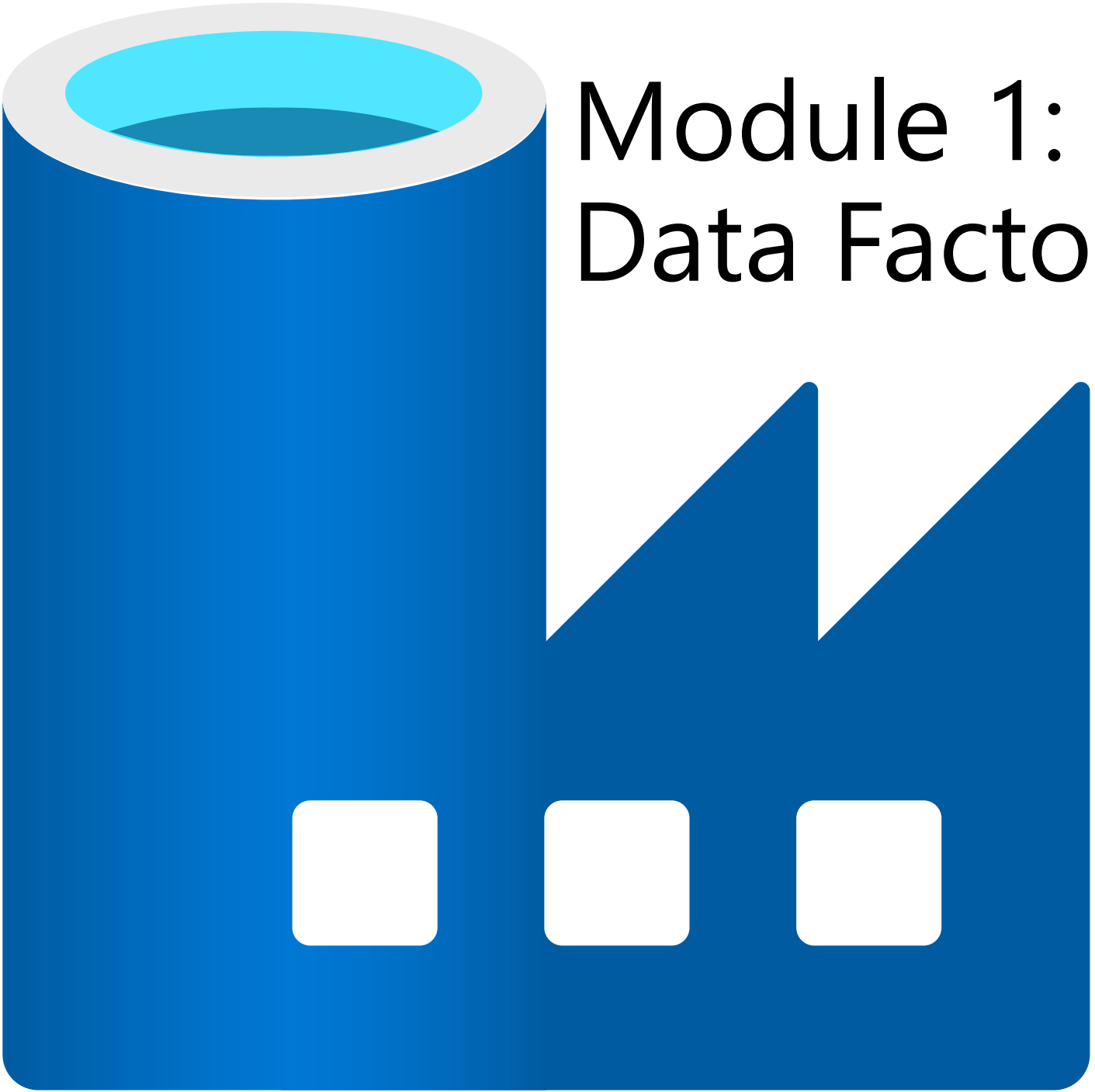
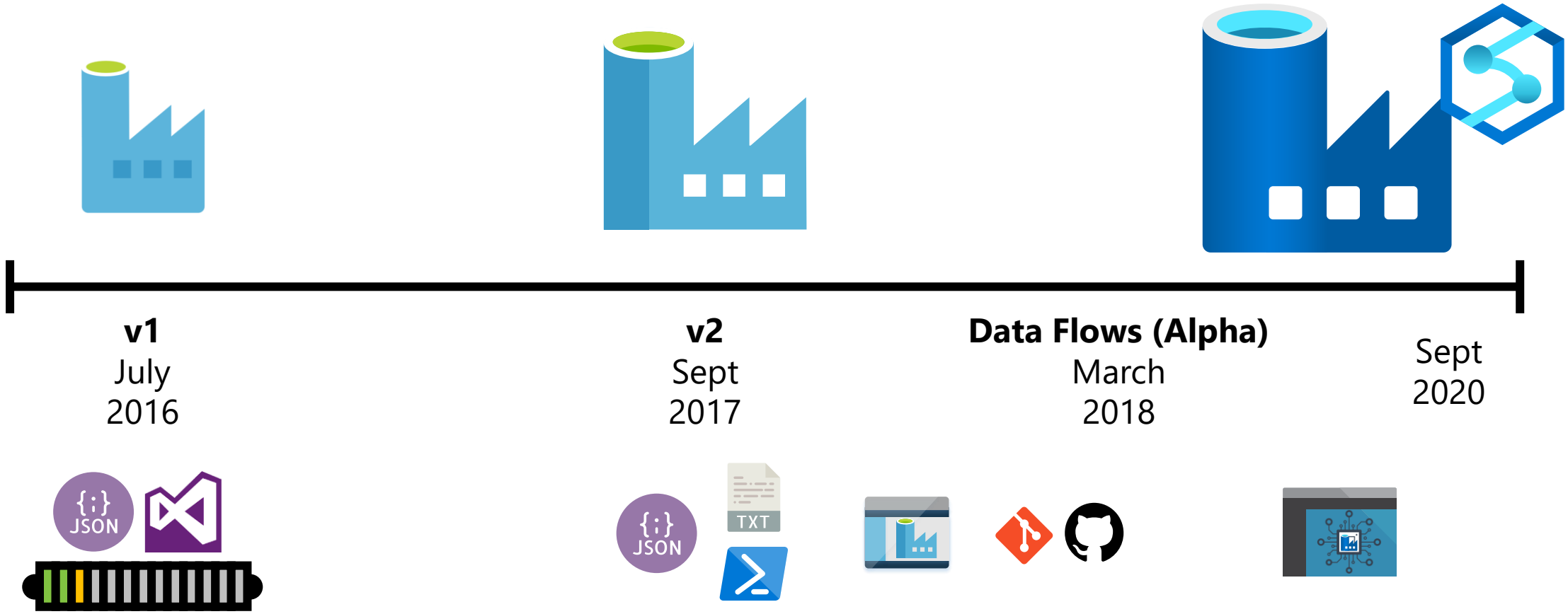


Module 1:
Data Factory Fundamentals

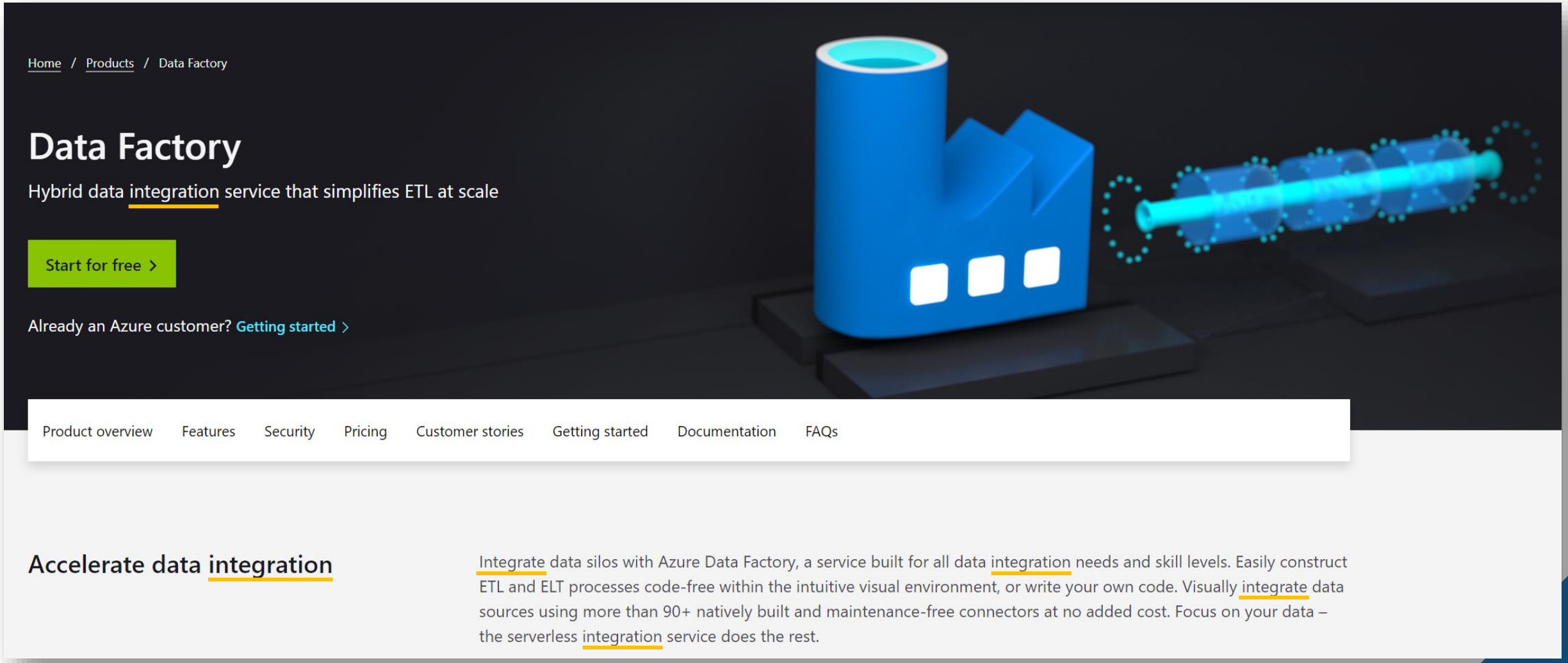
- What is it and why use it?
- Resource Components
- Common Activities
- Execution Dependencies

Azure Data Factory –
What is it?
Why use it?

A Quick History Lesson

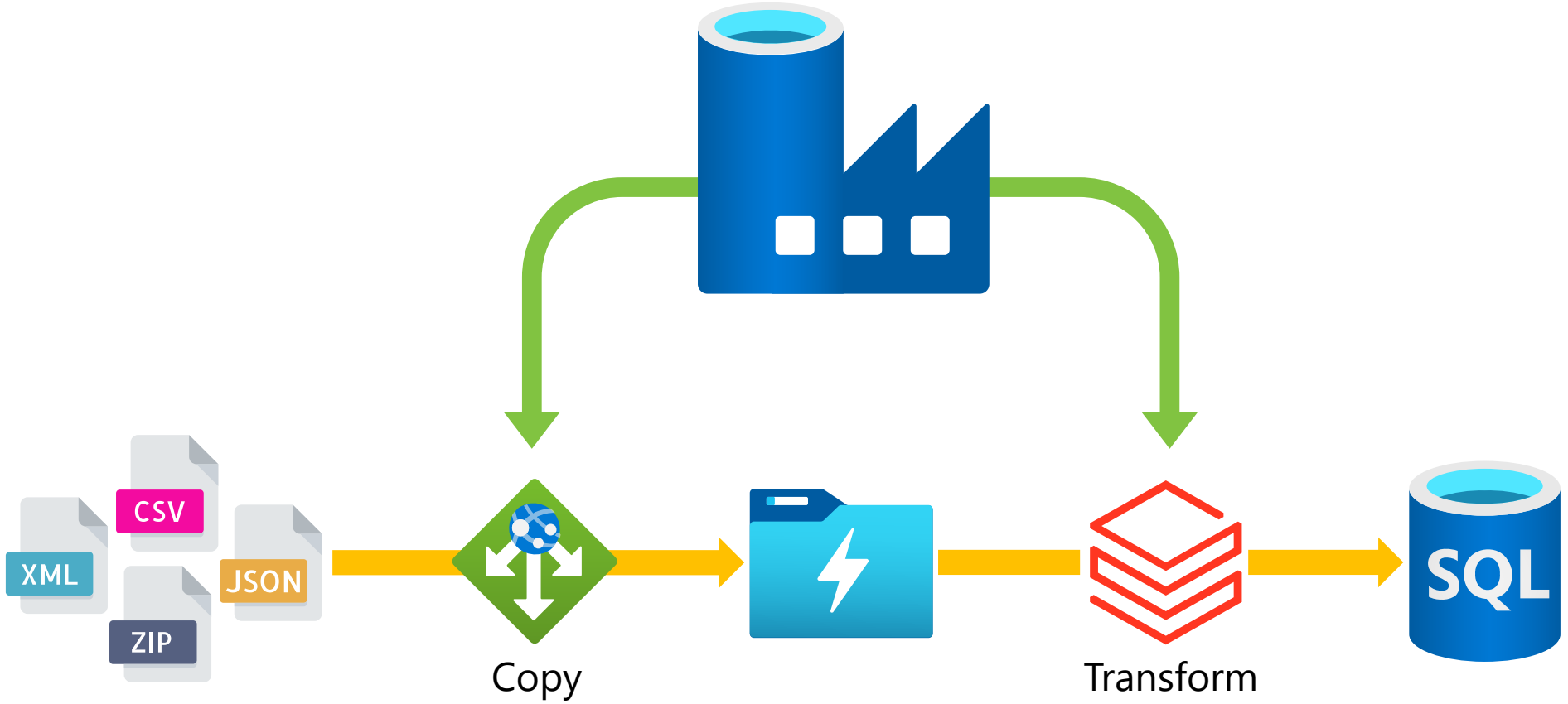


What is Azure Data Factory (ADF)?

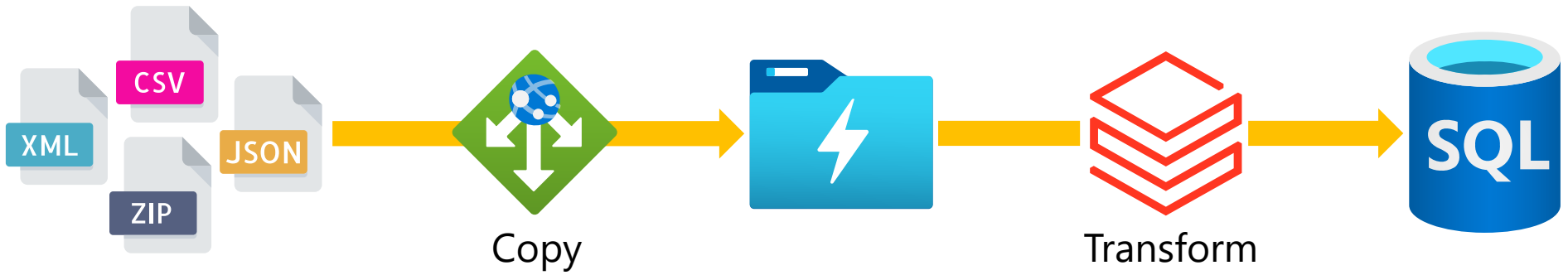


https://azure.microsoft.com/en-gb/services/data-factory

What is Azure Data Factory (ADF)?

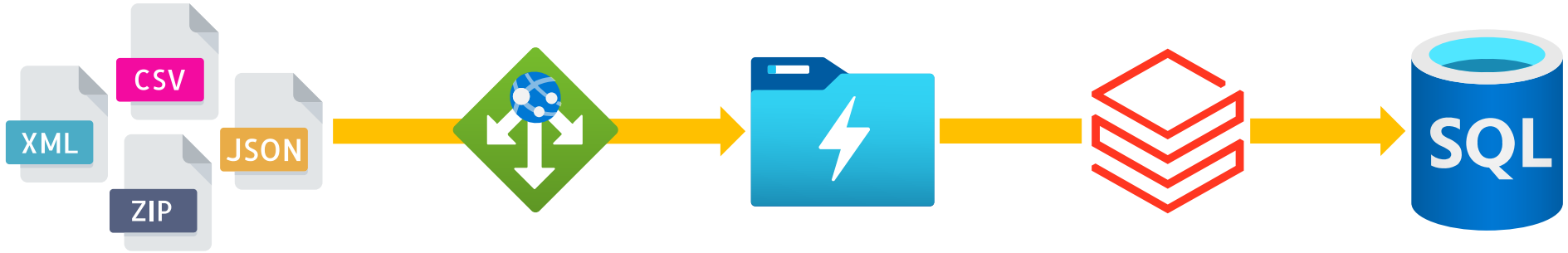


What is Azure Data Factory (ADF)?

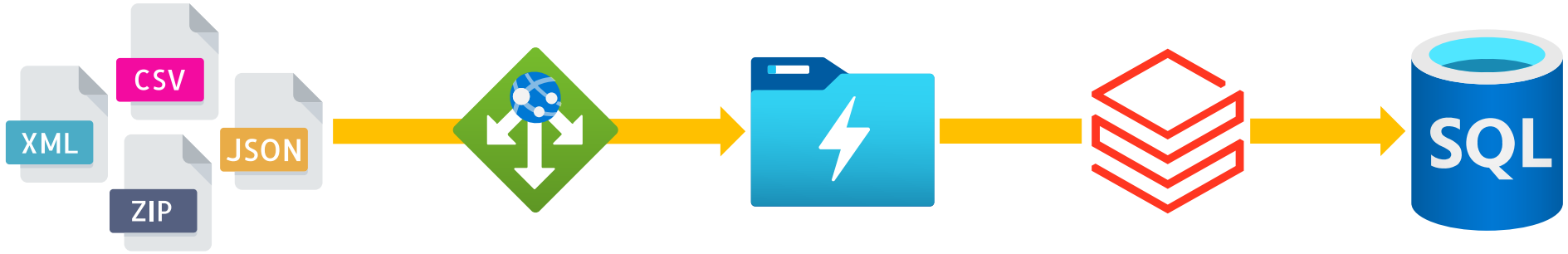


Resource Components

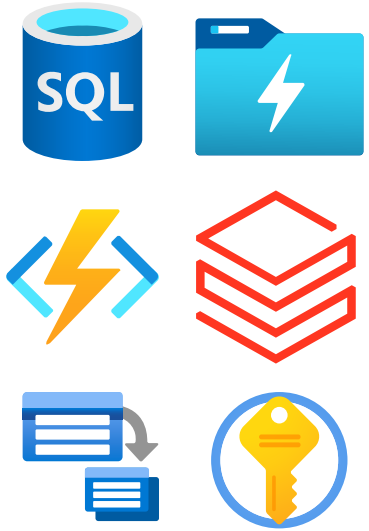
Data Factory Components

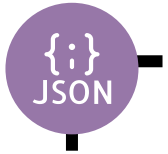


Data Factory Components



1 **Linked Services** – What to interact with and how?

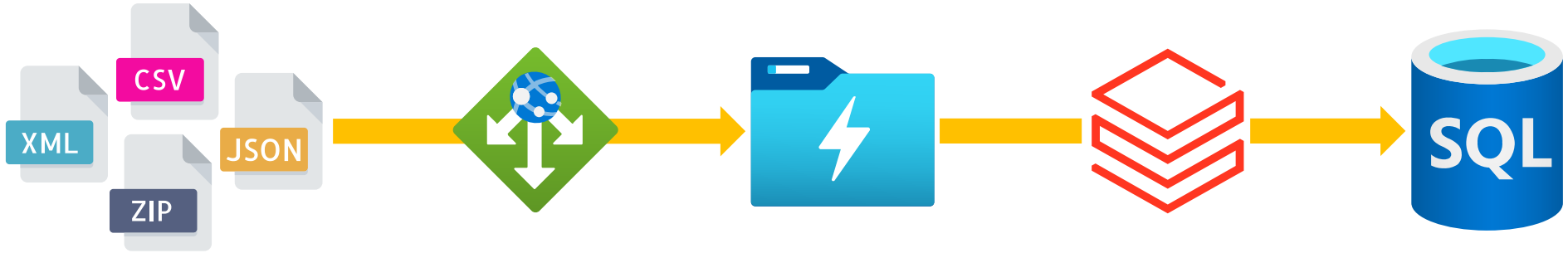




SQLDBLinkedService

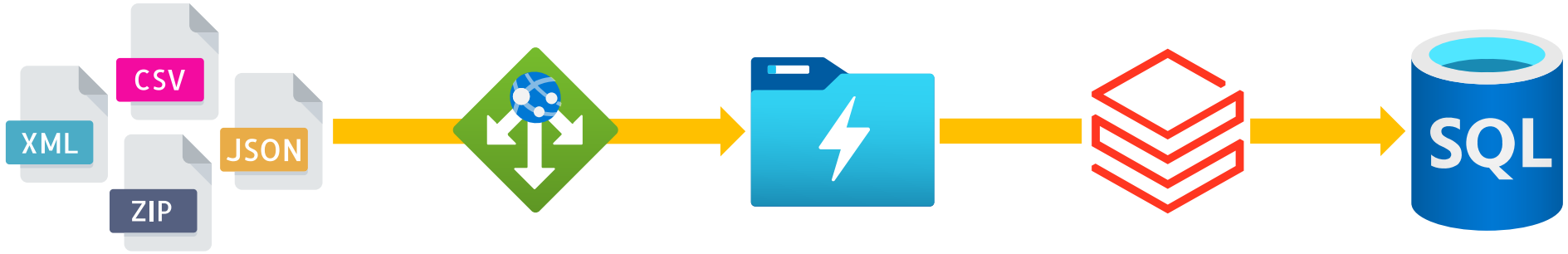
ConnectionString: *Server=MyServer;Database=myDataBase*
UserName: "MrPaulAndrew"
Password: **********

Data Factory Components



1 **Linked Services**

Data Factory Components

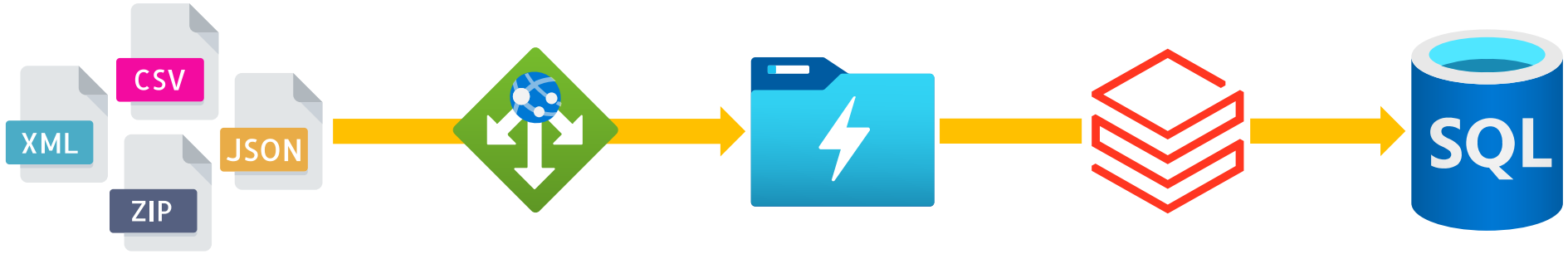


1. **Linked Services**
2. **Datasets** – Where is my data? What format? What file path/table do I need?

[dbo].[SalesOrders]

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

Data Factory Components



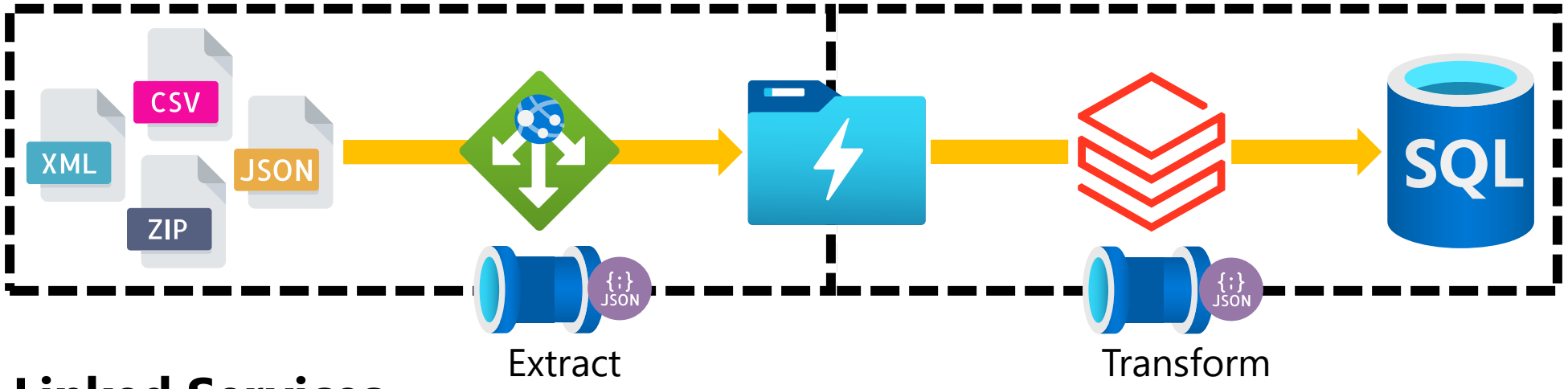
1. **Linked Services**
2. **Datasets**
3. **Activities** – What do we want to happen when we invoke a Linked Service? With what conditions?

JSON

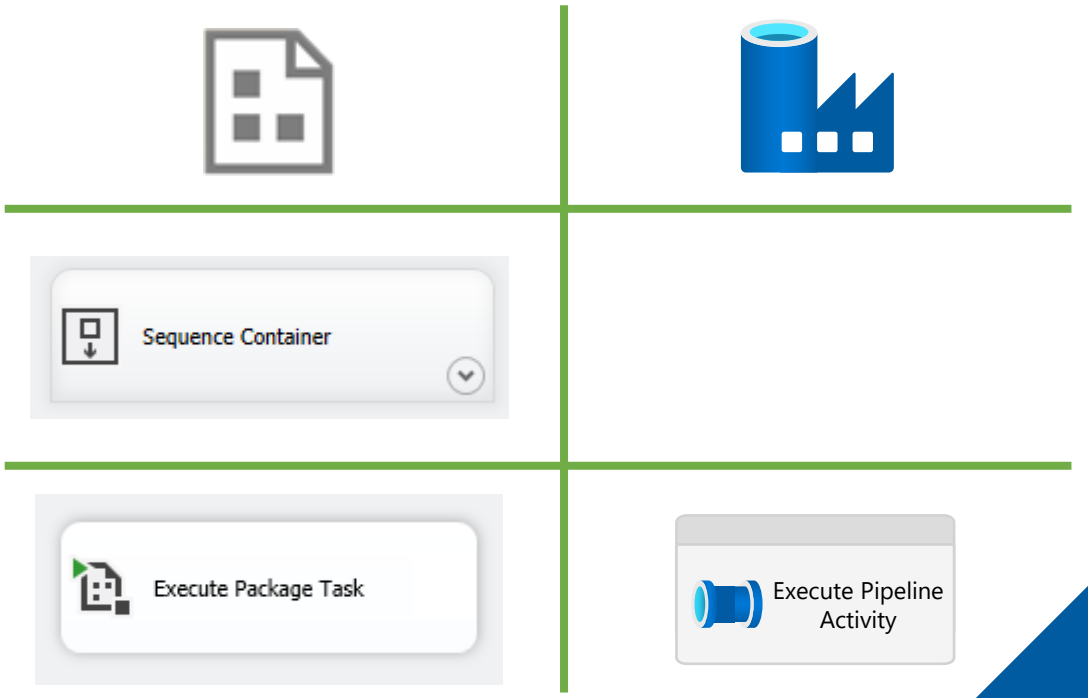
Databricks Notebook Activity

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

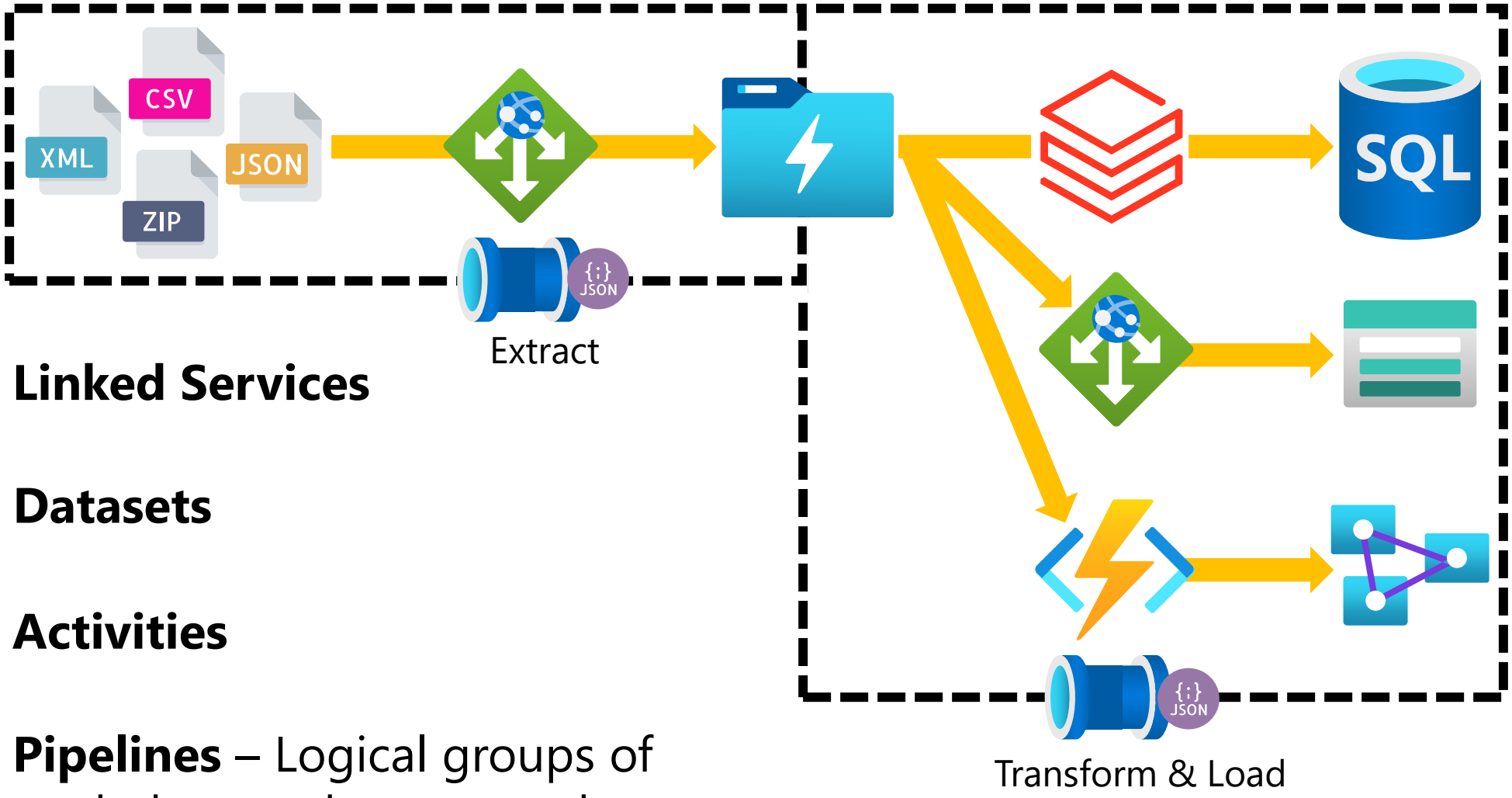
Data Factory Components



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

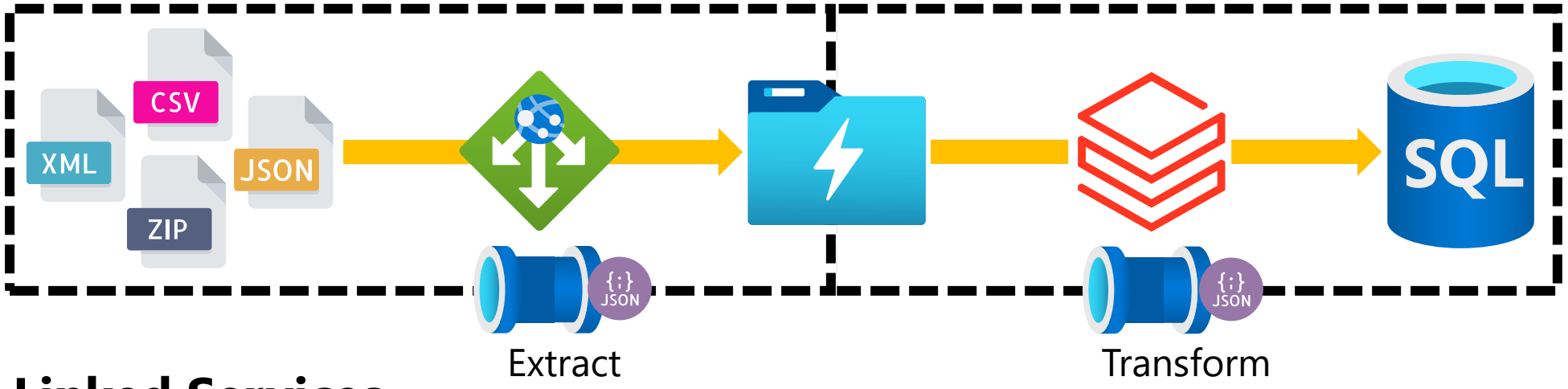


Data Factory Components

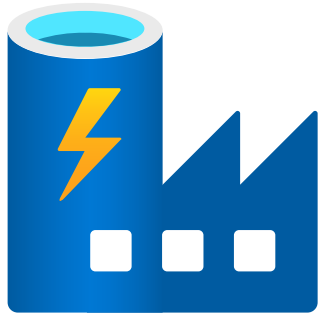


1. **Linked Services**
2. **Datasets**
3. **Activities**
4. **Pipelines** – Logical groups of work that can be executed.

Data Factory Components

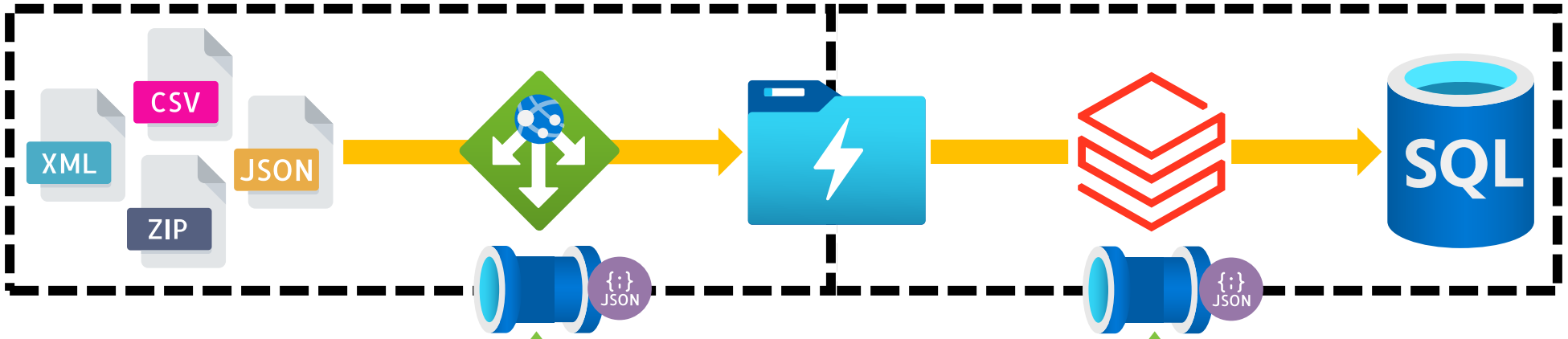


1. **Linked Services**
2. **Datasets**
3. **Activities**
4. **Pipelines**
5. **Triggers** – Telling our when pipelines to run.

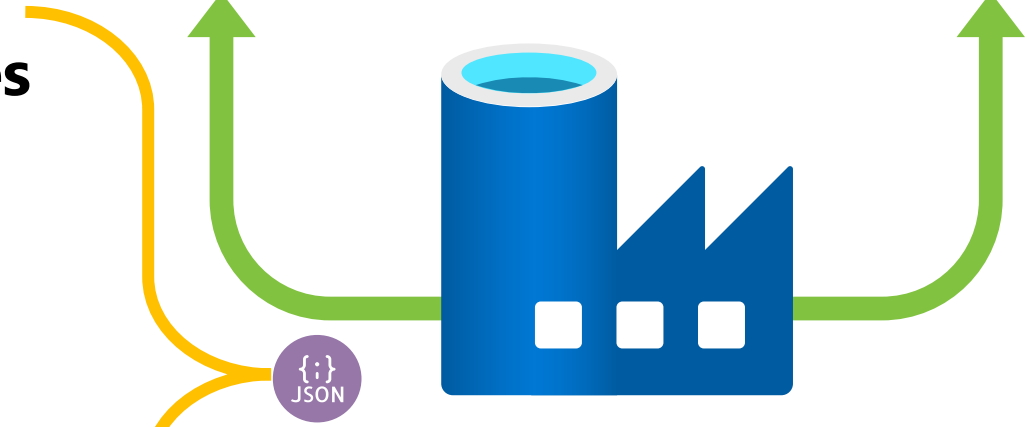


- Manually
- Programmatically
- Schedule
- Tumbling Windows
- File Event

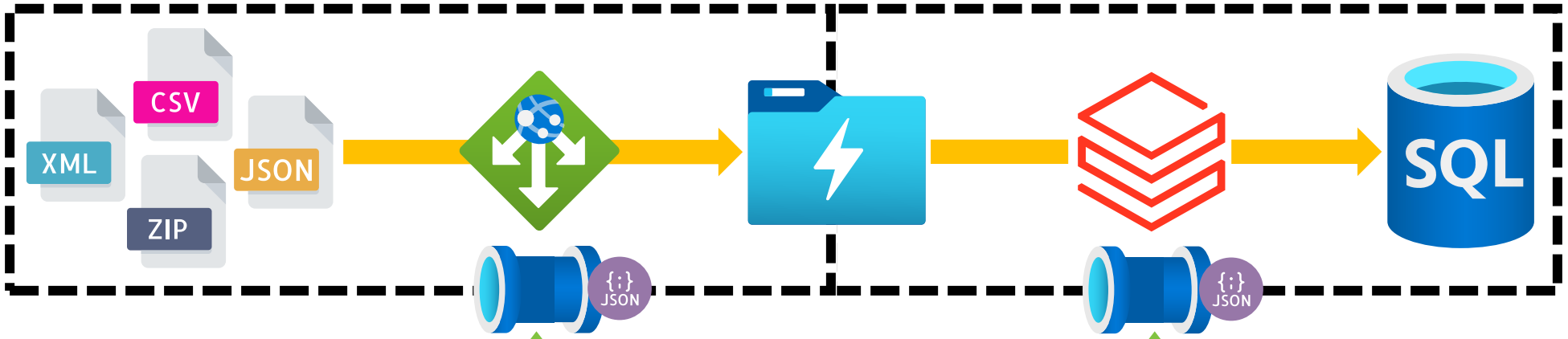
Data Factory Components



1. **Linked Services**
2. **Datasets**
3. **Activities**
4. **Pipelines**
5. **Triggers**



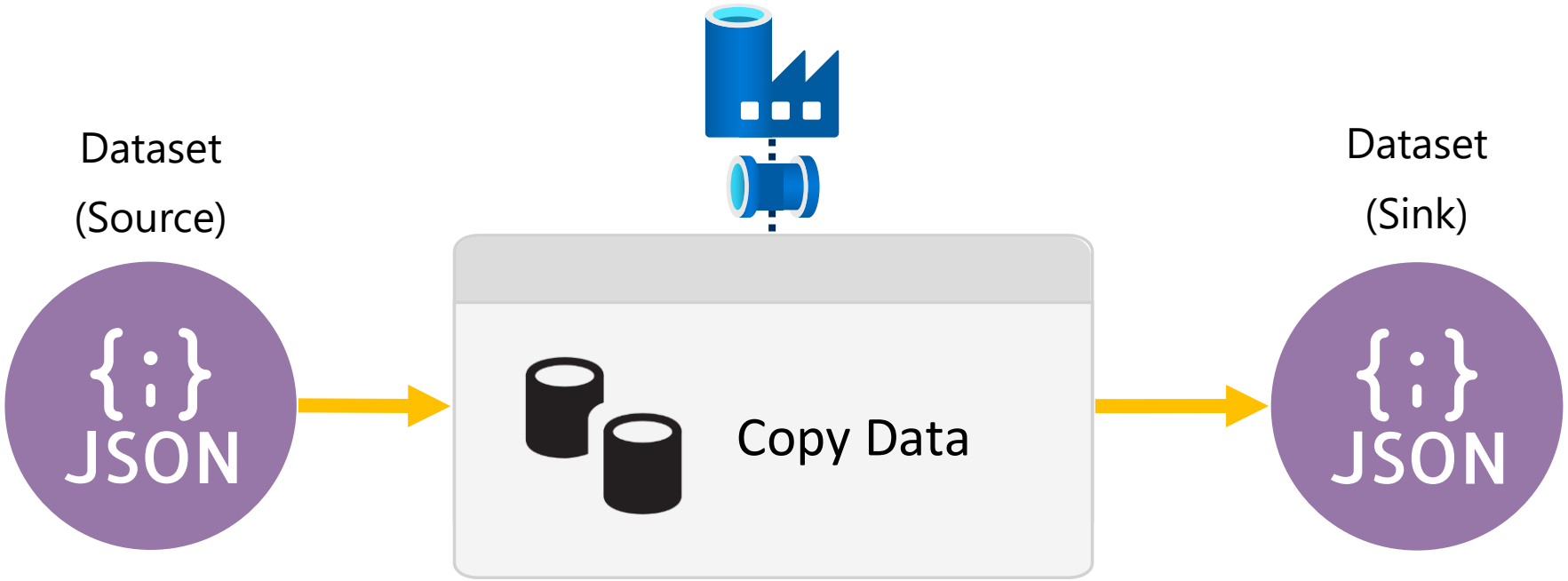
Data Factory Control Flow Components



1. **Linked Services**
2. **Datasets**
3. **Activities**
4. **Pipelines**
5. **Triggers**

Common Activities

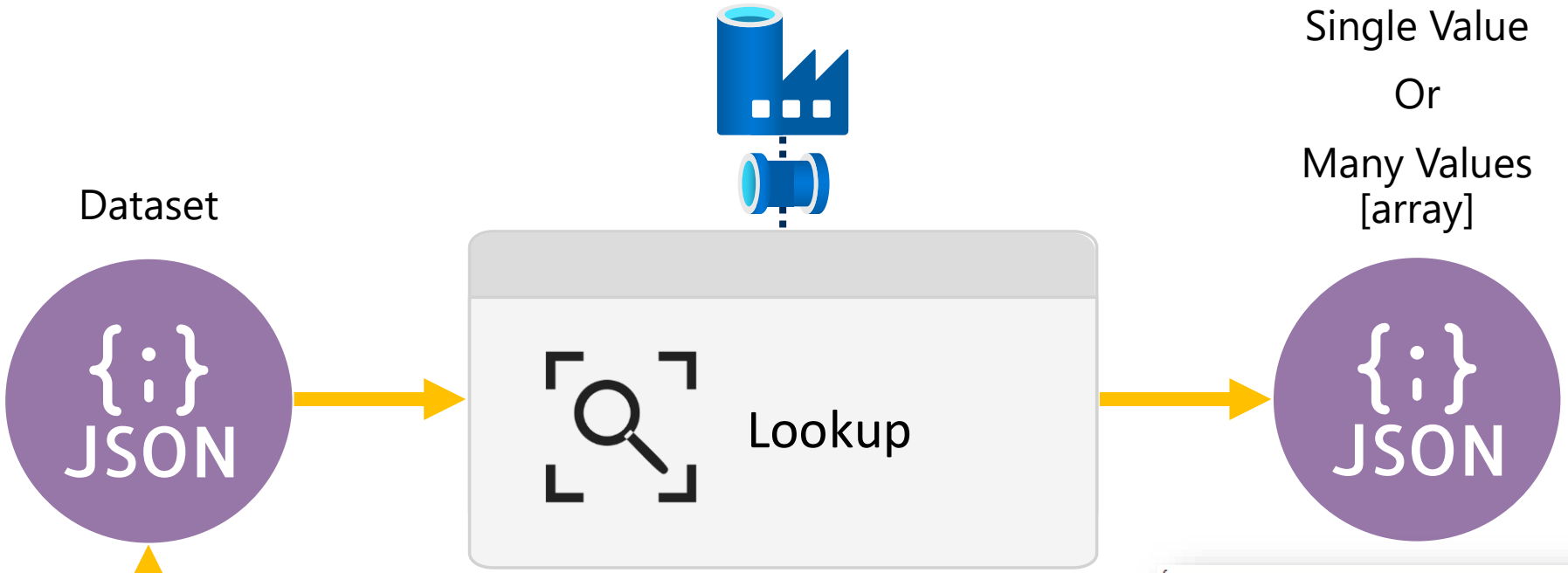
Copy



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

Lookup

Get Metadata to Support Other Control Flow Activities



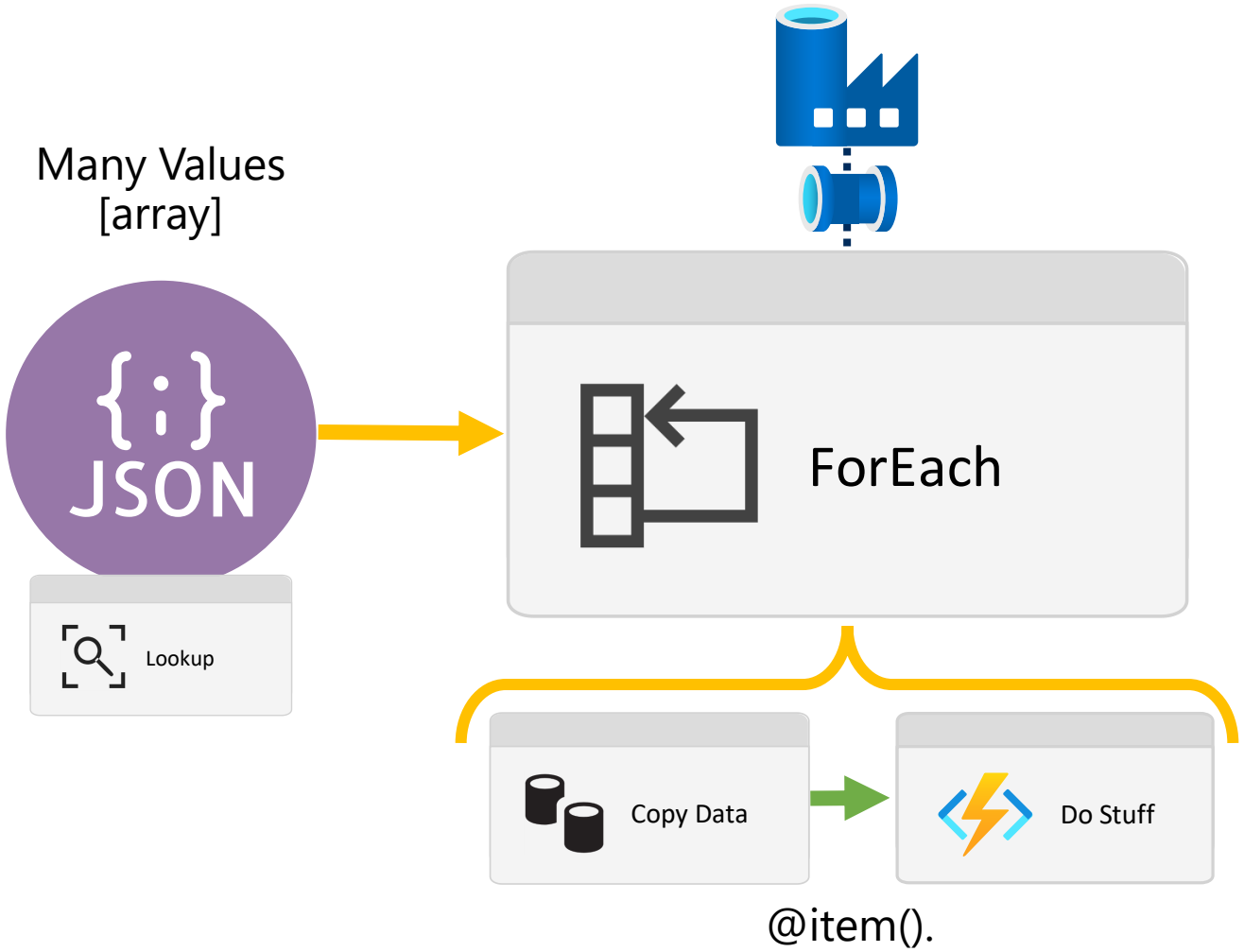

```
SELECT
    [SourceDIR],
    [TargetDIR],
    [FileName]
FROM
    [dbo].[FileList]
```


```
{
    "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"
        }
    ]
}
```

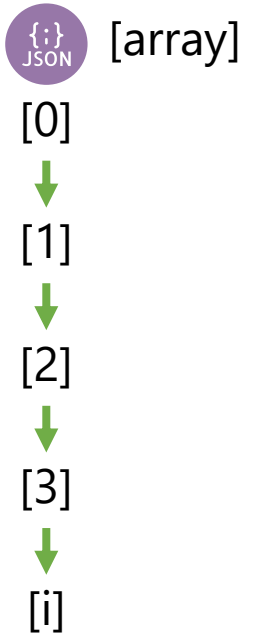

https://docs.microsoft.com/en-us/azure/data-factory/control-flow-lookup-activity

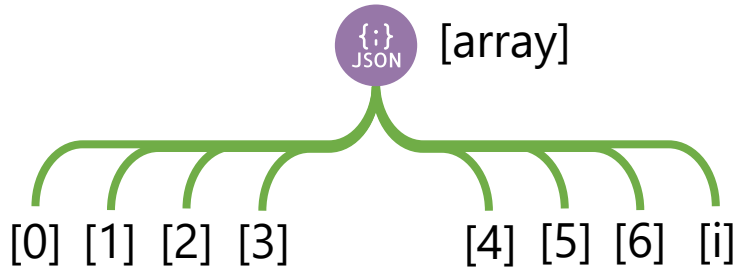
ForEach

Scaling Out Control Flow Activities



IsSequential: true



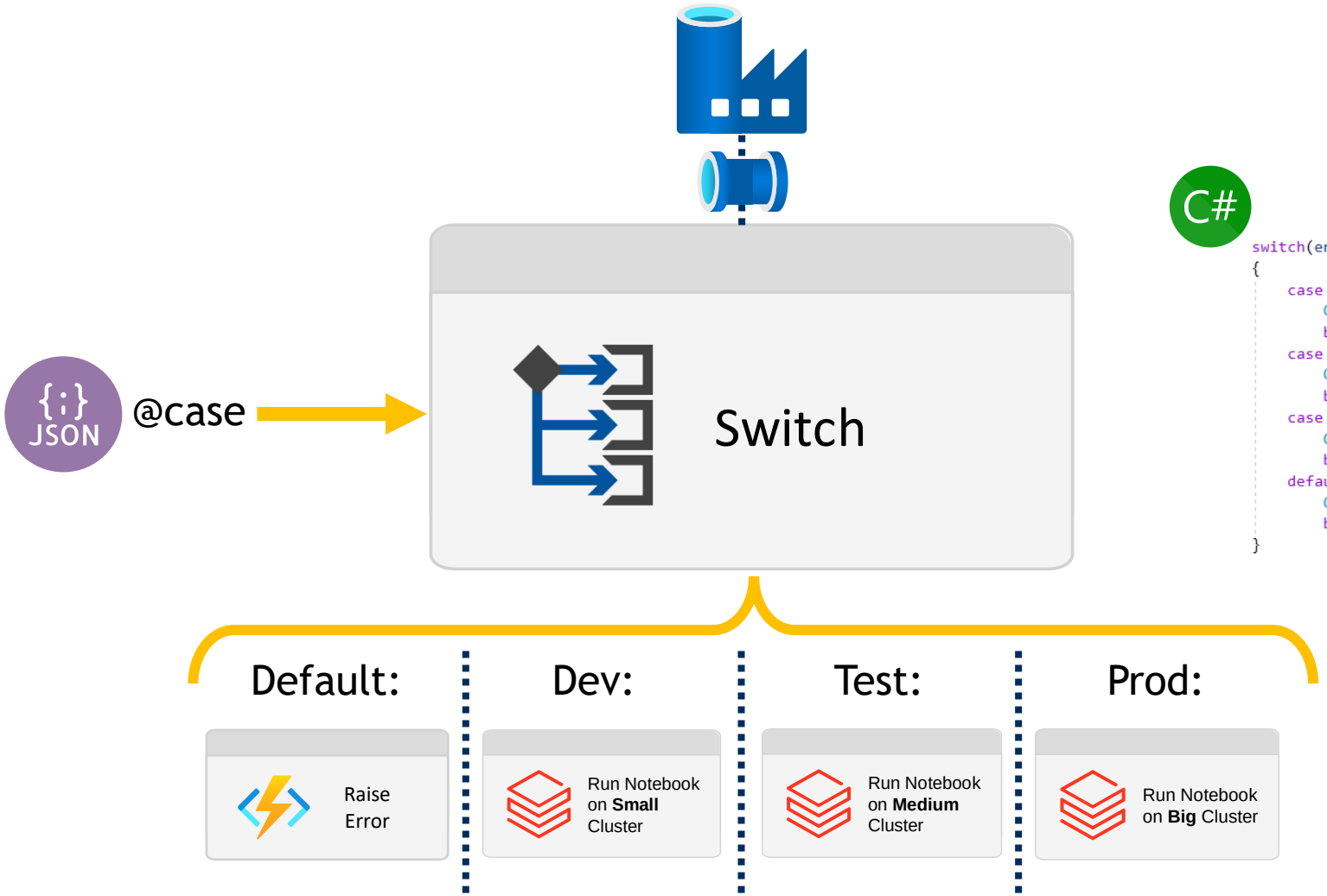


Batch Count Default: 20

Batch Count Max: 50

https://docs.microsoft.com/en-us/azure/data-factory/control-flow-for-each-activity

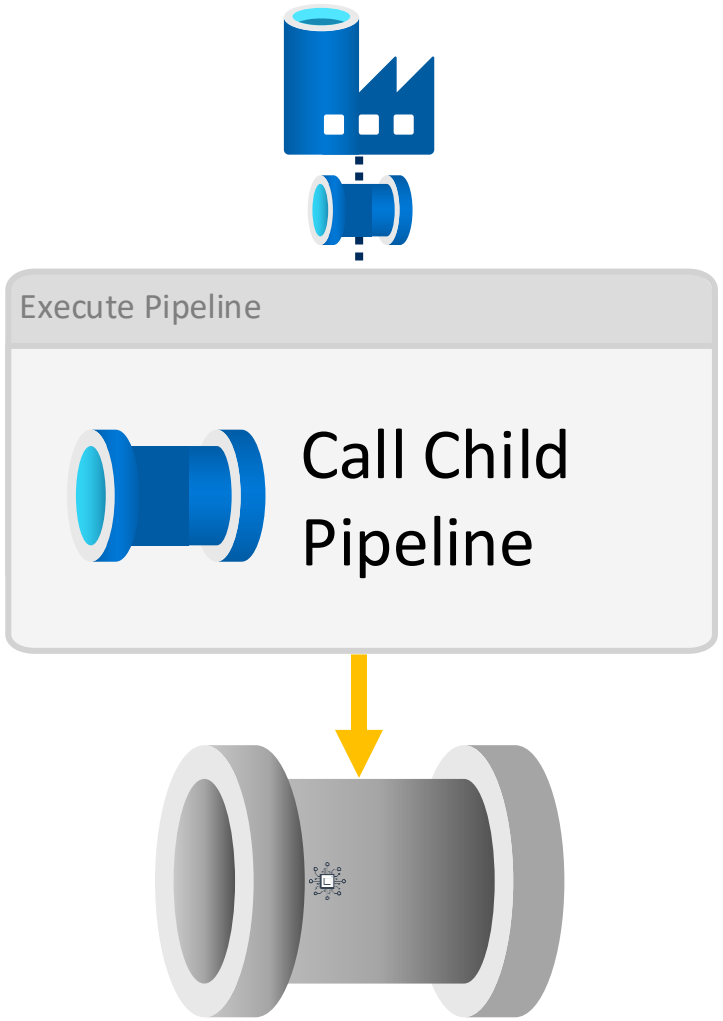
Switch




```csharp
switch(environmentValue)
{
    case "Dev":
        Console.WriteLine("Running in development envirnoment.");
        break;
    case "Test":
        Console.WriteLine("Running in testing envirnoment.");
        break;
    case "Prod":
        Console.WriteLine("Running in production envirnoment.");
        break;
    default:
        Console.WriteLine("Envirnoment unknown.");
        break;
}
```

https://mrpaulandrew.com/2020/01/22/using-the-azure-data-factory-switch-activity/

Execute Pipeline



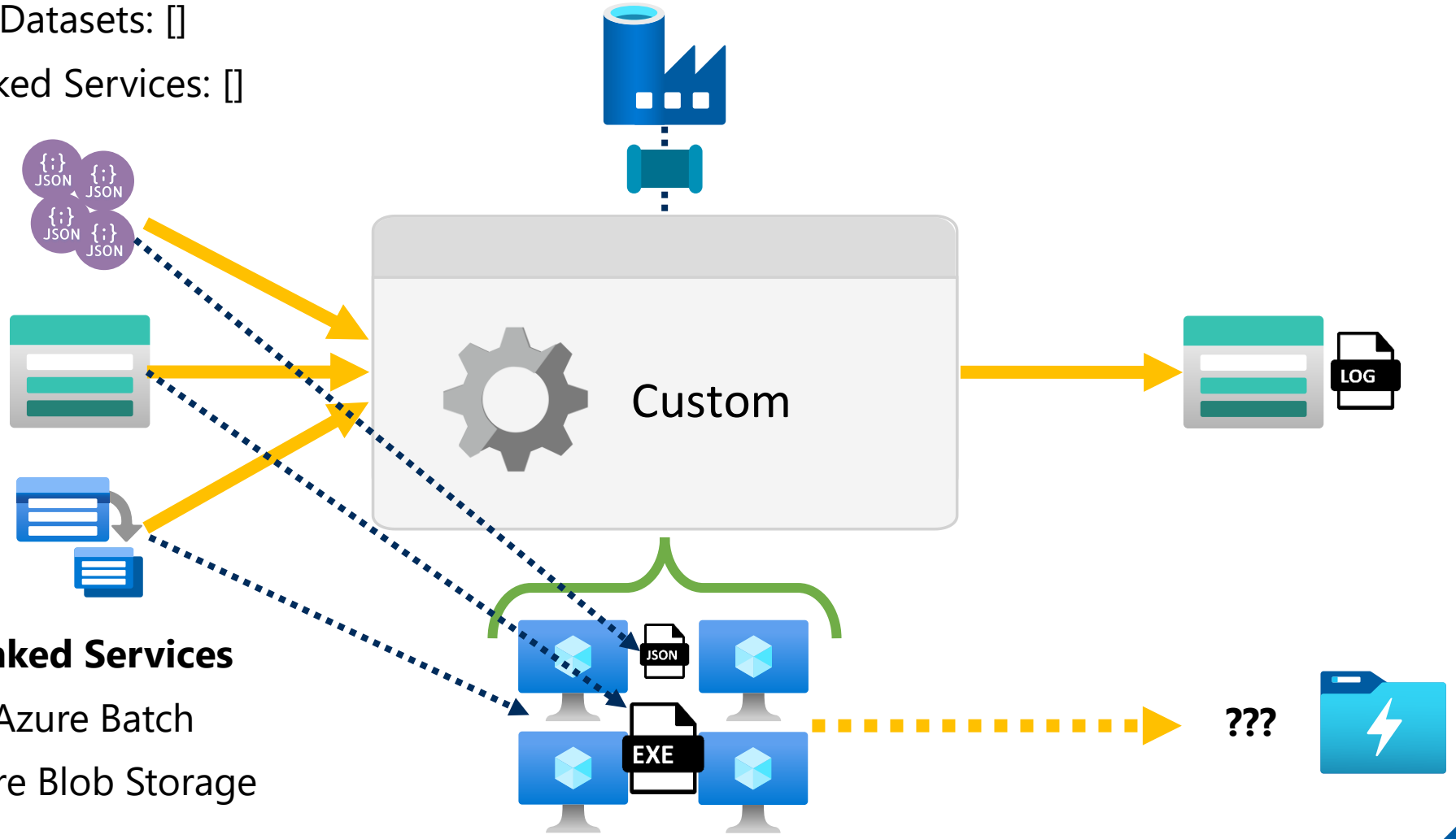
Custom

Extend Data Factory with Custom Code

References Objects

Datasets: []

Linked Services: []



Linked Services

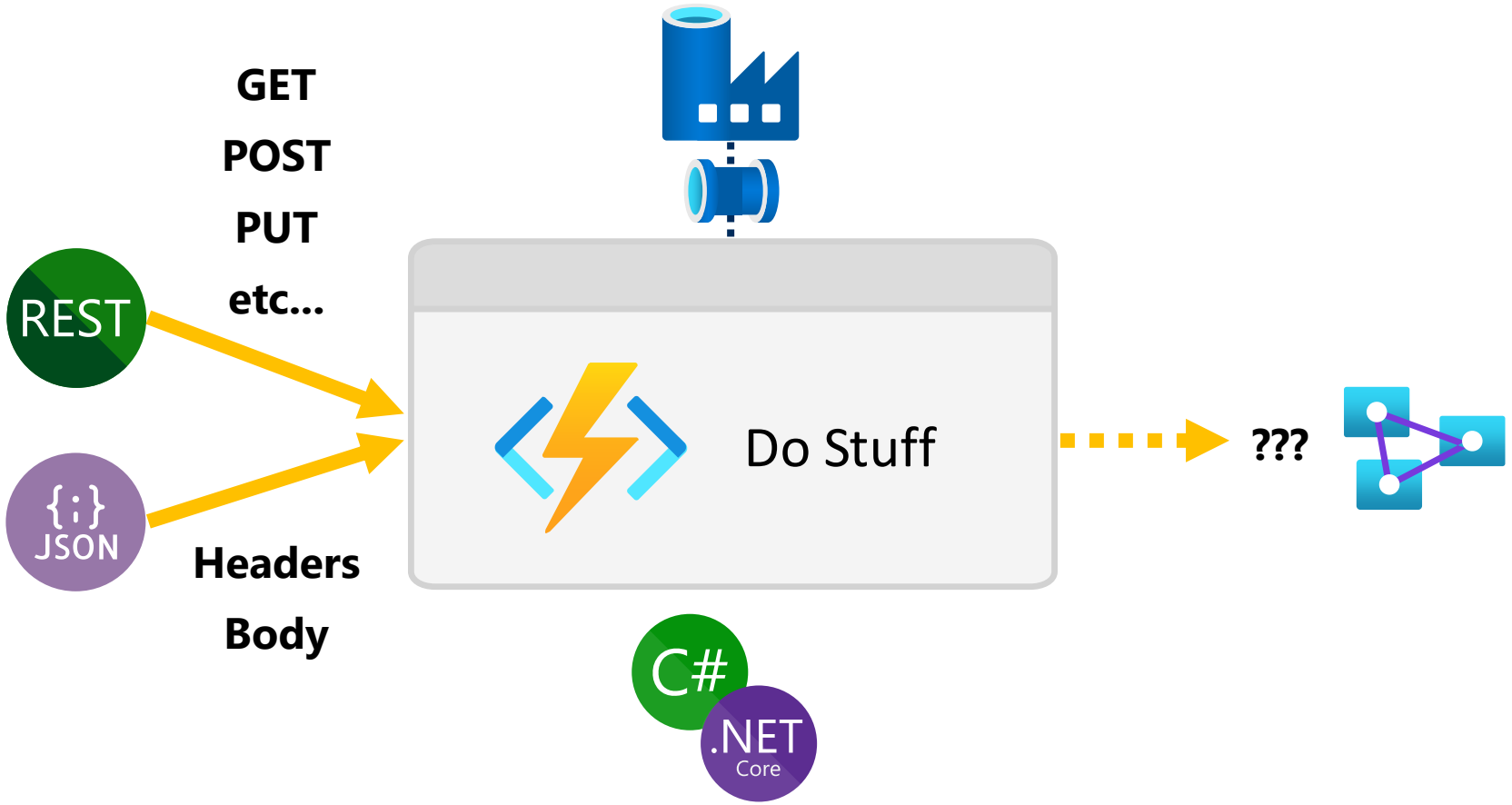
Azure Batch

Azure Blob Storage

https://mrpaulandrew.com/2018/11/12/creating-an-azure-data-factory-v2-custom-activity/

Azure Function

Extend Data Factory with Rest Calls



Web

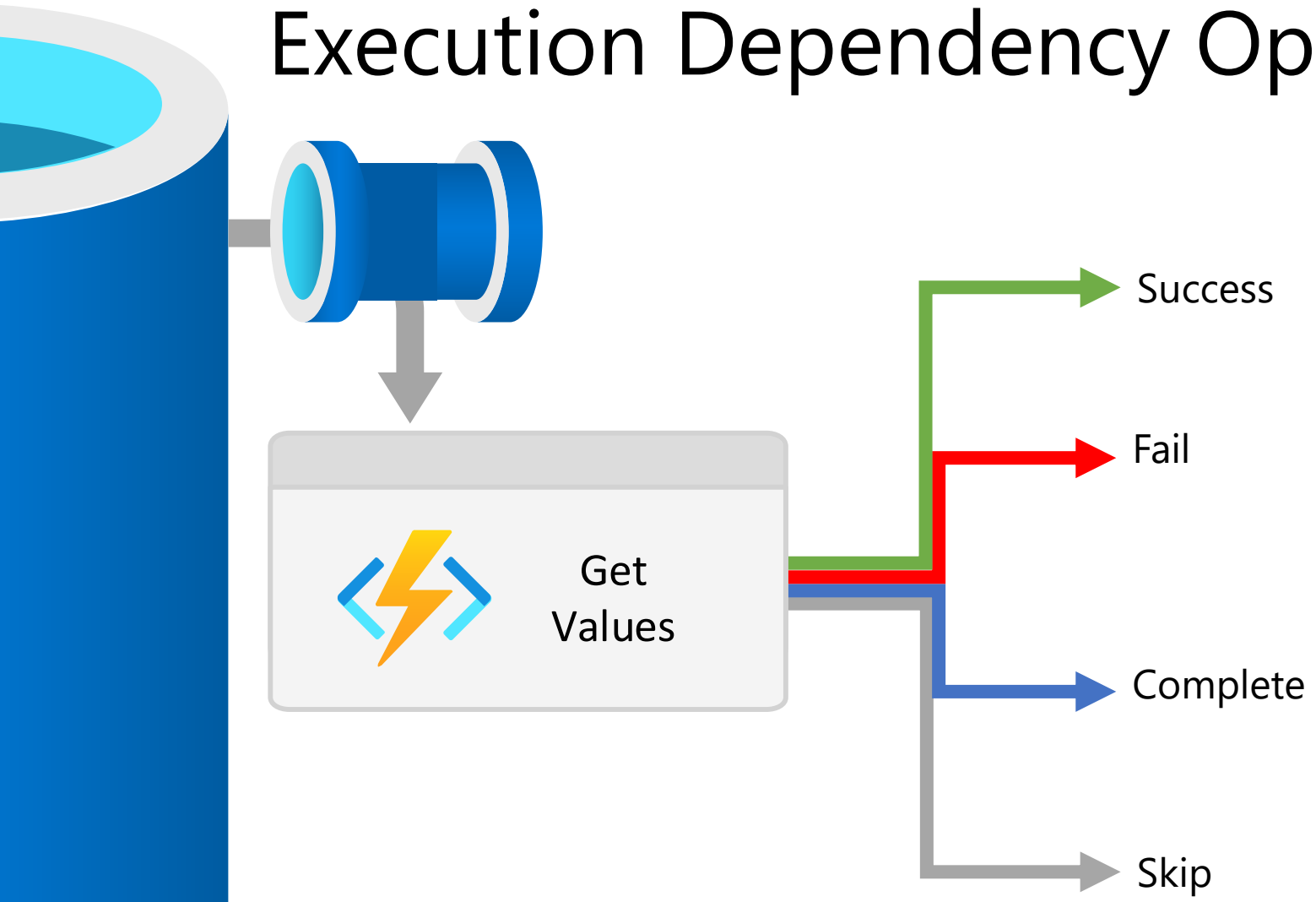
HTTP Do Stuff

Web Hook

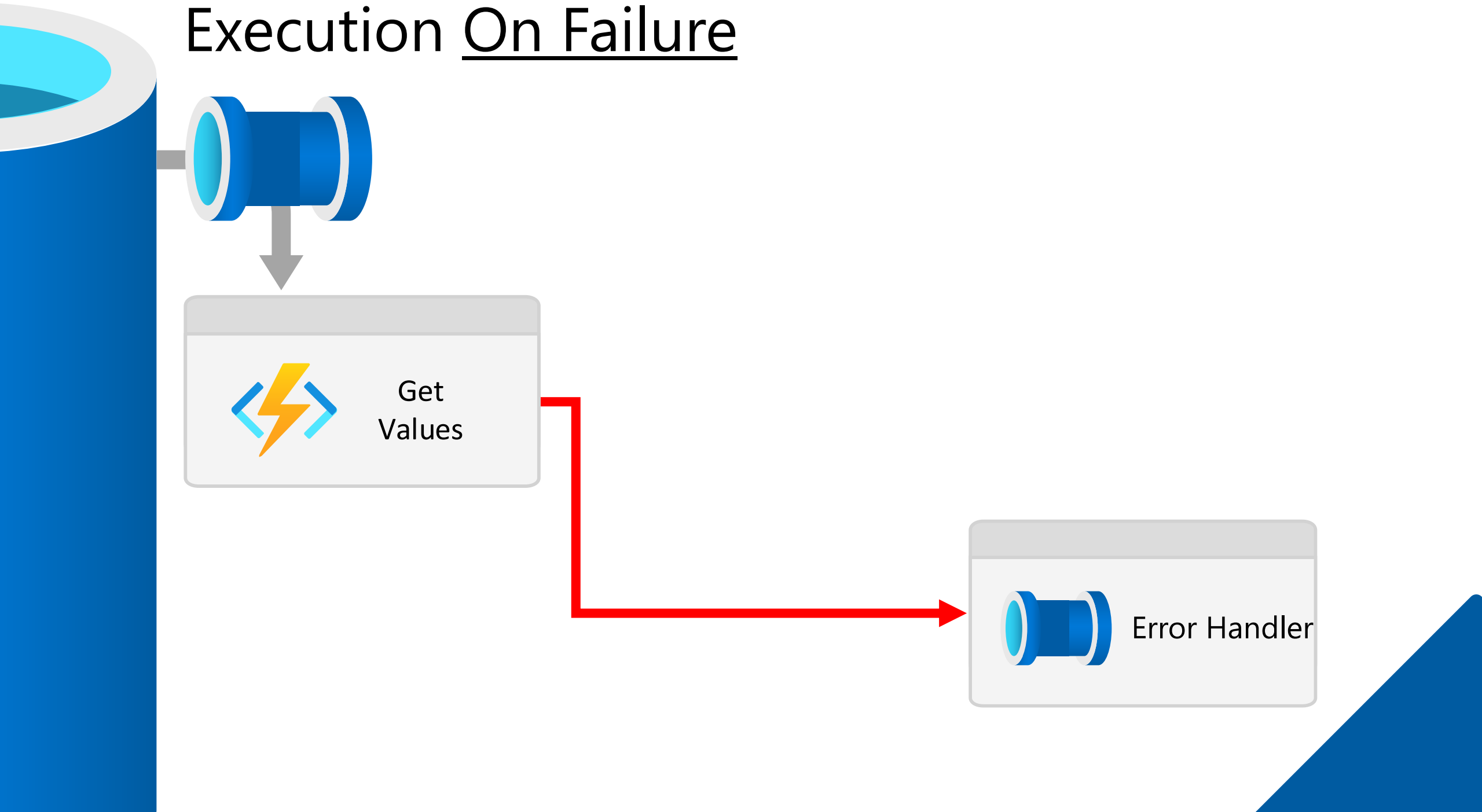
Do Stuff

Execution Dependencies

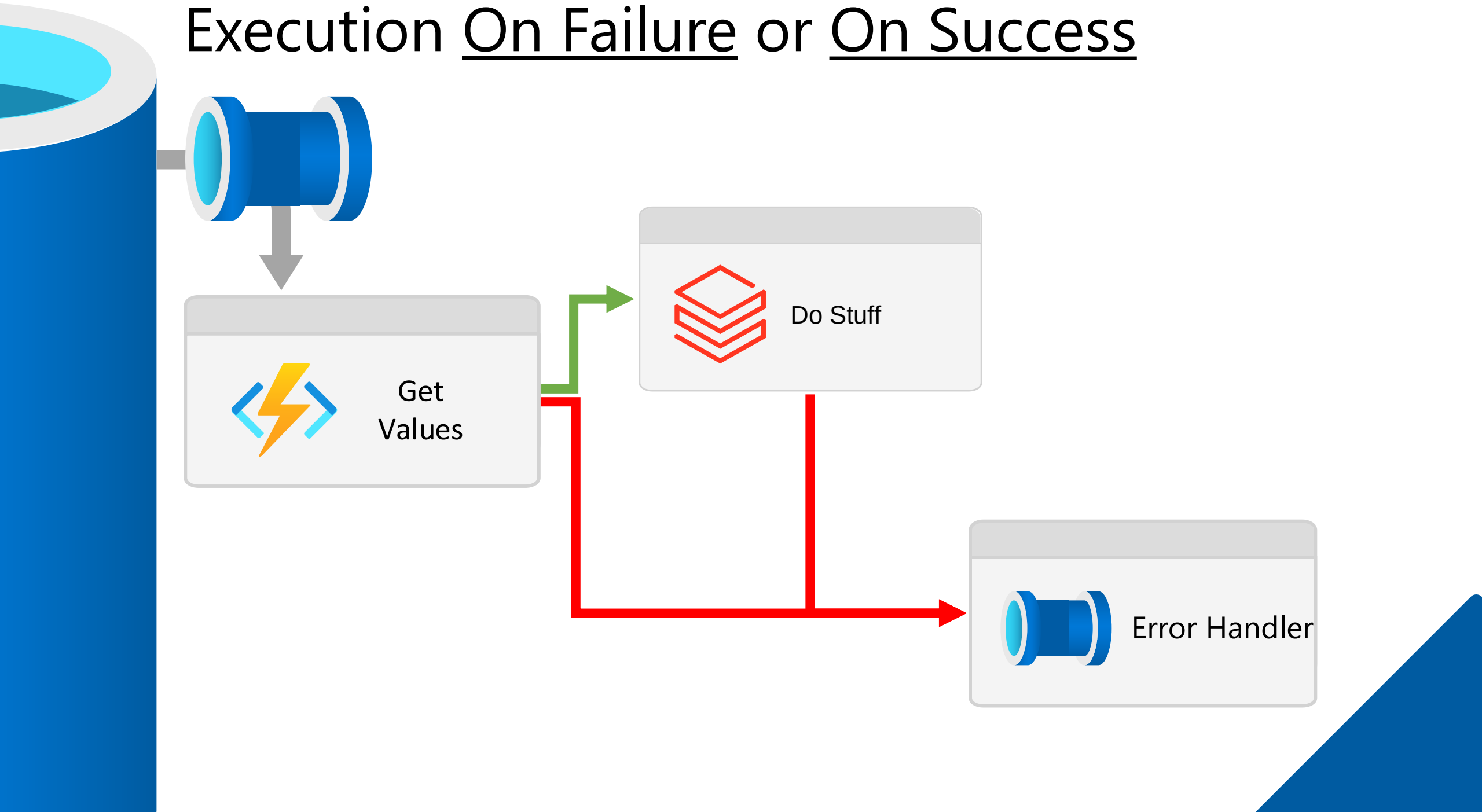
Execution Dependency Options



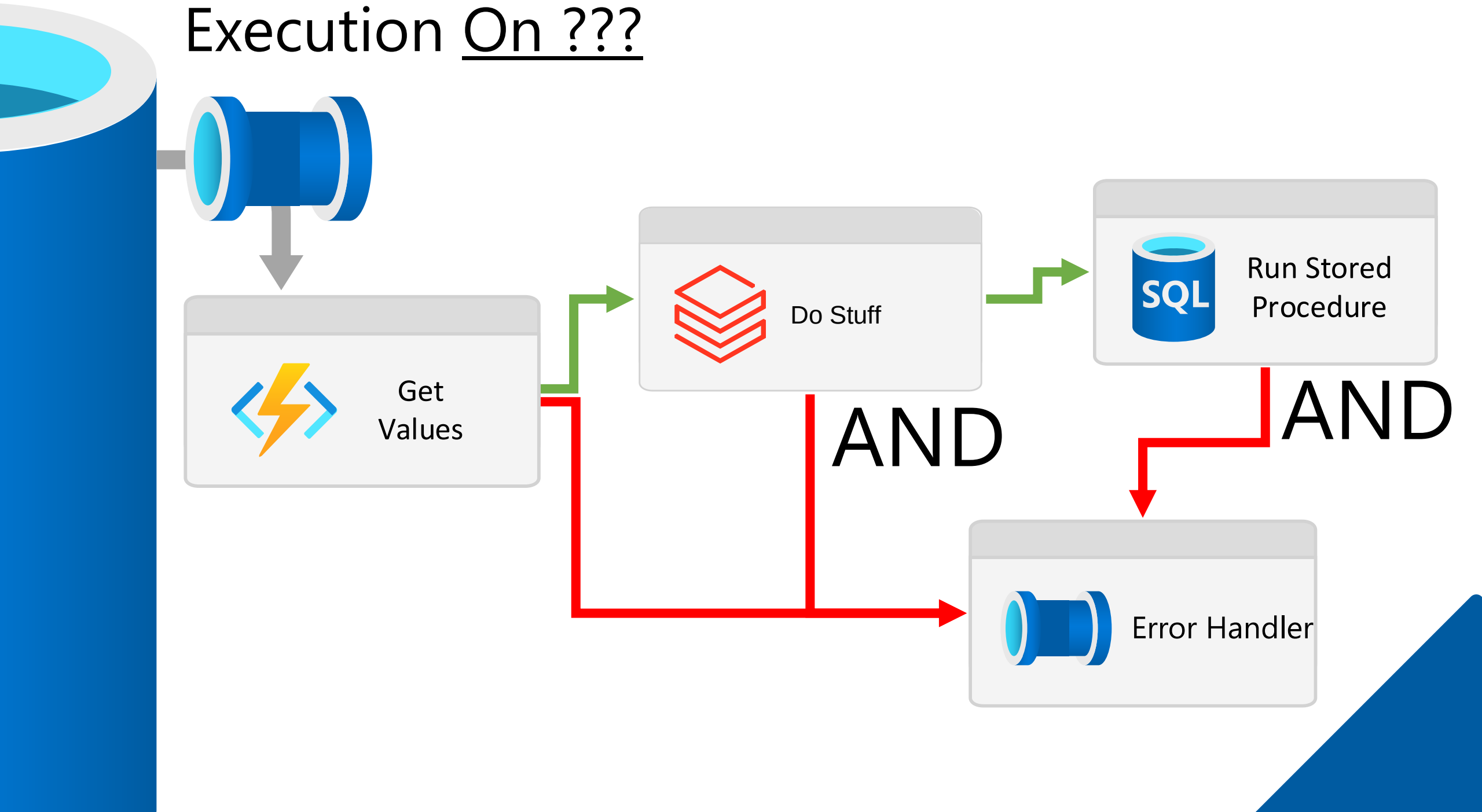
Execution On Failure



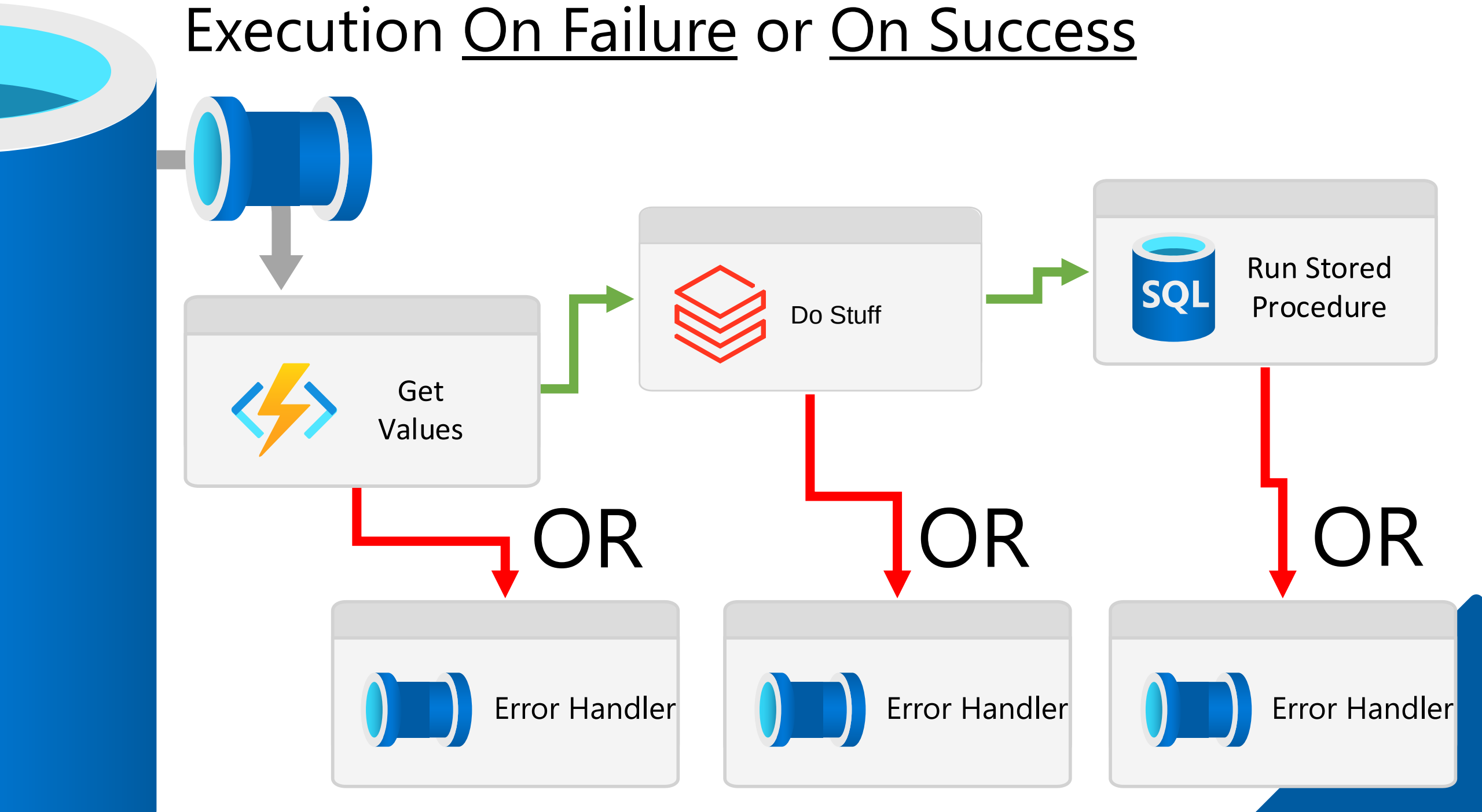
Execution On Failure or On Success

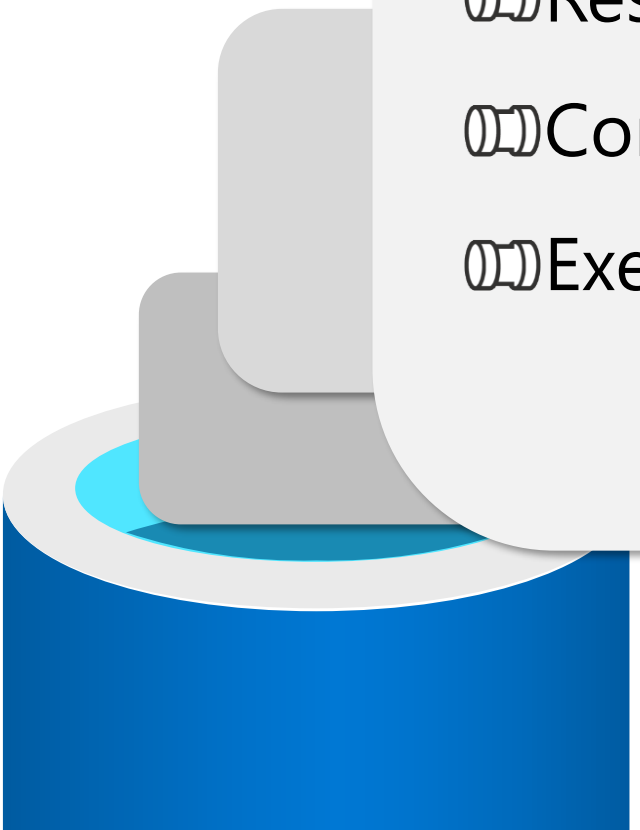


Execution On ???



Execution On Failure or On Success



Module 1:
Data Factory Fundamentals

- What is it and why use it? ✓
- Resource Components ✓
- Common Activities ✓
- Execution Dependencies ✓