Cloud Orchestration

PaaS SSIS & ADFV2 Patterns

Simon Whiteley | Adatis

08/02/2018

Agenda

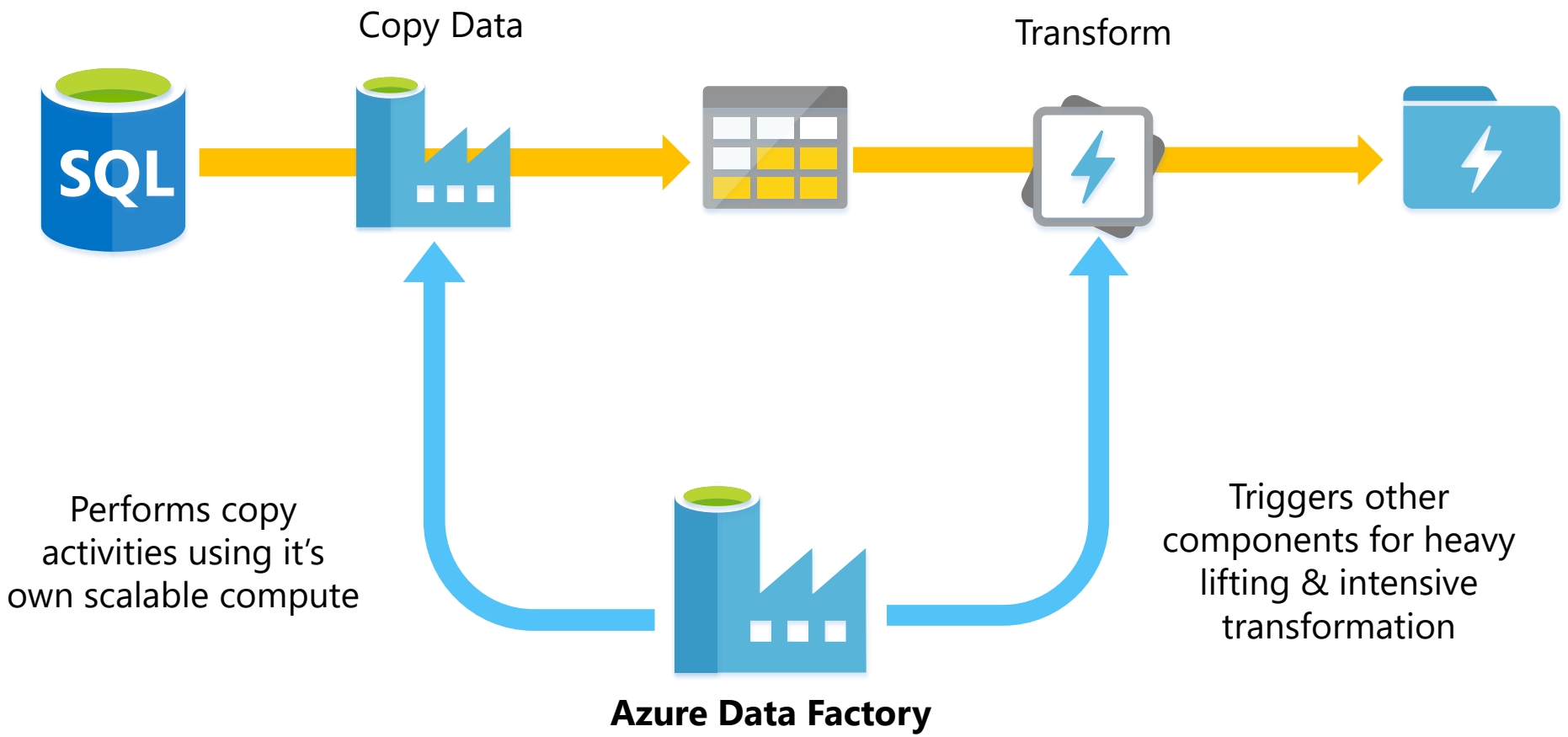
Data Factory Recap

ADFV2 Features Update

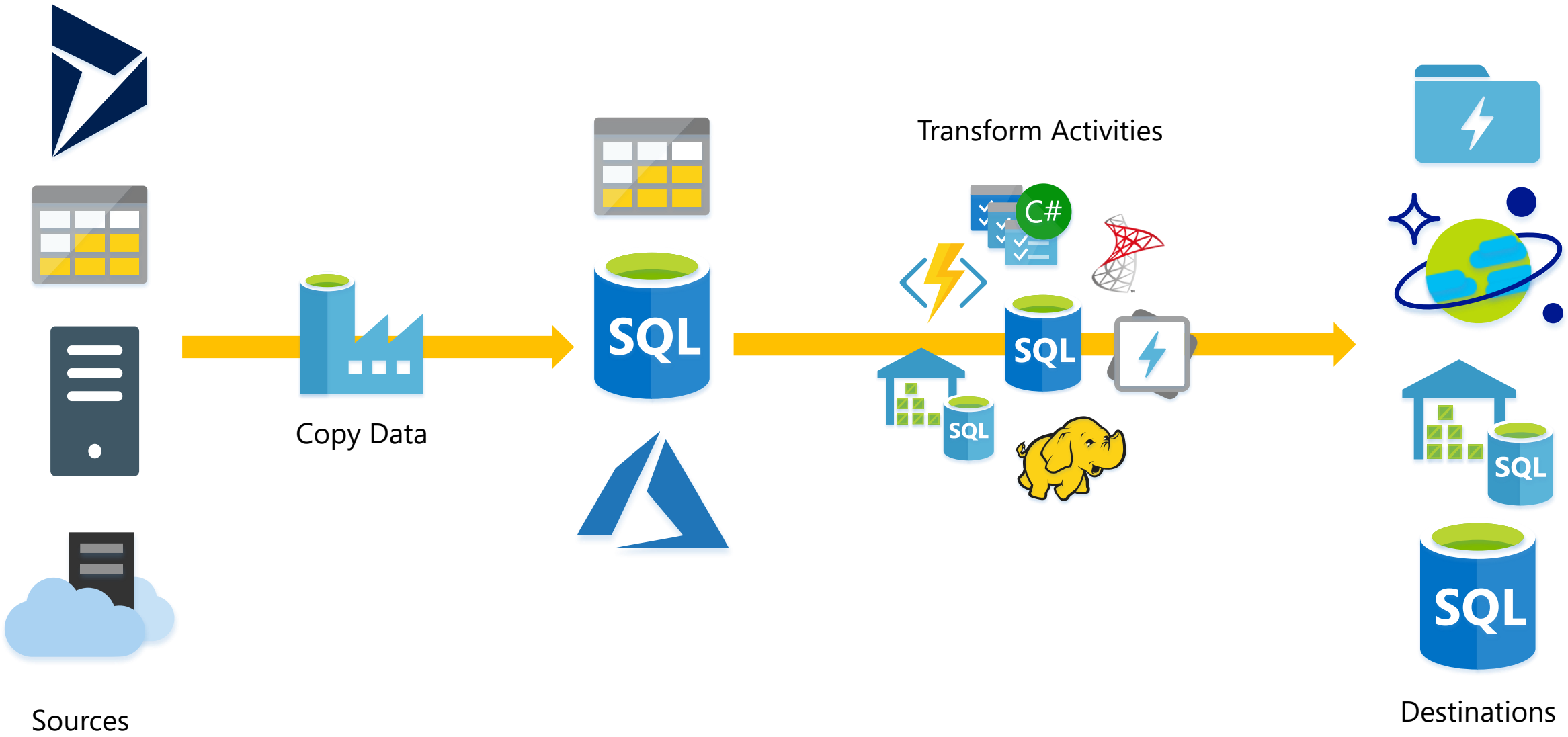
SSIS in Azure

Design Patterns & Conclusions

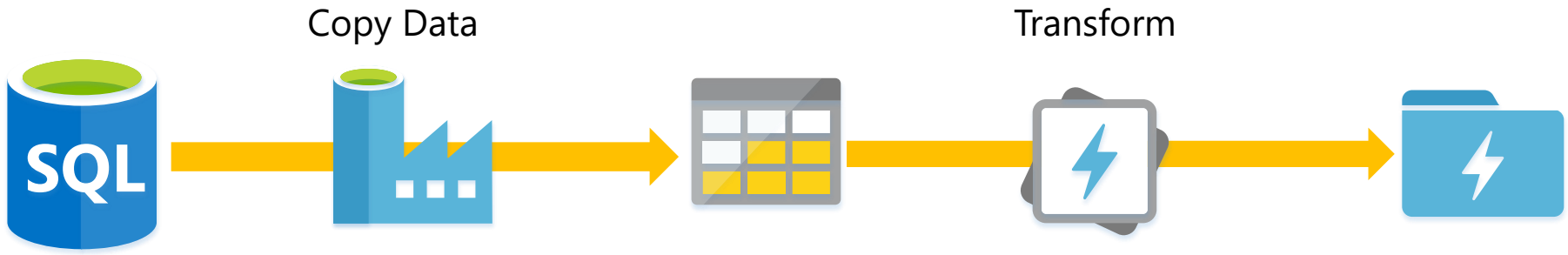
What is Azure Data Factory?



What is Azure Data Factory?



Data Factory Components



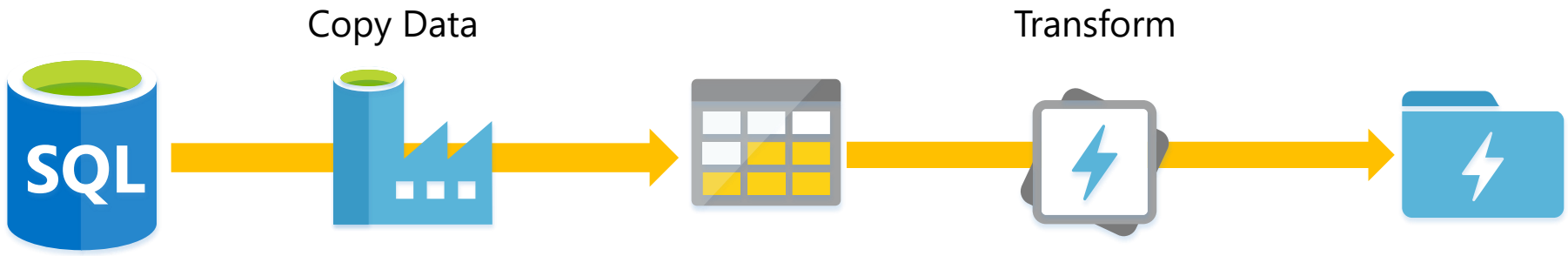
1 **Linked Services** – How do I connect?

Think SSIS Connection Managers!

SQLDBLinkedService

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

Data Factory Components

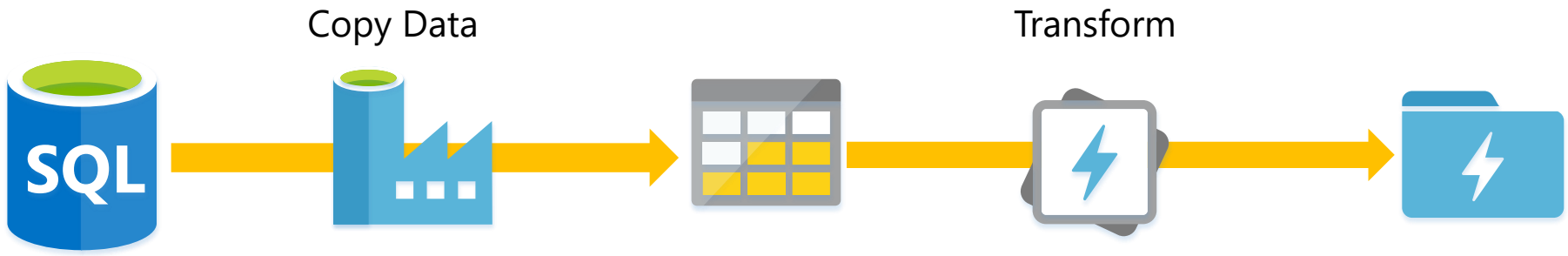


1. **Linked Services**
2. **Data Sets** – What slices/partitions does my data have?

dbo.DimCustomer

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

Data Factory Components

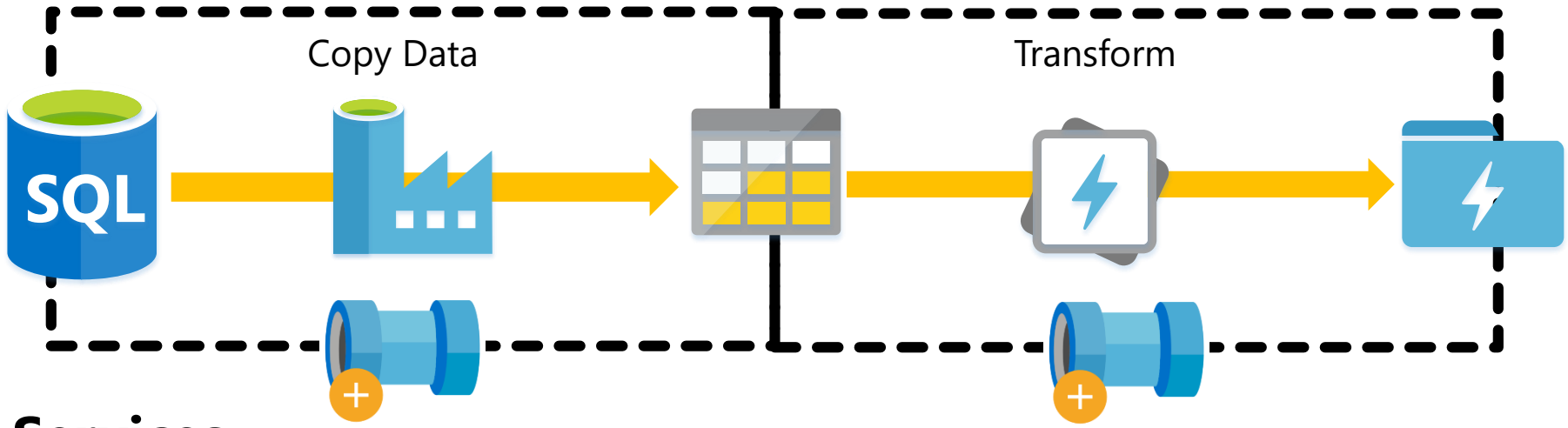


1. **Linked Services**
2. **Data Sets**
3. **Activities** – What actions am I taking?

U-SQL Activity

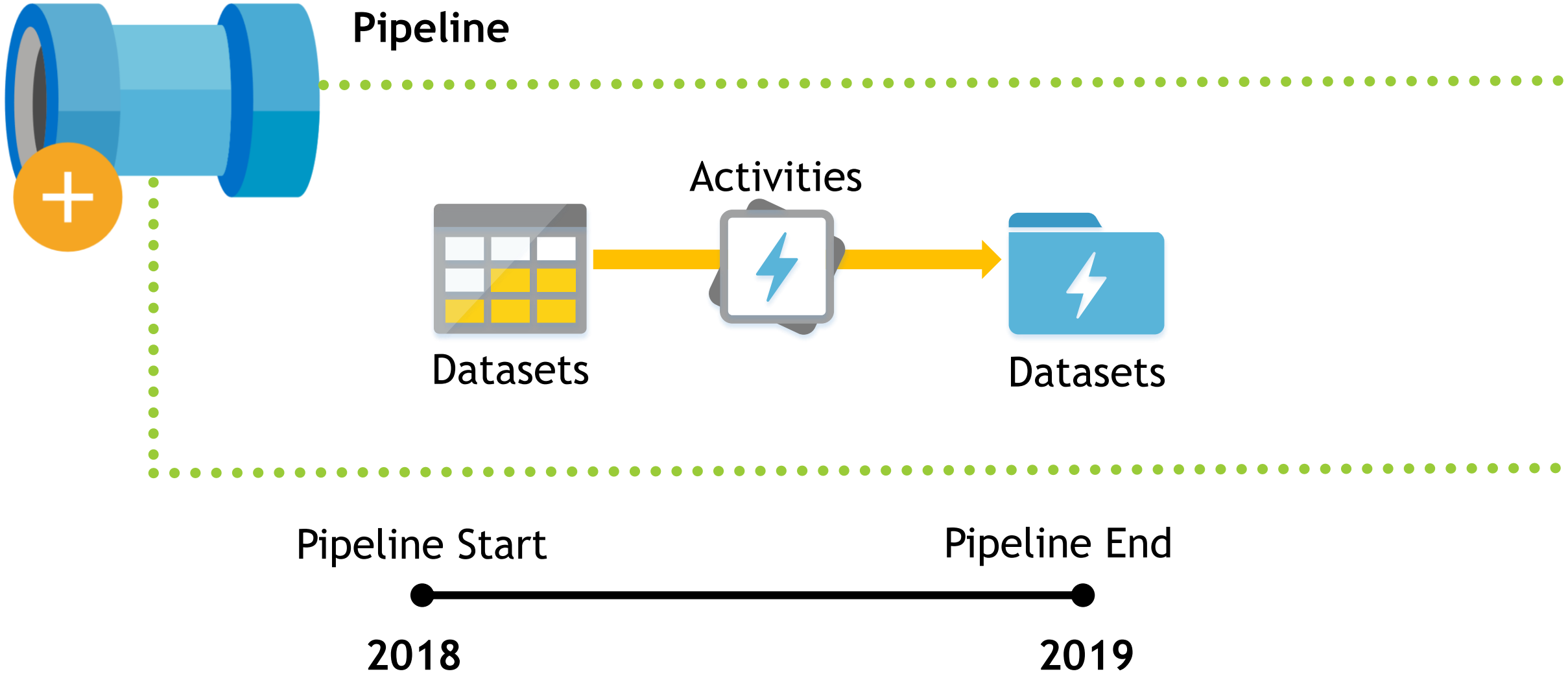
Script: *wasb//:myscripts/ProcessOrders.usql*
AUs: *5 units*
Priority: *1000*
Parameters: *@Output = "RAW/Orders/..."*

Data Factory Components

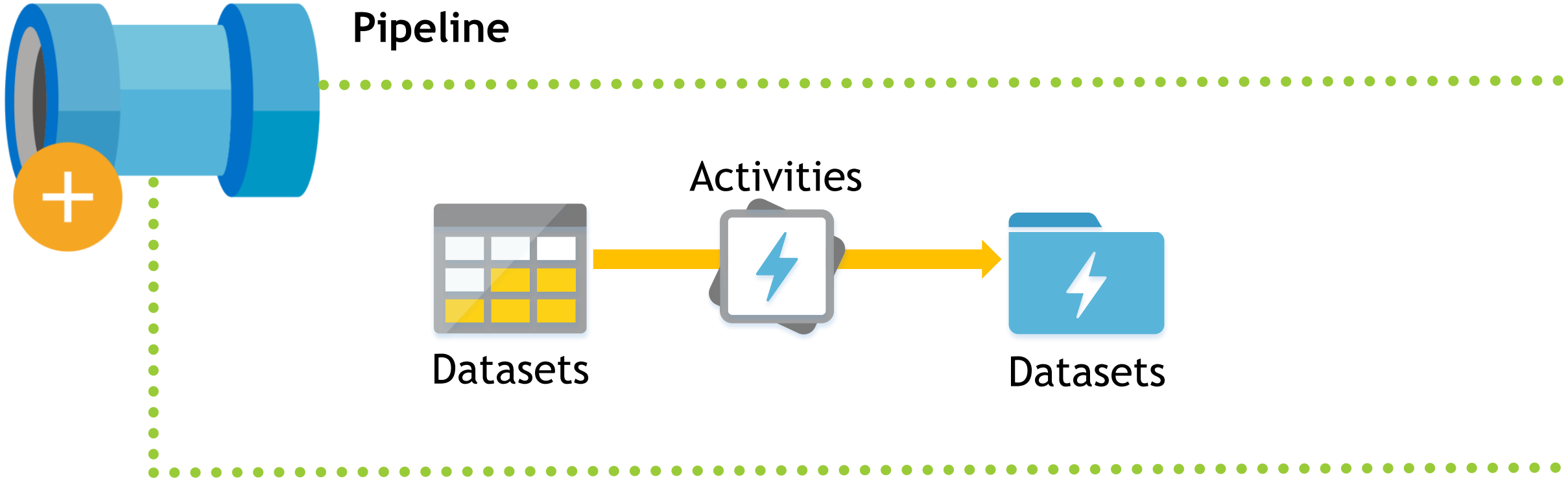


1. **Linked Services**
2. **Data Sets**
3. **Activities**
4. **Pipelines** – What groups of work do I want to do?

Azure Data Factory Concepts Continued

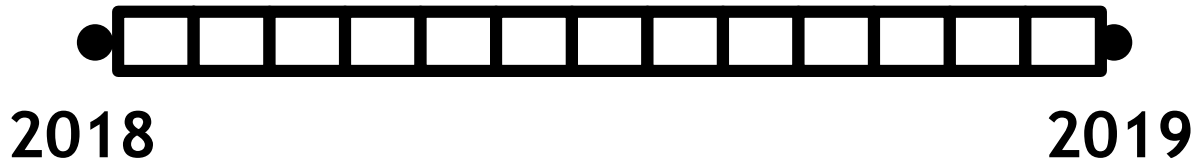


Azure Data Factory Concepts Continued

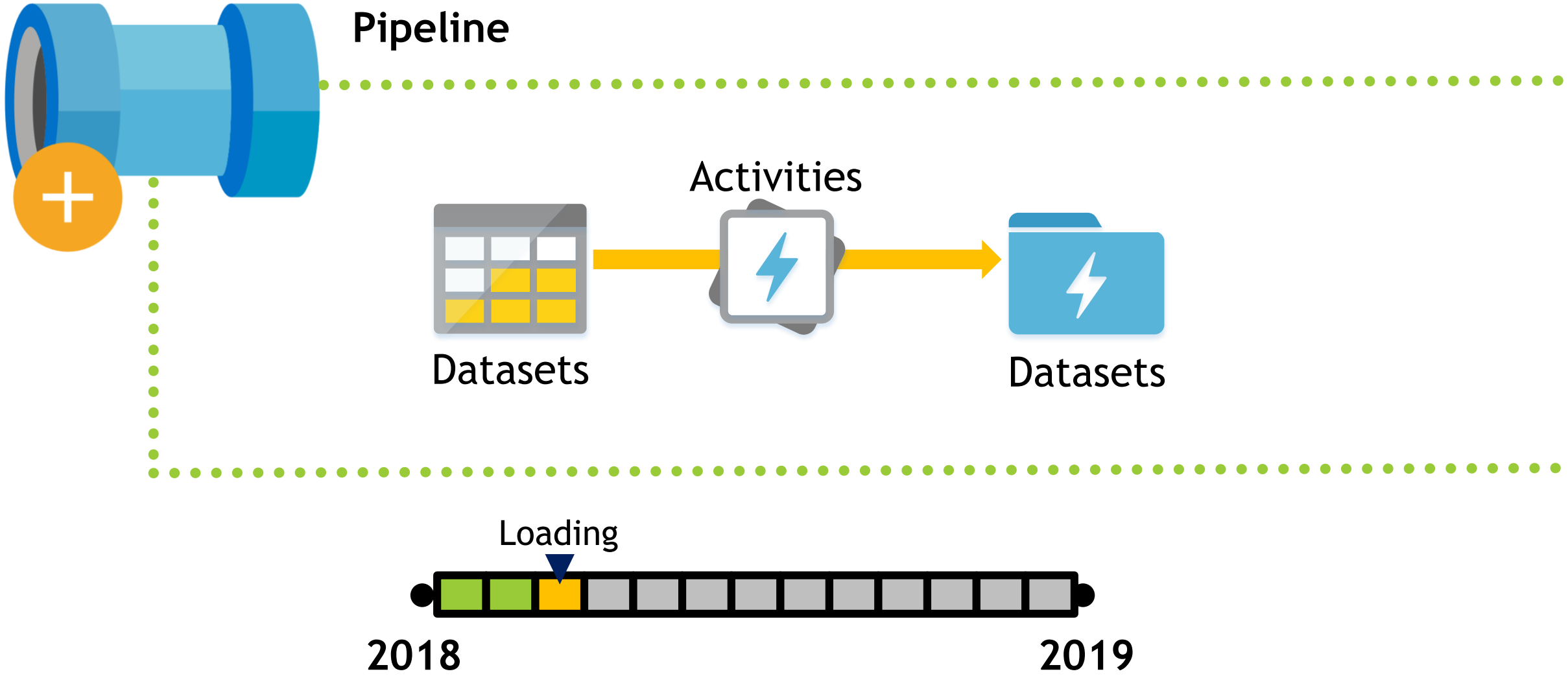


Interval: *Month*

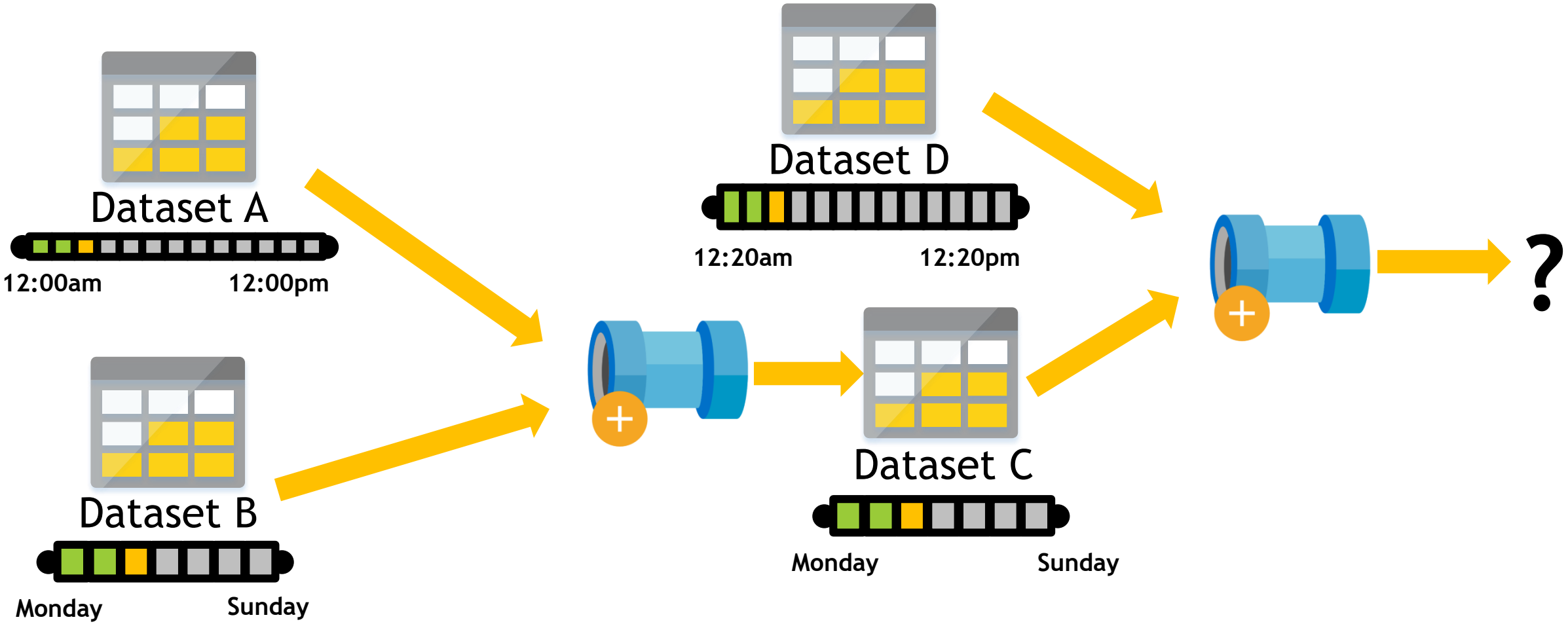
Frequency: *1*



Azure Data Factory Concepts Continued



Time Slice Problems…



Data Factory Issues & Limitations

- **Time Slices** – Complex, difficult to change & provision
- **Pricing** – High and low frequency
- **Control Flow** – Not conditional. Pass or fail
- **Developer Tools** – JSON templates
- **HardCoded Pipelines** – No dynamic values
- **C# Extensibility** – Often required for anything complex
- **Monitoring** – Times slice bound, focused on datasets
- **Connectivity** – Limited to Microsoft supported linked services

Agenda

Data Factory Recap

ADFV2 Features Update

SSIS in Azure

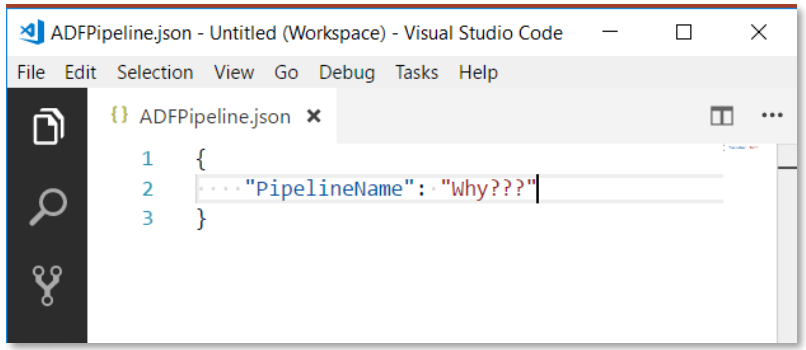
Design Patterns & Conclusions

So What's Changed?

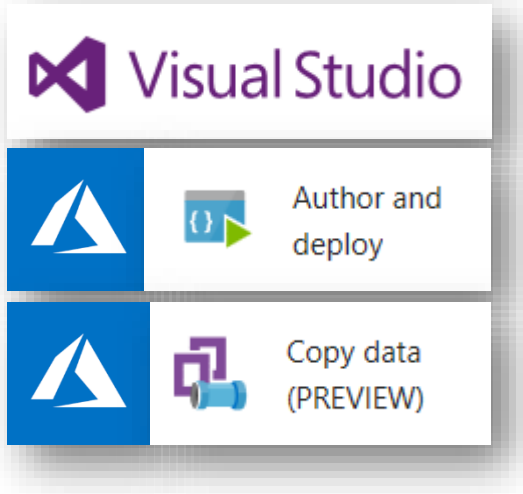
V1		V2
✓	Data Movement (Copy)	✓
✓	Activities	✓
✓	Pipelines	✓
✓	Datasets	✓
✓	Time Slices/Tumbling Windows	✓
✘	Event Triggers	✓
✘	Recurrent Schedules	✓
✘	Parameters	✓
✘	Expressions	✓
✘	Conditional Logic	✓
✘	Use of SSIS Packages	✓

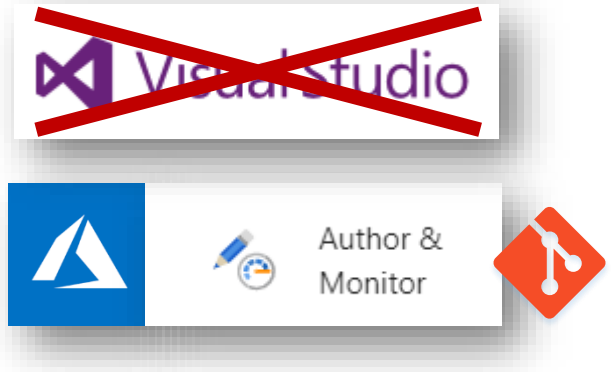
Development Environments

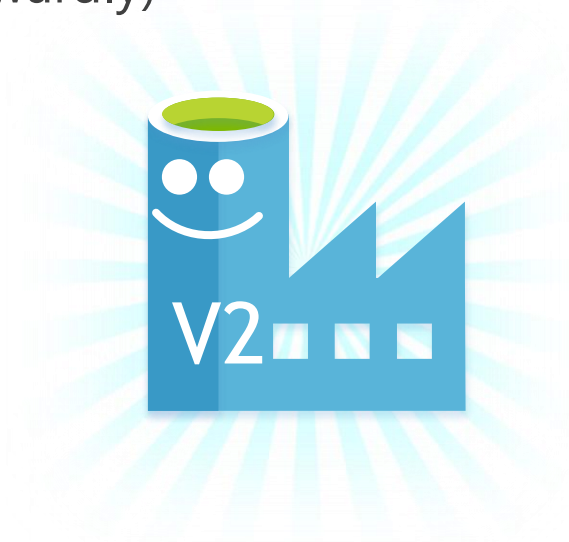
- JSON Templates
- Data Factory Wizard
- Reverse Engineer From Azure
- Deployment Wizard

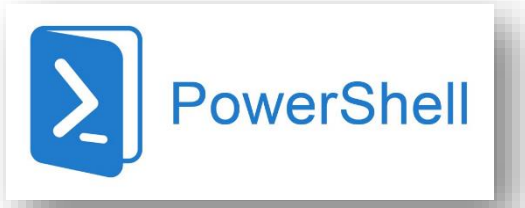


- No JSON Templates
- No Deployment Wizard
- Data Factory Wizard
- Reverse Engineer From Azure (awkwardly)







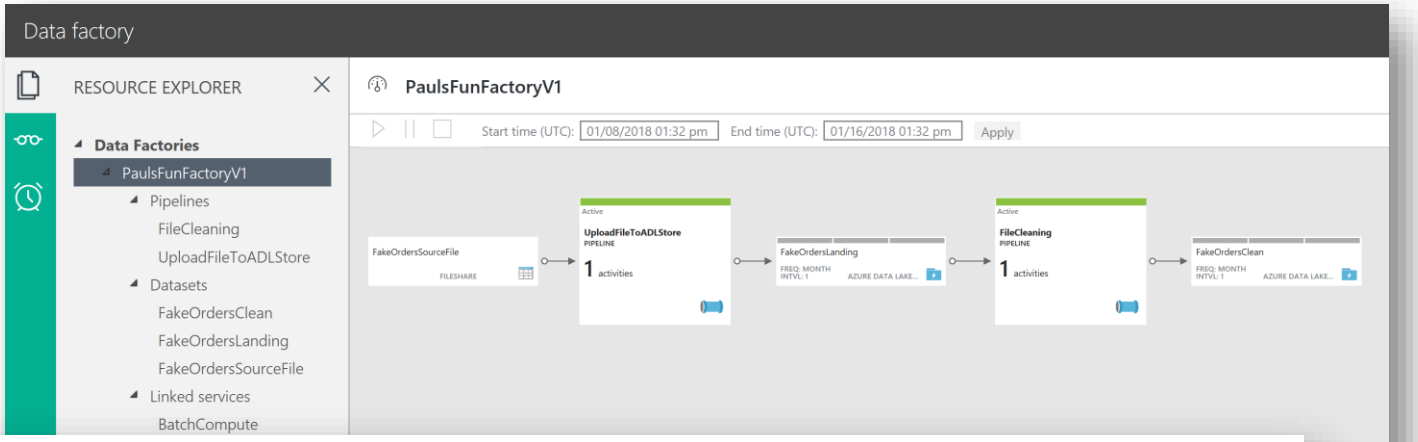


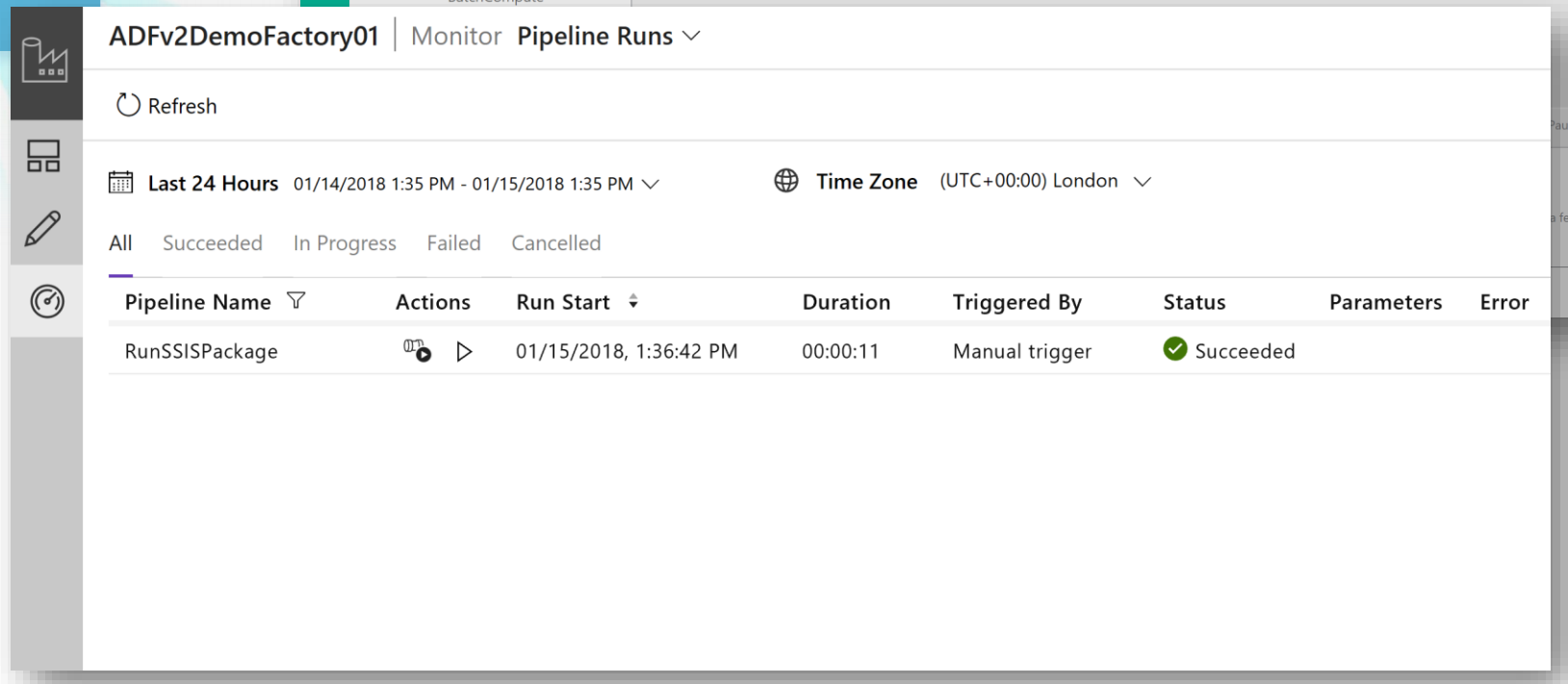
http://bit.ly/2tsyD90

Monitoring

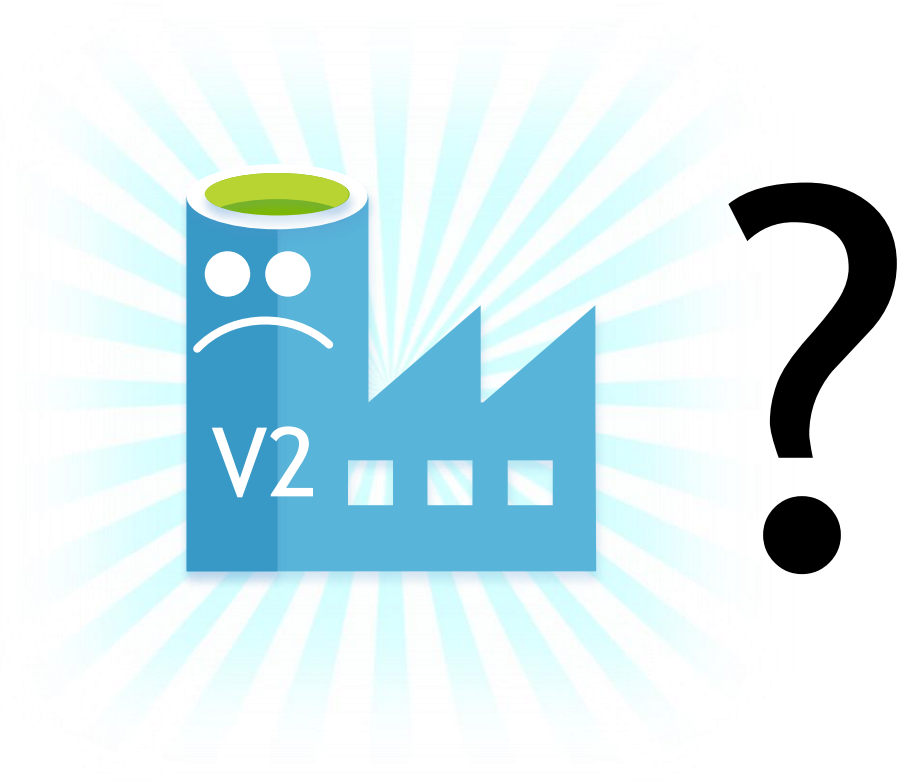
V2

V1

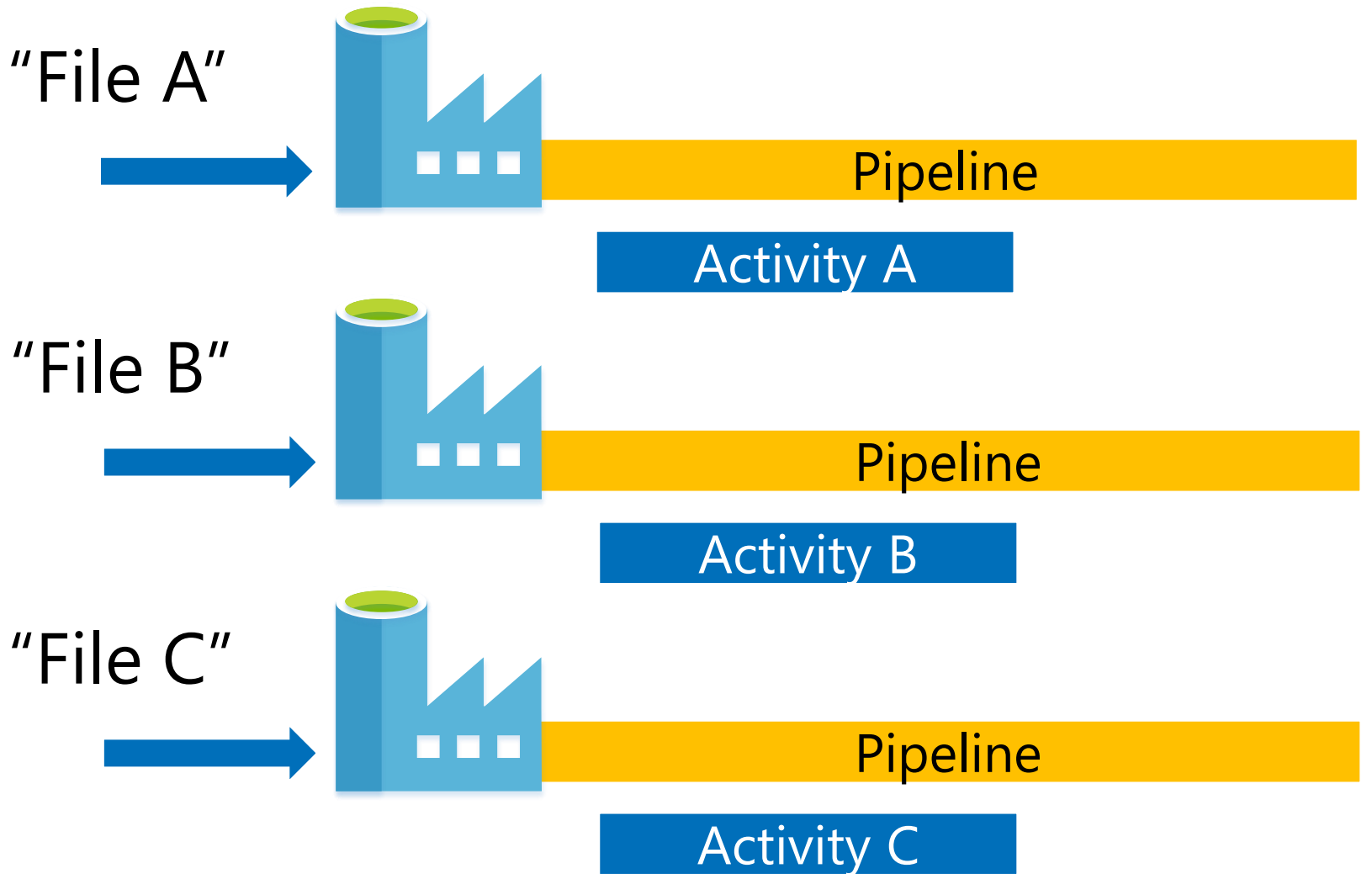

Data factory
RESOURCE EXPLORER
Data Factories
PaulsFunFactoryV1
Pipelines
FileCleaning
UploadFileToADLStore
Datasets
FakeOrdersClean
FakeOrdersLanding
FakeOrdersSourceFile
Linked services
BatchCompute
PaulsFunFactoryV1
Start time (UTC): 01/08/2018 01:32 pm
End time (UTC): 01/16/2018 01:32 pm
Apply
FakeOrdersSourceFile
UploadFileToADLStore
1 activities
FakeOrdersLanding
FileCleaning
1 activities
FakeOrdersClean


ADFv2DemoFactory01 | Monitor Pipeline Runs
Refresh
Last 24 Hours 01/14/2018 1:35 PM - 01/15/2018 1:35 PM
Time Zone (UTC+00:00) London
All
Succeeded
In Progress
Failed
Cancelled
Pipeline Name
Actions
Run Start
Duration
Triggered By
Status
Parameters
Error
RunSSISPackage
01/15/2018, 1:36:42 PM
00:00:11
Manual trigger
Succeeded

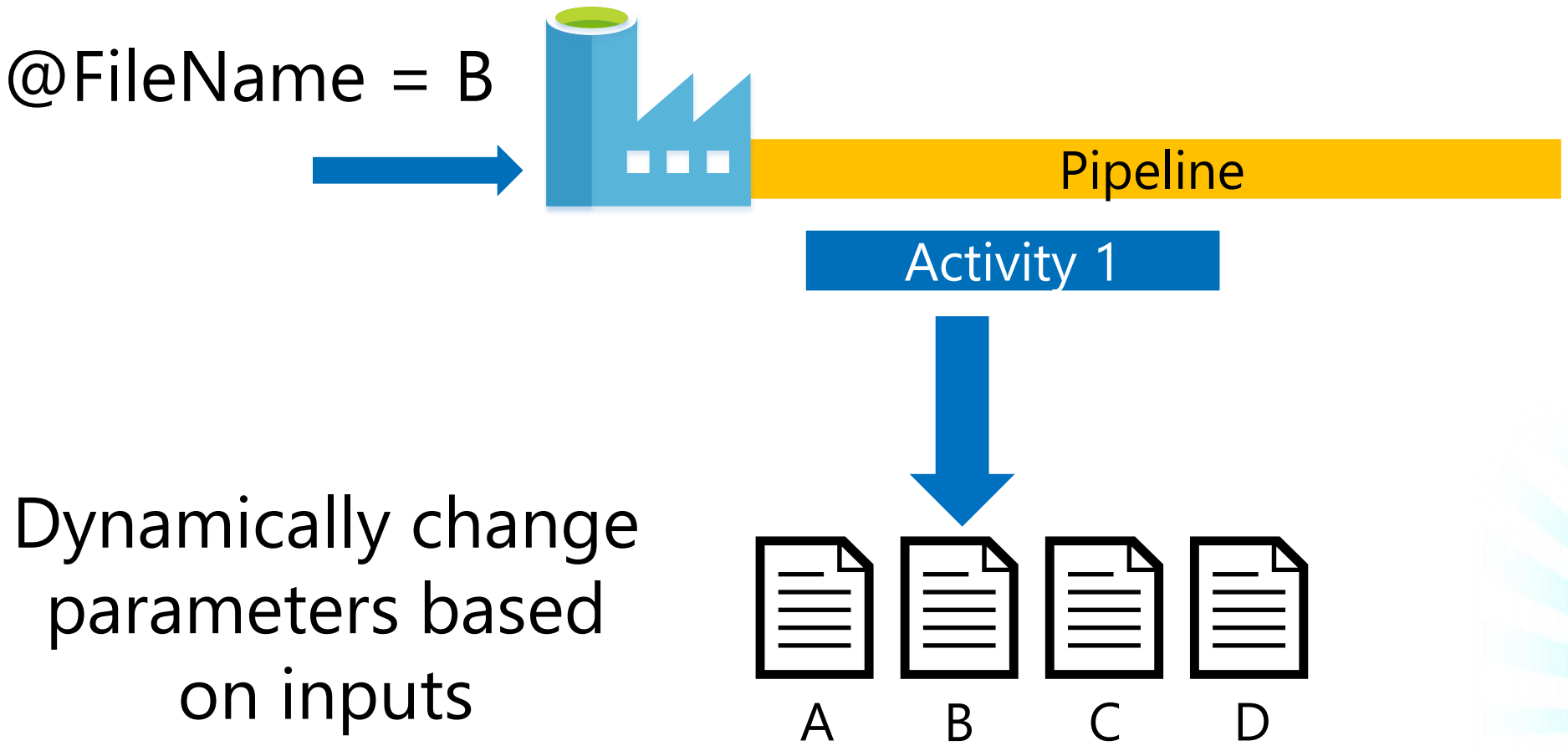
So what IS good?



Hardcoded Pipelines

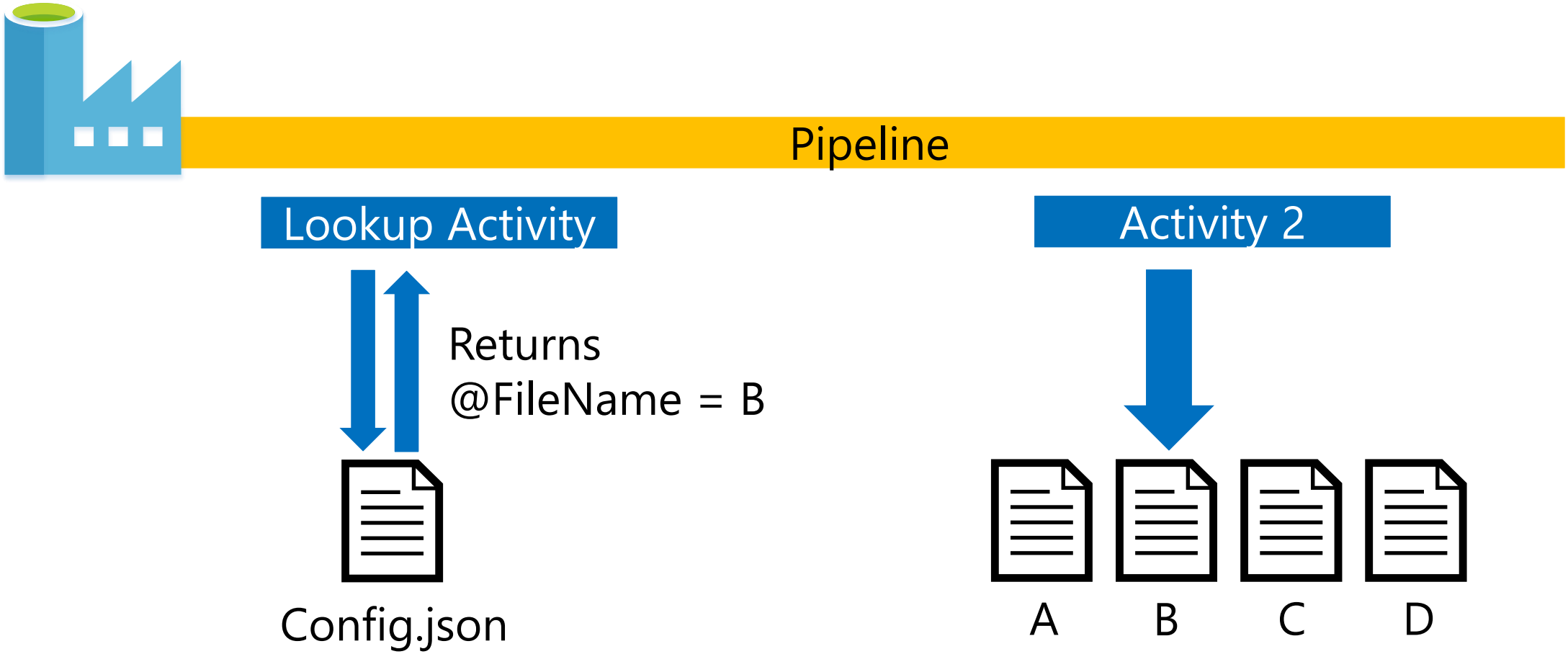


Dynamic Pipelines using Parameters



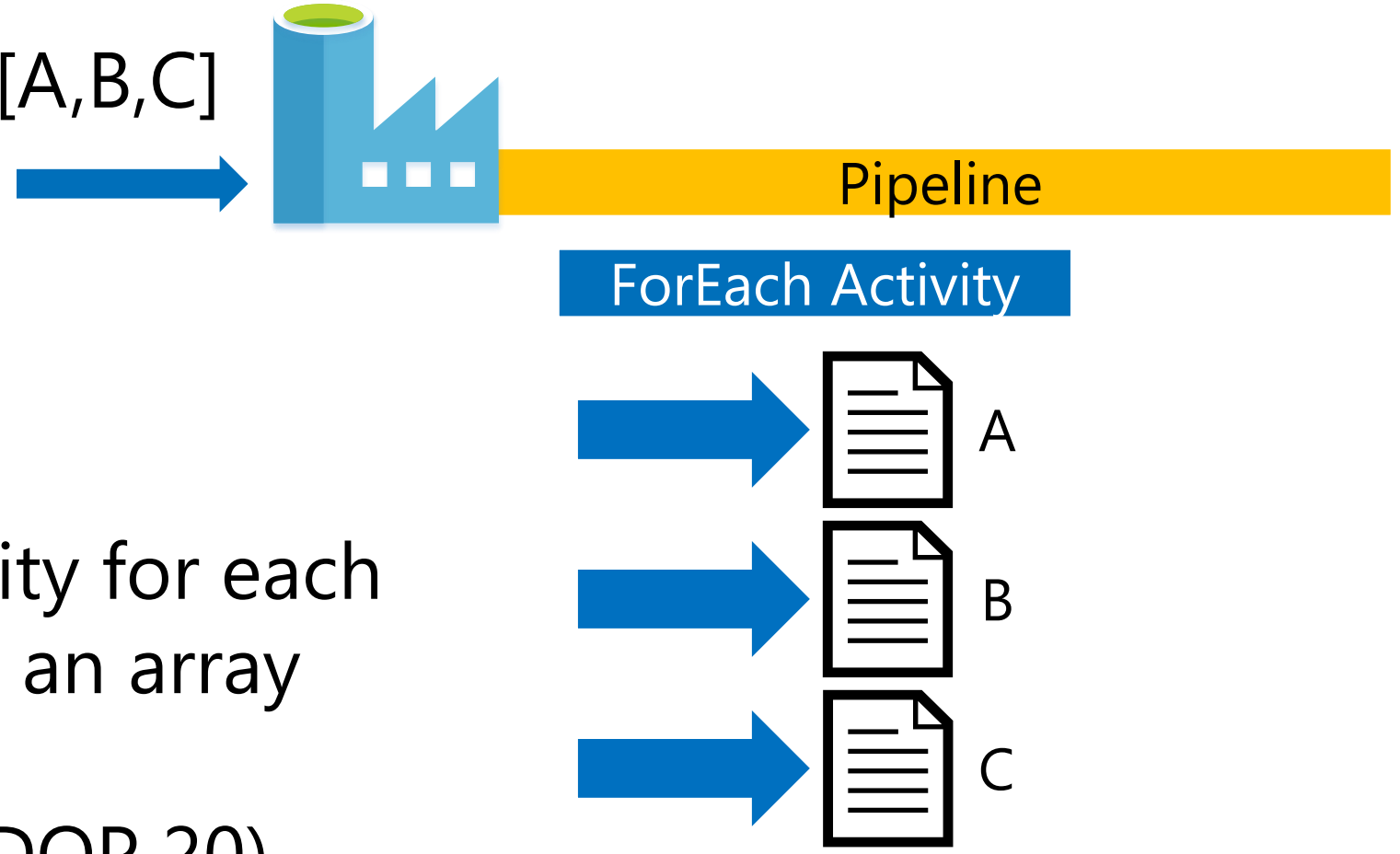
Dynamically change parameters based on inputs

Dynamic Pipelines using Lookup Activity



ForEach Pipelines

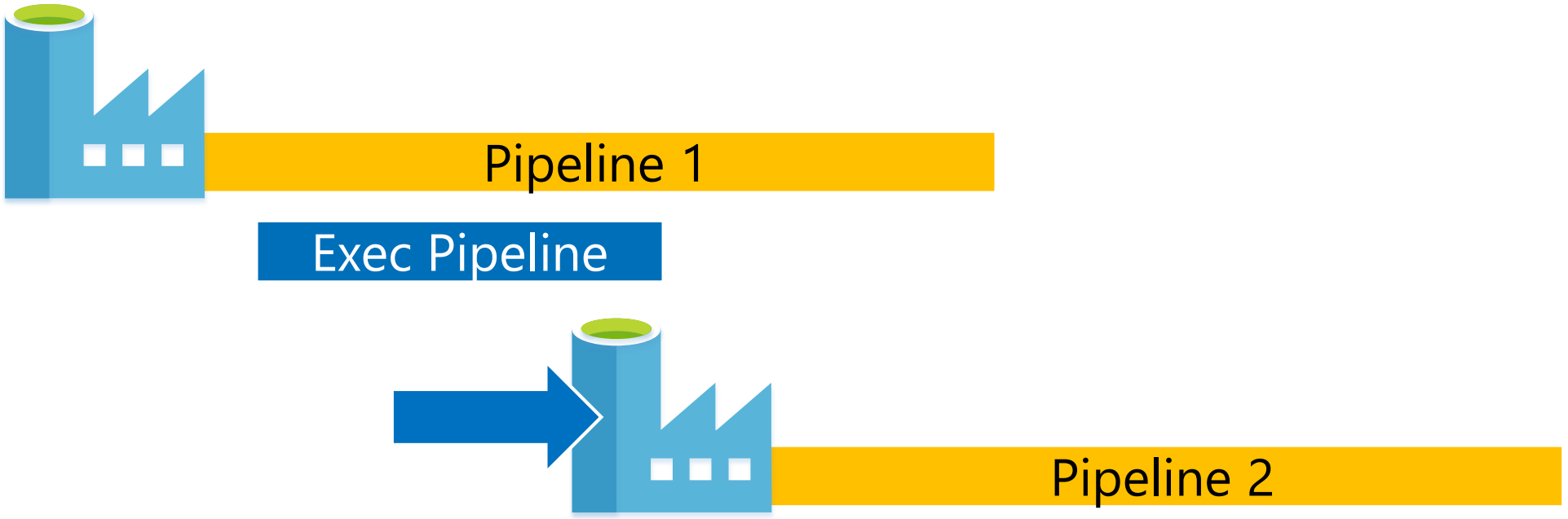
@Files = [A,B,C]



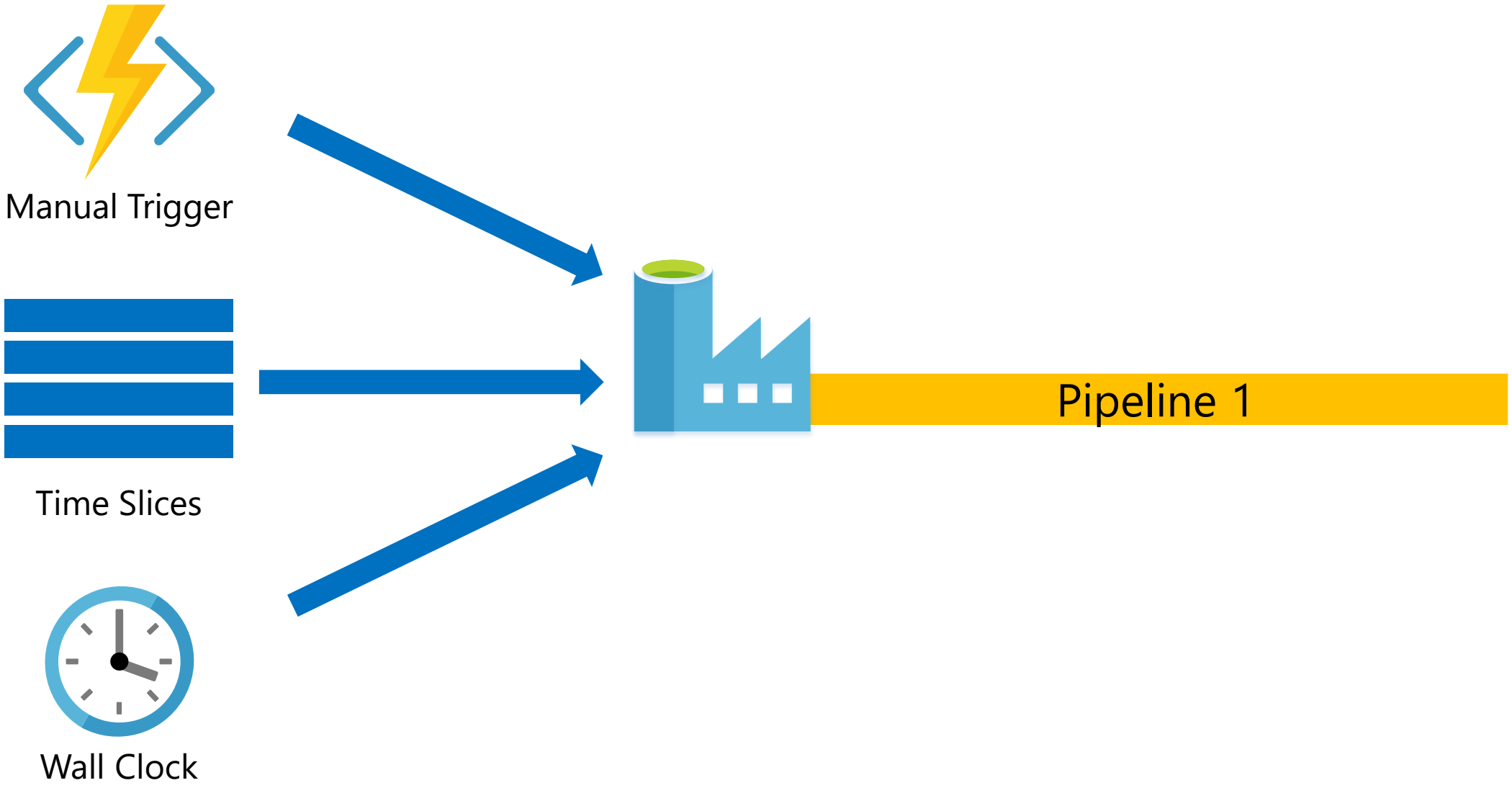
Run activity for each item in an array

(max DOP 20)

Execute Pipeline Activity



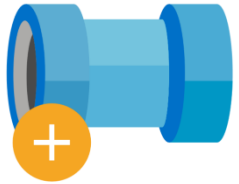
Pipeline Triggers



Integration Runtimes

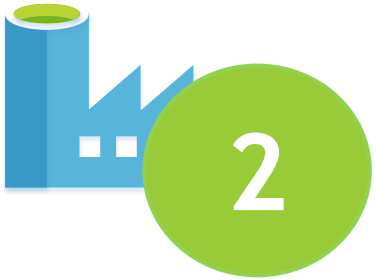
Default Integration Runtime

Movement Hours



Activity Orchestration

Flexible Region



SSIS Integration Runtime

SSIS Package Execution

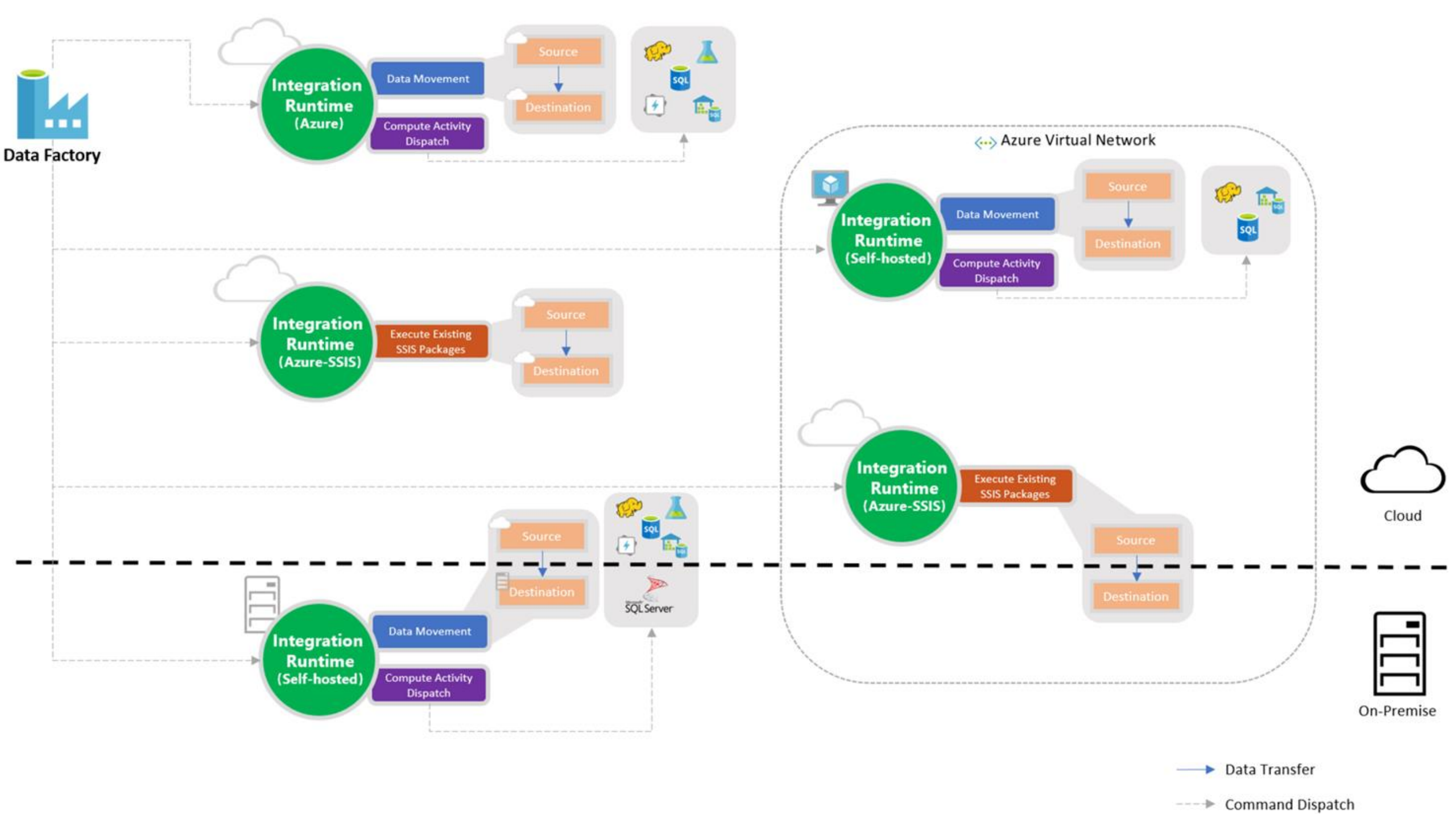
Specified Region

Self Hosted Integration Runtime

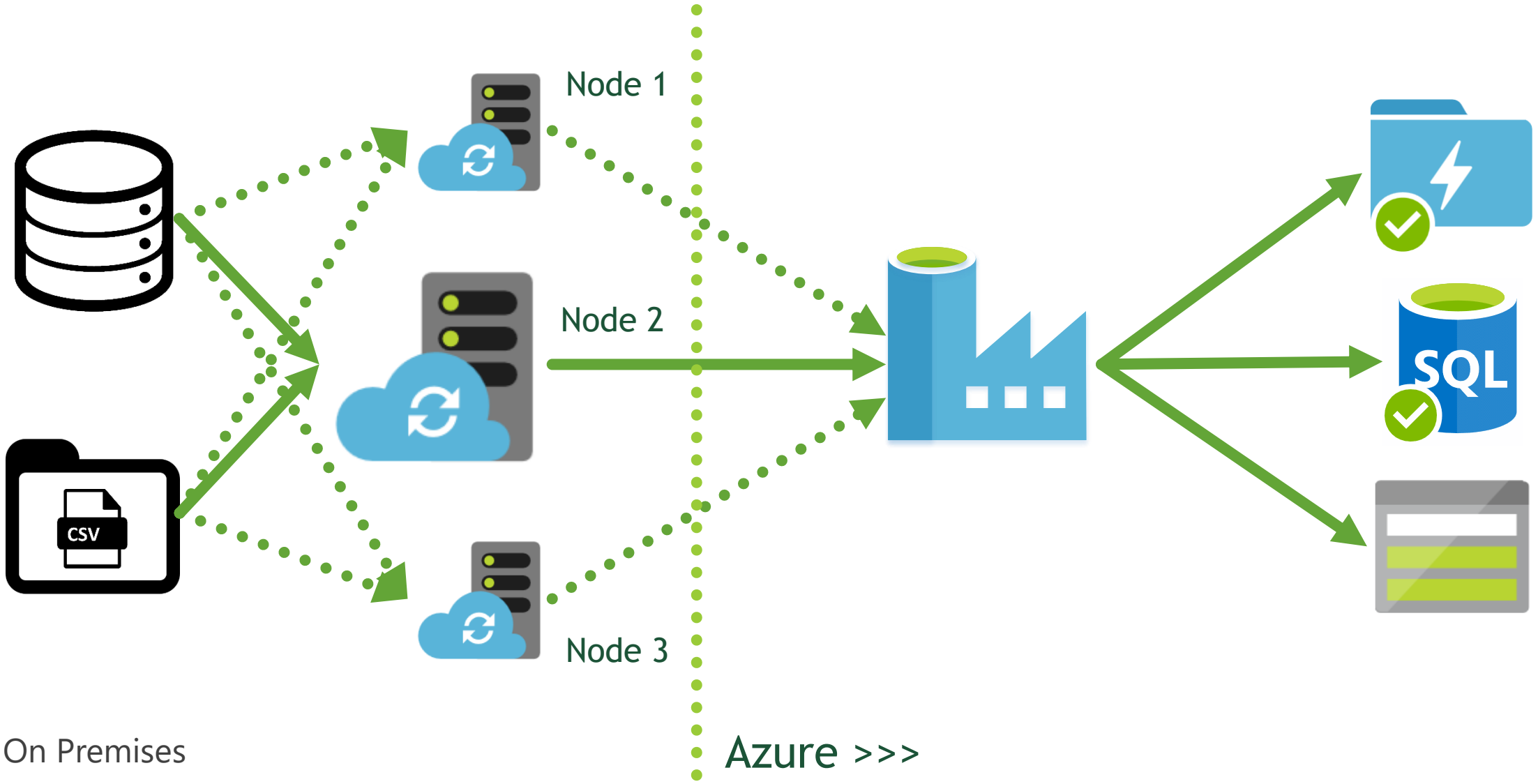
Local Compute

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)*



Demo: ADFV2

Agenda

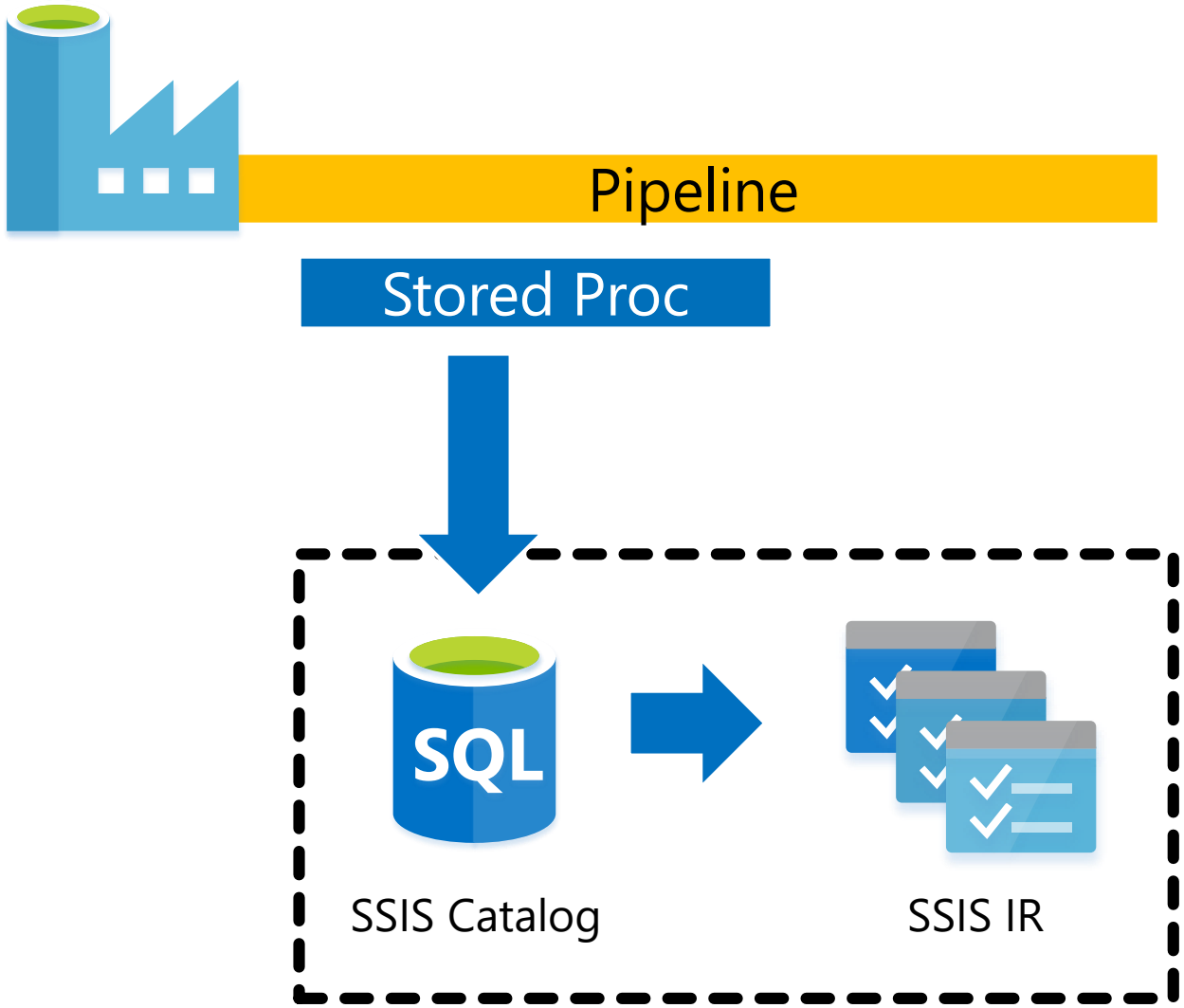
Data Factory Recap

ADFV2 Features Update

SSIS in Azure

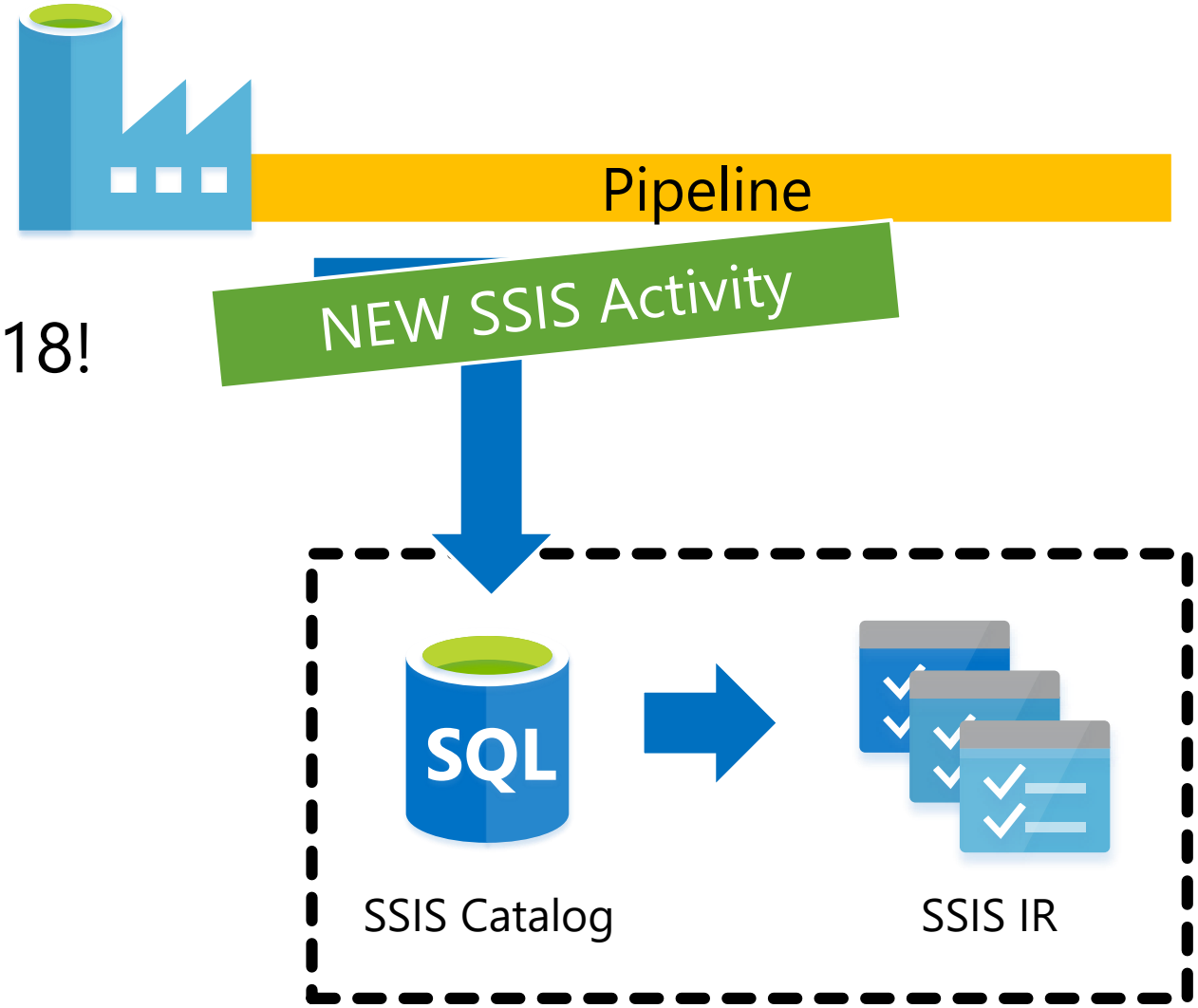
Design Patterns & Conclusions

SSIS Integration Runtimes

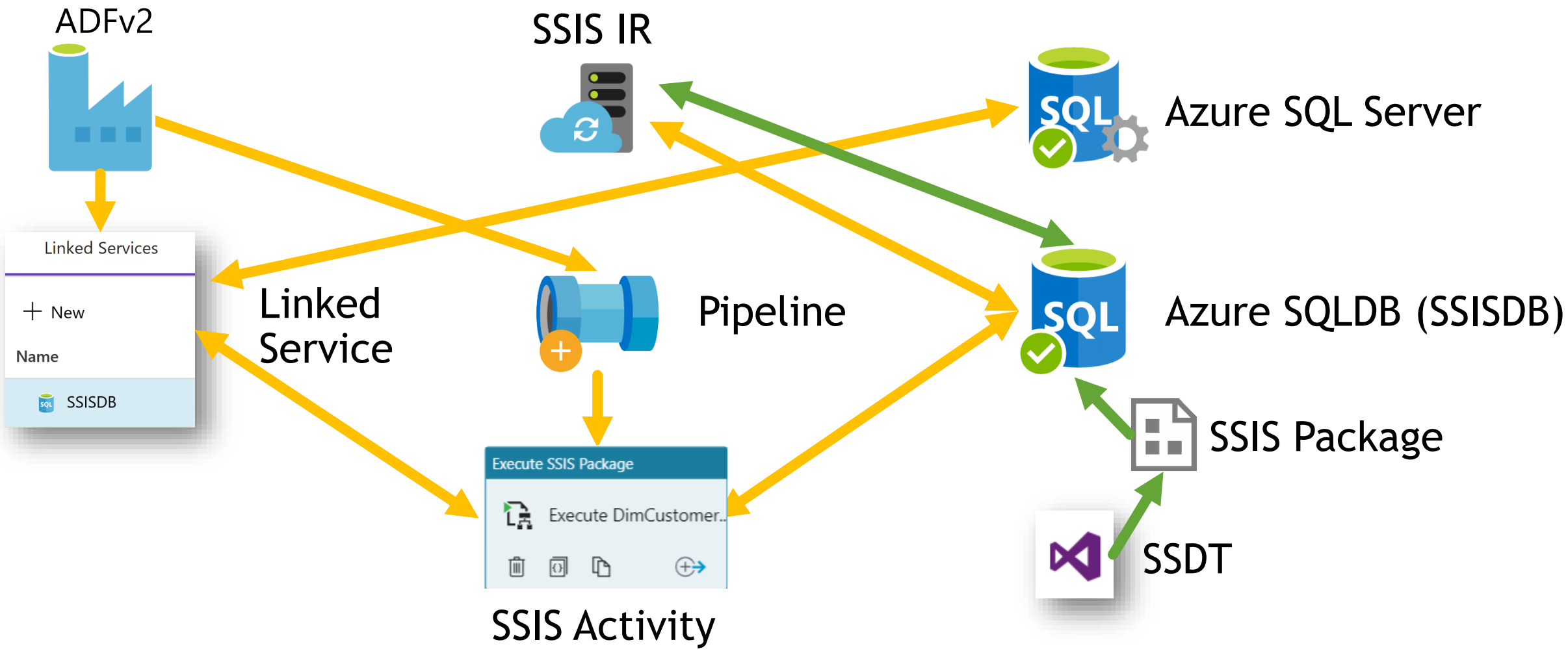


SSIS Integration Runtimes

Released 18th April 2018!



Configuring SSIS in Azure



Demo: SSIS

Agenda

Data Factory Recap

ADFV2 Features Update

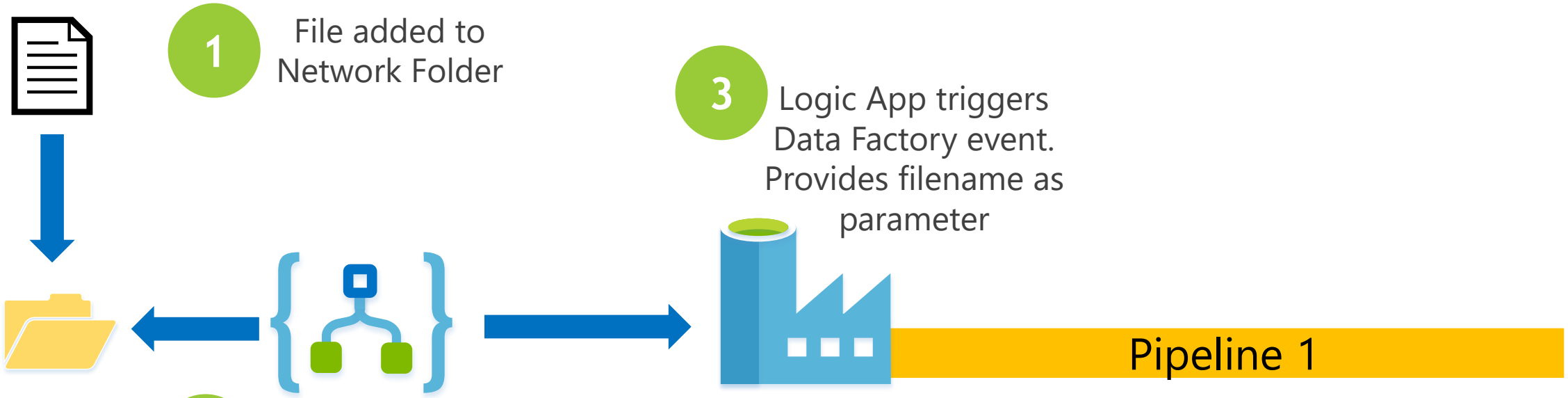
SSIS in Azure

Design Patterns & Conclusions

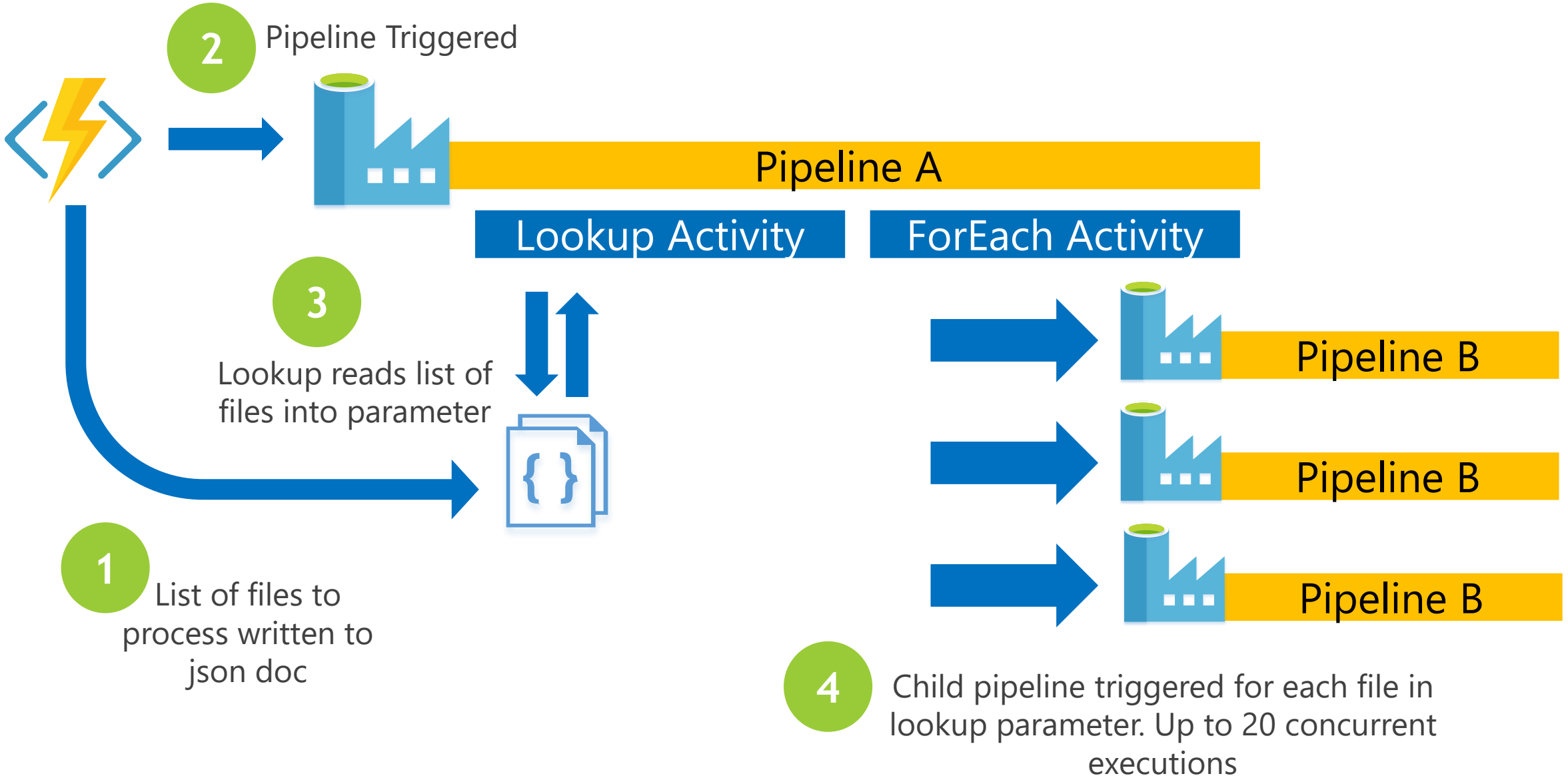
Back to our Problems…

- **Time Slices** – **Scheduled triggers**
- **Pricing** – **Pool of executions. Cheaper the more you use**
- **Control Flow** – **Conditional activities and iterative behaviour**
- **Developer Tools** – **Full browser based UI that creates the JSON**
- **HardCoded Pipelines** – **Expressions and parameters**
- **C# Extensibility** – **New activities include REST API calls & SSIS**
- **Connectivity** – **New activities include REST API calls so you can connect to anything!**

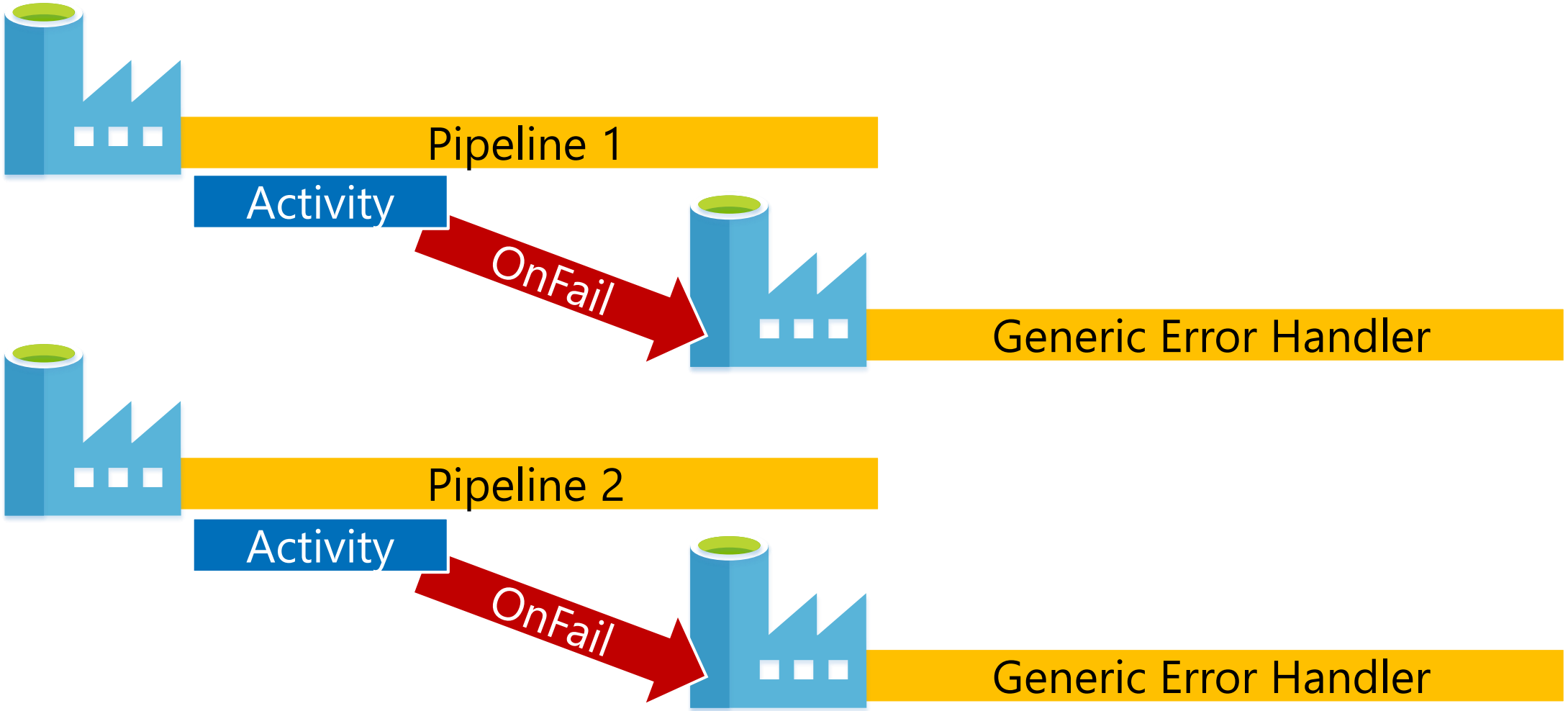
Reactive Data Processing



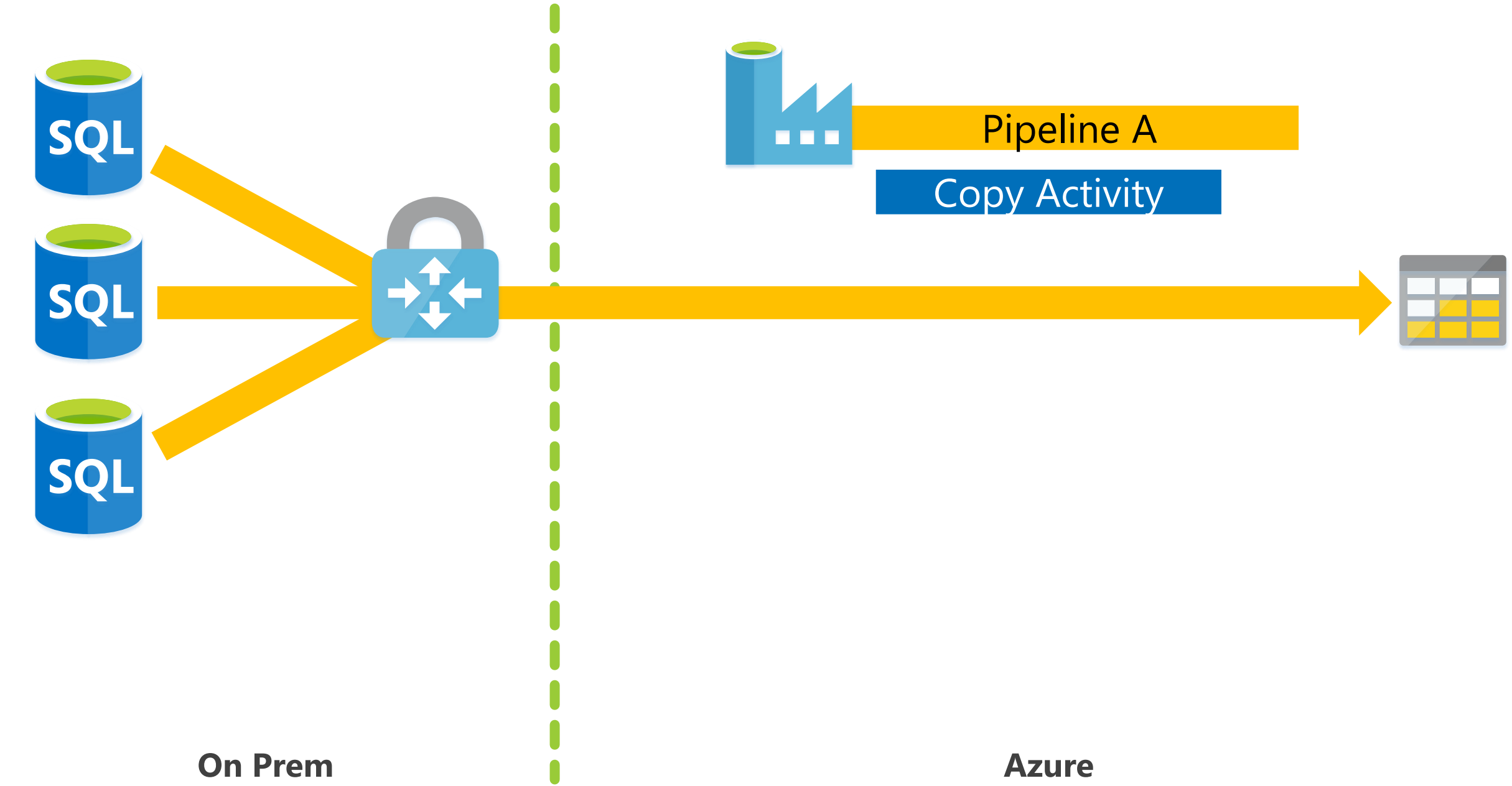
Batch Queuing



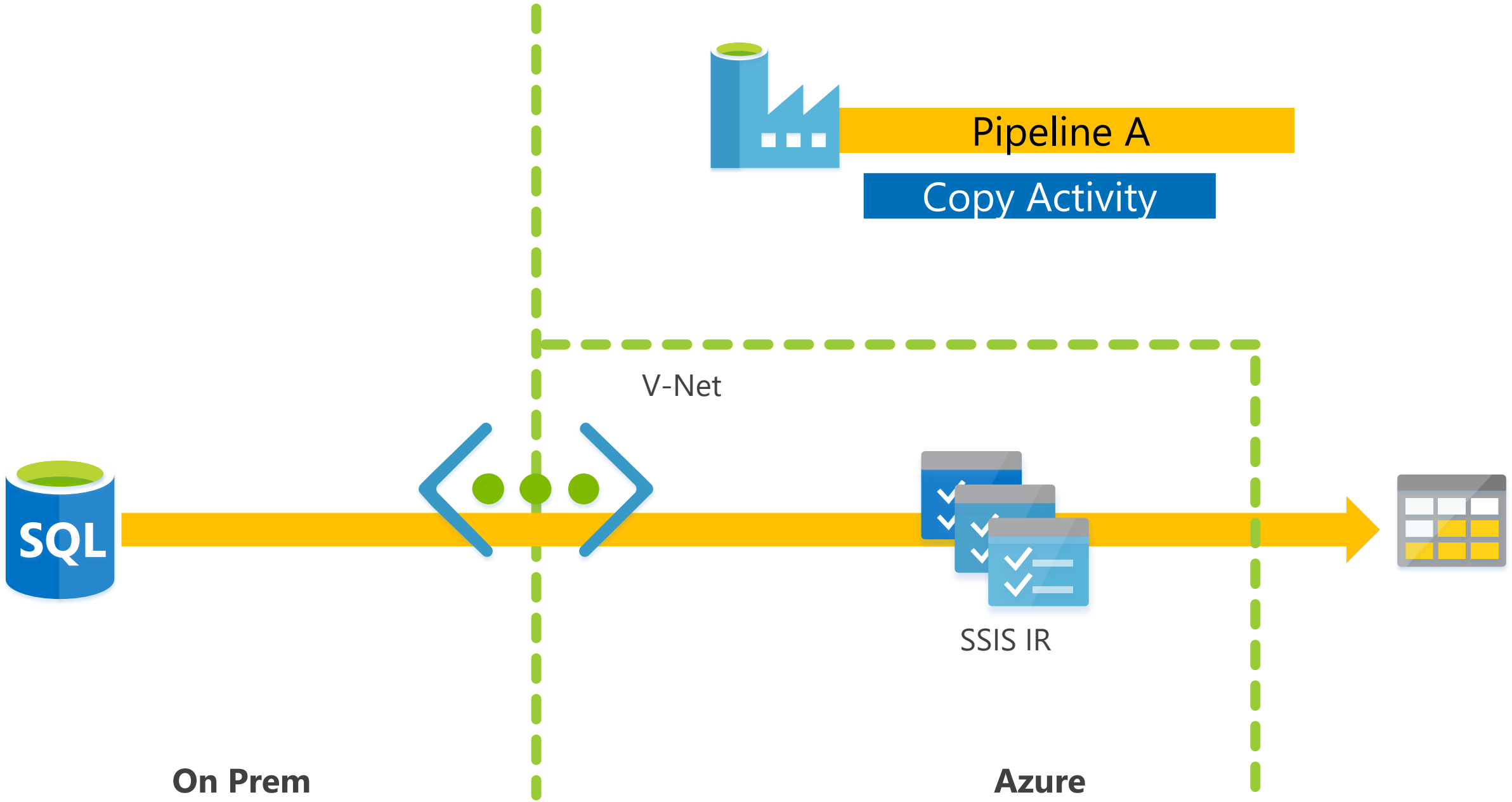
Reusable Pipelines



The SSIS Gateway



The SSIS Gateway



Thanks for Listening

Simon Whiteley

@MrSiWhiteley

http://blogs.adatis.co.uk