

Azure Data Factory v2

SSIS Data Flows & Custom Extensibility

Paul Andrew | Senior Consultant

08/09/2018

 @MrPaulAndrew



Gold Data Analytics
Gold Data Platform
Gold Cloud Platform



Thank you to our sponsors!

sqlcloud

Coeeo
Making SQL sense

dox42[®]
automate your documents
integrate your data

IDERA

Quest

LightningTools

 adatis

 happit

WEBCON[®]
BUSINESS APPLICATIONS

 SOLUTIONS
2SHARE

 redgate

**Lawrence
HARVEY**

 AvePoint[®]

 Microsoft

 PASS



<https://github.com/mrpaulandrew>

CommunityEvents

Demo code, content and slides from various community events.

● C++

[{Event/Location}-{Month}-{Year}](#)

Agenda

Data Factory

Concepts

Components

Why use it?

Data Factory Extensibility

SSIS, Functions,
Custom Activities

Conclusions

Design Patterns
ETL/ELT in Azure

Coming Soon!

Data Flows with
Data Bricks

Agenda

Data Factory

Concepts

Components

Why use it?

Data Factory Extensibility

SSIS, Functions,
Custom Activities

Conclusions

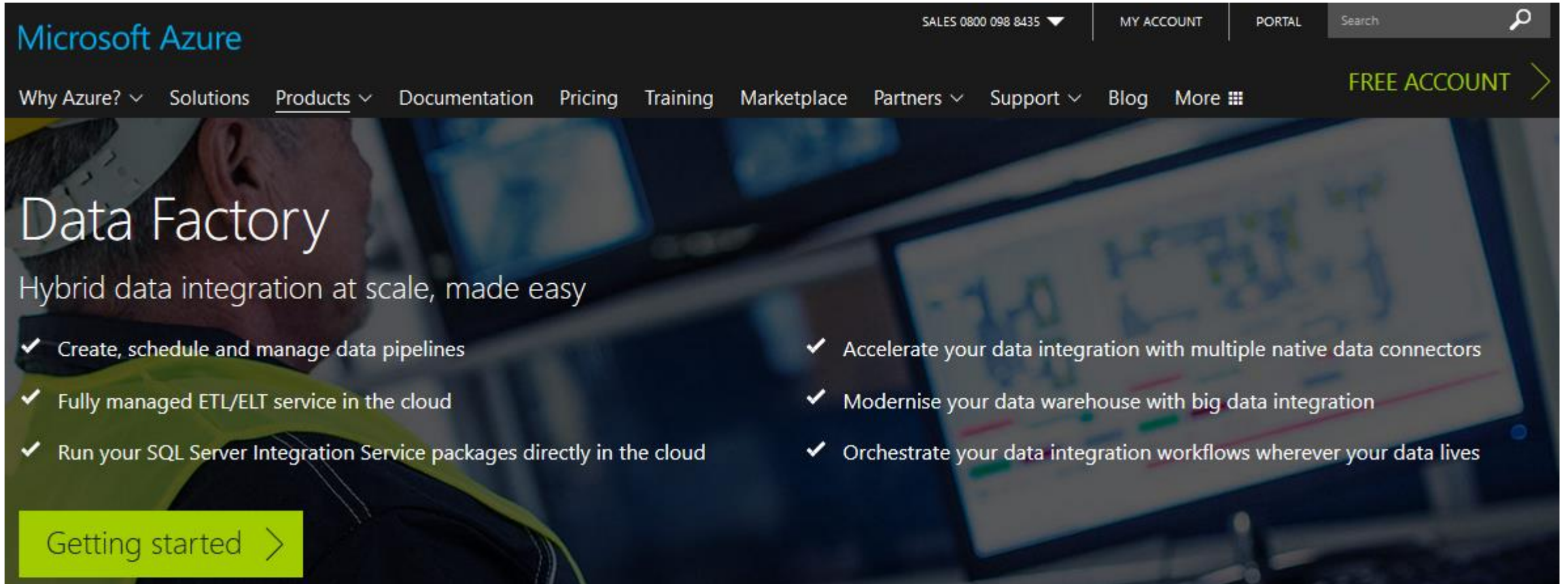
Design Patterns
ETL/ELT in Azure

Coming Soon!

Data Flows with
Data Bricks

What is Azure Data Factory?

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



Microsoft Azure

SALES 0800 098 8435 ▼ | MY ACCOUNT | PORTAL | Search 🔍

Why Azure? ▾ Solutions Products ▾ Documentation Pricing Training Marketplace Partners ▾ Support ▾ Blog More ☰

FREE ACCOUNT ➤

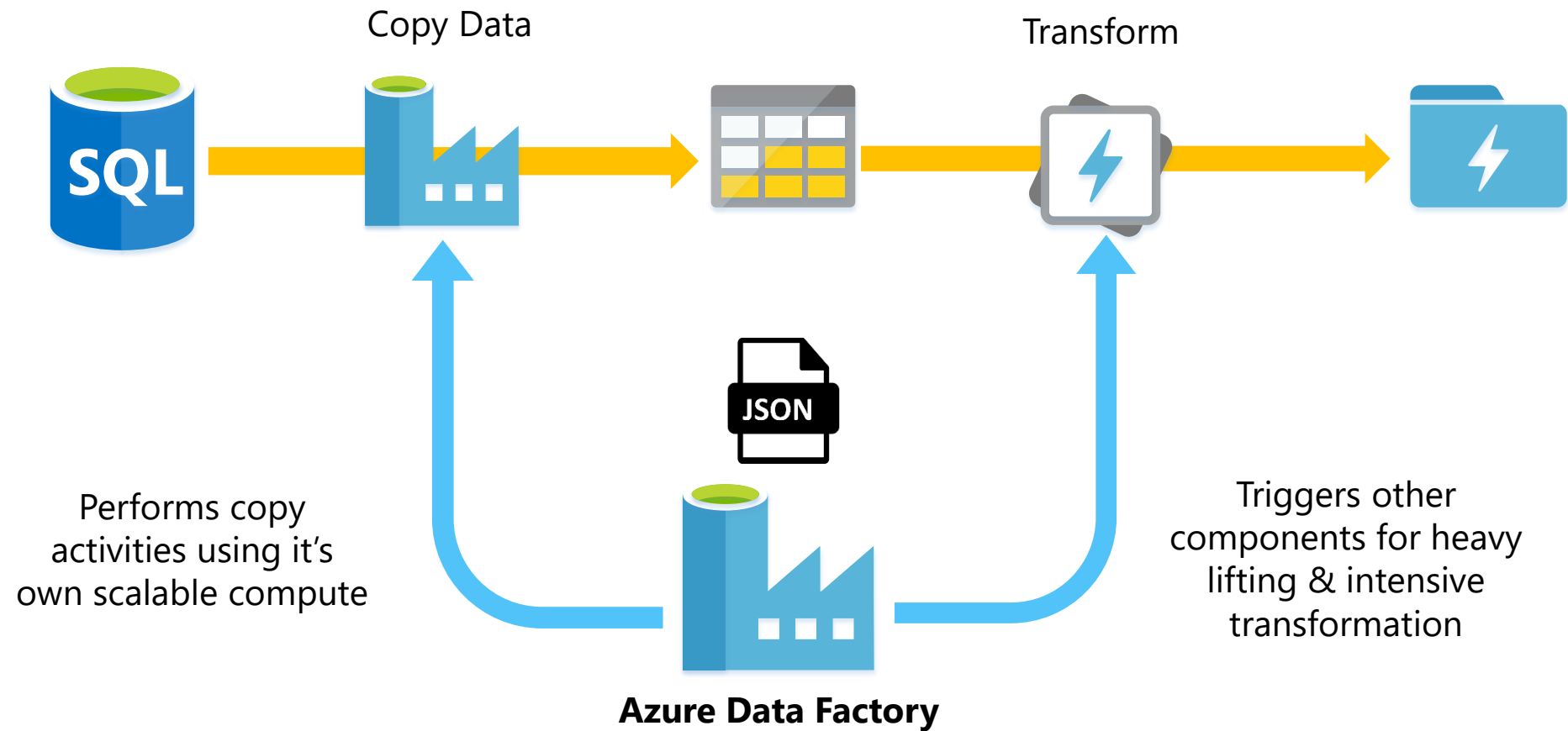
Data Factory

Hybrid data integration at scale, made easy

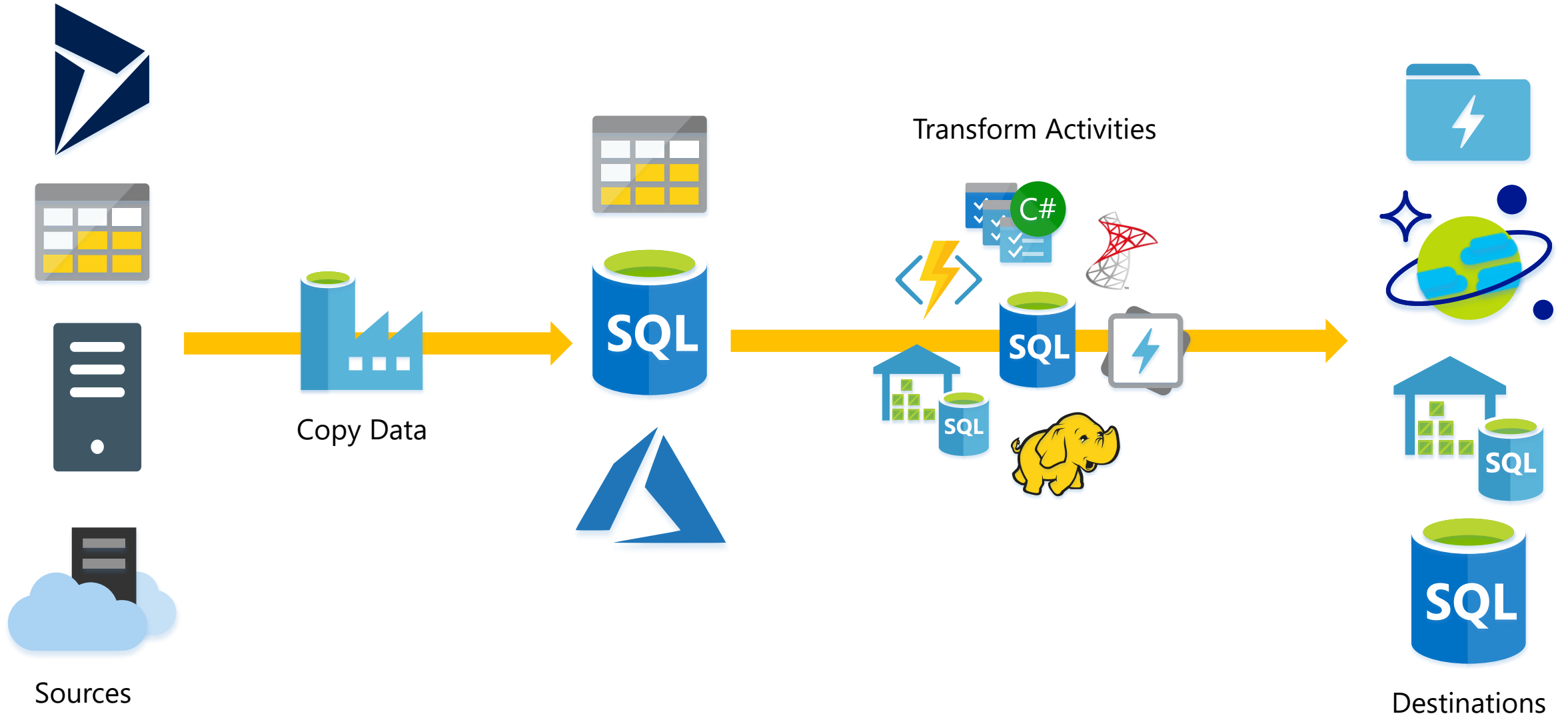
- ✓ Create, schedule and manage data pipelines
- ✓ Fully managed ETL/ELT service in the cloud
- ✓ Run your SQL Server Integration Service packages directly in the cloud
- ✓ Accelerate your data integration with multiple native data connectors
- ✓ Modernise your data warehouse with big data integration
- ✓ Orchestrate your data integration workflows wherever your data lives

Getting started ➤

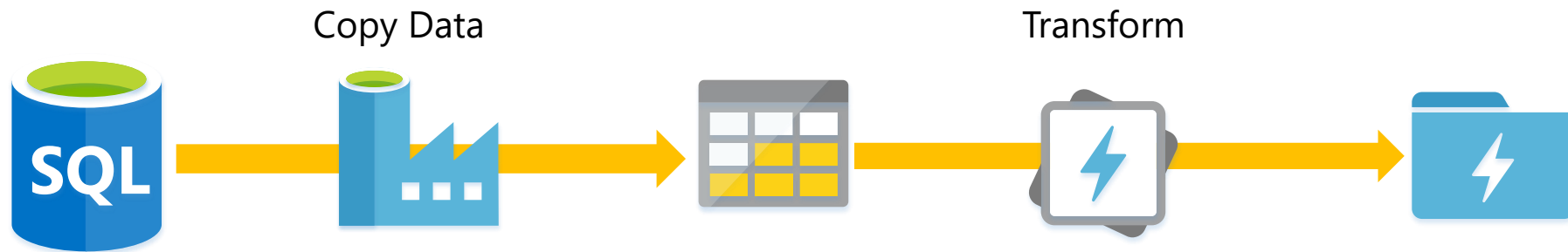
What is Azure Data Factory?



What does Azure Data Factory do?



Data Factory Components

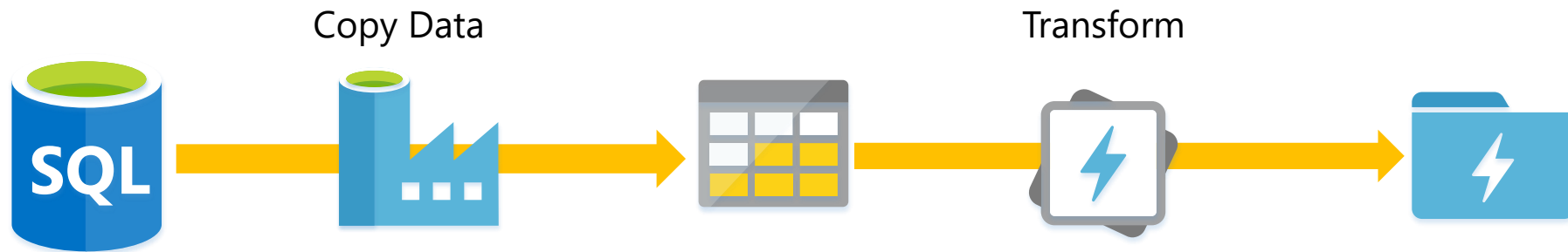


1 Linked Services – How do I connect?

Like the SSIS Connection Manager!



Data Factory Components



1

Linked Services

2

Data Sets – What slices/partitions does my data have?

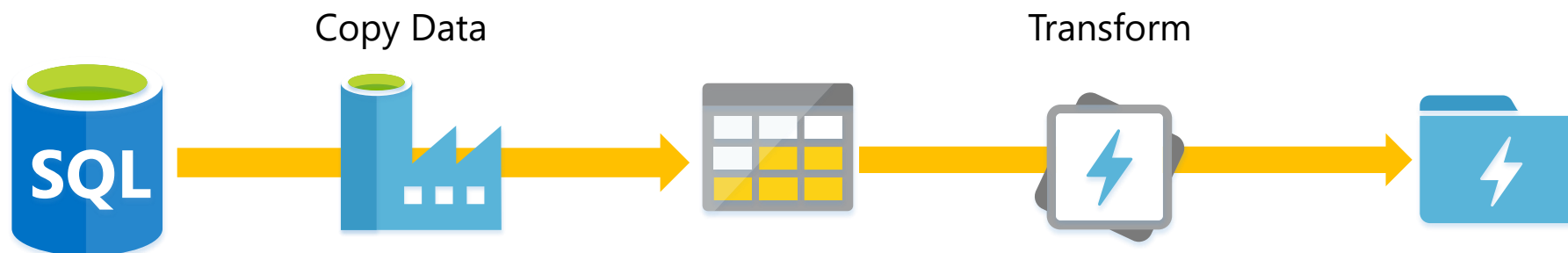


dbo.DimCustomer



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

Data Factory Components



1

Linked Services

2

Data Sets

3

Activities – What do we want to happen?
With what conditions?



U-SQL Activity

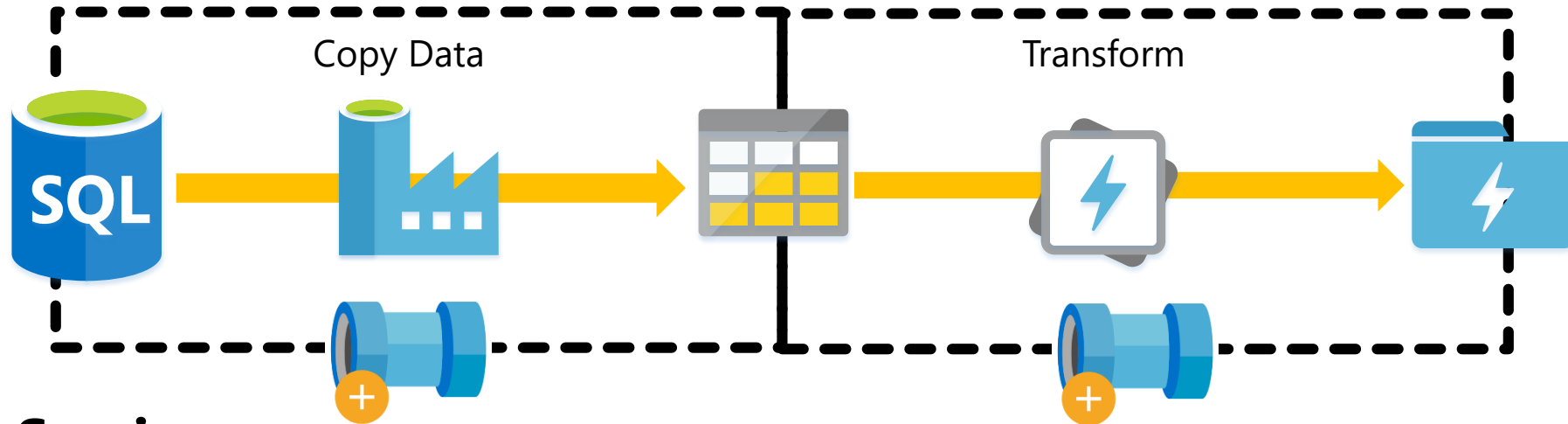
Script: *wasb://:myscripts/ProcessOrders.usql*

AUs: *5 units*

Priority: *1000*

Parameters: *@Output = "RAW/Orders/..."*

Data Factory Components



1

Linked Services

2

Data Sets

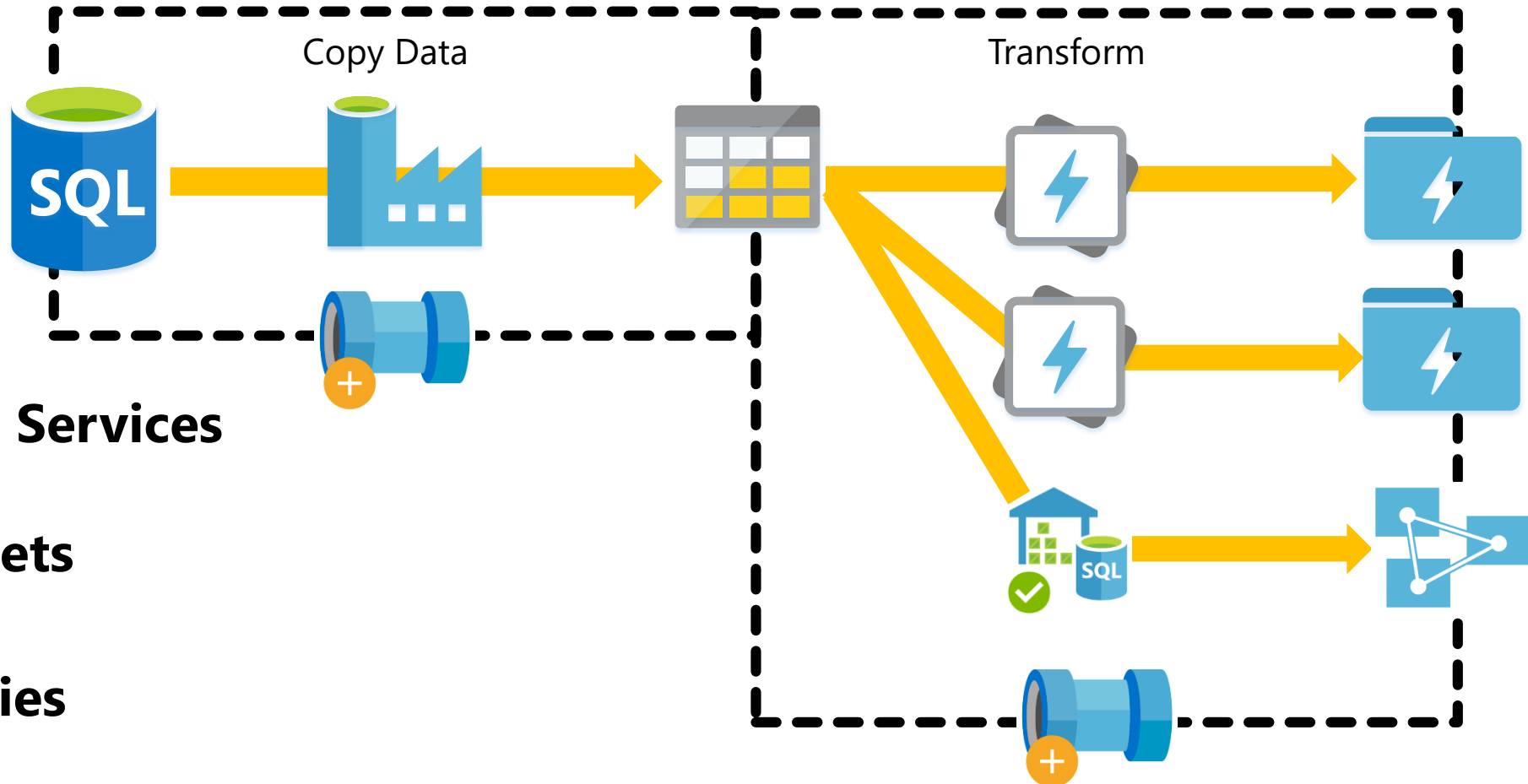
3

Activities

4

Pipelines – What groups of work do I want to do?

Data Factory Components



1

Linked Services

2

Data Sets

3

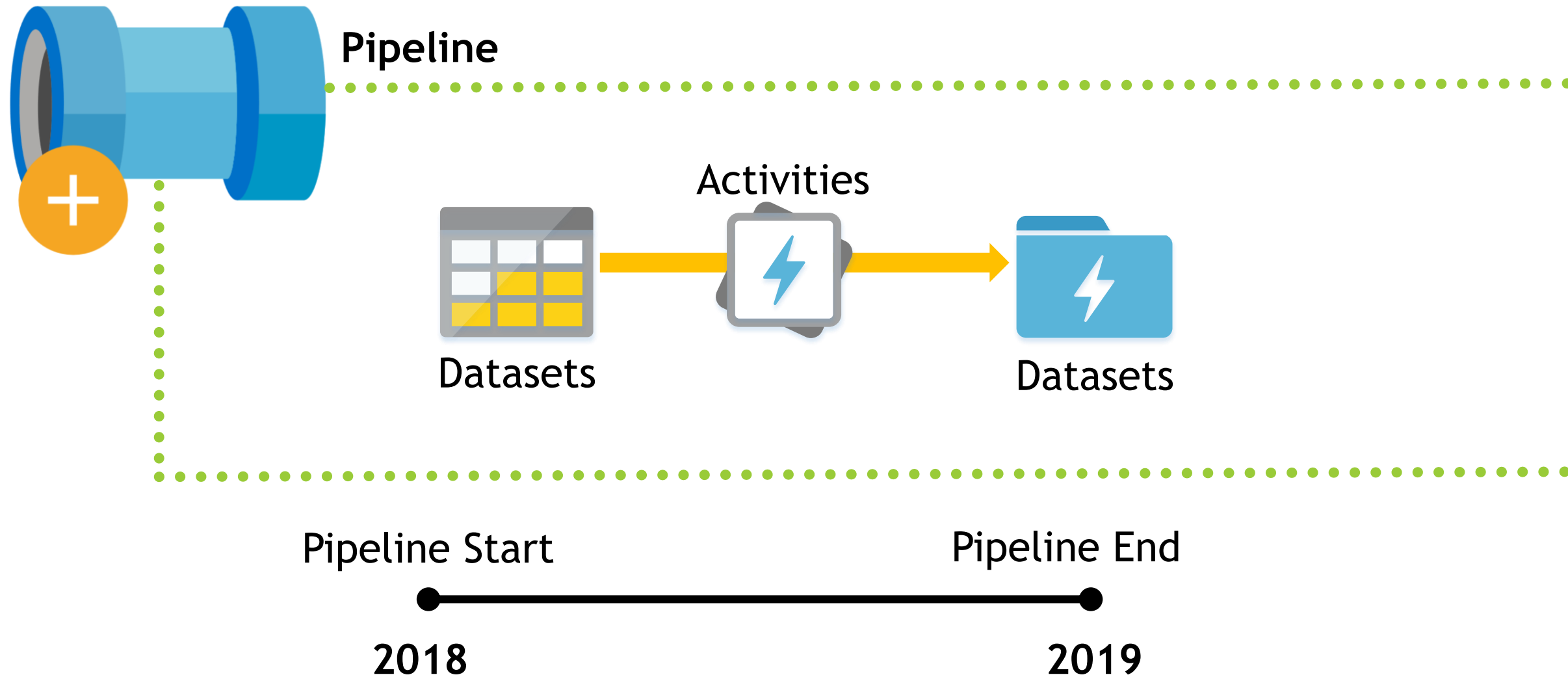
Activities

4

Pipelines – What groups of work do I want to do?

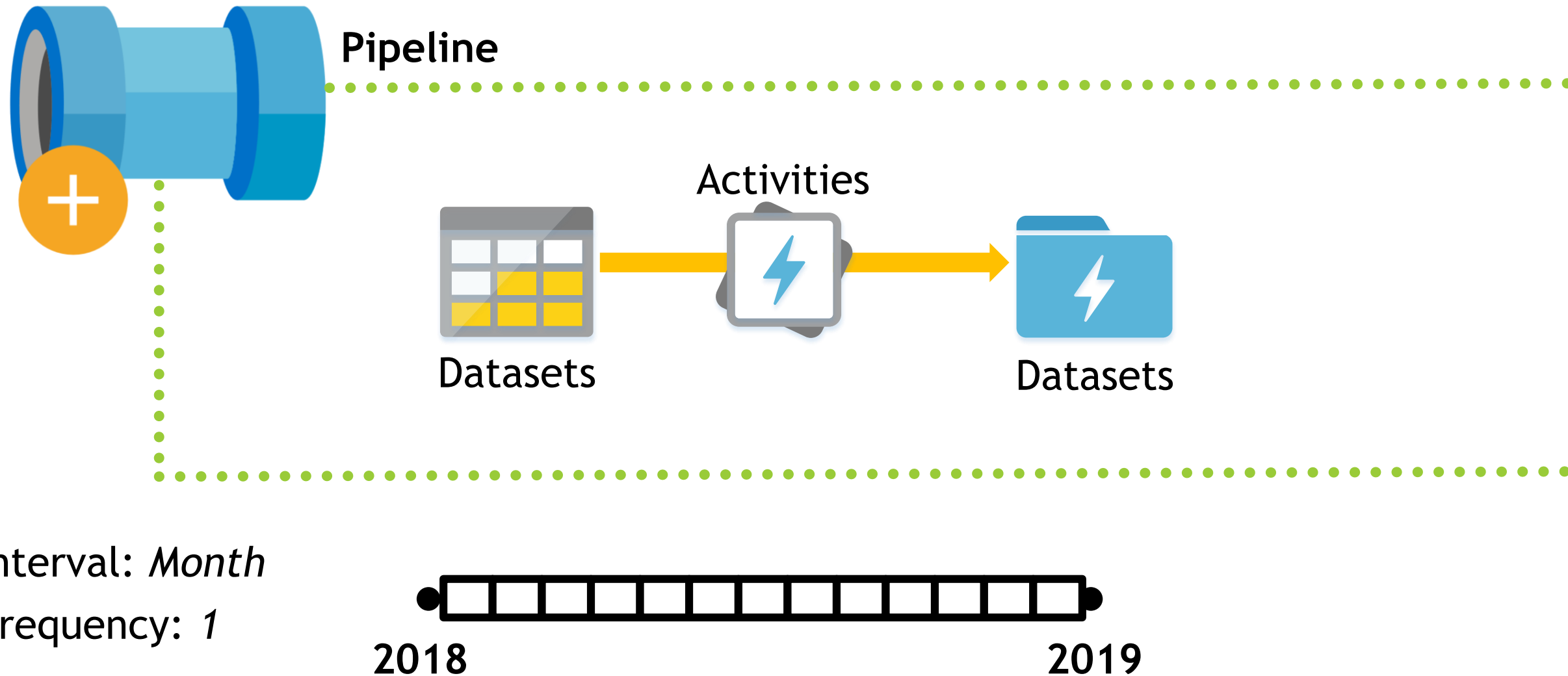
Azure Data Factory Concepts

Time Slices – triggering an activity execution.



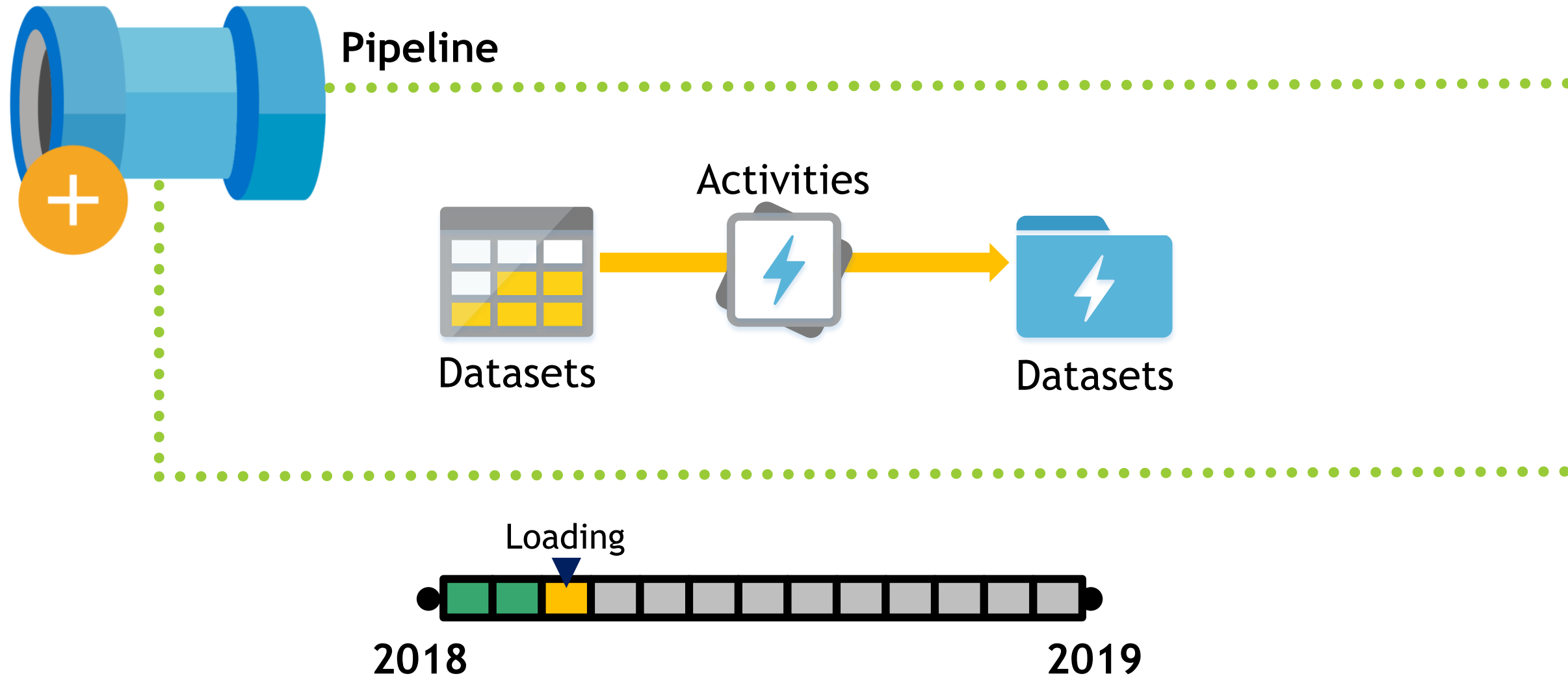
Azure Data Factory Concepts Continued

Time Slices – triggering an activity execution.

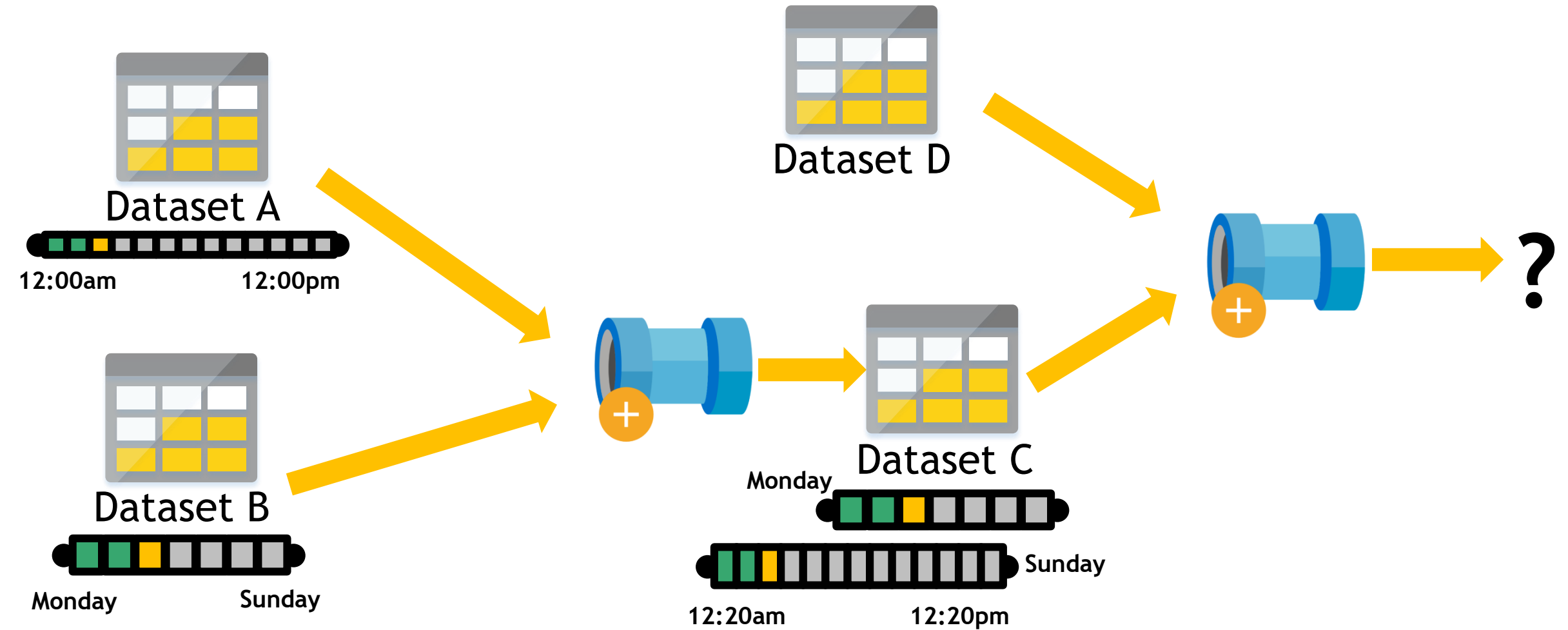


Azure Data Factory Concepts Continued

Time Slices – triggering an activity execution.



Time Slice Problems...



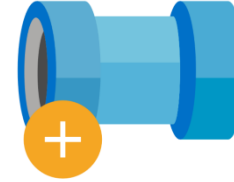
Integration Runtimes



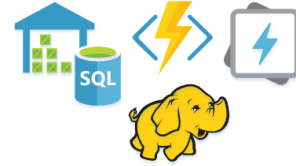
1

Azure
Integration Runtime

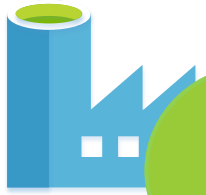
Movement Hours



Activity
Orchestration



Flexible Region



2

SSIS
Integration Runtime

SSIS Package
Execution



Specified Region



3

Self Hosted
Integration Runtime

Local Compute



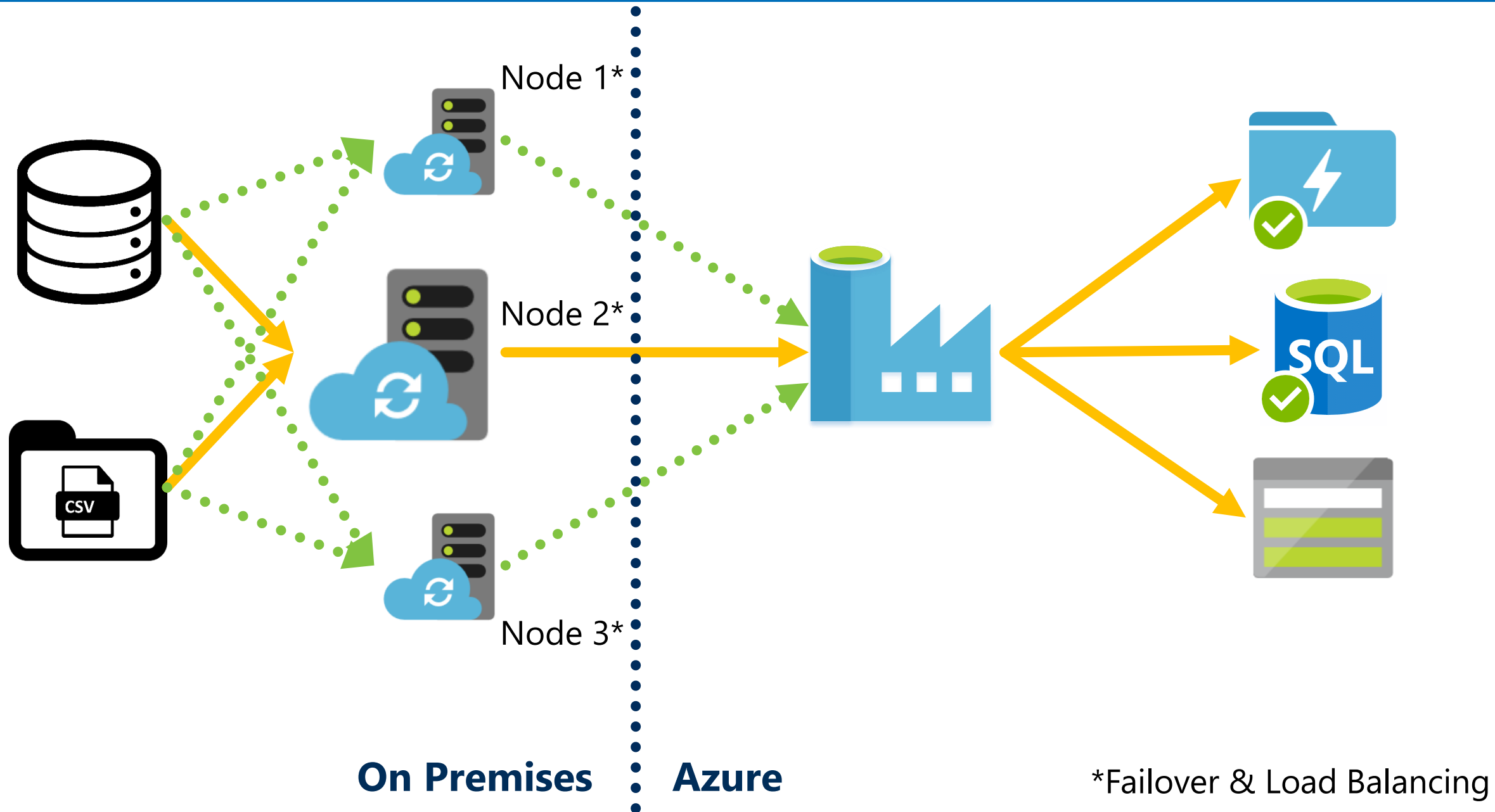
Activity
Orchestration



On-Prem Server



The Integration Runtime (AKA The Data Management Gateway)



Azure Data Factory Concepts & Components Recap

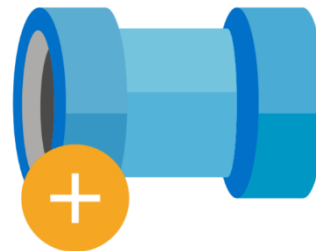


1 **Linked Services**

2 **Data Sets**

3 **Activities**

4 **Pipelines**



Time Slices



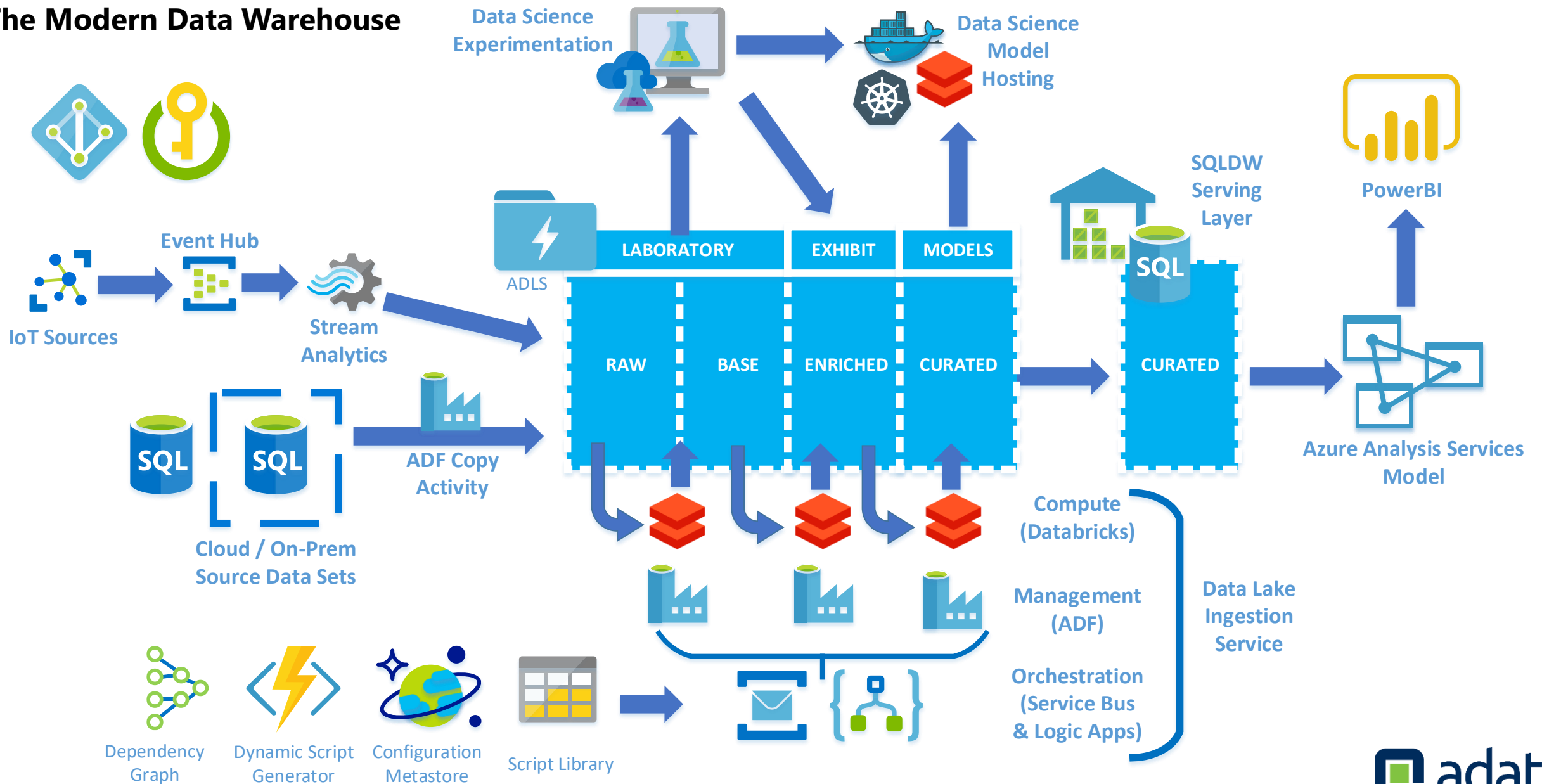
1 **Azure**
Integration Runtime

2 **SSIS**
Integration Runtime

3 **Self Hosted**
Integration Runtime

Why use Azure Data Factory?

The Modern Data Warehouse



Agenda

Data Factory

Concepts

Components

Why use it?

Data Factory Extensibility

SSIS, Functions,
Custom Activities

Conclusions

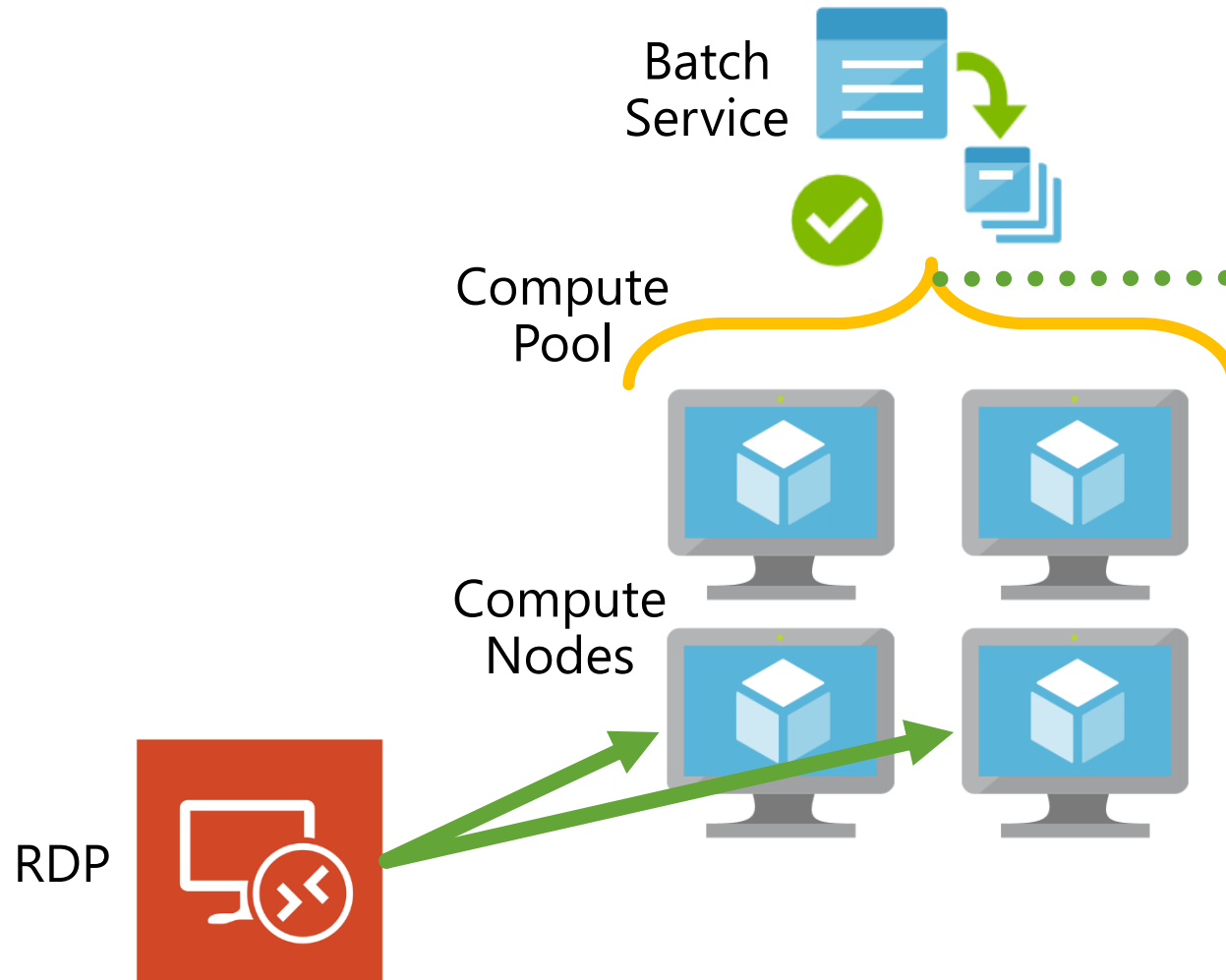
Design Patterns
ETL/ELT in Azure

Coming Soon!

Data Flows with
Data Bricks

1

Custom Activities – A .Net Console App Executed Using Azure Batch Service



VM node size set per compute pool:

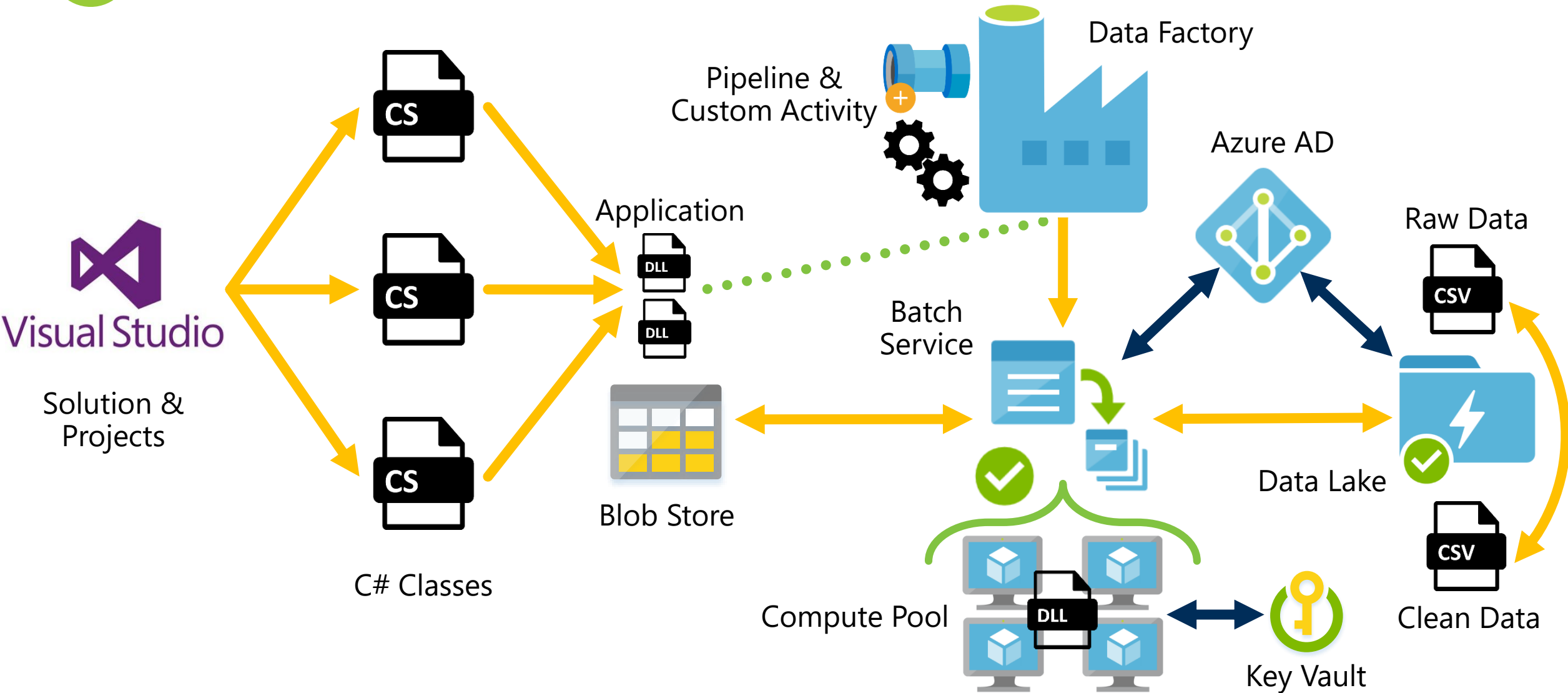
A1 Standard	A2 Standard	A3 Standard
1 Cores	2 Cores	4 Cores
1.8 GB	3.5 GB	7 GB
1 TB OS disk size	1 TB OS disk size	1 TB OS disk size
70 GB Resource disk size	135 GB Resource disk size	285 GB Resource disk size
2 Max data disk	4 Max data disk	8 Max data disk
Unable to display pricing	Unable to display pricing	Unable to display pricing

- ▶ 1 compute node = 1 virtual machine.
- ▶ 1 job per compute node.
- ▶ Max of 4 tasks per node.
- ▶ OS on D drive, not C.
- ▶ Special environment variables.

ADF Extensibility Continued

1

Custom Activities – A .Net Console App Executed Using Azure Batch Service

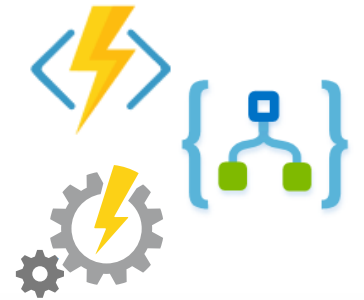


ADF Extensibility Continued

1 **Custom Activities** – A .Net Console App Executed Using Azure Batch Service

2 **Rest API Calls** – Eg. Web Activities Calling:

Azure Functions
Azure Logic Apps
Azure Automation



General Settings² Parameters Advanced

Name * Web1

Description

Timeout 7.00:00:00

Retry 0

Retry interval 20

General Settings² Parameters Advanced

URL *

Method * Select API method...
Select API method...
GET
POST
PUT

Headers

General Settings² Parameters Advanced

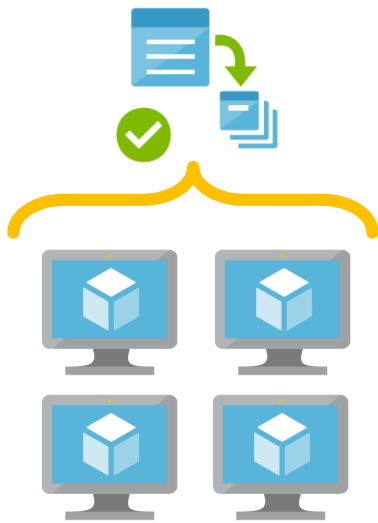
Use [expressions, functions](#) or refer to [system variables](#) in the 'value' column.

Parameterizable properties ⓘ

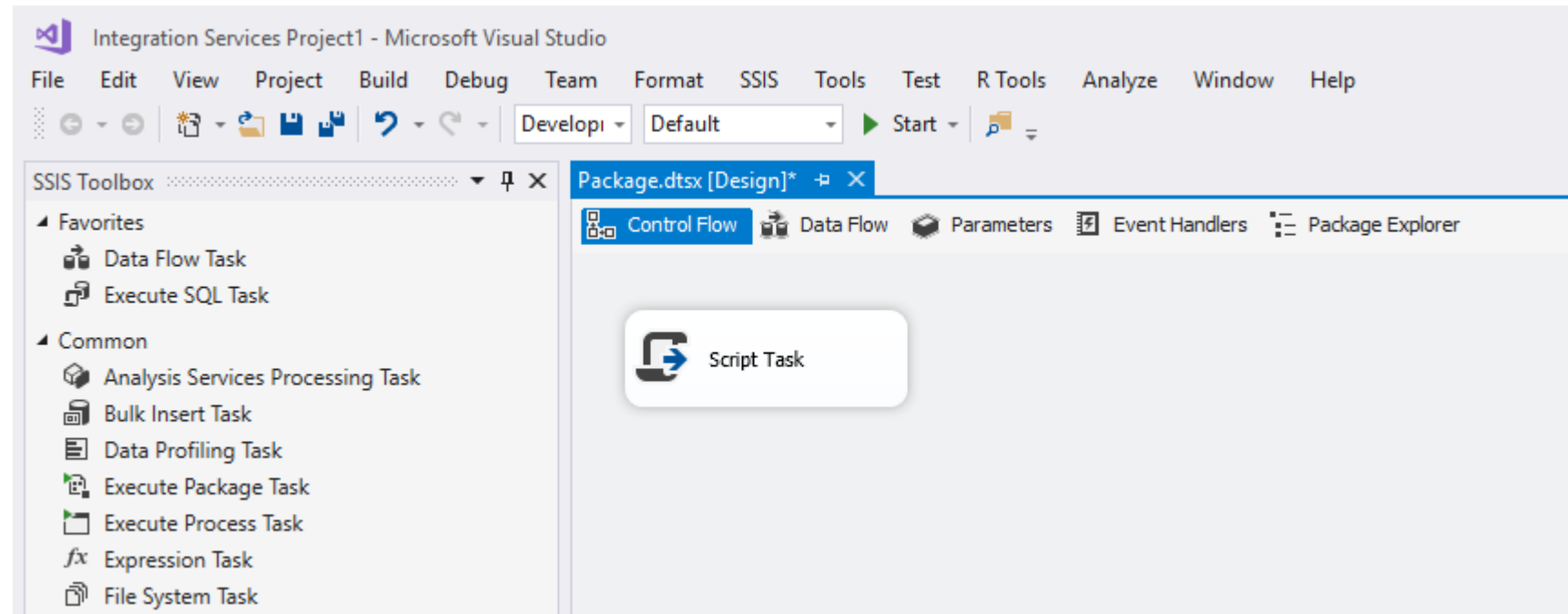
NAME	VALUE
url	<input type="text" value="Value"/>
body	<input type="text" value="Value"/>
Timeout	<input type="text" value="Value"/>
Retry	<input type="text" value="Value"/>

ADF Extensibility Continued

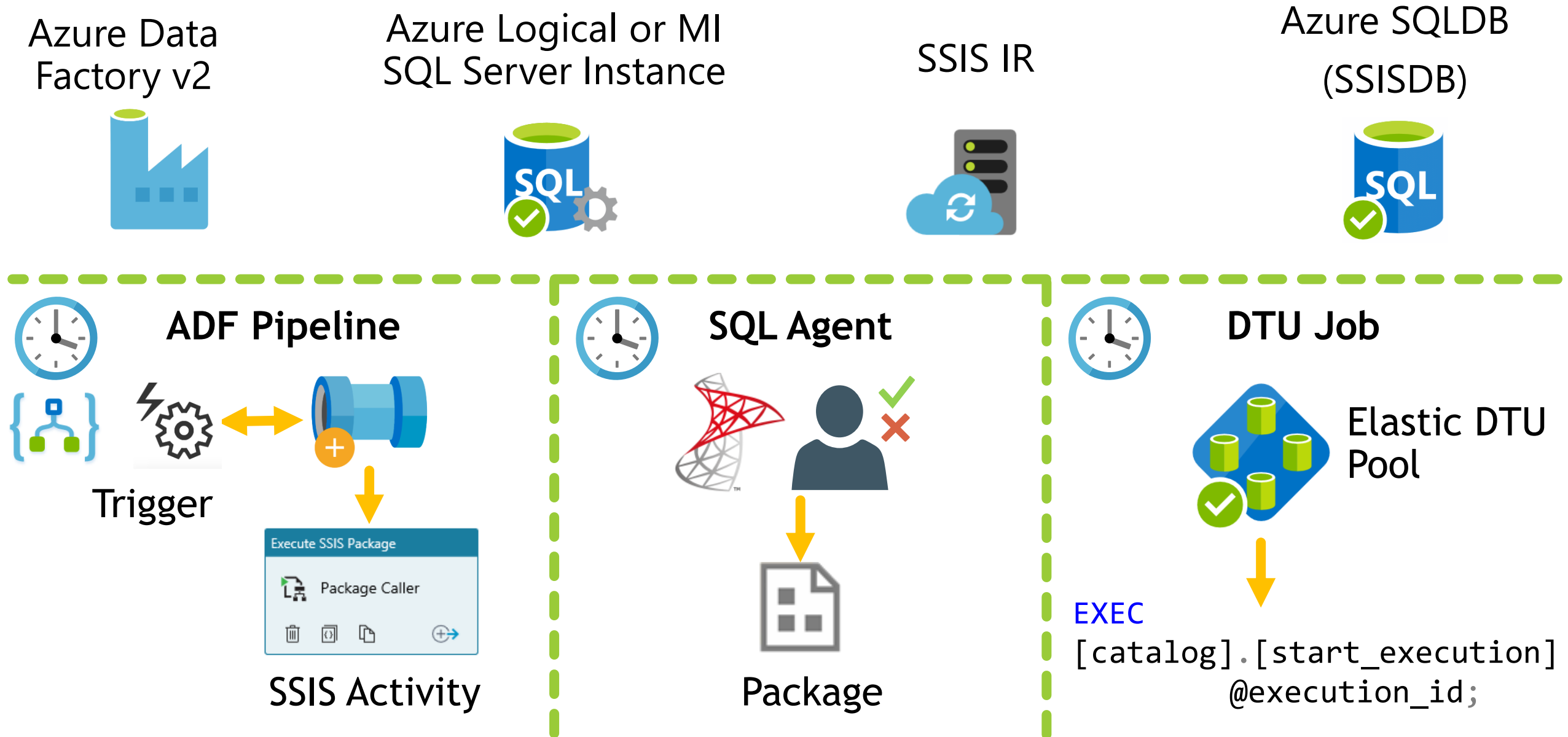
- 1 **Custom Activities**
- 2 **Rest API Calls**
- 3 **SSIS** – Packages with Control Flows and Data Flows



ADF SSIS IR



How do we schedule an SSIS Package in Azure?





Agenda

Data Factory

Concepts

Components

Why use it?

Data Factory Extensibility

SSIS, Functions,
Custom Activities

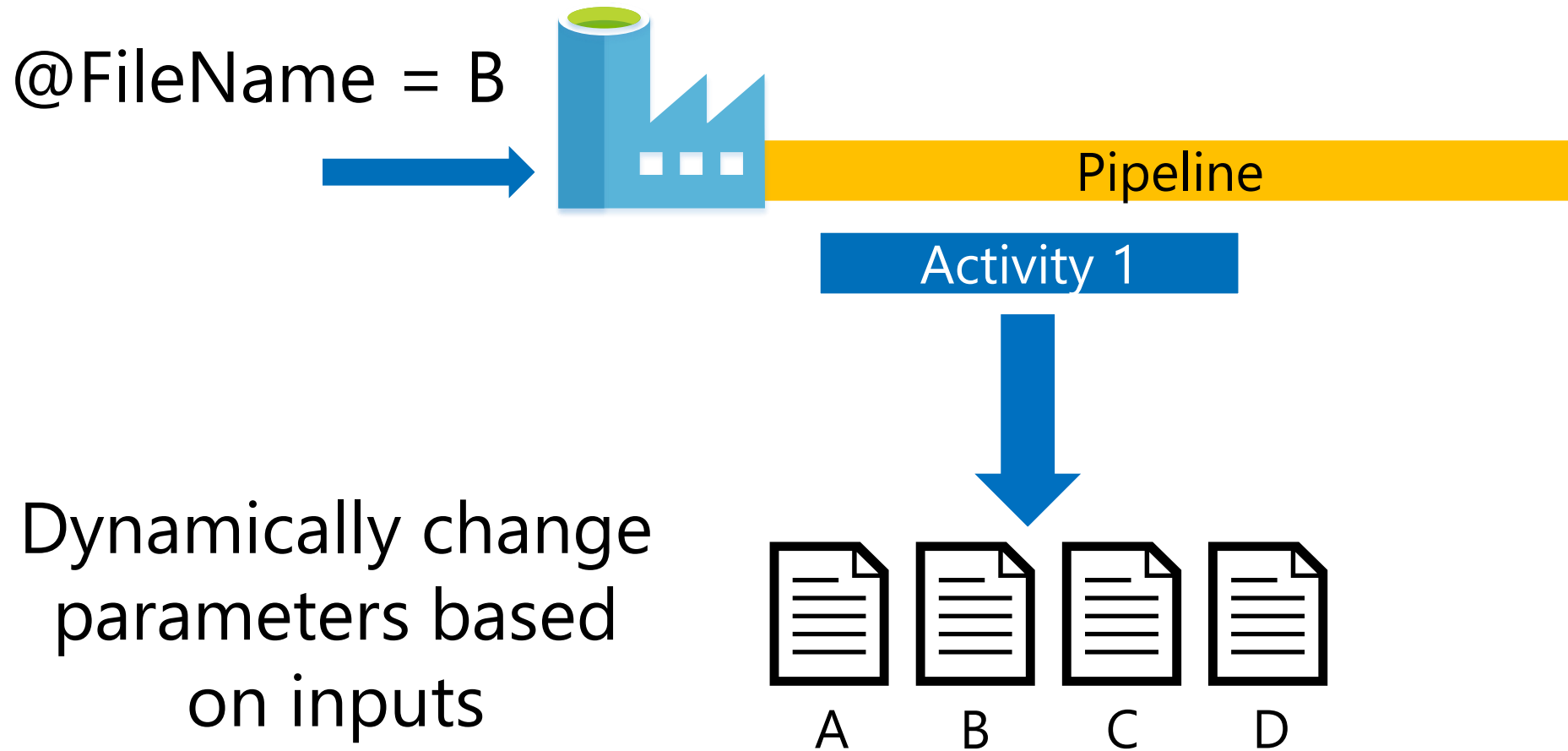
Conclusions

Design Patterns
ETL/ELT in Azure

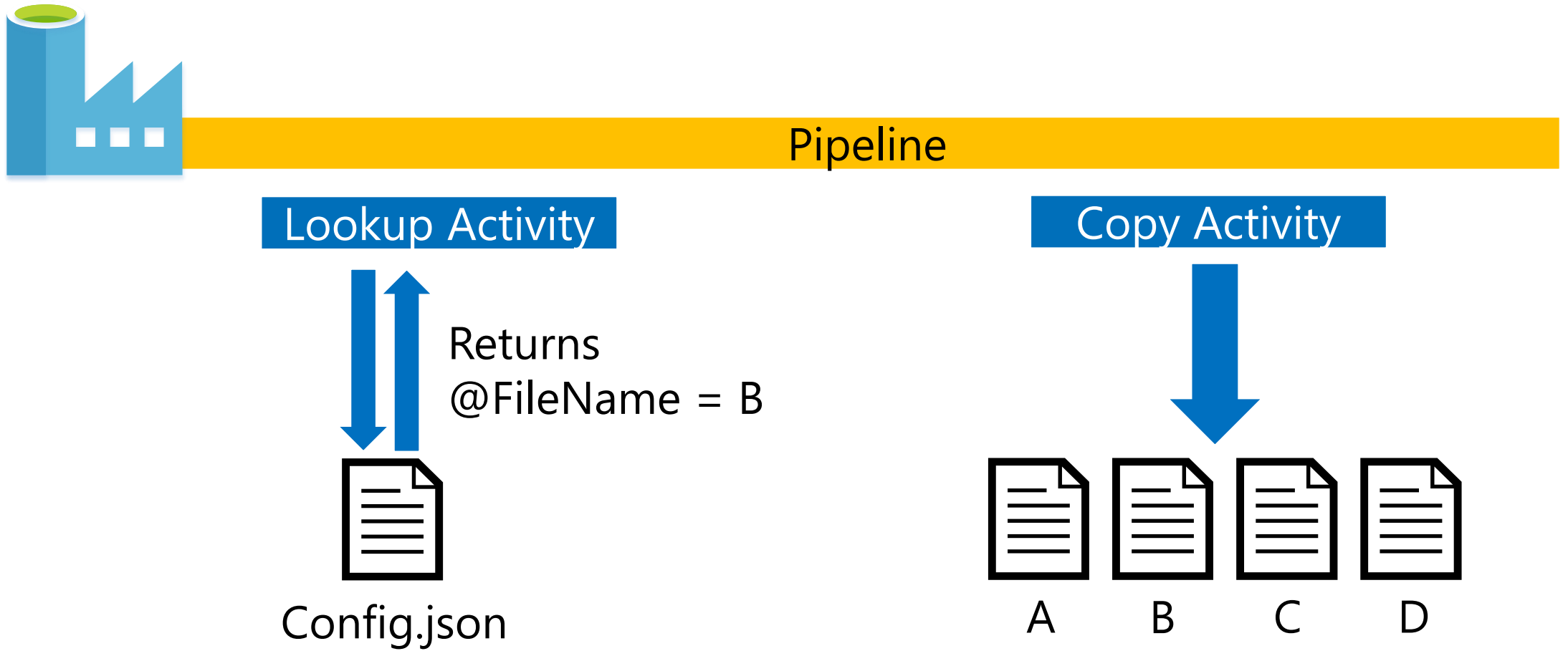
Coming Soon!

Data Flows with
Data Bricks

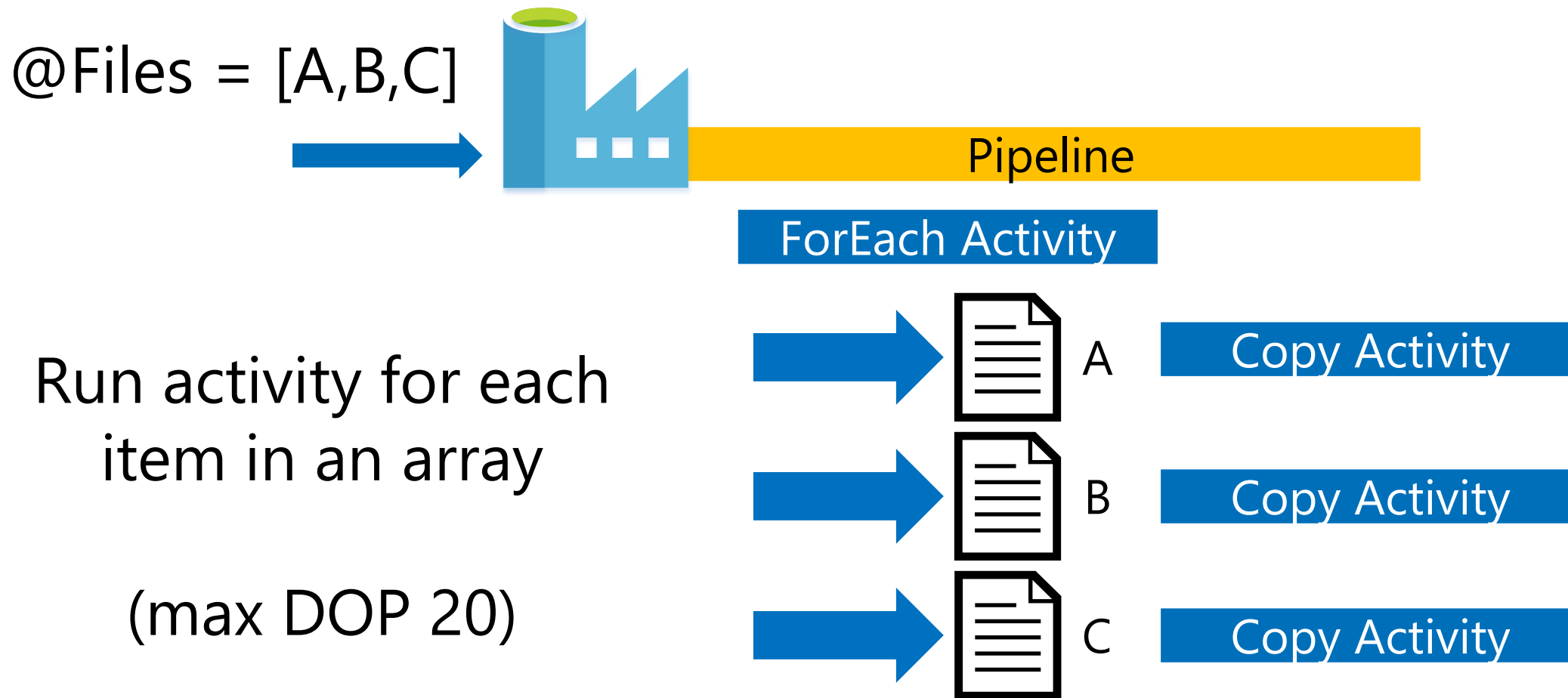
Dynamic Pipelines using Parameters & Expressions



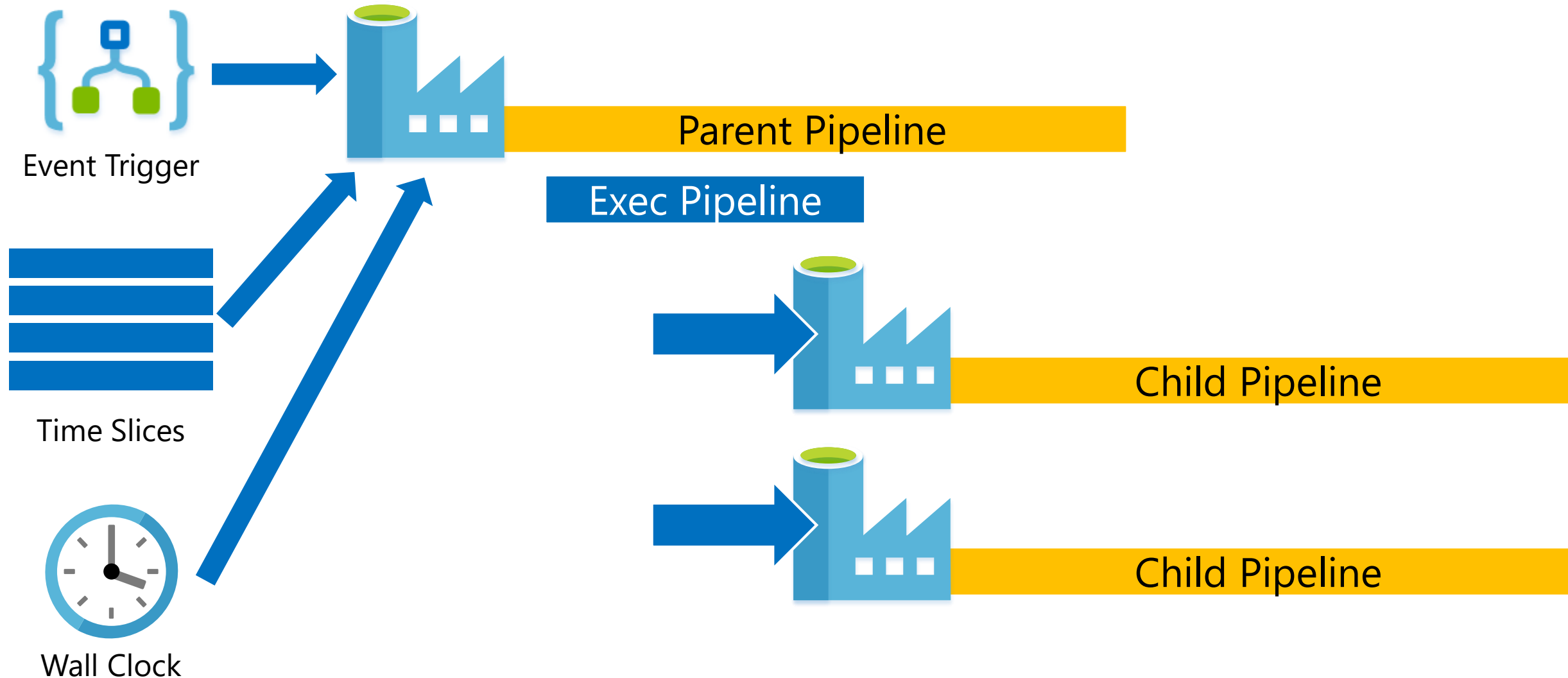
Dynamic Pipelines using Lookup Activity



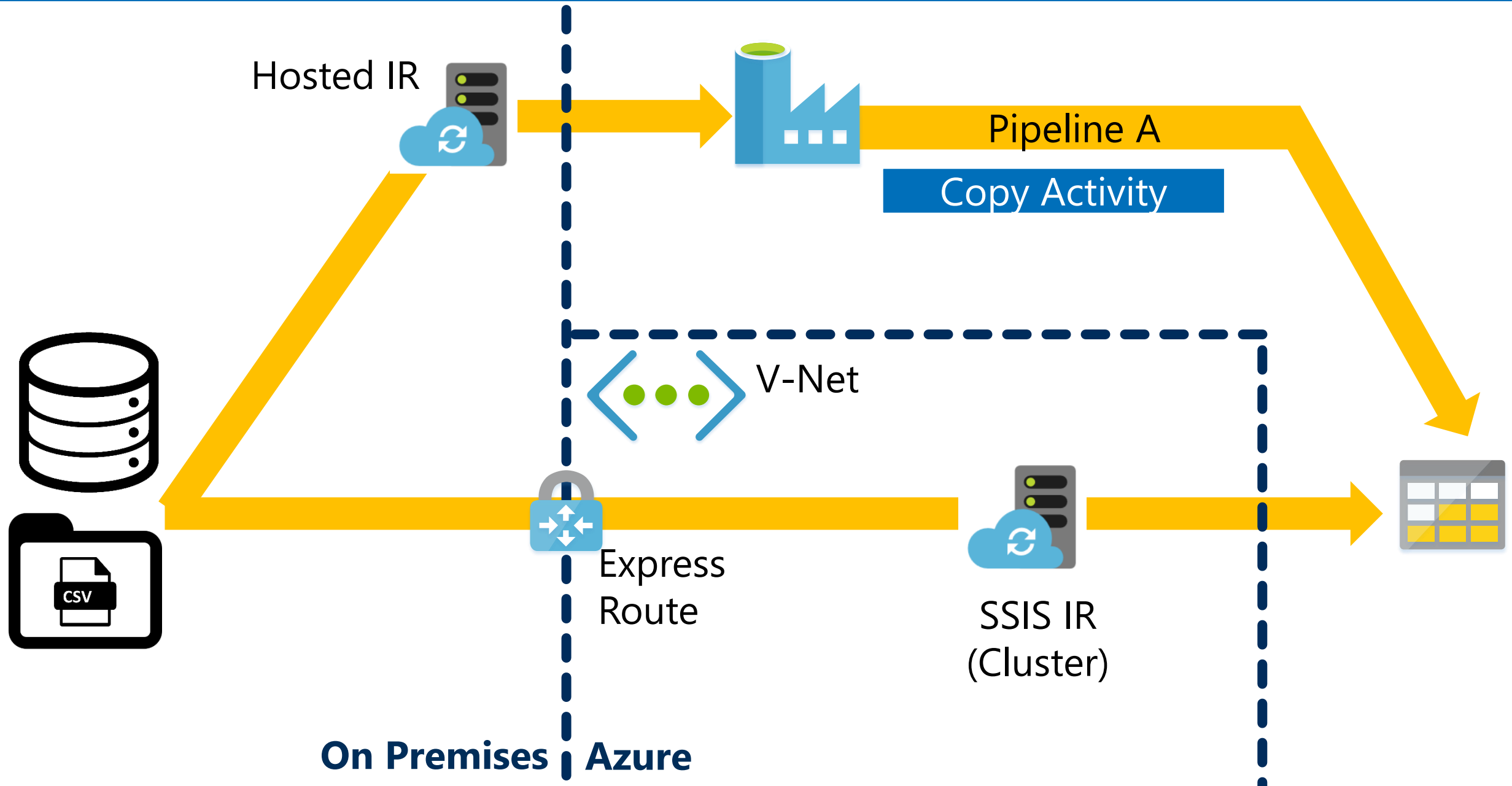
For Each Pipelines



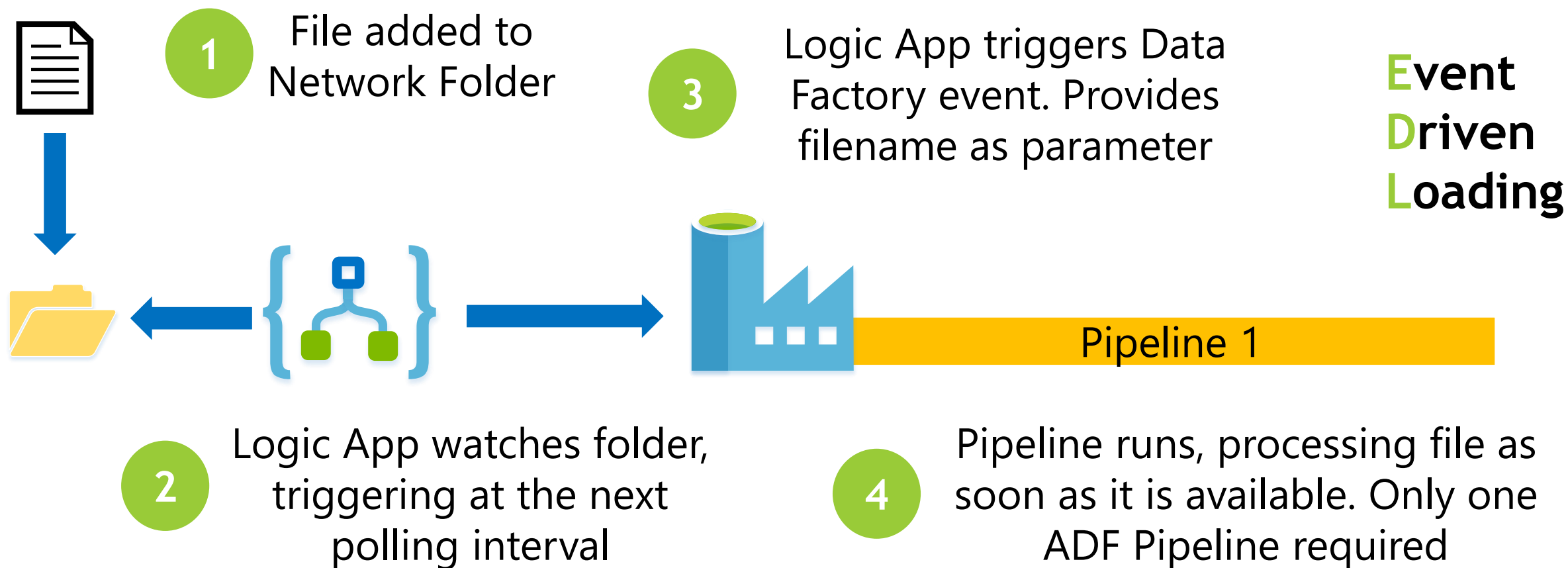
Parent/Child Pipelines & Triggering



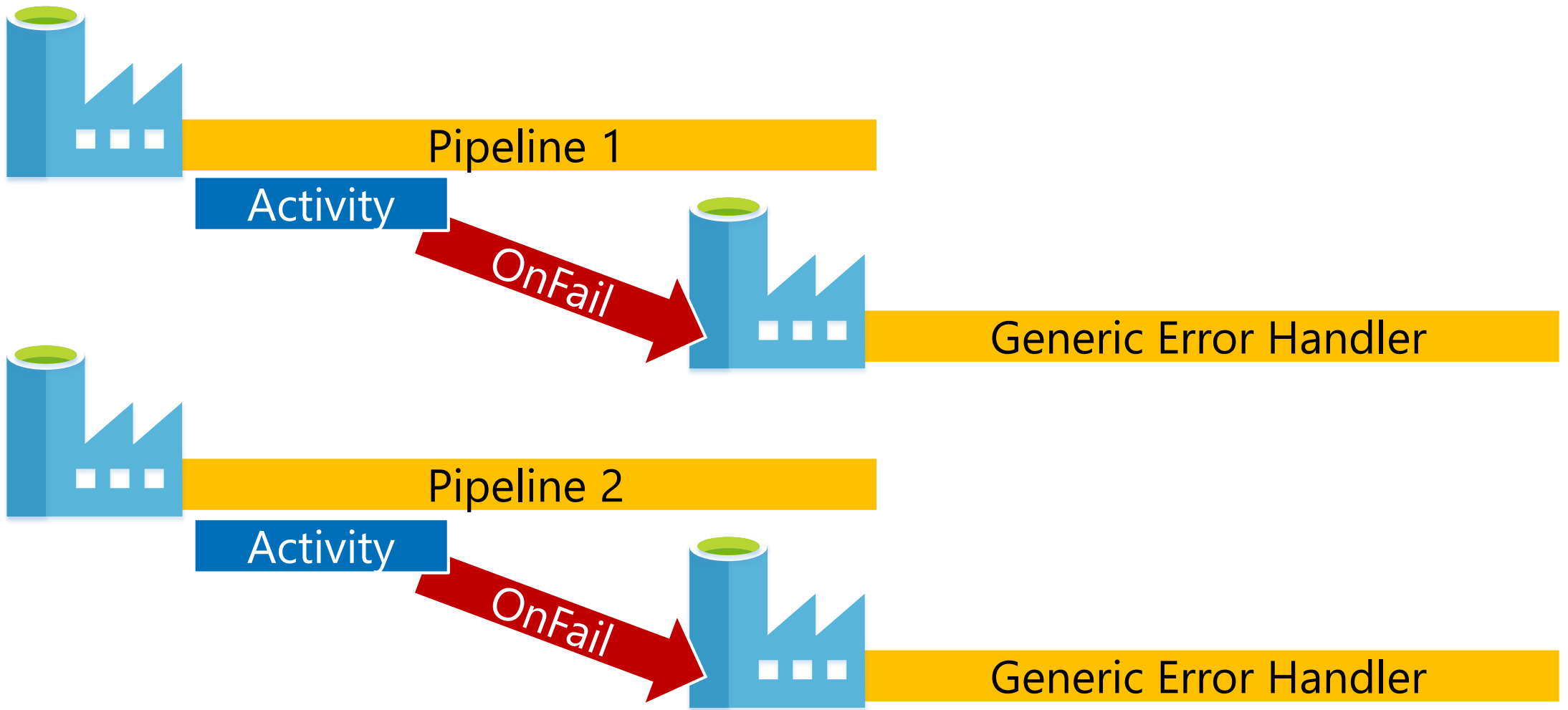
The SSIS IR with Azure V-Net Access



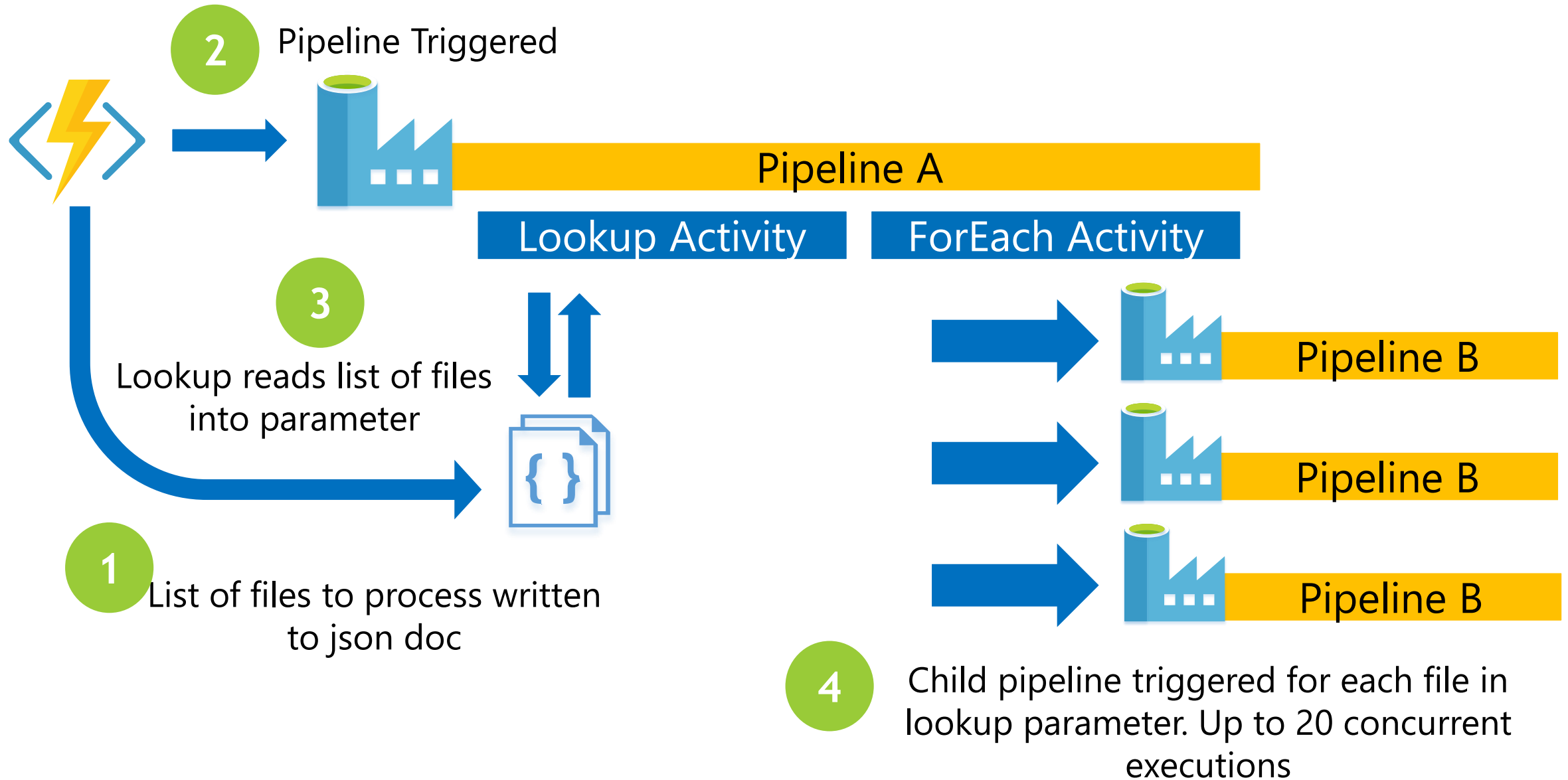
Event Driven Loading



Reusable Pipelines with Conditional Logic



Design Pattern Combinations



Is ADF the right tool for our data integration & orchestration in Azure?



Maybe, limited use.



Yes, definitely.

Agenda

Data Factory

Concepts

Components

Why use it?

Data Factory Extensibility

SSIS, Functions,
Custom Activities

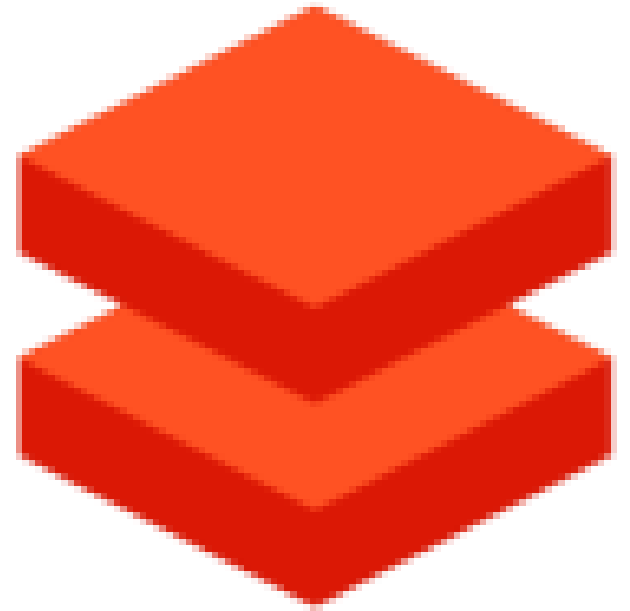
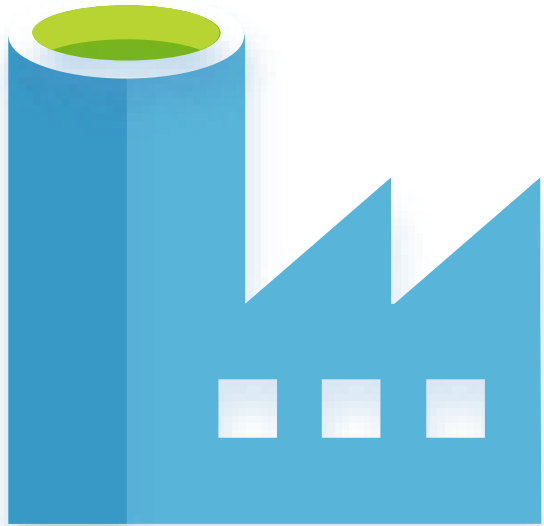
Conclusions

Design Patterns
ETL/ELT in Azure

Coming Soon!

Data Flows with
Data Bricks


Did you know...



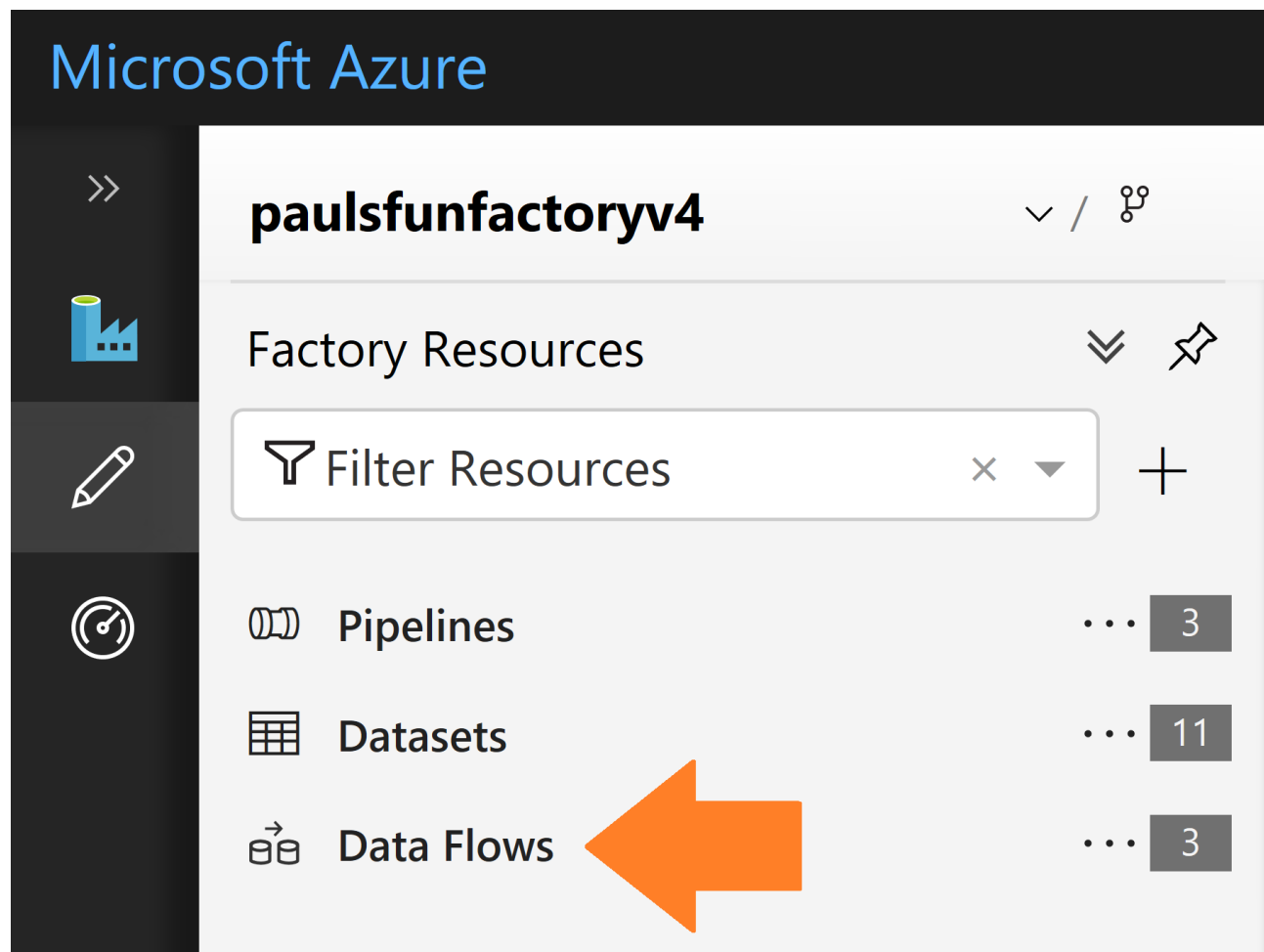
Azure Data Factory with Azure Data Bricks

Currently Data Bricks Activities



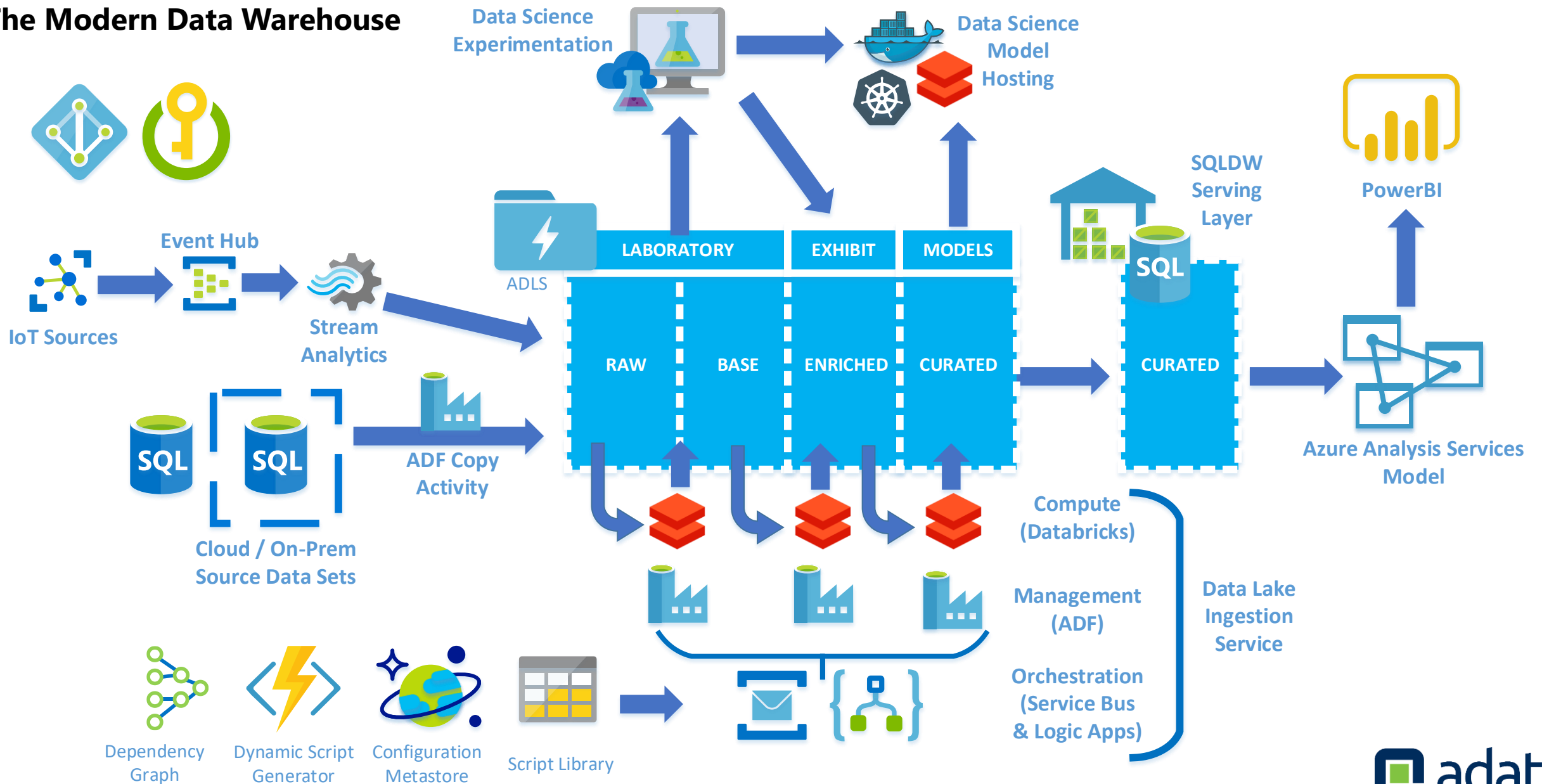
-  Notebook
-  Jar
-  Python

Coming Soon ADF Data Flow Engine



Why use Azure Data Factory?

The Modern Data Warehouse





Thanks for Listening

Simon Whiteley



@MrSiWhiteley



Blog: <http://mrpaulandrew.com>

Email: paul@mrpaulandrew.com