Cloud Orchestration

PaaS SSIS & ADFV2 Patterns

Simon Whiteley | Adatis

Microsoft MVP Most Valuable Professional

Microsoft Partner
Microsoft
Gold Data Analytics
Gold Data Platform
Gold Cloud Platform

adatis

Agenda

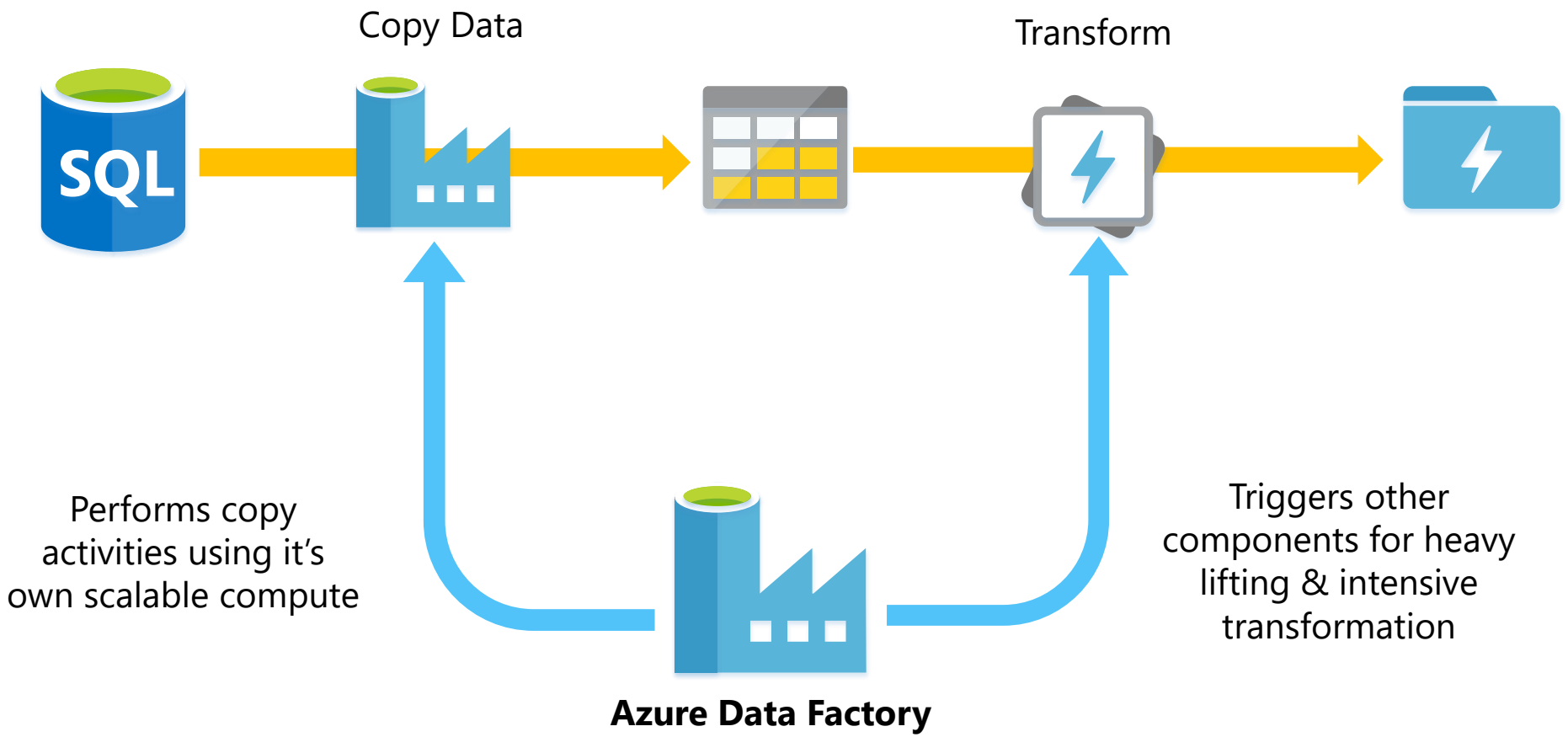
Data Factory Recap

ADFV2 Features Update

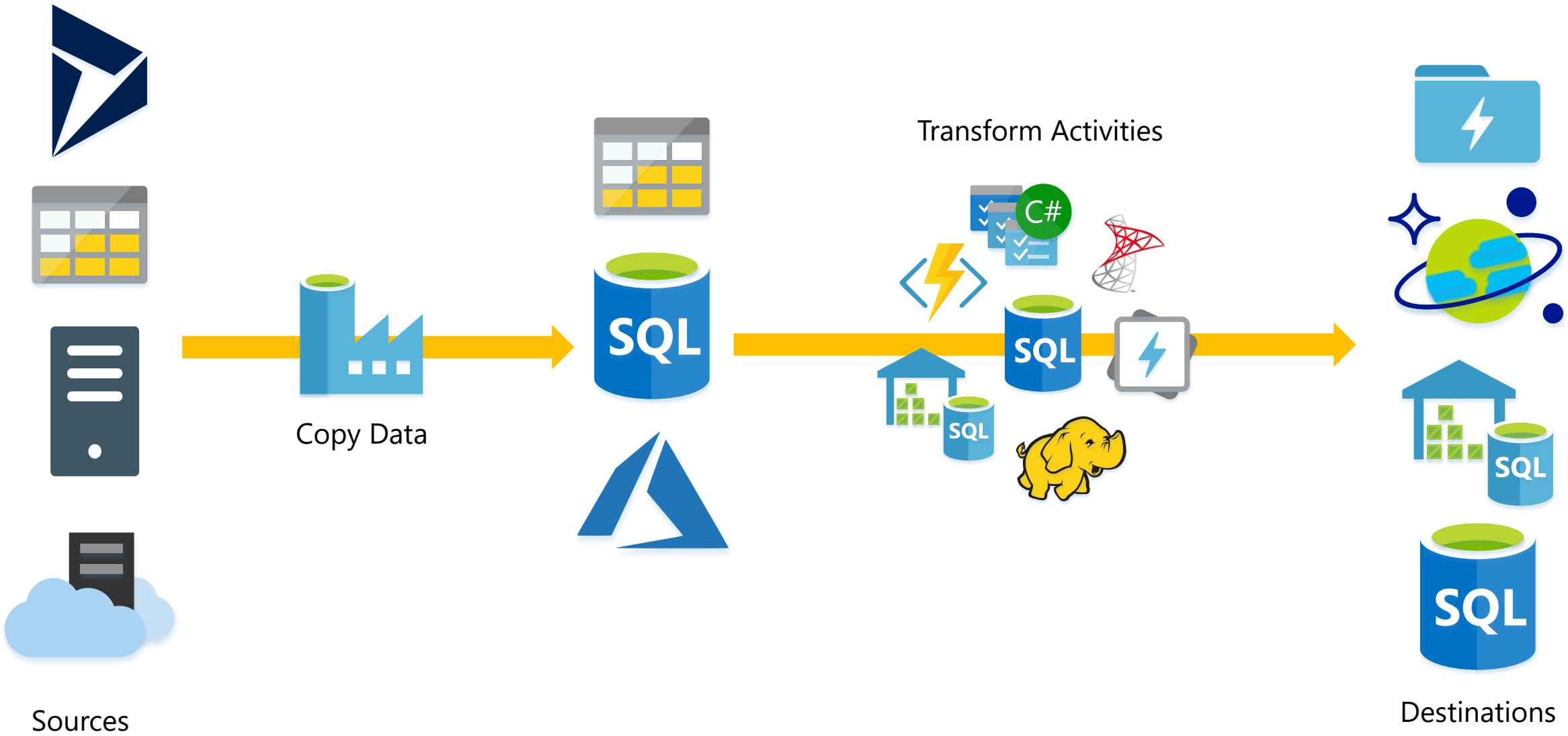
SSIS in Azure

Design Patterns & Conclusions

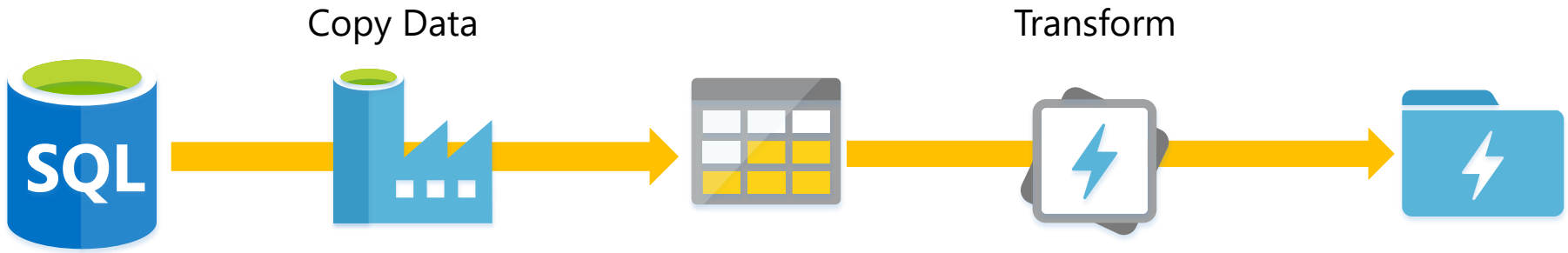
What is Azure Data Factory?



What is Azure Data Factory?



Data Factory Components



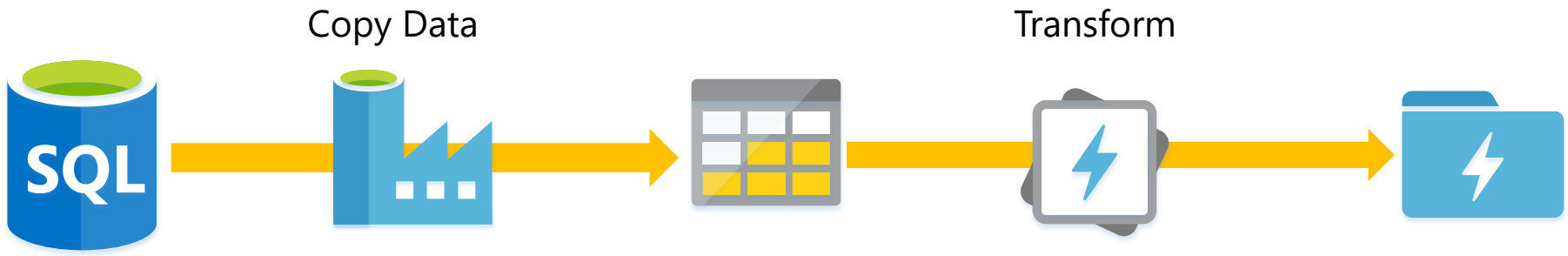
1 **Linked Services** – How do I connect?

Think SSIS Connection Managers!

SQLDBLinkedService

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

Data Factory Components

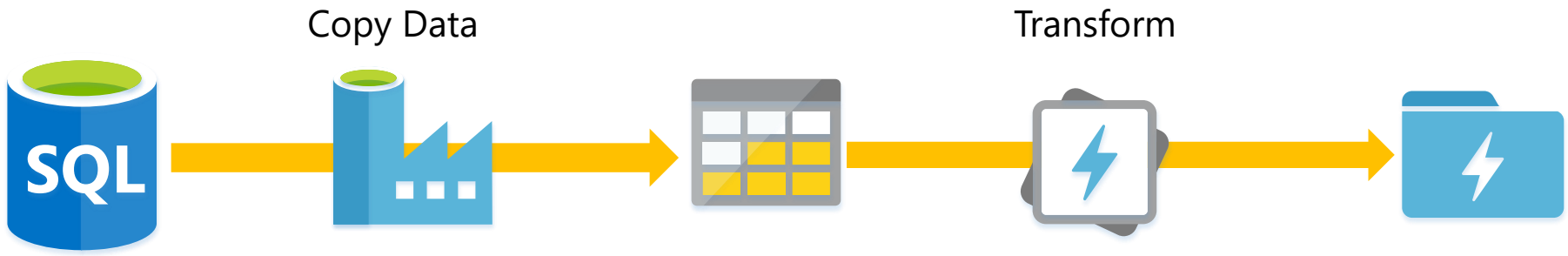


1. **Linked Services**
2. **Data Sets** – What slices/partitions does my data have?

dbo.DimCustomer

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

Data Factory Components

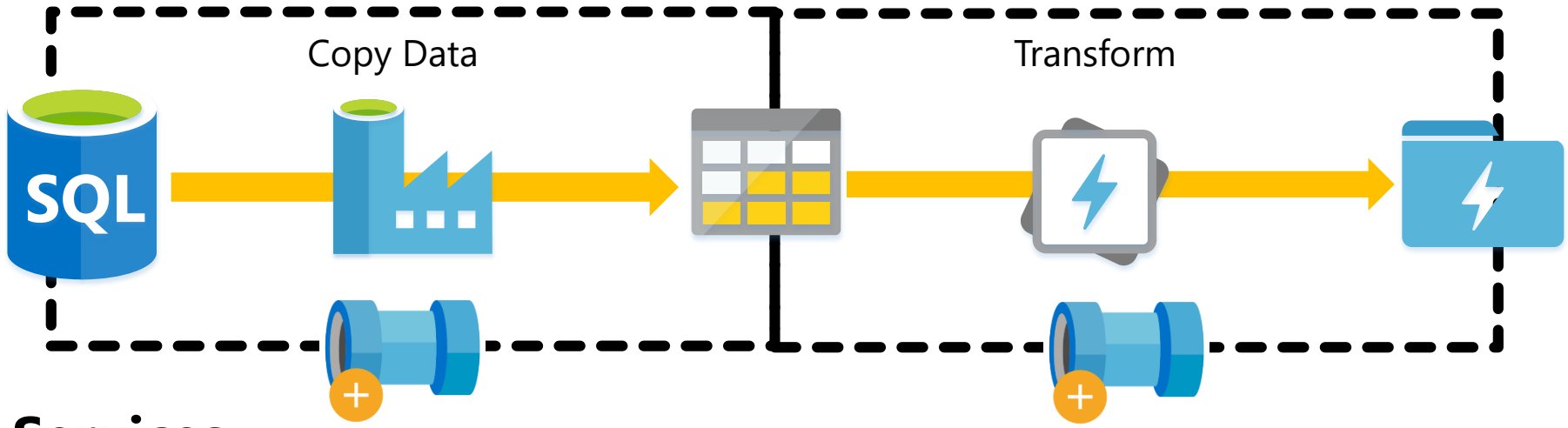


1. **Linked Services**
2. **Data Sets**
3. **Activities** – What actions am I taking?

U-SQL Activity

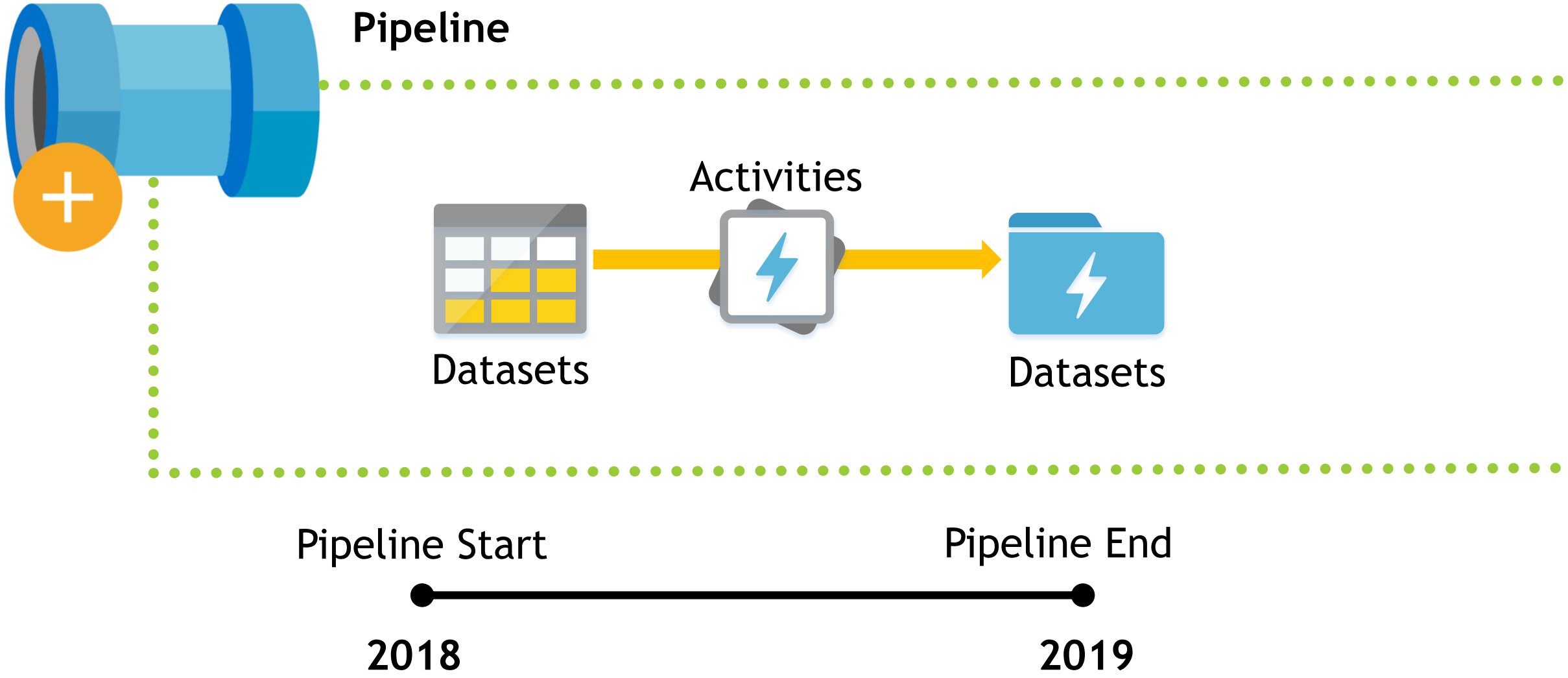
Script: *wasb//:myscripts/ProcessOrders.usql*
AUs: *5 units*
Priority: *1000*
Parameters: *@Output = "RAW/Orders/..."*

Data Factory Components

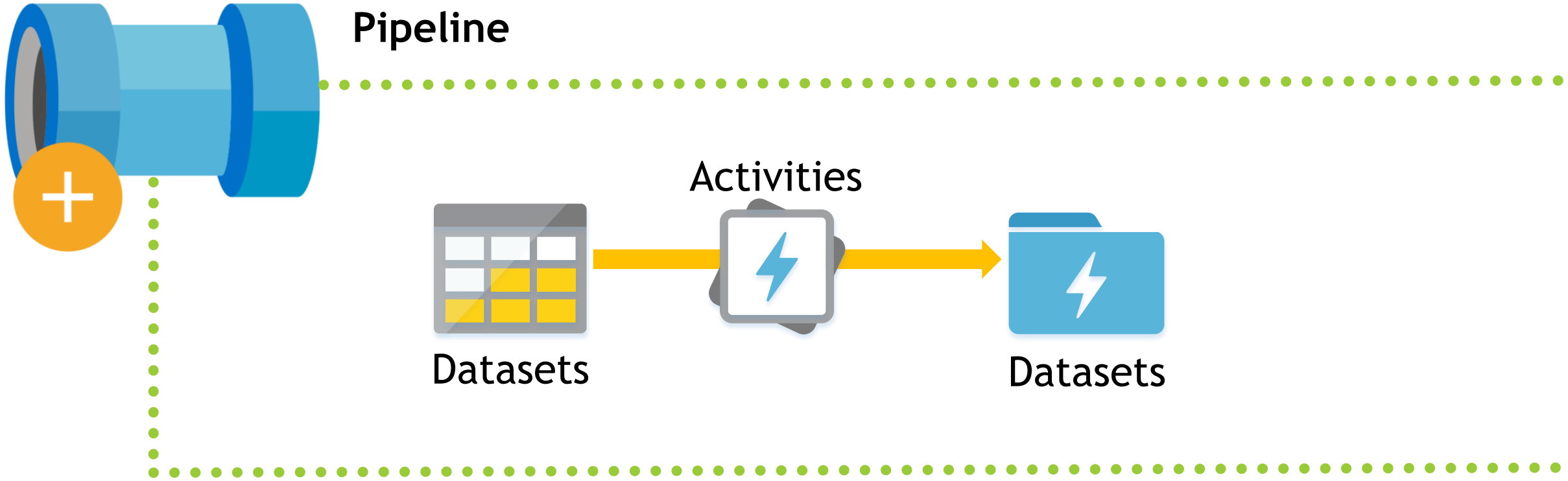


1. **Linked Services**
2. **Data Sets**
3. **Activities**
4. **Pipelines** – What groups of work do I want to do?

Azure Data Factory Concepts Continued

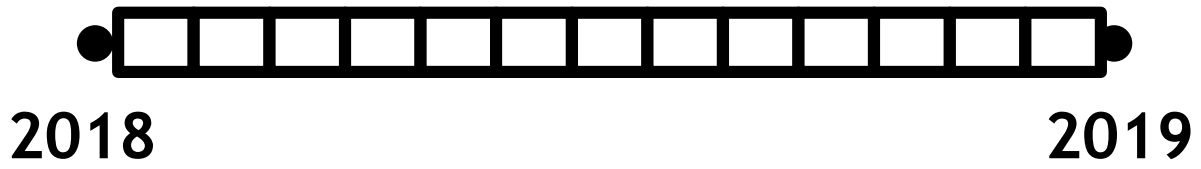


Azure Data Factory Concepts Continued

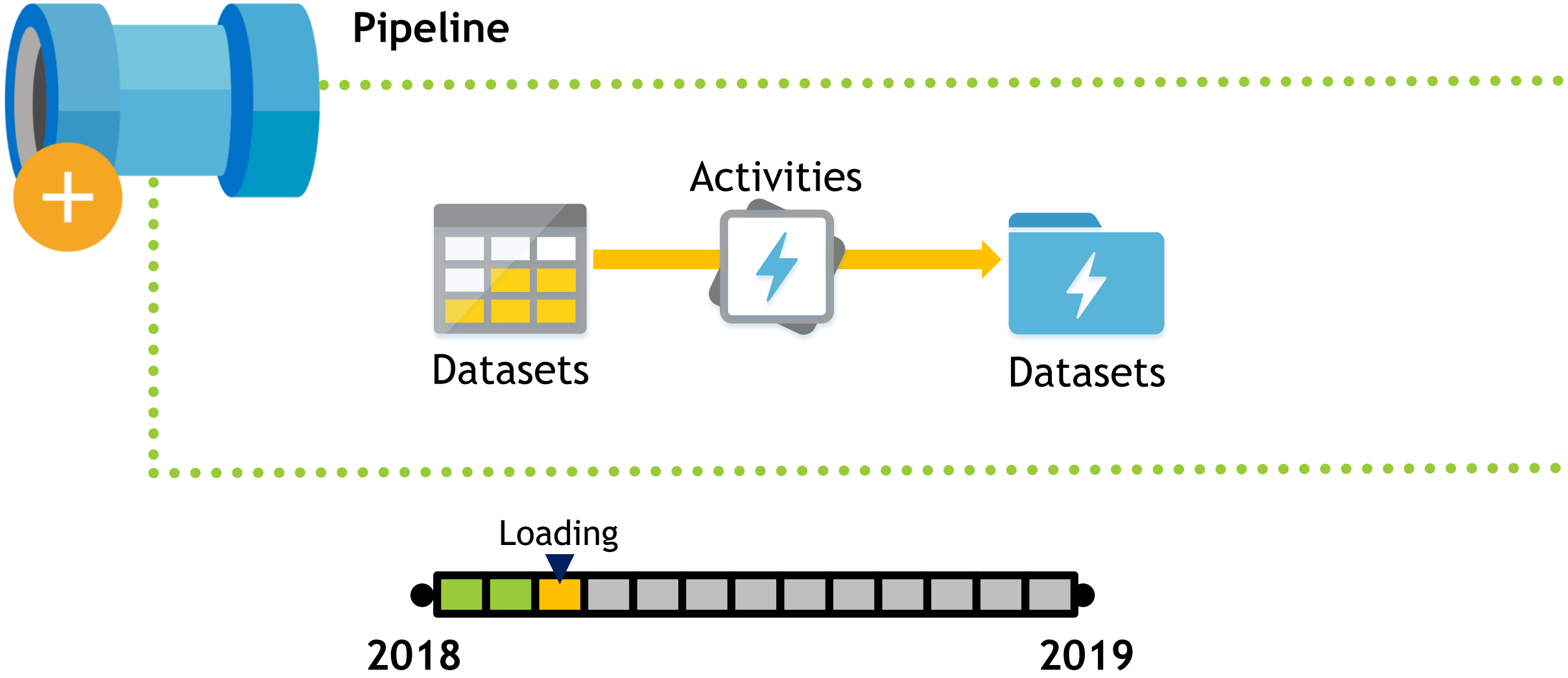


Interval: *Month*

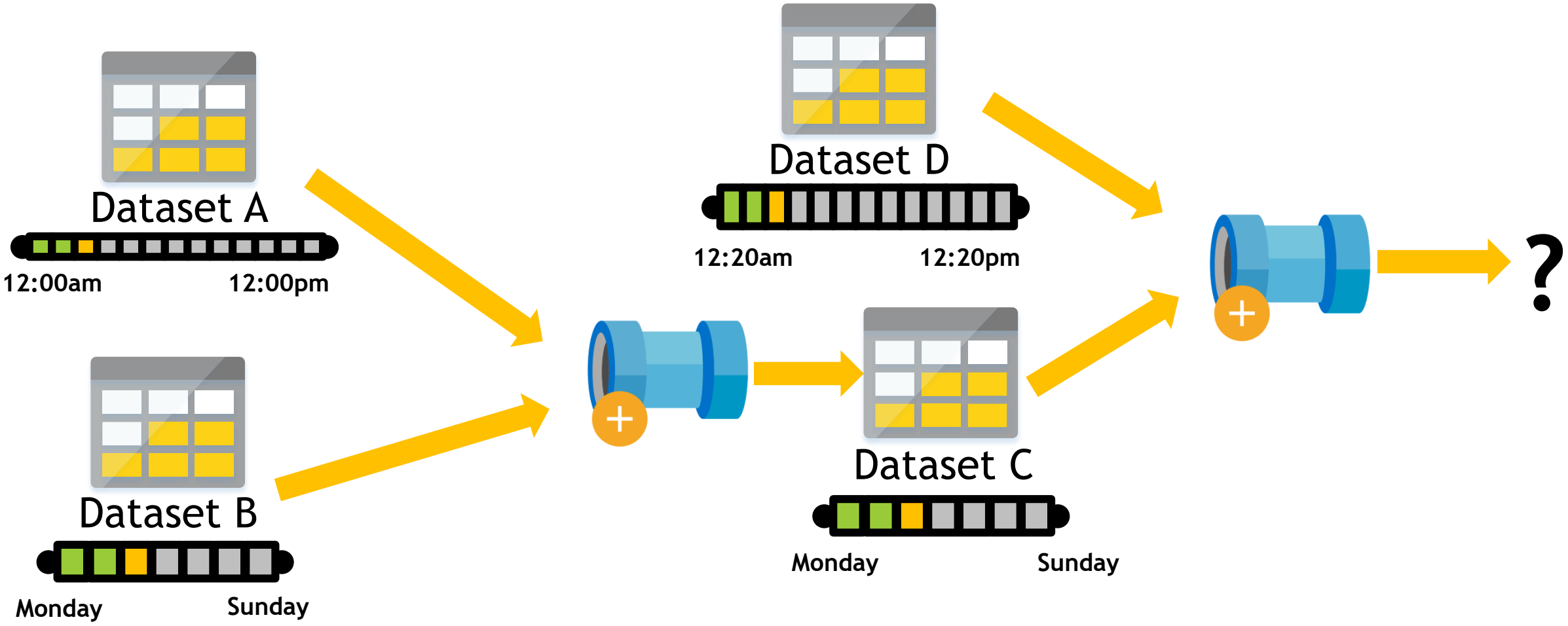
Frequency: *1*



Azure Data Factory Concepts Continued



Time Slice Problems…



Agenda

Data Factory Recap

ADFV2 Features Update

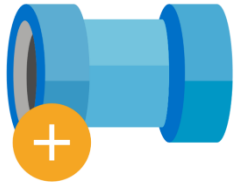
SSIS in Azure

Design Patterns & Conclusions

Integration Runtimes

1 Default Integration Runtime

Movement Hours



Activity Orchestration

Flexible Region

2 SSIS Integration Runtime

SSIS Package Execution

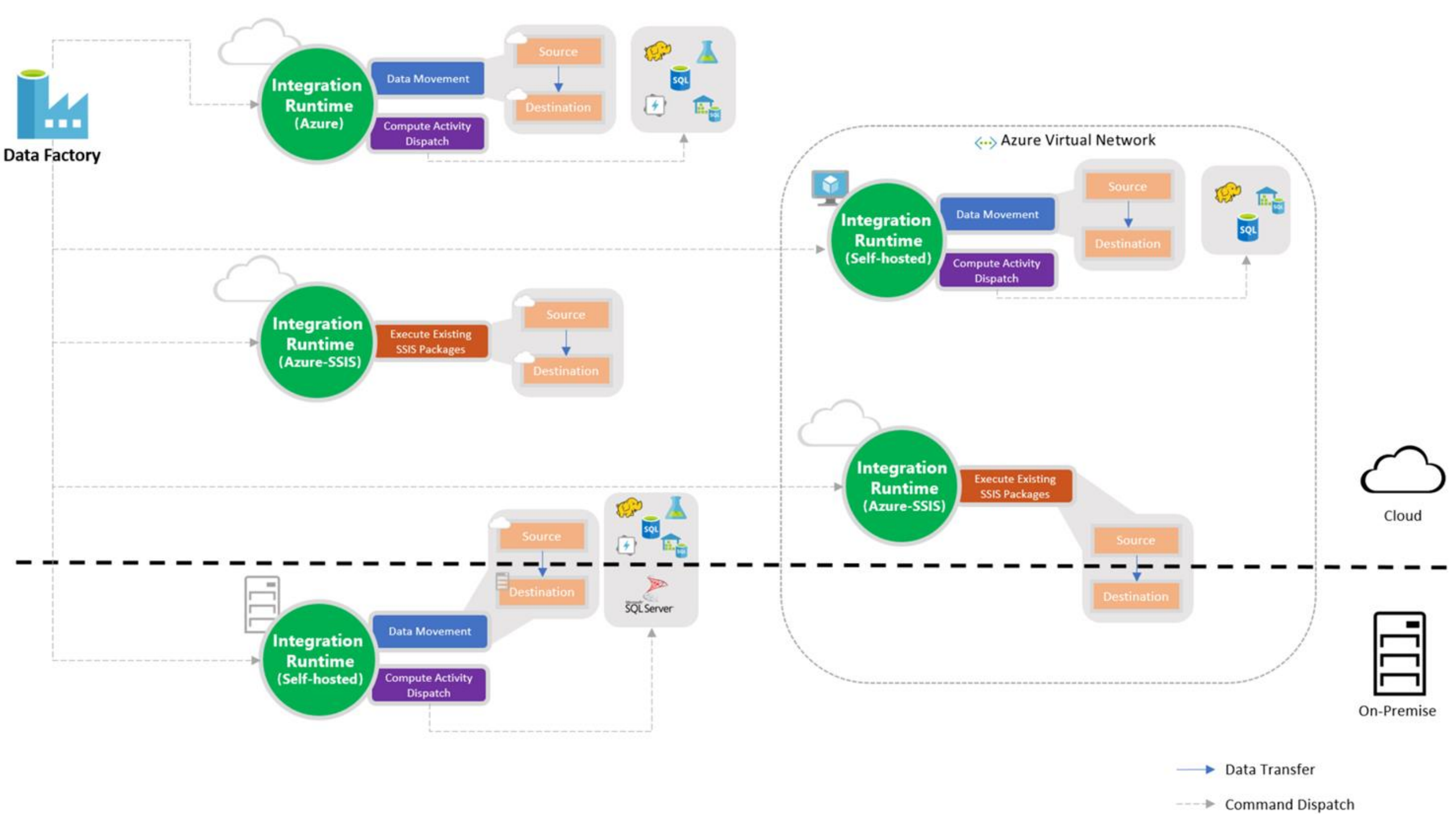
Specified Region

3 Self Hosted Integration Runtime

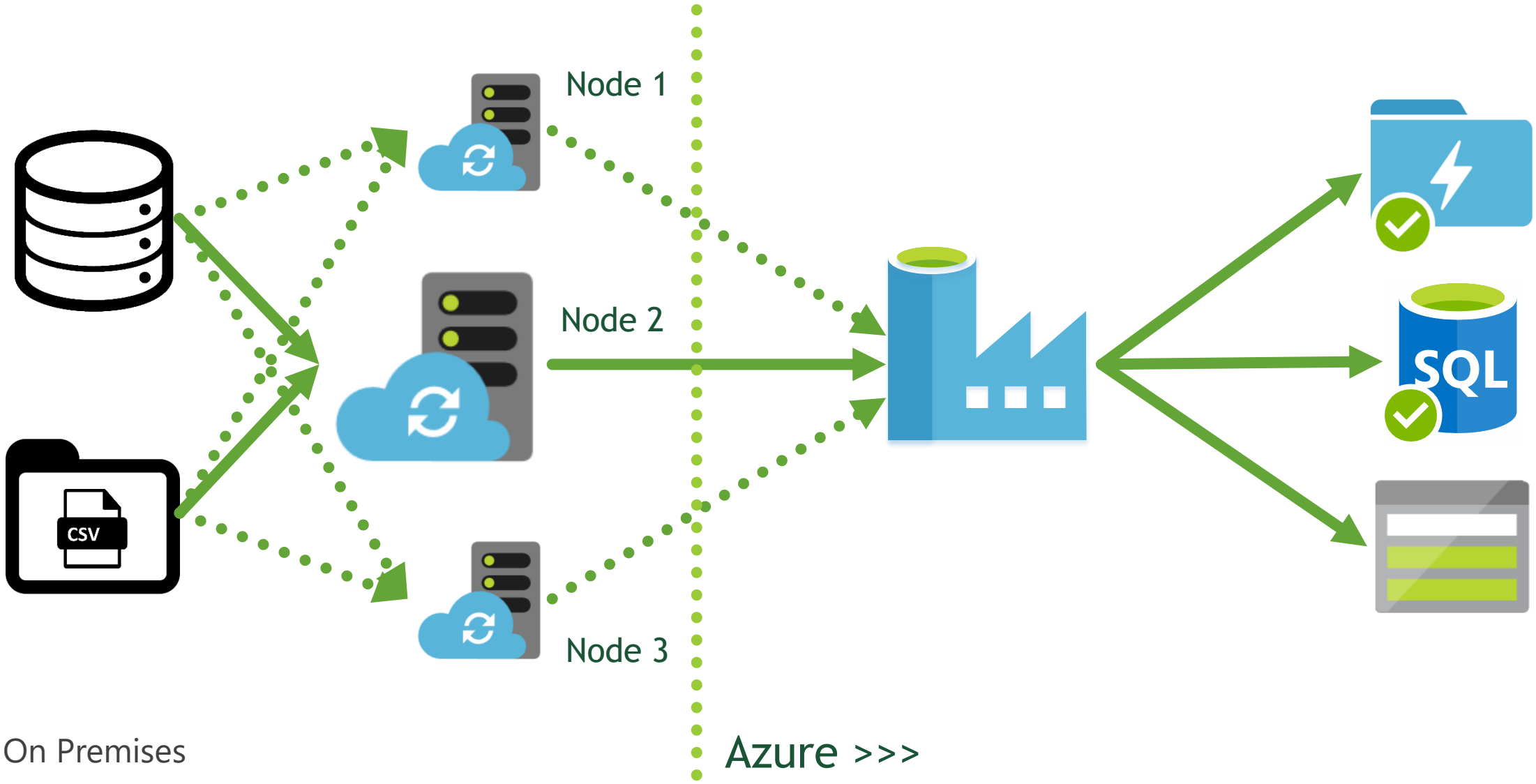
Local Compute

Activity Orchestration

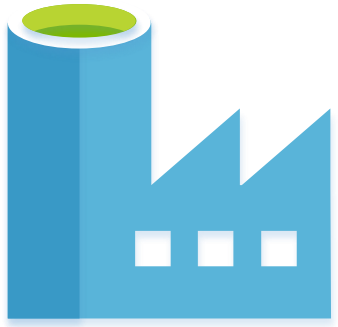
On-Prem Server


Data Factory
Integration Runtime (Azure)
Data Movement
Compute Activity Dispatch
Source
Destination
Azure Virtual Network
Integration Runtime (Self-hosted)
Data Movement
Compute Activity Dispatch
Source
Destination
Integration Runtime (Azure-SSIS)
Execute Existing SSIS Packages
Source
Destination
Integration Runtime (Azure-SSIS)
Execute Existing SSIS Packages
Source
Destination
Cloud
Integration Runtime (Self-hosted)
Data Movement
Compute Activity Dispatch
Source
Destination
SQL Server
On-Premise
Data Transfer
Command Dispatch

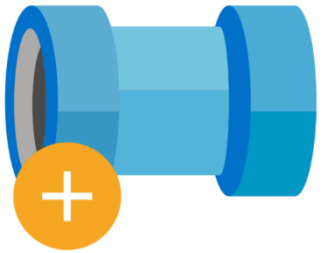
The Self-Hosted Integration Runtime *(AKA The Data Management Gateway)*



Azure Data Factory Concepts & Components Recap



1. **Linked Services**
2. **Data Sets**
3. **Activities**
4. **Pipelines**

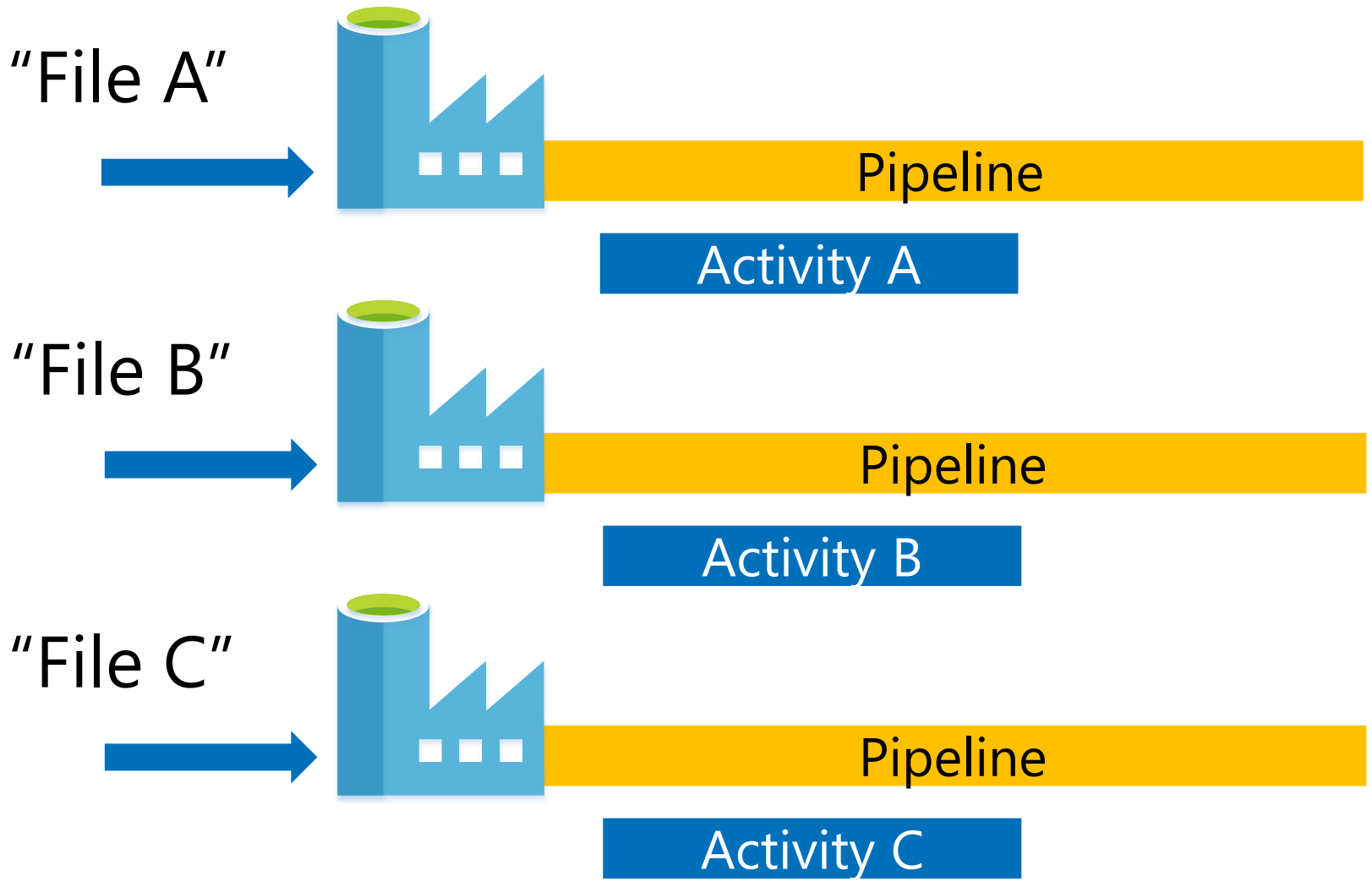


Time Slices

1. **Azure** Integration Runtime
2. **SSIS** Integration Runtime
3. **Self Hosted** Integration Runtime

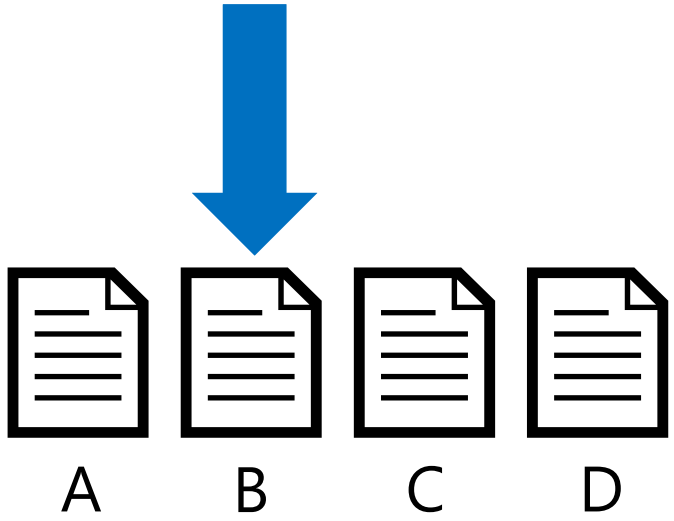
Demo: ADFV2

Hardcoded Pipelines

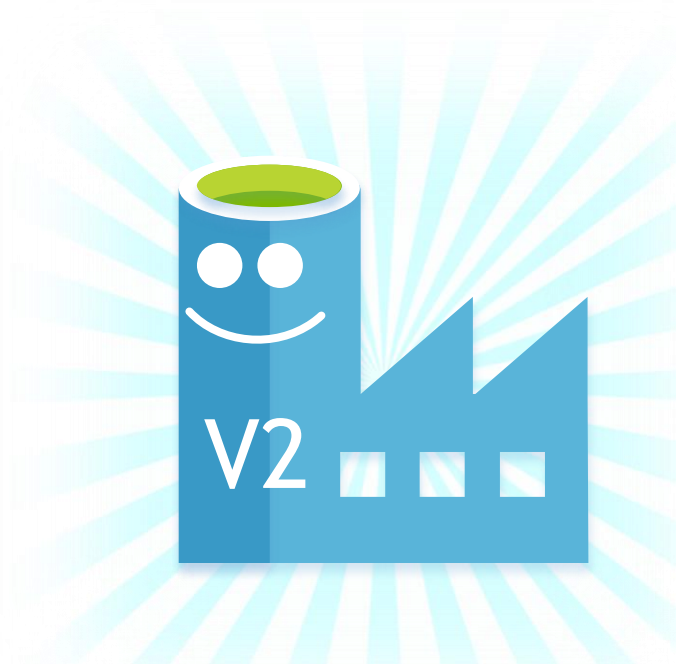


Dynamic Pipelines using Parameters

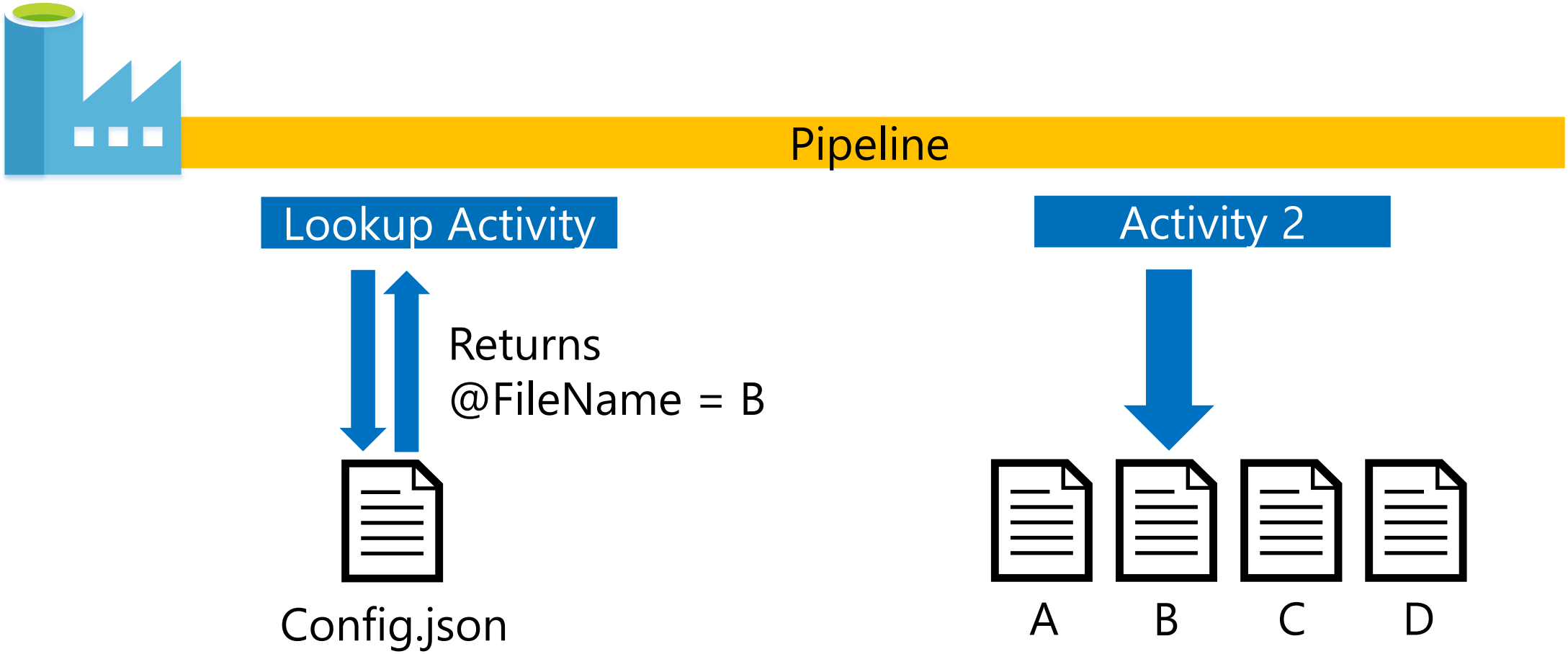
@FileName = B



Dynamically change parameters based on inputs



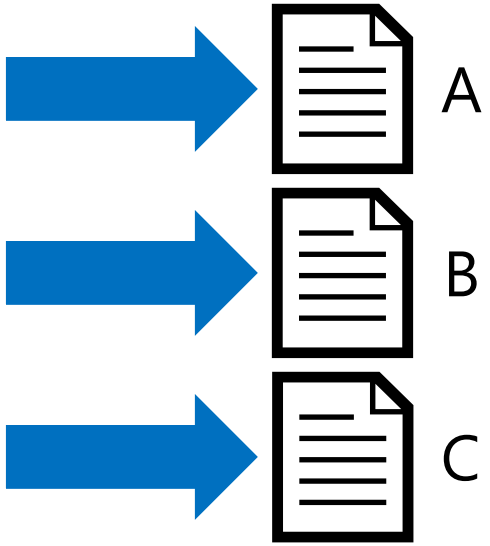
Dynamic Pipelines using Lookup Activity



ForEach Pipelines

@Files = [A,B,C]

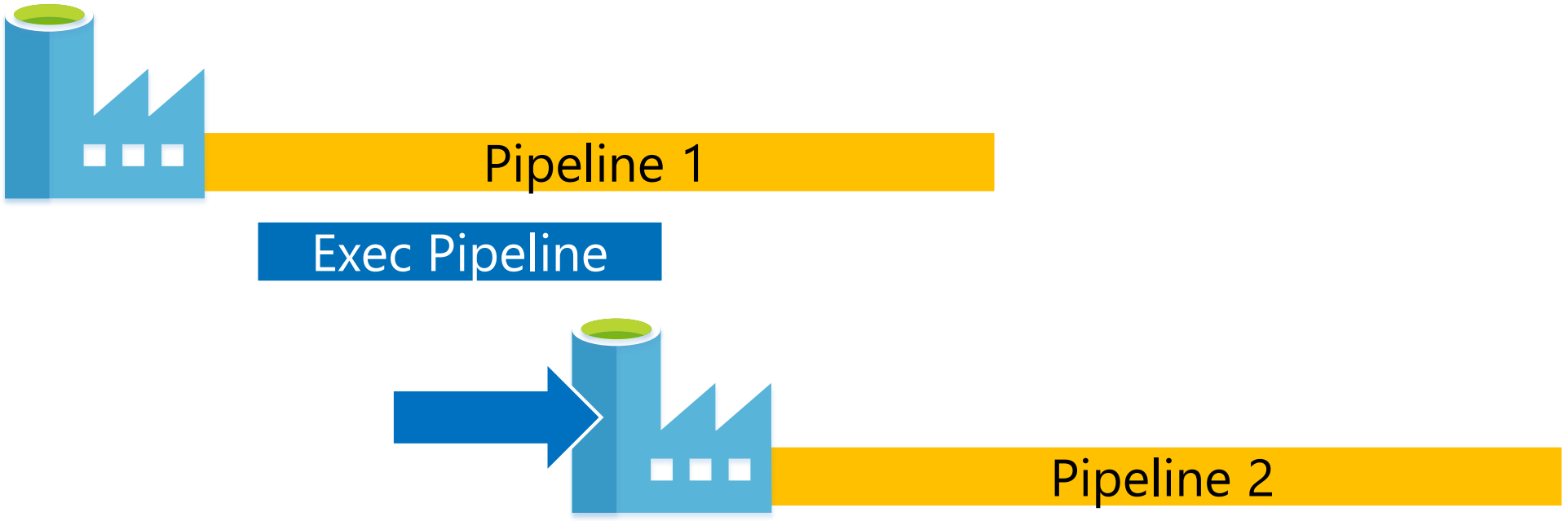
ForEach Activity



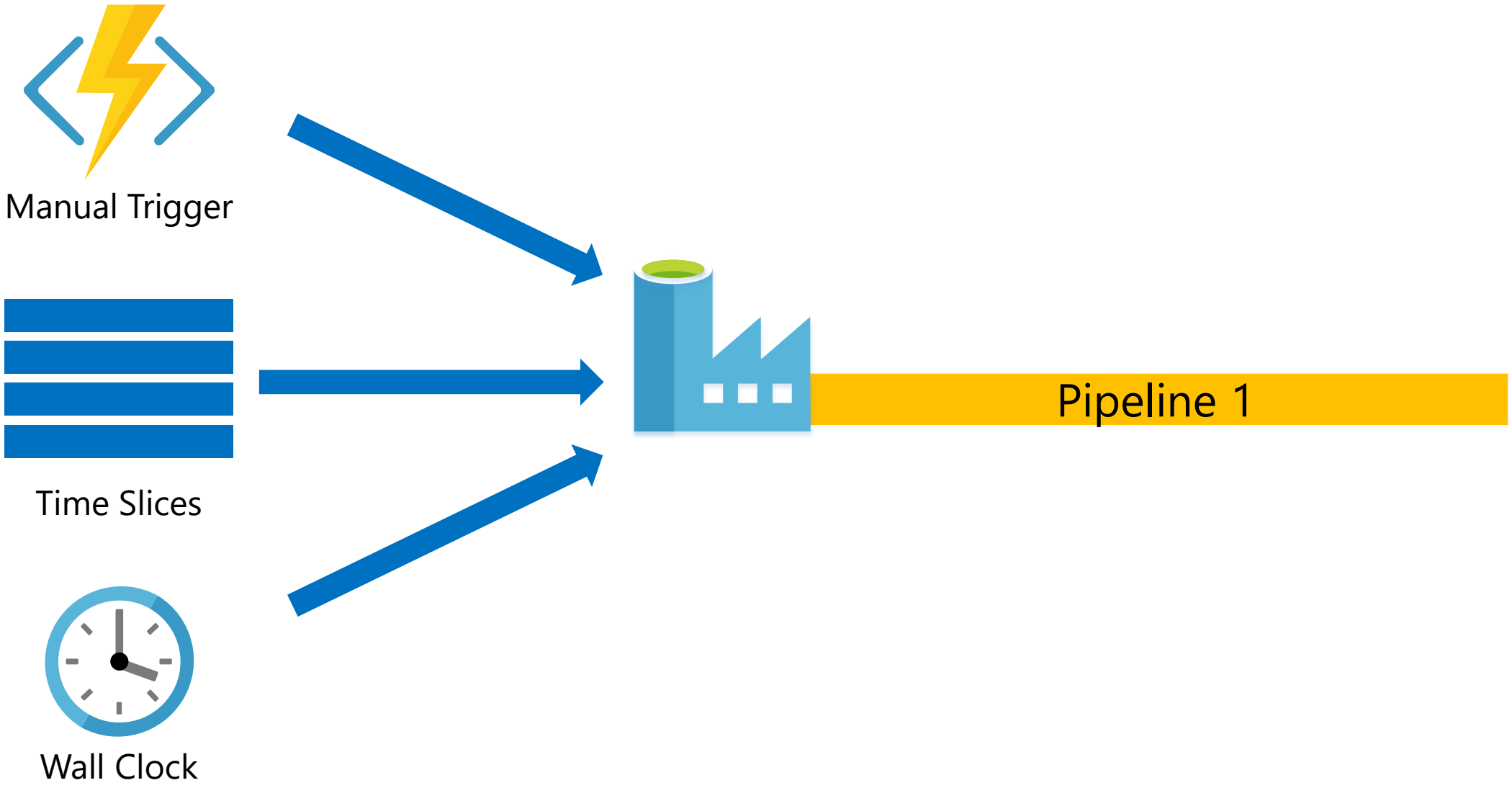
Run activity for each item in an array

(max DOP 20)

Execute Pipeline Activity



Pipeline Triggers



Demo: Dynamic Pipelines

Agenda

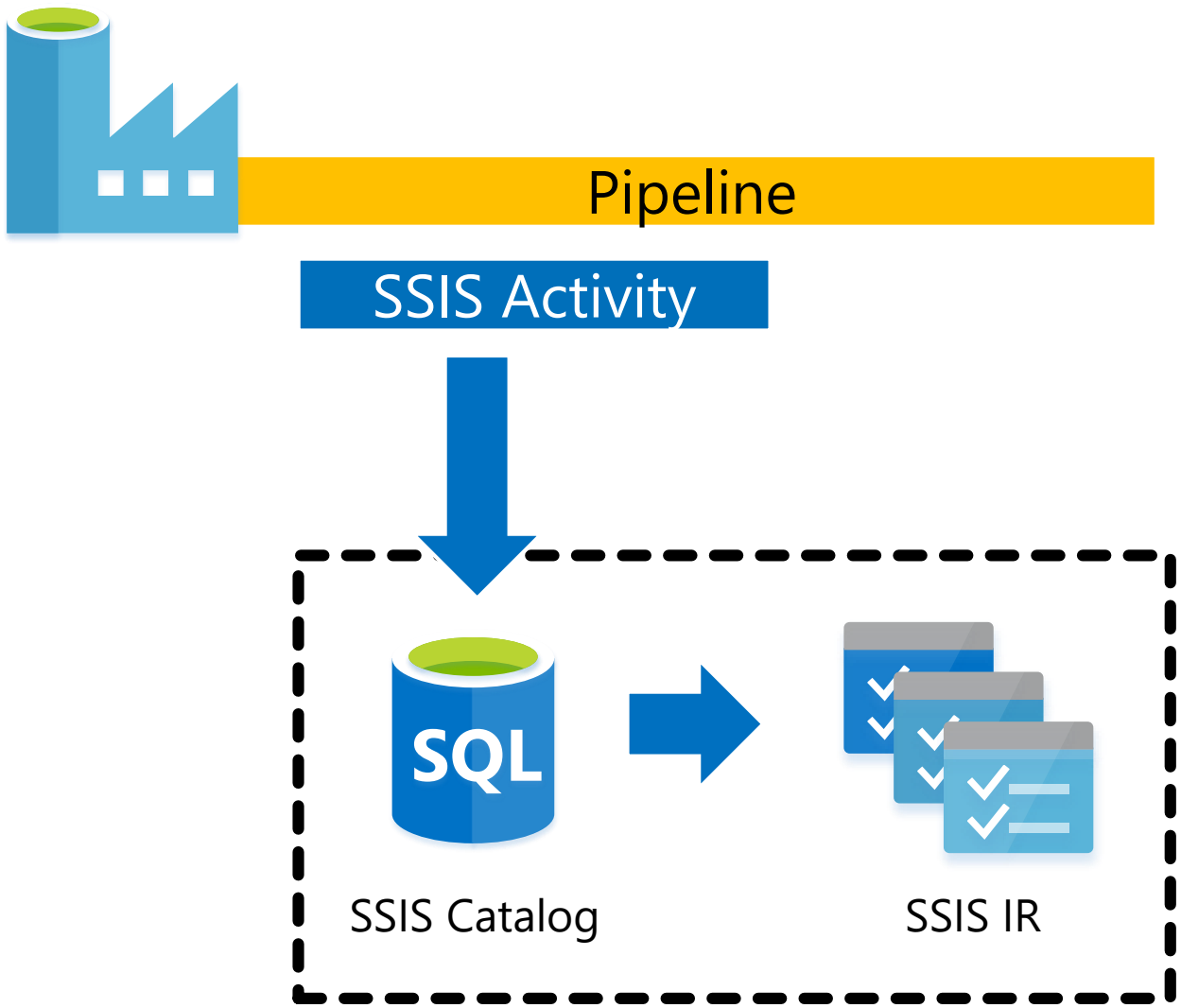
Data Factory Recap

ADFV2 Features Update

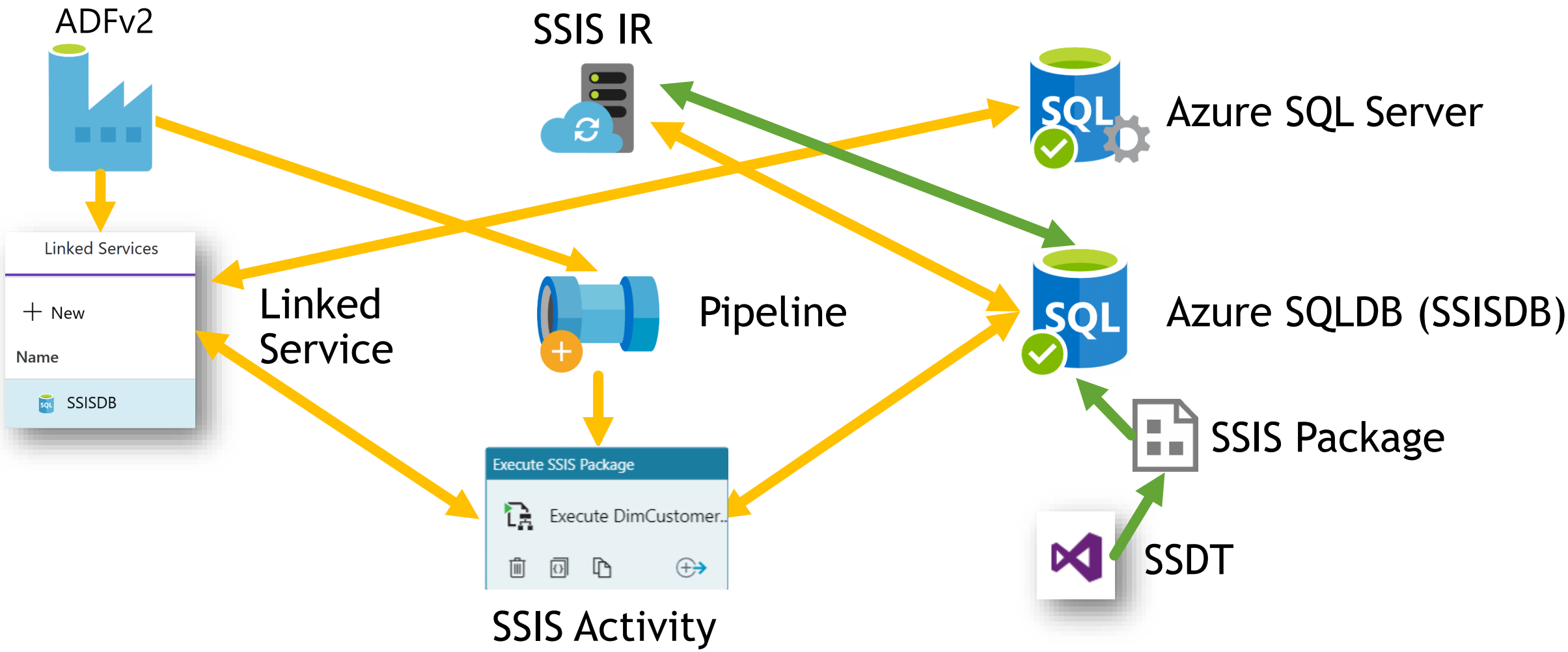
SSIS in Azure

Design Patterns & Conclusions

SSIS Integration Runtimes

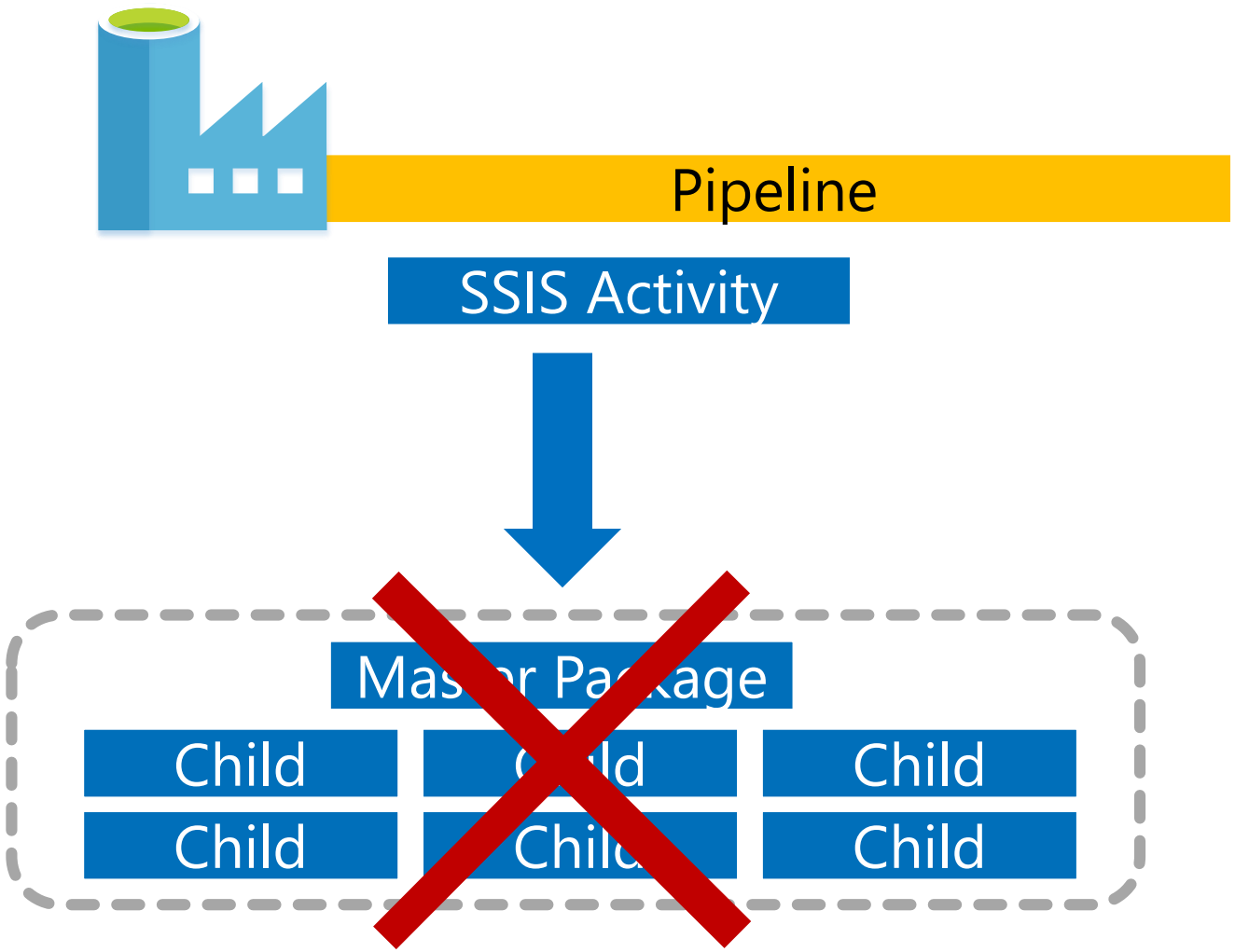


Configuring SSIS in Azure



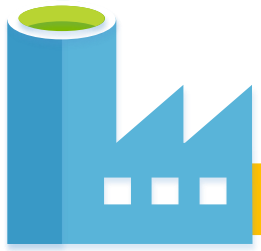
Demo: SSIS

Child Executions



This will execute the packages on a single node, rather than scale out across multiple!

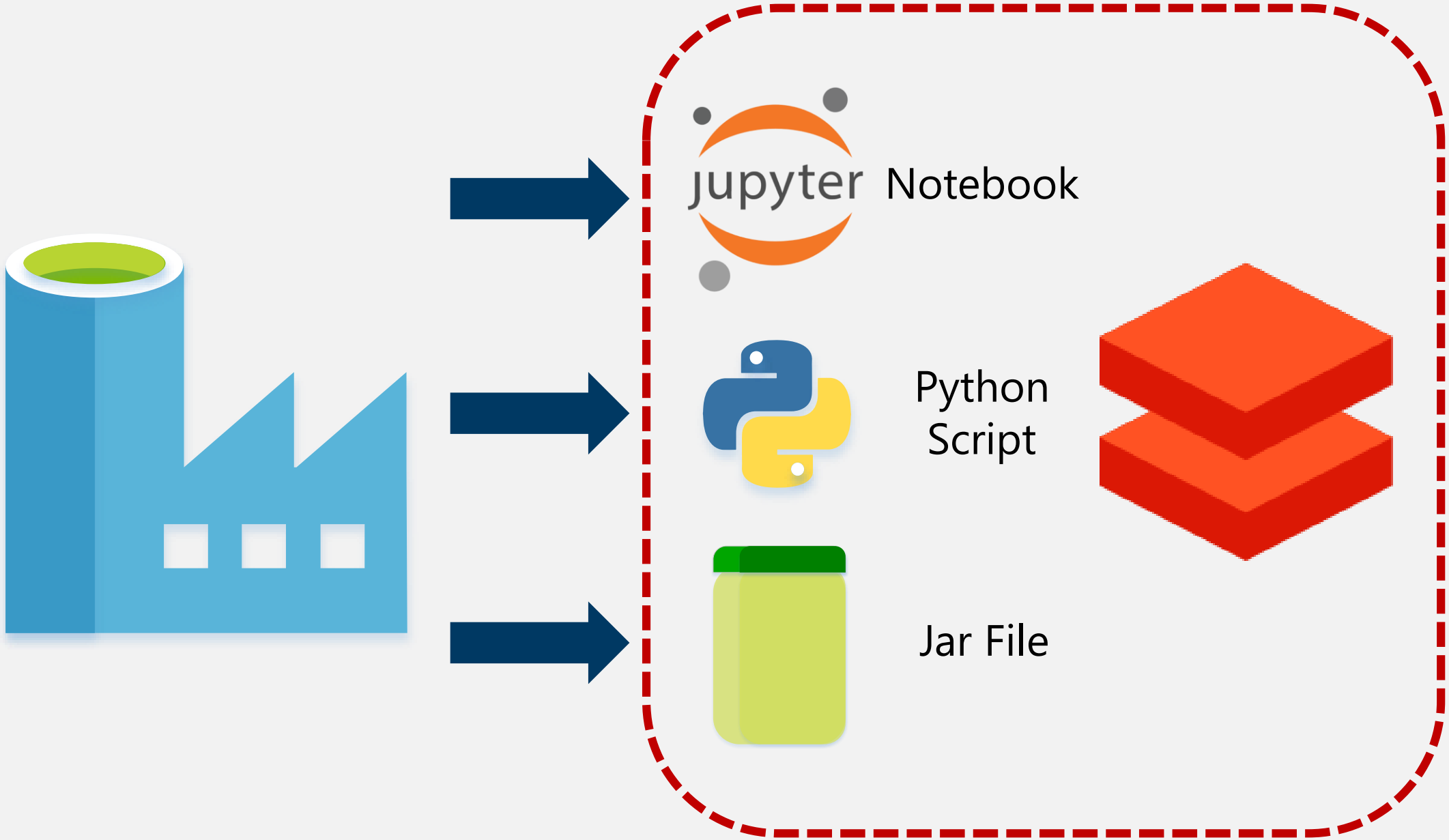
Child Executions



Pipeline

ForEach Activity

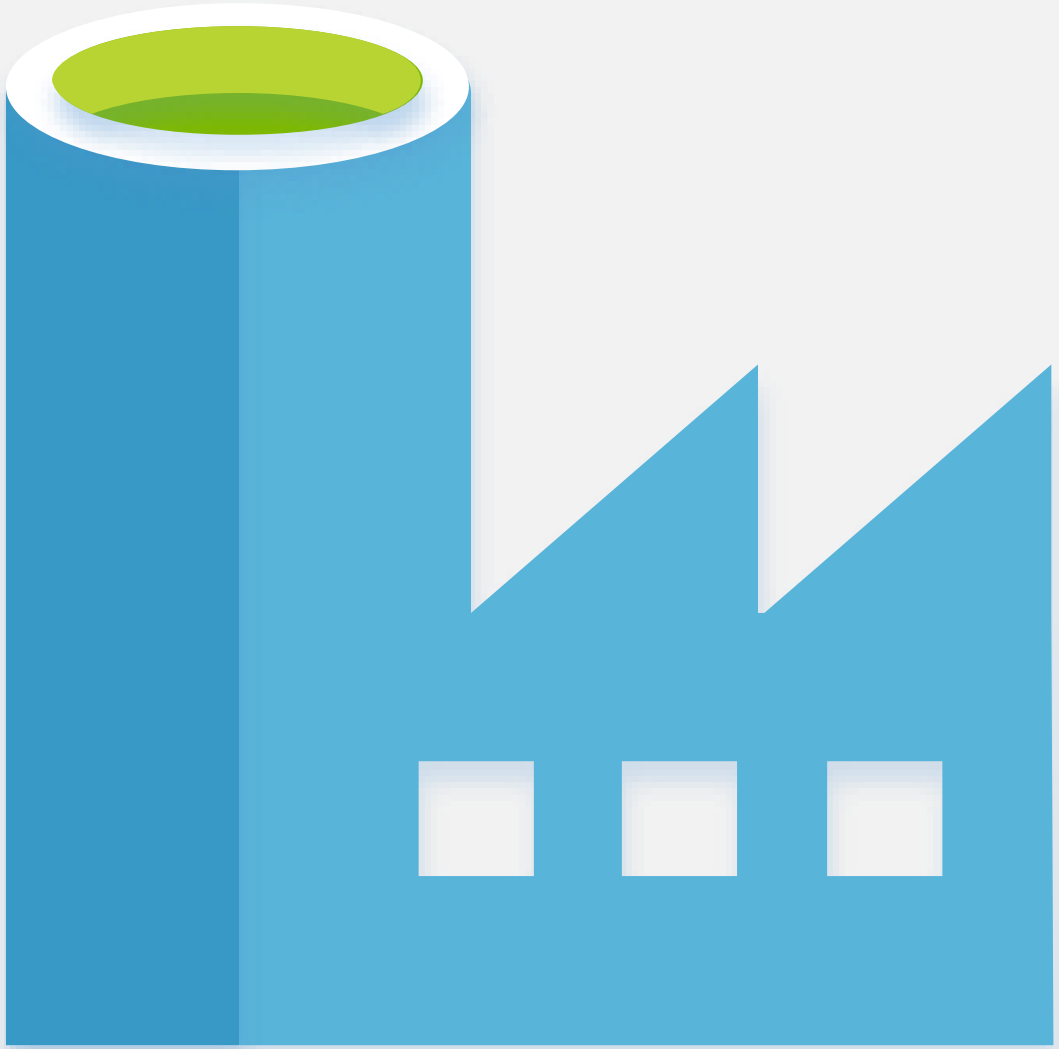
SSIS Activity	→	Child
SSIS Activity	→	Child
SSIS Activity	→	Child
SSIS Activity	→	Child
SSIS Activity	→	Child
SSIS Activity	→	Child



adatis

But what if I don’t want to write any code?

adatis

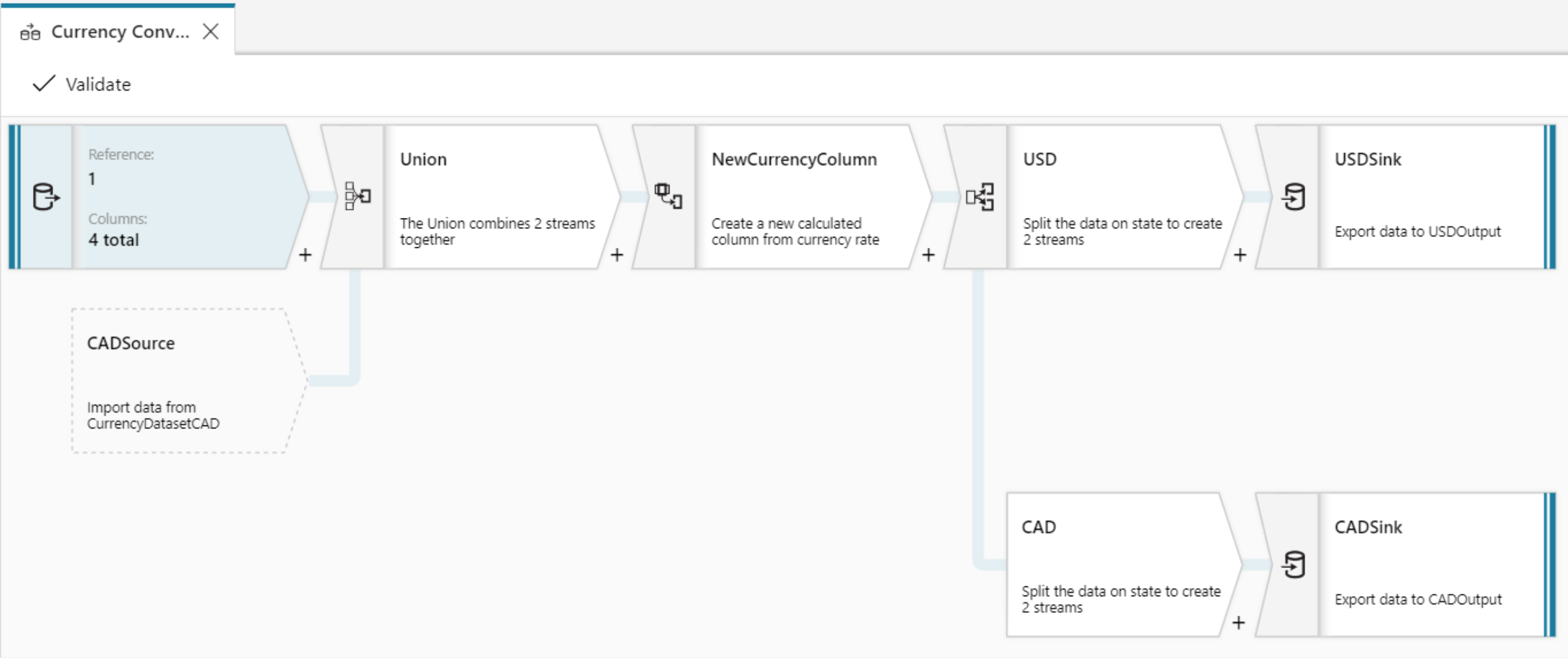


Azure Data Factory

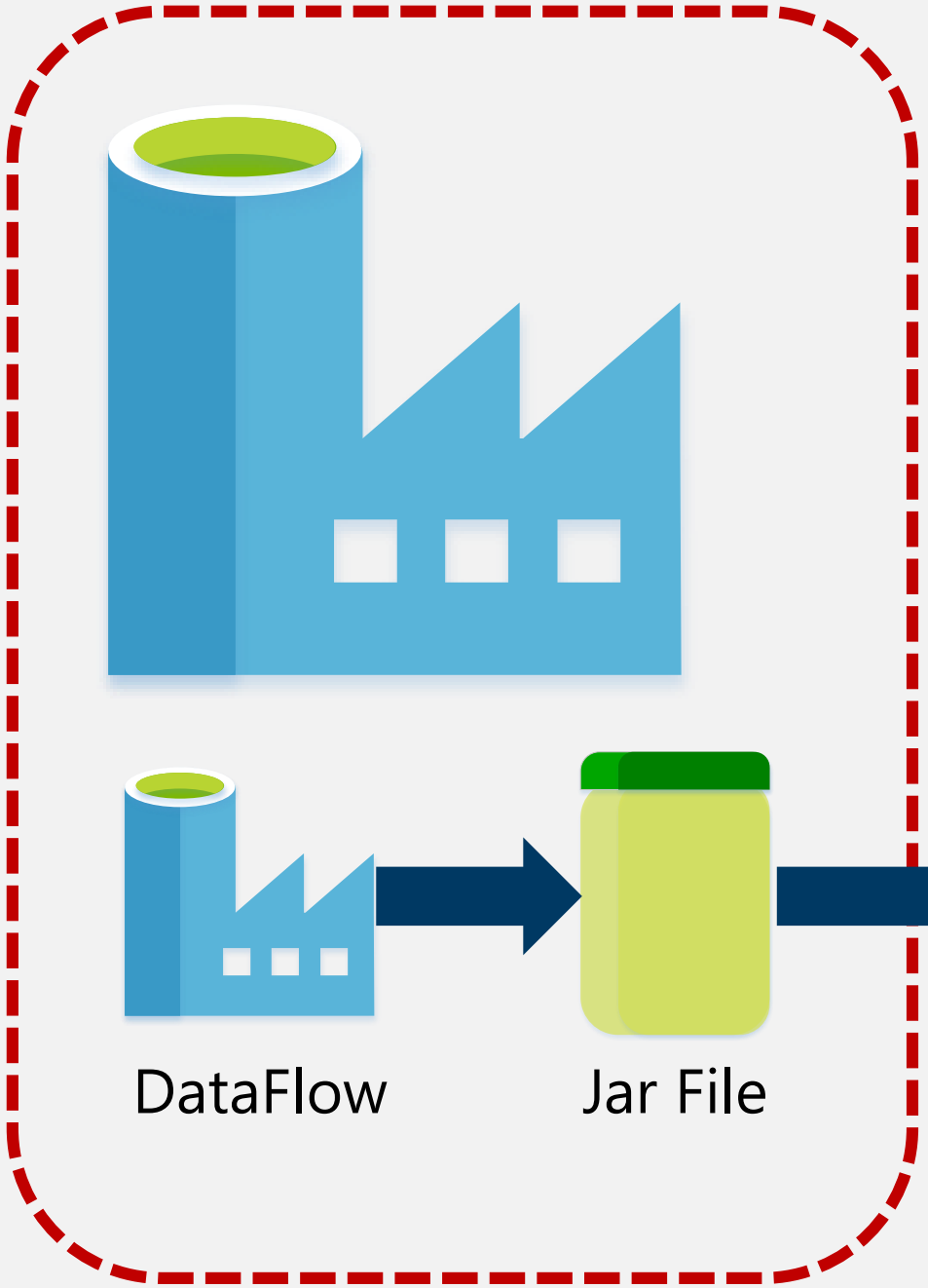
Data Flows

adatis

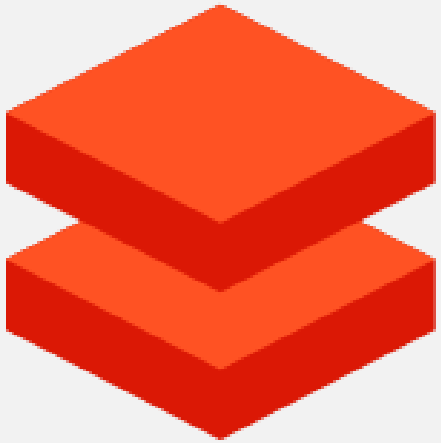
New Data Factory DataFlows can write Databricks processing packages for you!!



adatis



Dataflows will compile down to a JAR file which will be sent to the Databricks cluster for execution



This means it uses Scala!

adatis

Demo: DataFlows

Agenda

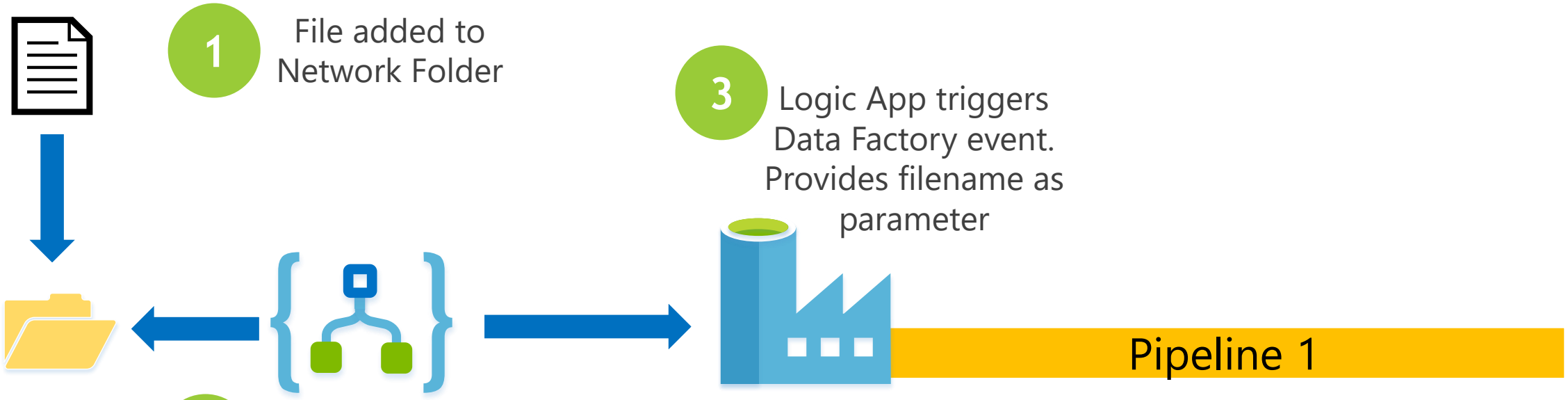
Data Factory Recap

ADFV2 Features Update

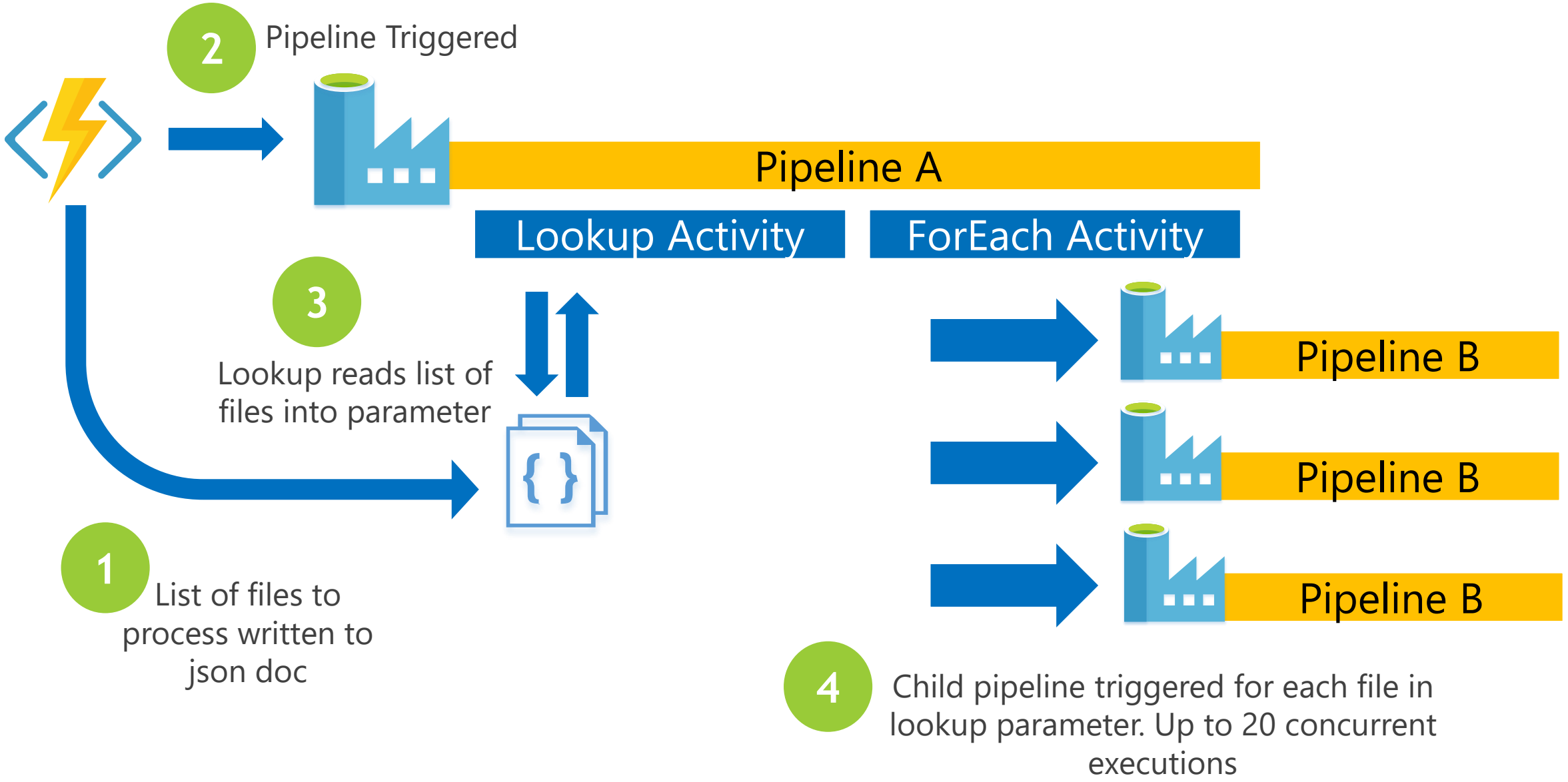
SSIS in Azure

Design Patterns & Conclusions

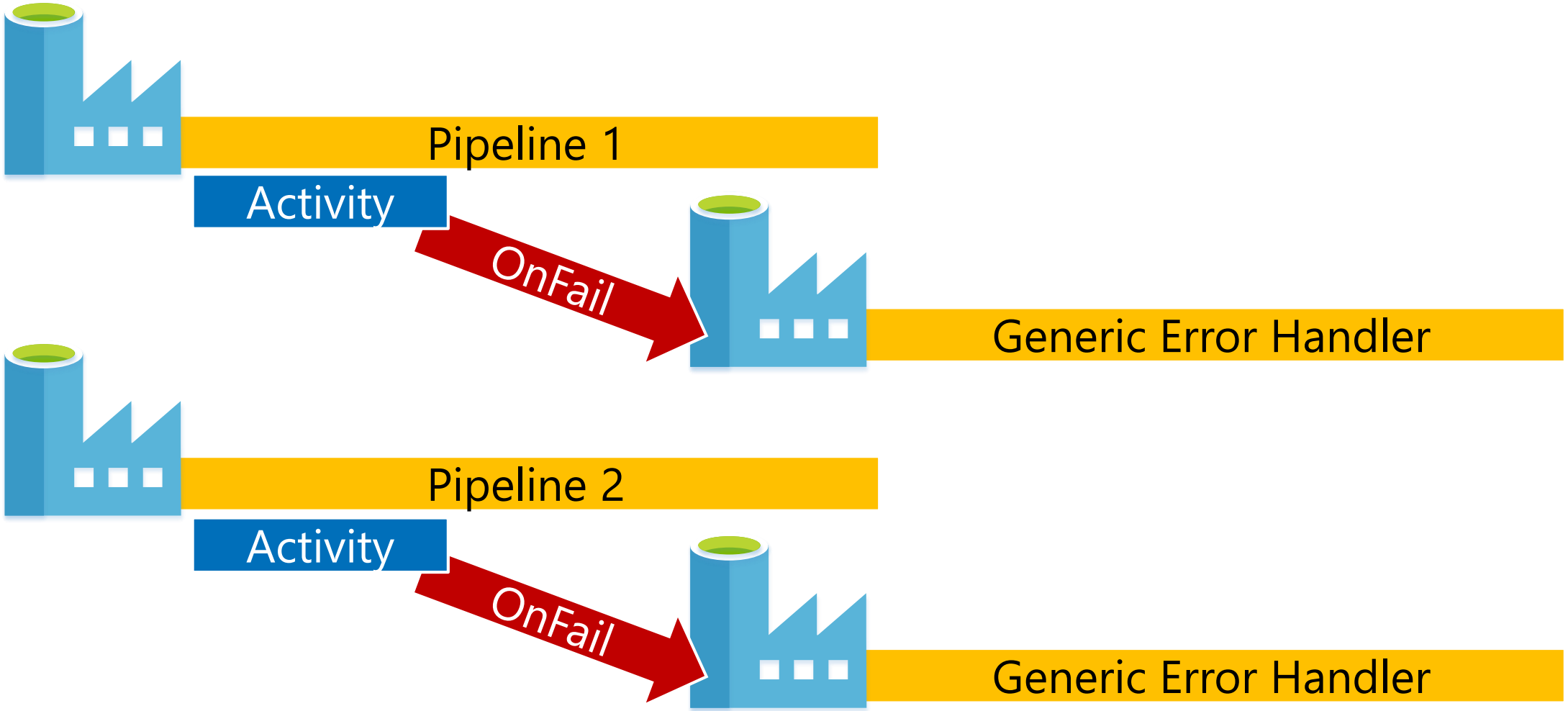
Reactive Data Processing



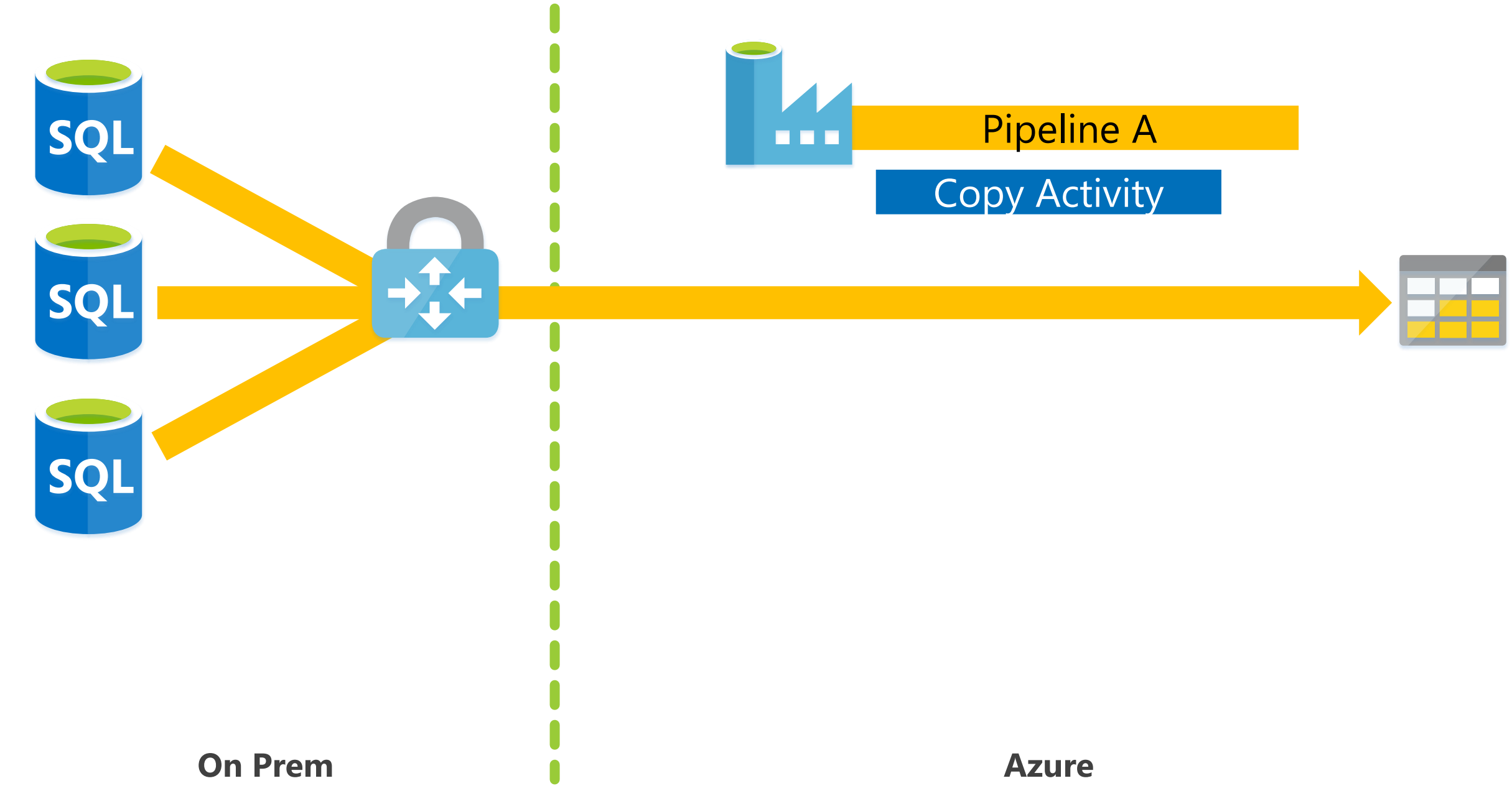
Batch Queuing



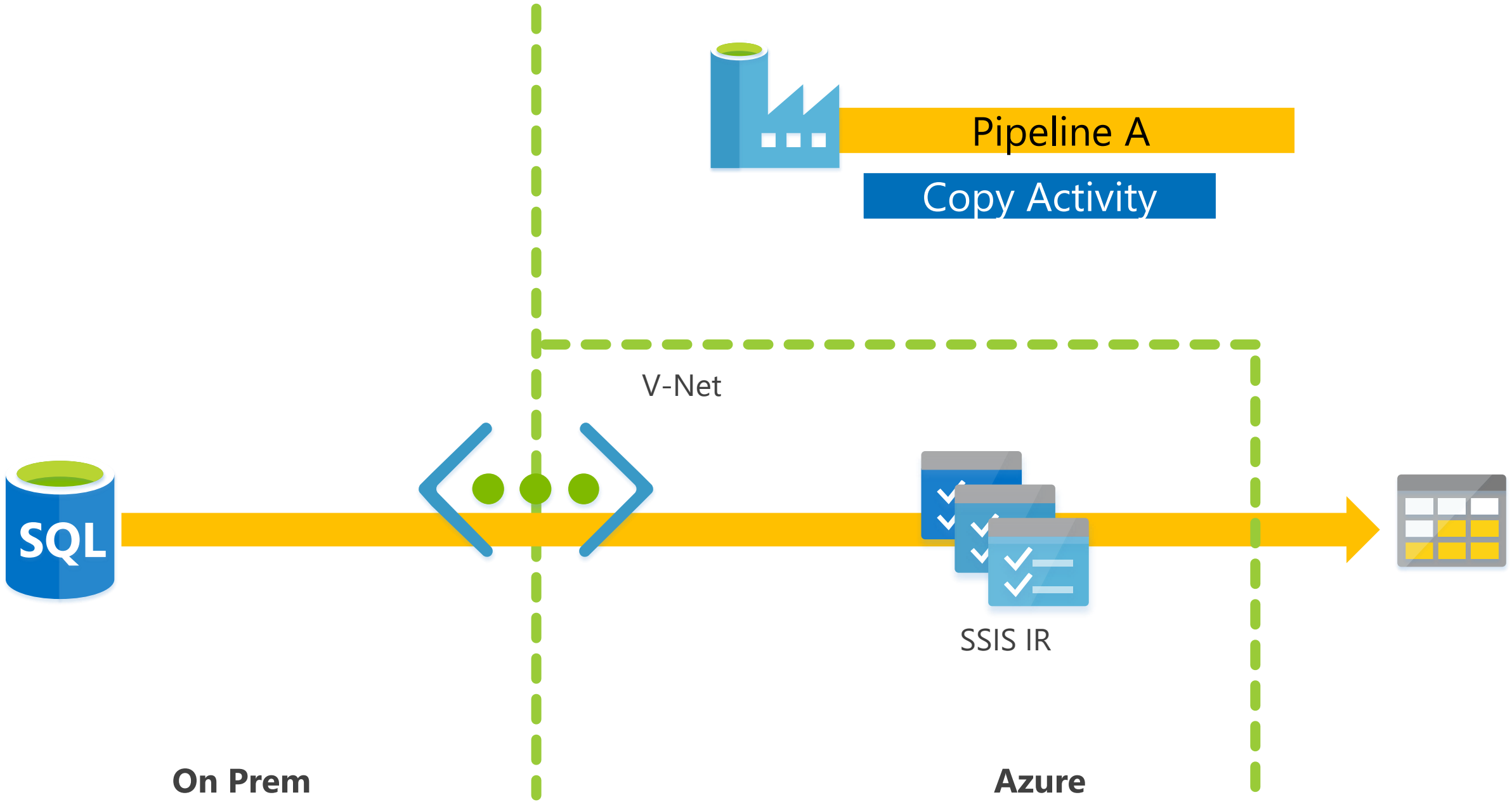
Reusable Pipelines



The SSIS Gateway

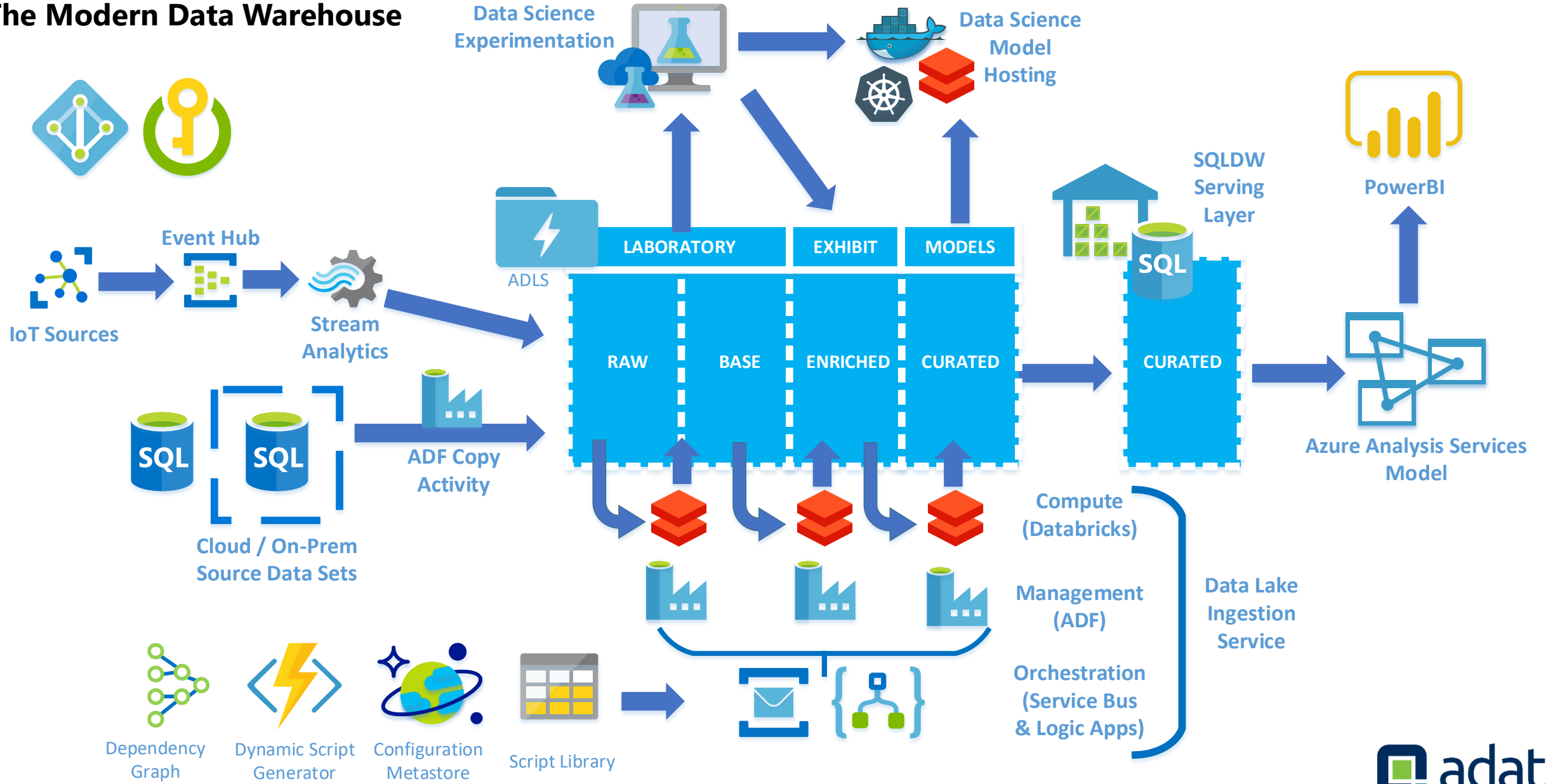


The SSIS Gateway



Why use Azure Data Factory?

The Modern Data Warehouse



adatis

Thanks for Listening

Simon Whiteley

@MrSiWhiteley

http://blogs.adatis.co.uk