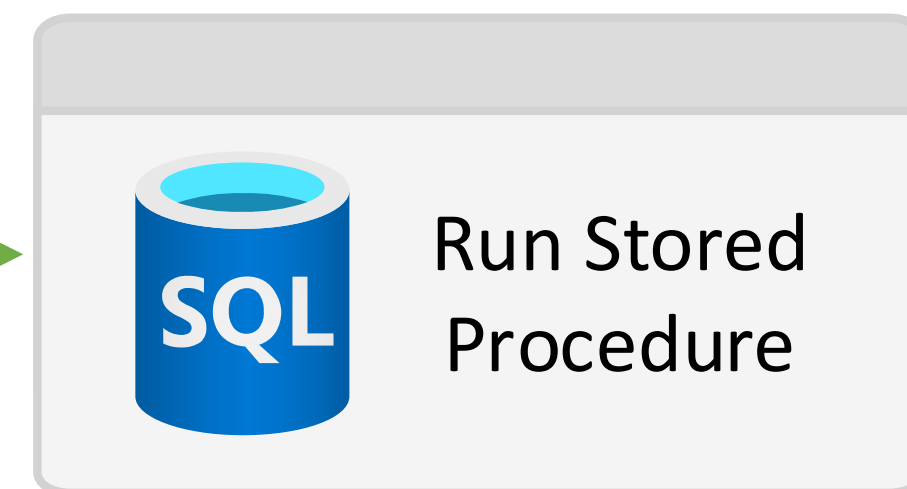
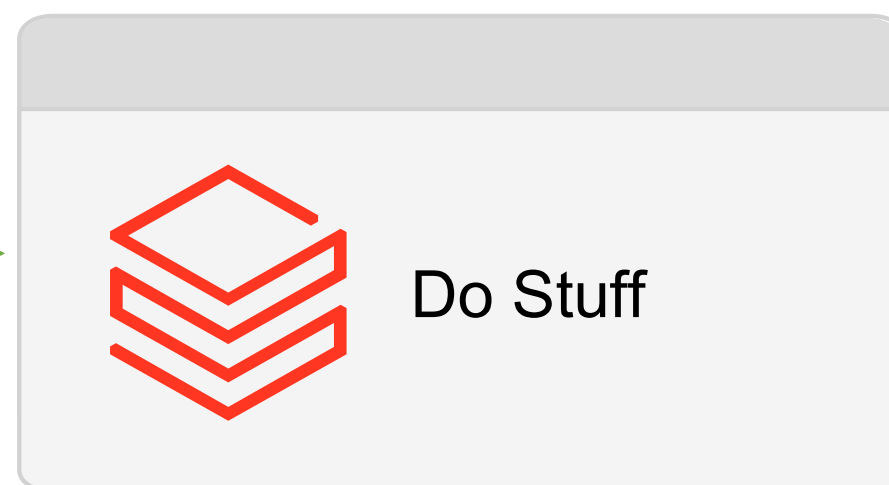
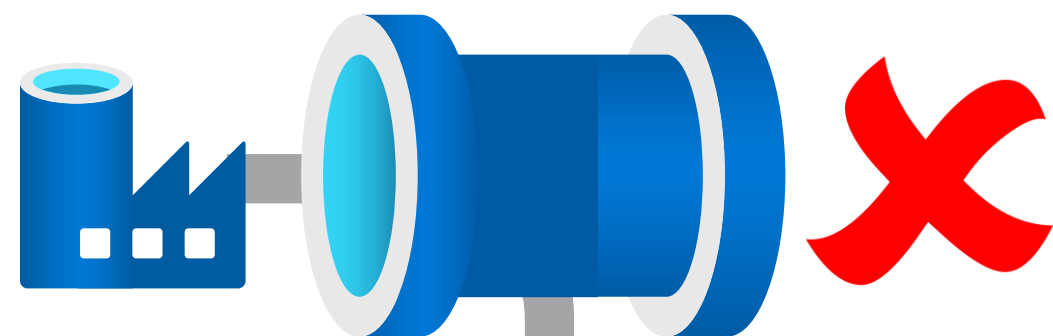
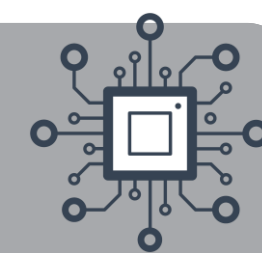


Execution On Failure or On Success



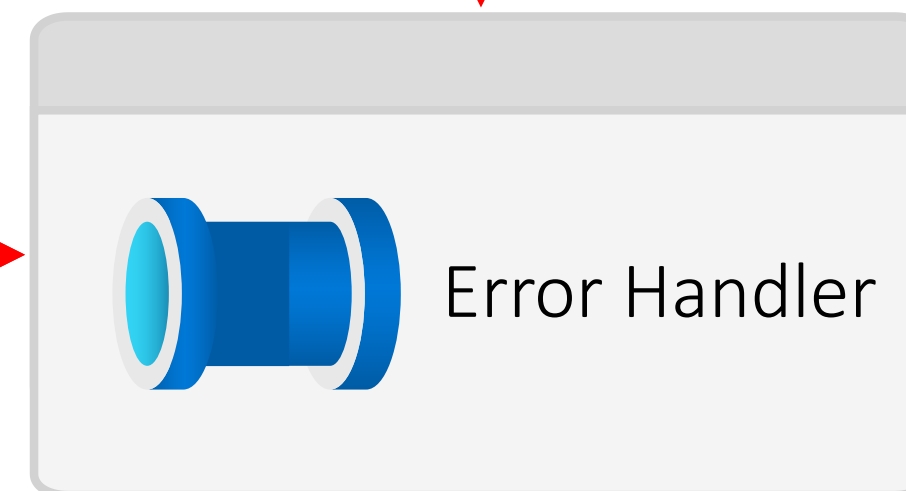


Execution On ???



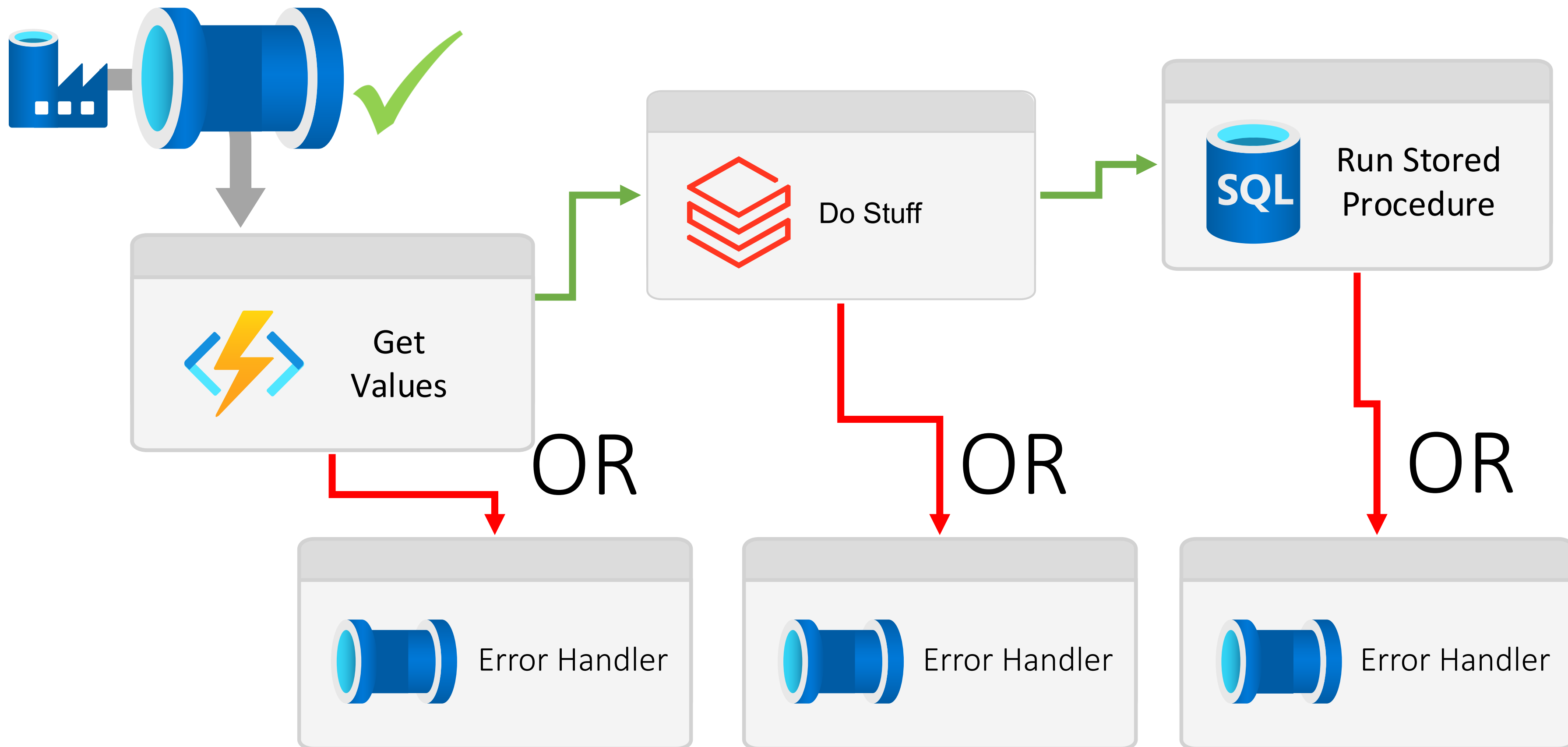
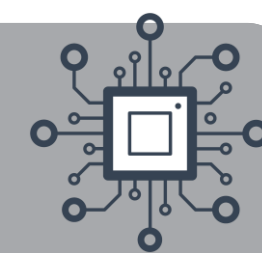
AND

AND



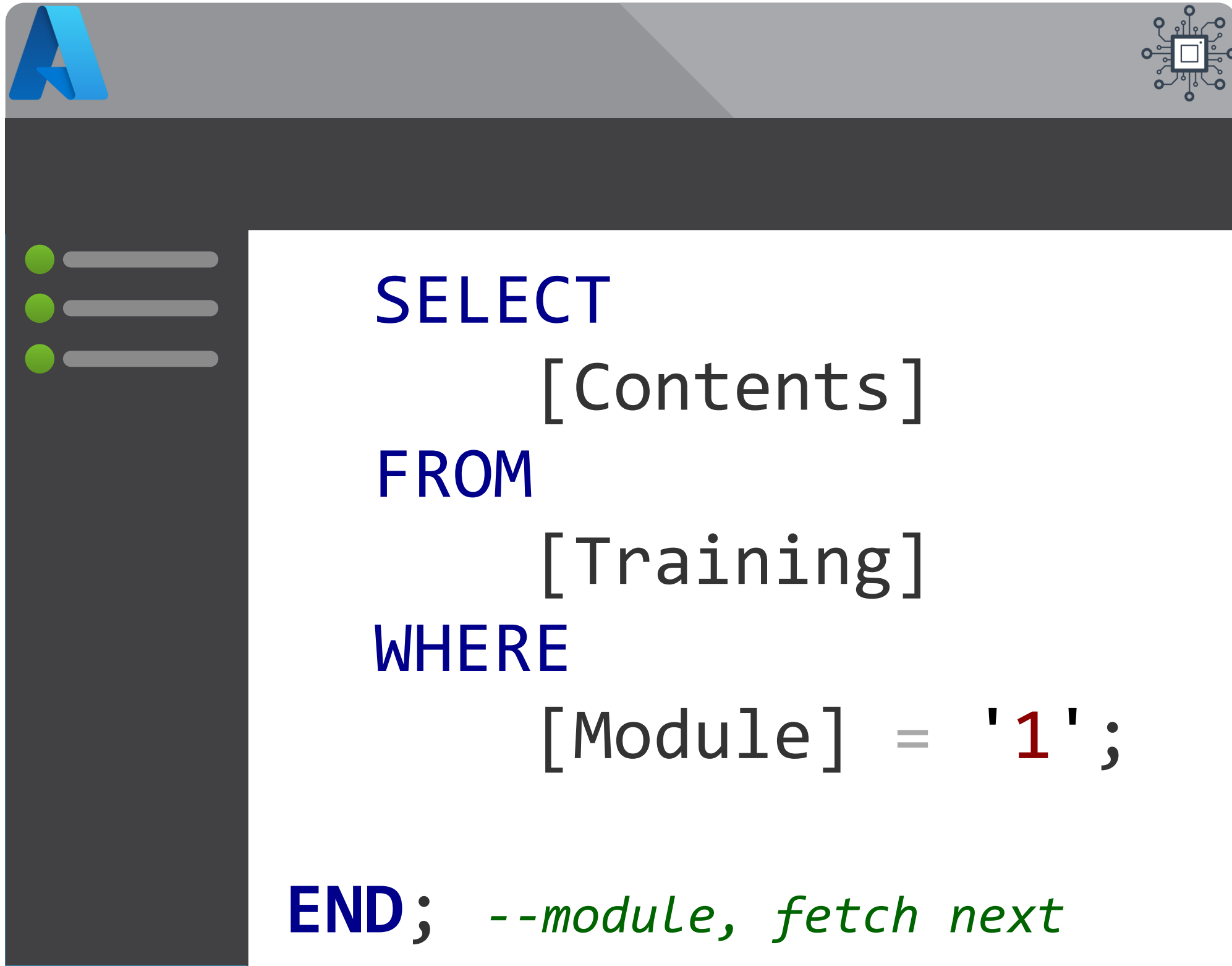


Execution On Failure or On Success



Module 1

Pipeline Fundamentals



- The History of Azure Orchestration
- Synapse Analytics vs Data Factory
- Integration Components
- Common Activities
- Execution Dependencies