




# Module 1: What and why Azure Data Factory?

- Service Components
  - Integration Runtimes
  - Source Control
- 

What is Azure Data Factory?

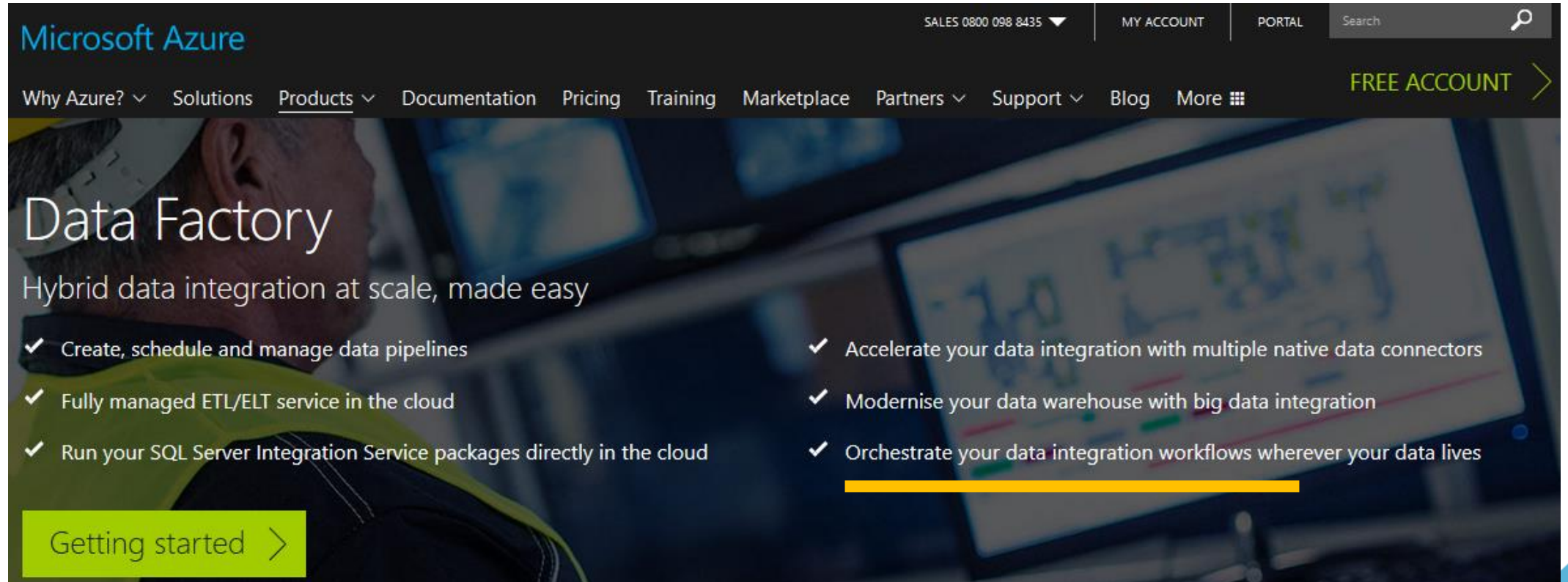
# The History of Azure Data Factory



Azure Data Factory is SQL Server Integration Services in the cloud.

# What is Azure Data Factory?

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



The screenshot shows the Microsoft Azure Data Factory landing page. At the top, the Microsoft Azure logo is on the left, and navigation links for 'SALES 0800 098 8435', 'MY ACCOUNT', 'PORTAL', and a search bar are on the right. Below this is a secondary navigation bar with links for 'Why Azure?', 'Solutions', 'Products', 'Documentation', 'Pricing', 'Training', 'Marketplace', 'Partners', 'Support', 'Blog', and 'More'. A 'FREE ACCOUNT' button with a right arrow is also present. The main section features a large background image of a person in a hard hat looking at a computer screen displaying a data pipeline diagram. The text 'Data Factory' is prominently displayed, followed by the tagline 'Hybrid data integration at scale, made easy'. Below this, there are two columns of bullet points, each starting with a checkmark. The first column lists: 'Create, schedule and manage data pipelines', 'Fully managed ETL/ELT service in the cloud', and 'Run your SQL Server Integration Service packages directly in the cloud'. The second column lists: 'Accelerate your data integration with multiple native data connectors', 'Modernise your data warehouse with big data integration', and 'Orchestrate your data integration workflows wherever your data lives'. A yellow horizontal bar is positioned under the last bullet point in the second column. At the bottom left, there is a green button with the text 'Getting started' and a right arrow.

Microsoft Azure

SALES 0800 098 8435 ▼ MY ACCOUNT PORTAL Search

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

FREE ACCOUNT >

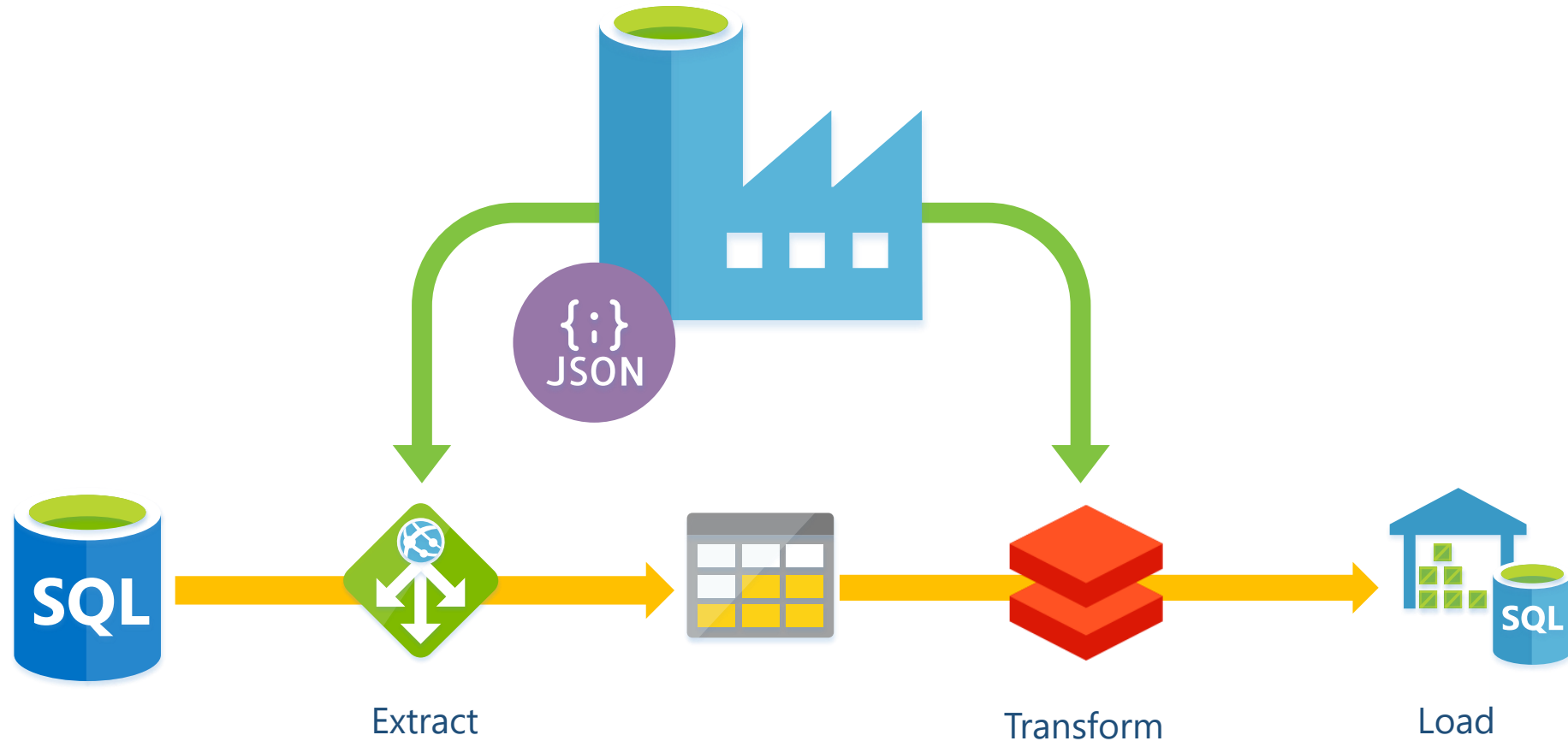
## Data Factory

Hybrid data integration at scale, made easy

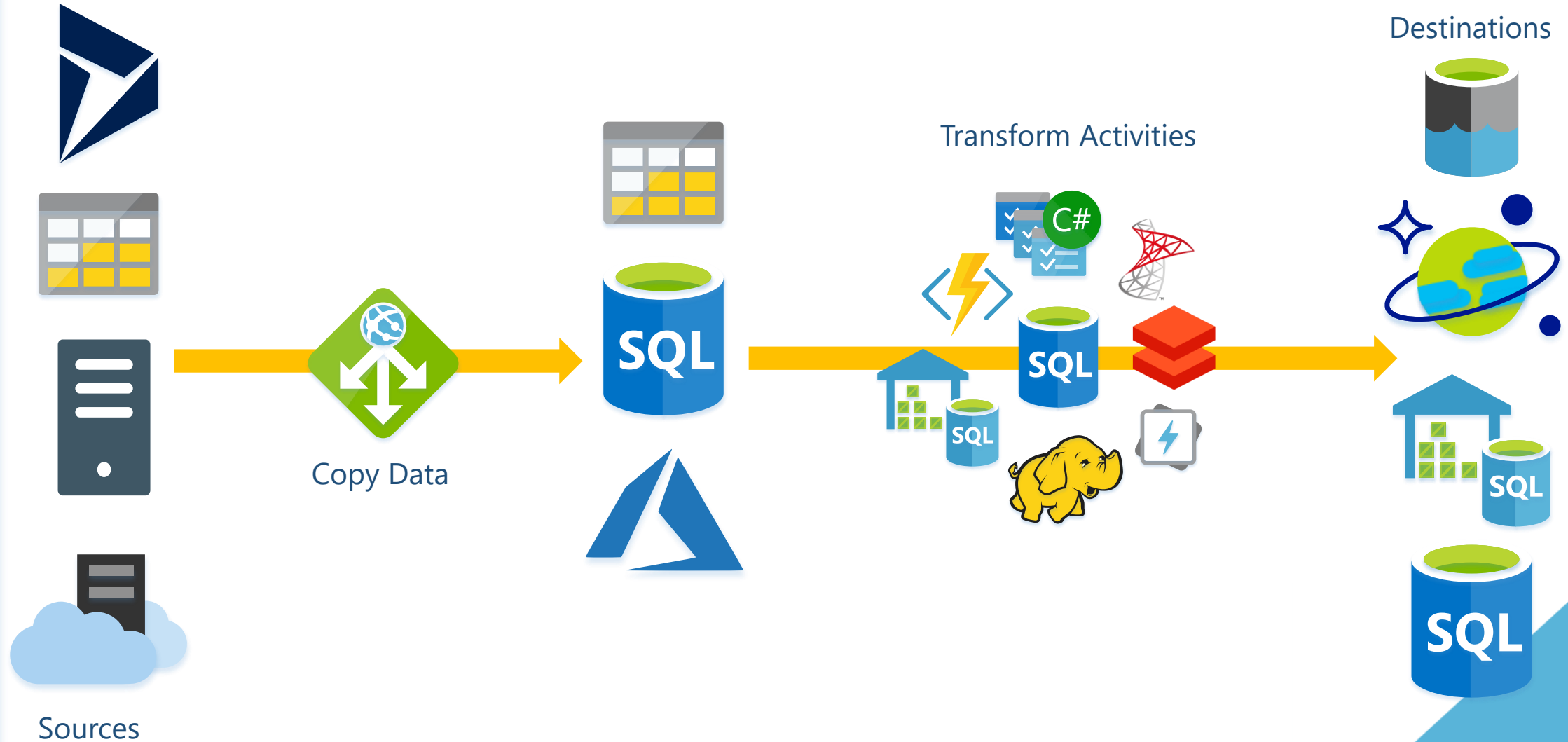
- ✓ Create, schedule and manage data pipelines
- ✓ Fully managed ETL/ELT service in the cloud
- ✓ Run your SQL Server Integration Service packages directly in the cloud
- ✓ Accelerate your data integration with multiple native data connectors
- ✓ Modernise your data warehouse with big data integration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

Getting started >

# What is Azure Data Factory?

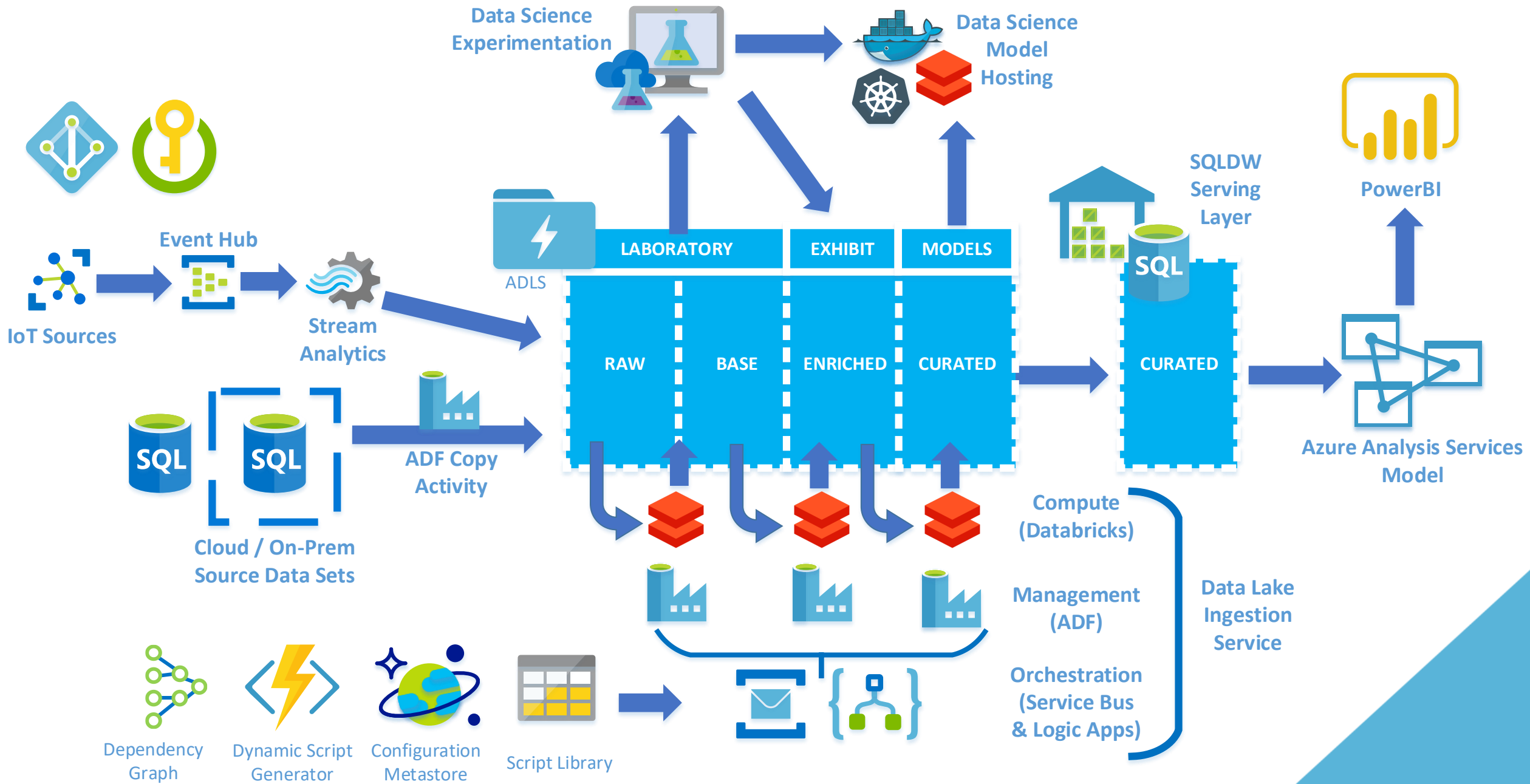


# What does Azure Data Factory do?



Why use it?

# The Modern Data Warehouse

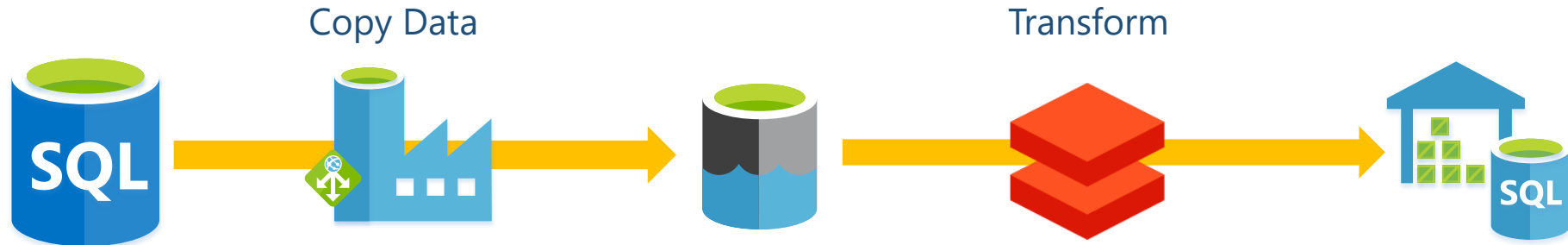




# Service Components

A large blue triangle is positioned in the bottom right corner of the slide, pointing towards the top right.

# Data Factory Components



## 1 Linked Services – How do I connect?

Like the SSIS Connection Manager!



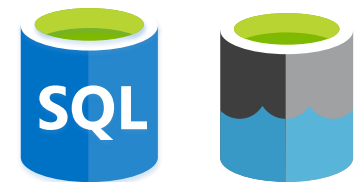
SQLDBLinkedService

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

# Data Factory Components

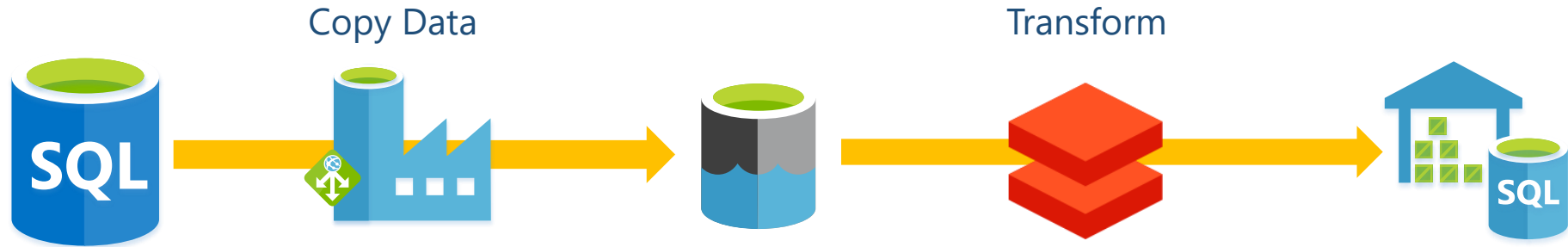
1

## Linked Services –



 Amazon Marketplace Web Service (Preview)	 Amazon Redshift	 Amazon S3	 HDFS	 HTTP	 Hive	 Netezza	 ODBC	 OData	 Azure Batch	 Azure Data Lake Analytics	 Azure Databricks
 Apache Impala (Preview)	 Azure Blob Storage	 Azure Cosmos DB (MongoDB API)	 HubSpot (Preview)	 Informix	 Jira (Preview)	 Office 365 (Preview)	 Oracle	 Oracle Eloqua (Preview)	 Azure Function	 Azure HDInsight	 Azure ML
 Azure Cosmos DB (SQL API)	 Azure Data Explorer (Kusto)	 Azure Data Lake Storage Gen1	 Magento (Preview)	 MariaDB	 Marketo (Preview)	 Oracle Responsys (Preview)	 Oracle Service Cloud (Preview)	 Paypal (Preview)	 ServiceNow	 Shopify (Preview)	 Spark
 Azure Data Lake Storage Gen2 (Preview)	 Azure Database for MariaDB	 Azure Database for MySQL	 Microsoft Access	 MongoDB	 MySQL	 Phoenix	 PostgreSQL	 Presto (Preview)	 Square (Preview)	 Sybase	 Teradata
 Azure Database for PostgreSQL	 Azure File Storage	 Azure Key Vault	 DB2	 Drill (Preview)	 Dynamics 365	 QuickBooks (Preview)	 REST	 SAP BW Open Hub	 Vertica	 Web Table	 Xero (Preview)
 Azure SQL Data Warehouse	 Azure SQL Database	 Azure SQL Database Managed Instance	 Dynamics AX (Preview)	 Dynamics CRM	 FTP	 SAP BW via MDX	 SAP Cloud For Customer	 SAP ECC	 Zoho (Preview)		
 Azure Search	 Azure Table Storage	 Cassandra	 File System	 Google AdWords (Preview)	 Google BigQuery	 SAP HANA	 SFTP	 SQL Server			
 Common Data Service for Apps	 Concur (Preview)	 Couchbase (Preview)	 Google Cloud Storage (S3 API)	 Greenplum	 HBase	 Salesforce	 Salesforce Marketing Cloud (Preview)	 Salesforce Service Cloud			

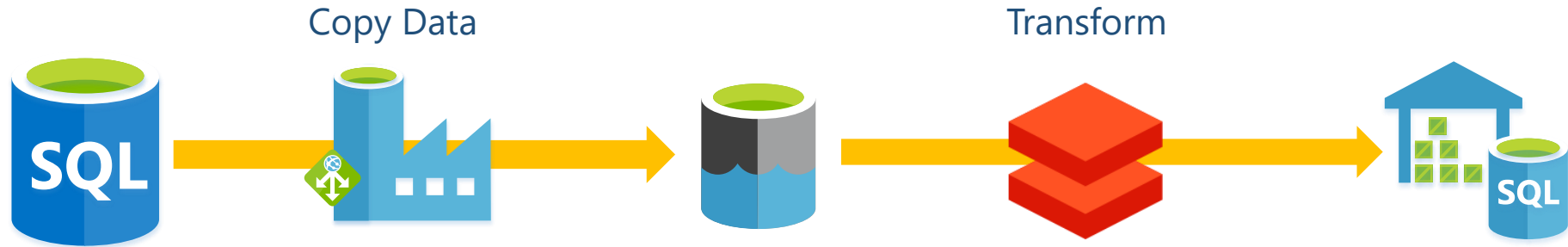
# Data Factory Components



1

## Linked Services

# Data Factory Components



1

## Linked Services

2

**Data Sets** – Where is my data? What format? What file path/table do I need?

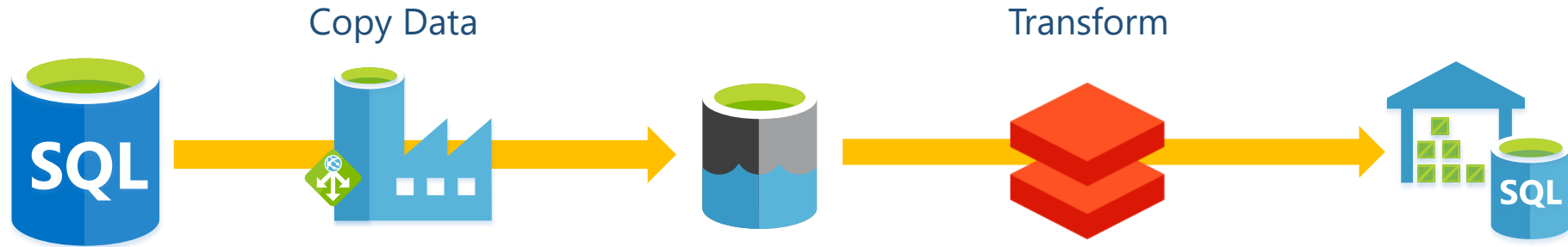


dbo.DimCustomer



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

# Data Factory Components



1 **Linked Services**

2 **Data Sets**

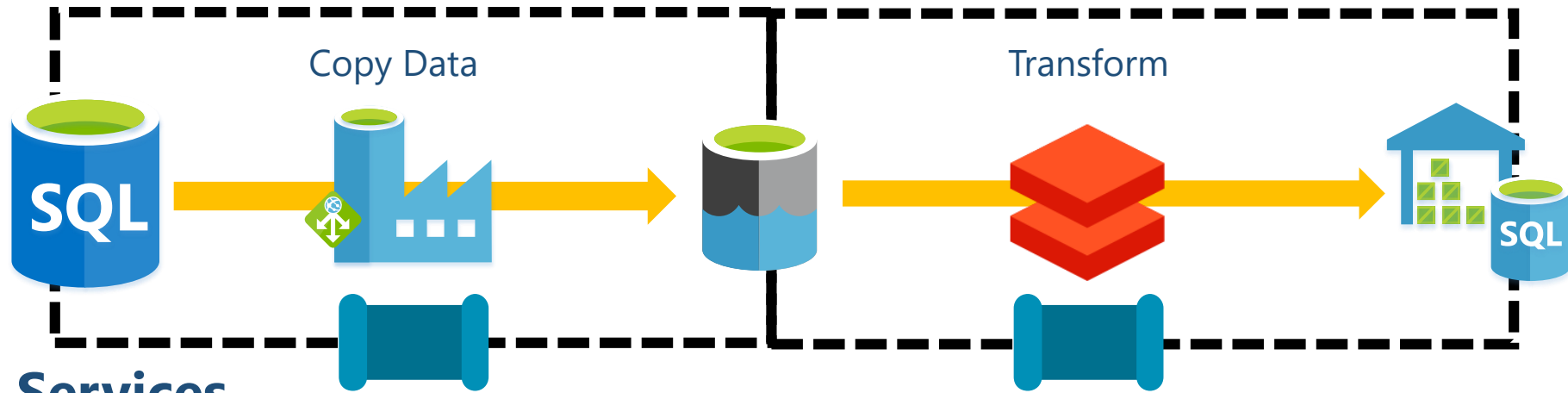
3 **Activities** – What do we want to happen?  
With what conditions?



## Databricks Notebook Activity

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

# Data Factory Components



1

**Linked Services**

2

**Data Sets**

3

**Activities**

4

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



Sequence Container

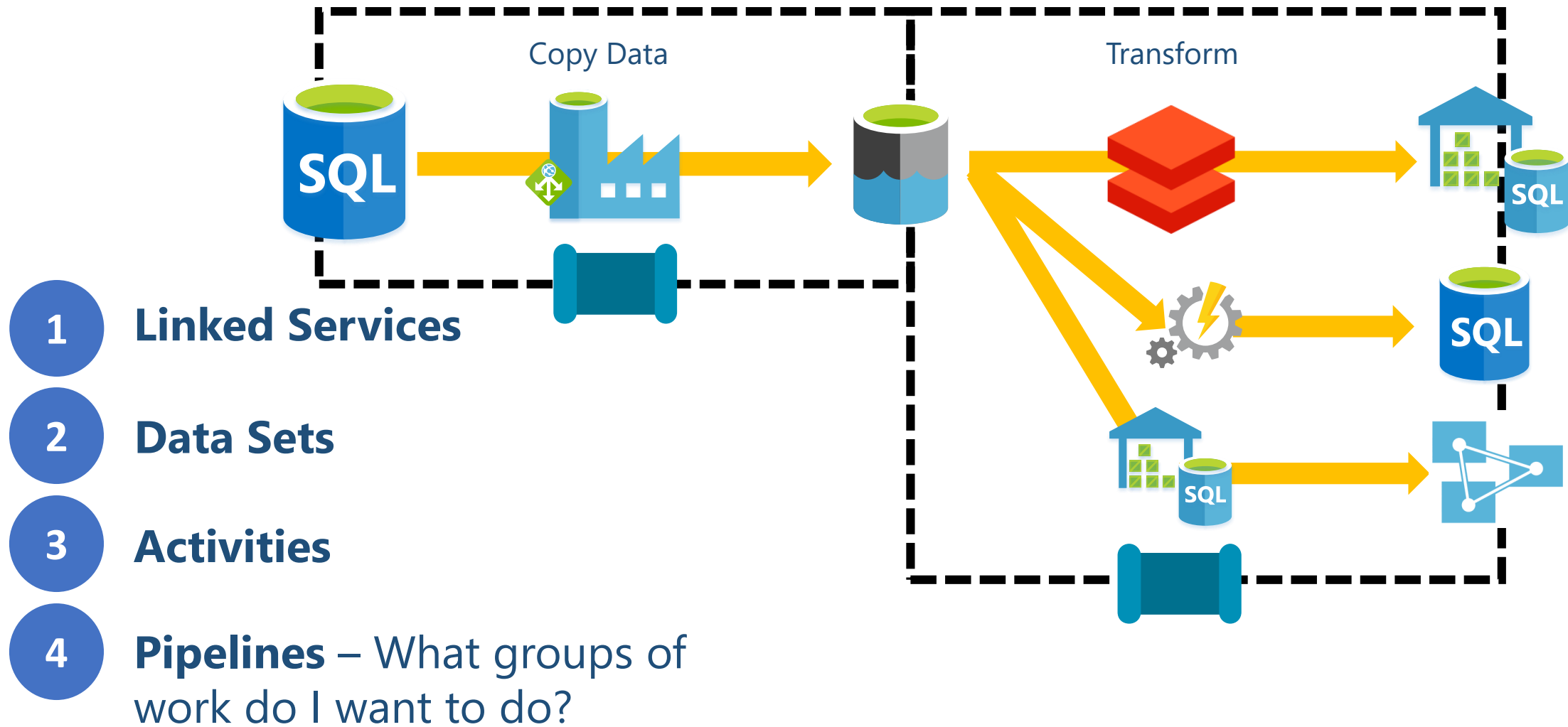


Execute Package Task



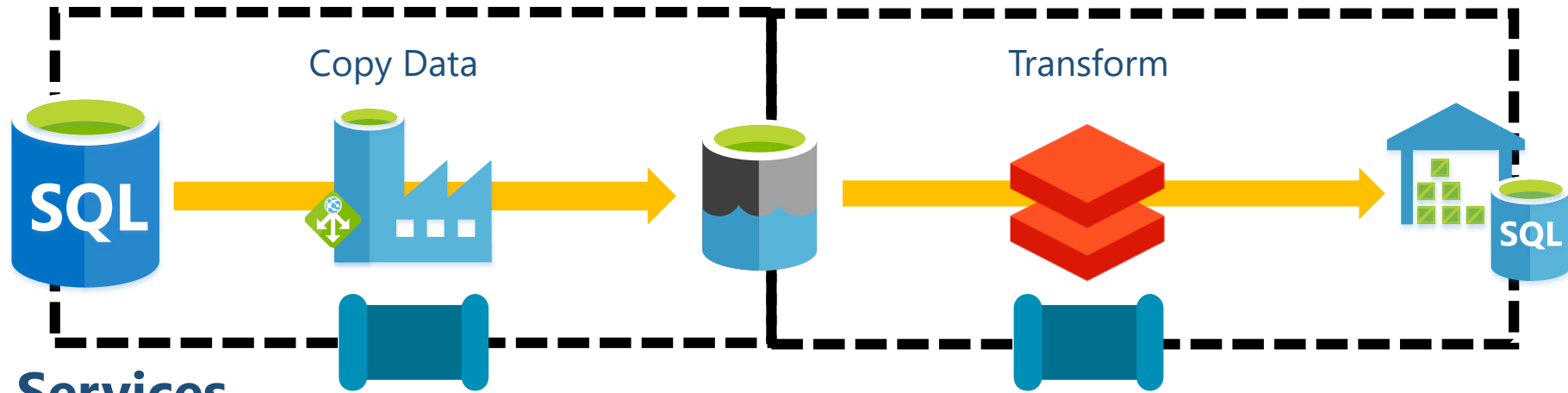
Execute Pipeline Activity

# Data Factory Components





# Data Factory Components



1

**Linked Services**

2

**Data Sets**

3

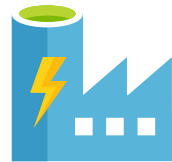
**Activities**

4

**Pipelines**

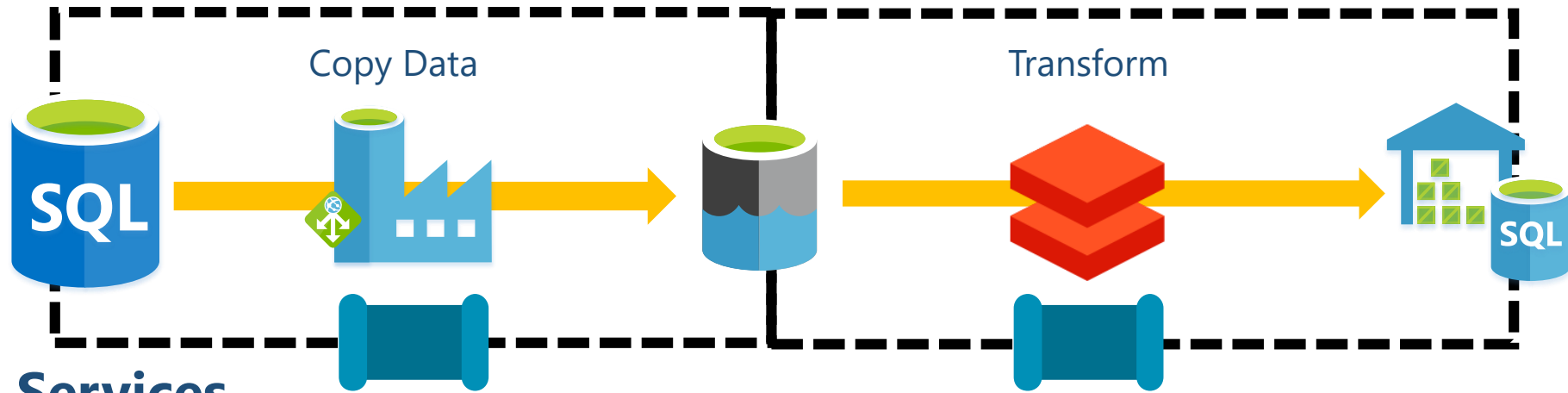
5

**Triggers** – How are we going to tell our pipeline(s) to execute?



- Manual via UI
- Tumbling Windows
- Scheduled
- Blob File Events
- Logic App Calls

# Data Factory Components



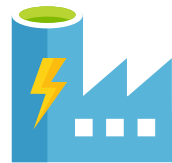
1 Linked Services

2 Data Sets

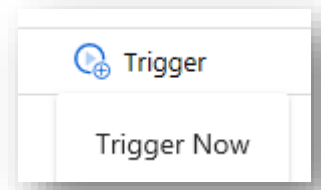
3 Activities

4 Pipelines

5 Triggers

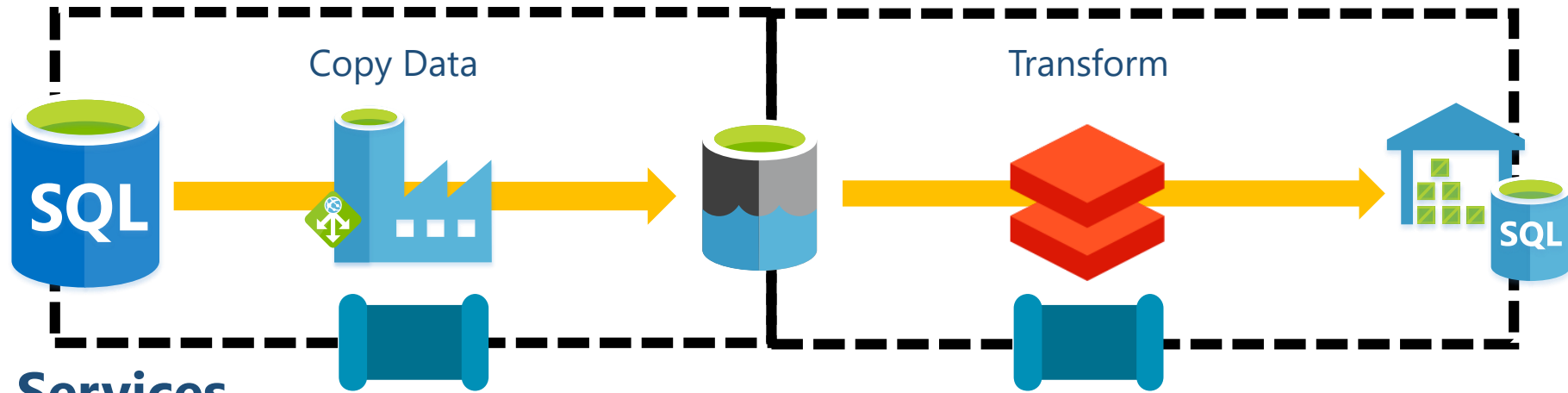


- **Manual**
- Tumbling Windows
- Scheduled
- Blob File Events
- Logic App Calls



`Invoke-AzureRmDataFactoryV2Pipeline`  
`-DataFactoryName $dataFactoryName`  
`-ResourceGroupName $resourceGroupName`  
`-PipelineName $pipelineName`

# Data Factory Components



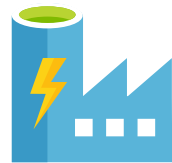
1 Linked Services

2 Data Sets

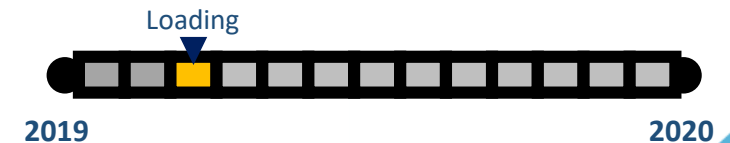
3 Activities

4 Pipelines

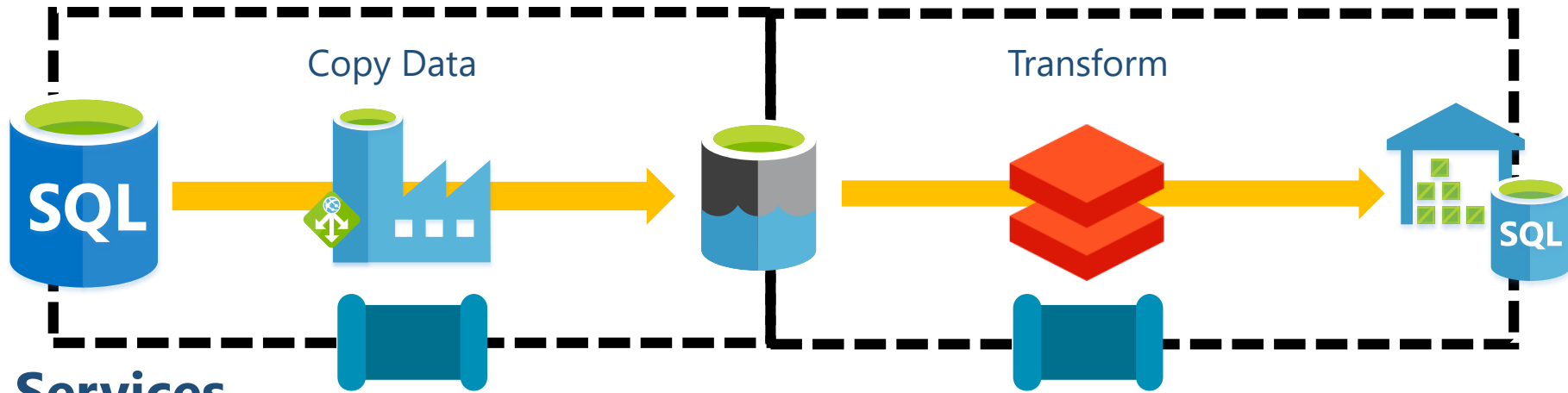
5 Triggers



- Manual via UI
- **Tumbling Windows** - AKA Time Slices
- Scheduled
- Blob File Events
- Logic App Calls



# Data Factory Components



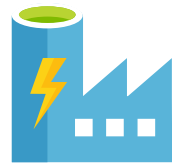
1 Linked Services

2 Data Sets

3 Activities

4 Pipelines

5 Triggers

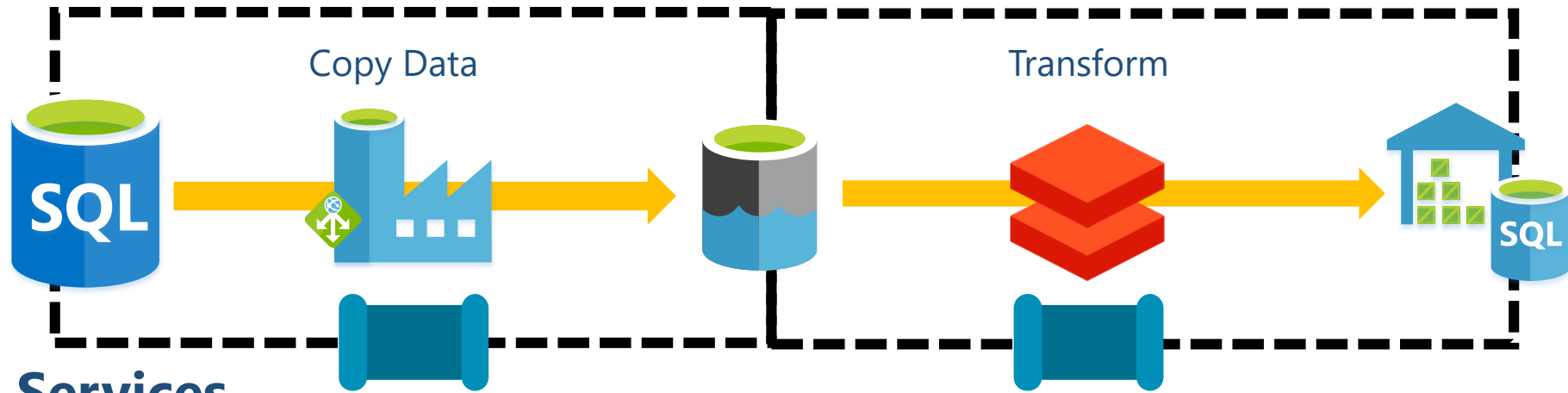


- Manual via UI
- Tumbling Windows
- **Scheduled**
- Blob File Events
- Logic App Calls



- Every 1 minute.
- UTC

# Data Factory Components



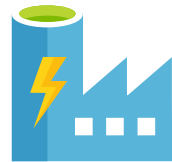
1 Linked Services

2 Data Sets

3 Activities

4 Pipelines

5 Triggers

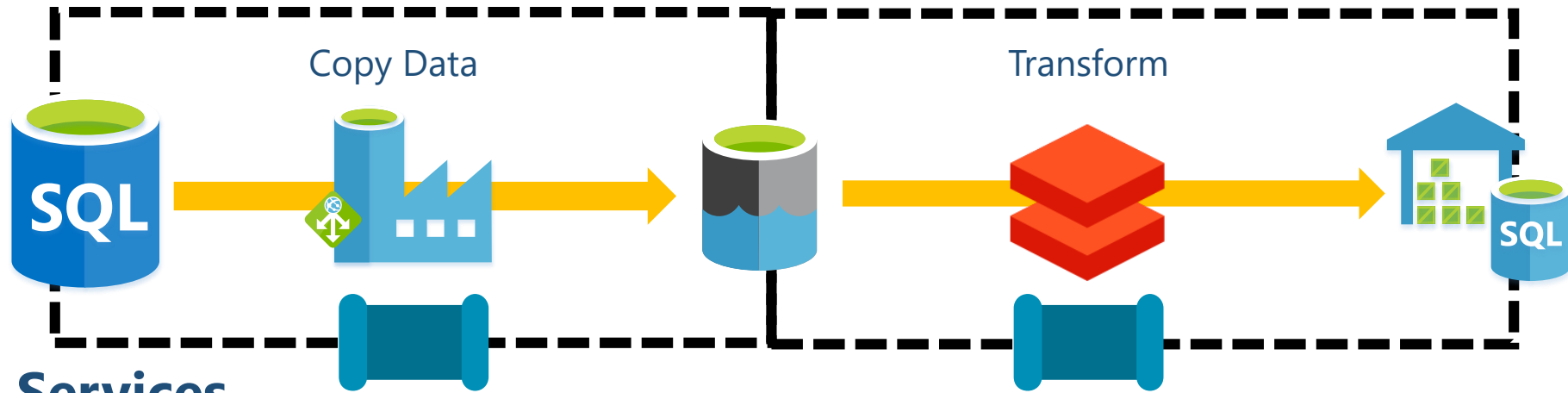


- Manual via UI
- Tumbling Windows
- Scheduled
- **Blob File Events**
- Logic App Calls



{Path} Created  
{Path} Deleted

# Data Factory Components



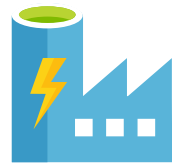
1 Linked Services

2 Data Sets

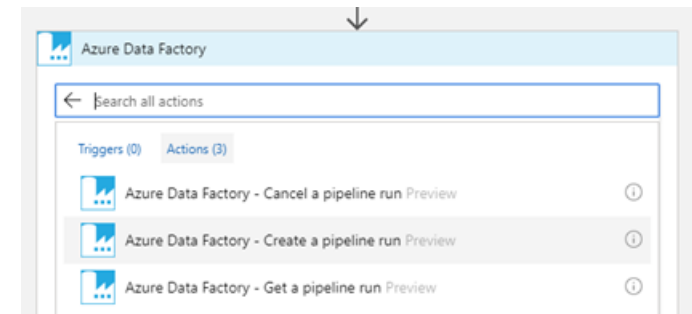
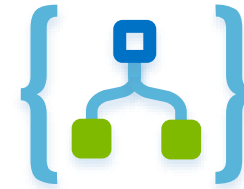
3 Activities

4 Pipelines

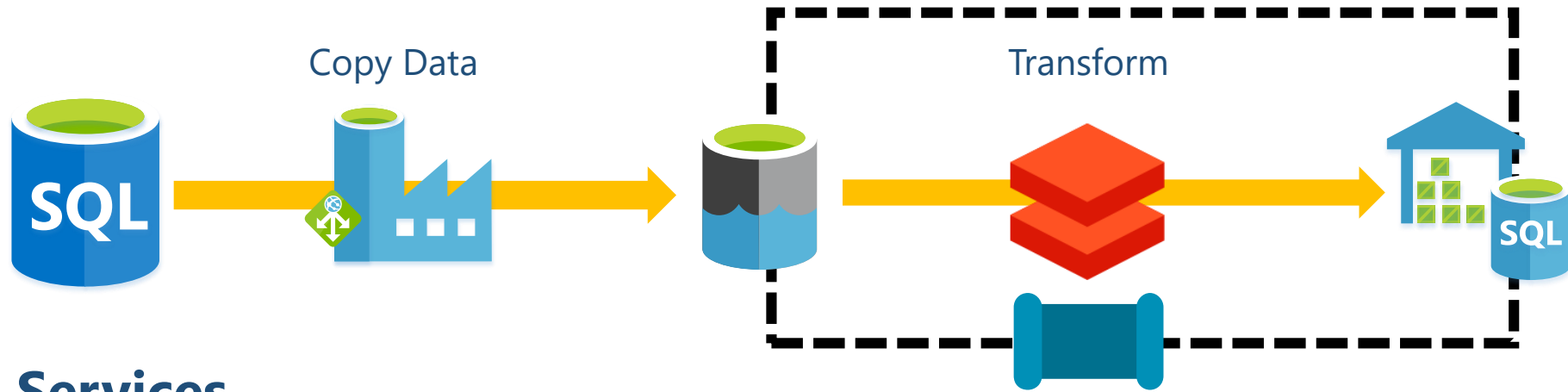
5 Triggers



- Manual via UI
- Tumbling Windows
- Scheduled
- Blob File Events
- **Logic App Calls**



# Data Factory Components – Recap



1 **Linked Services**

2 **Data Sets**

3 **Activities**

4 **Pipelines**

5 **Triggers**

# Integration Runtimes

A large, solid blue triangle is positioned in the bottom right corner of the slide, pointing towards the top right.



# Integration Runtimes



1

**Azure**  
Integration Runtime

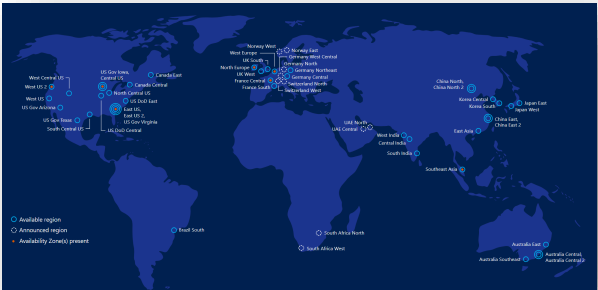
Movement Hours



Activity  
Orchestration



Flexible Region



Integration Runtimes / AutoResolveIntegrationRuntime / Resource Monitor (Details)

Refresh

Edit

Details

Activities

STATUS	TYPE	REGION
Running	Azure	<u>Auto Resolve</u>

**Orchestrator**

**Worker**

# Integration Runtimes



# Integration Runtimes – Azure



# Integration Runtimes



**Azure**  
Integration Runtime

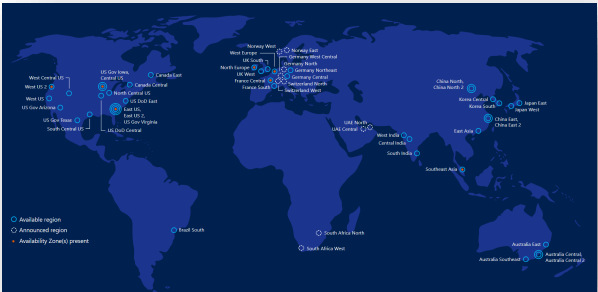
Movement Hours



Activity  
Orchestration



Flexible Region



# Integration Runtimes



1

**Azure**  
Integration Runtime

Movement Hours



Activity  
Orchestration







2

**SSIS**  
Integration Runtime

SSIS Package  
Execution









3

**Self Hosted**  
Integration Runtime

Gateway Access



Activity  
Orchestration





# Integration Runtimes

3

**Self Hosted**  
Integration Runtime

Gateway Access

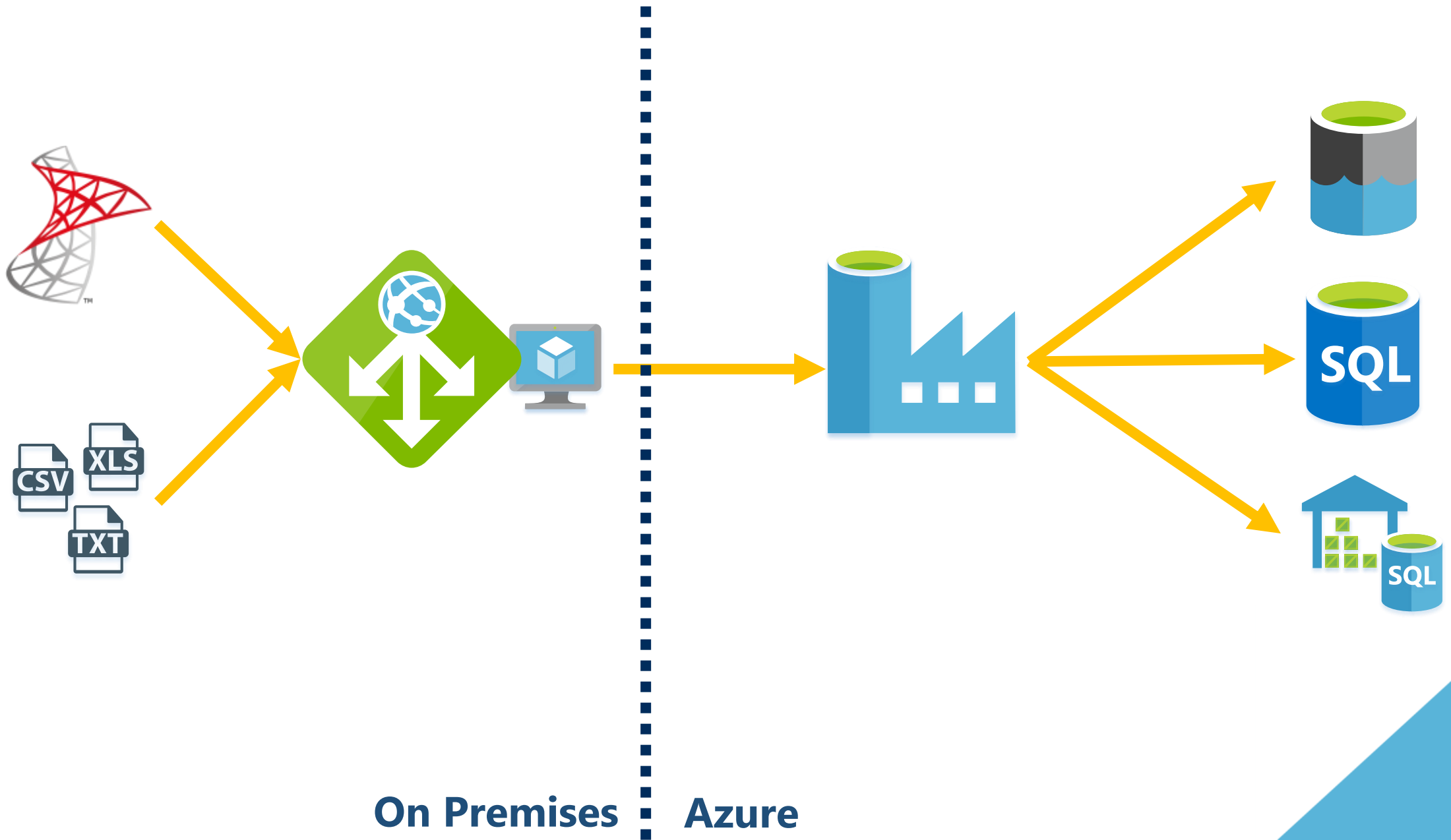
Activity  
Orchestration



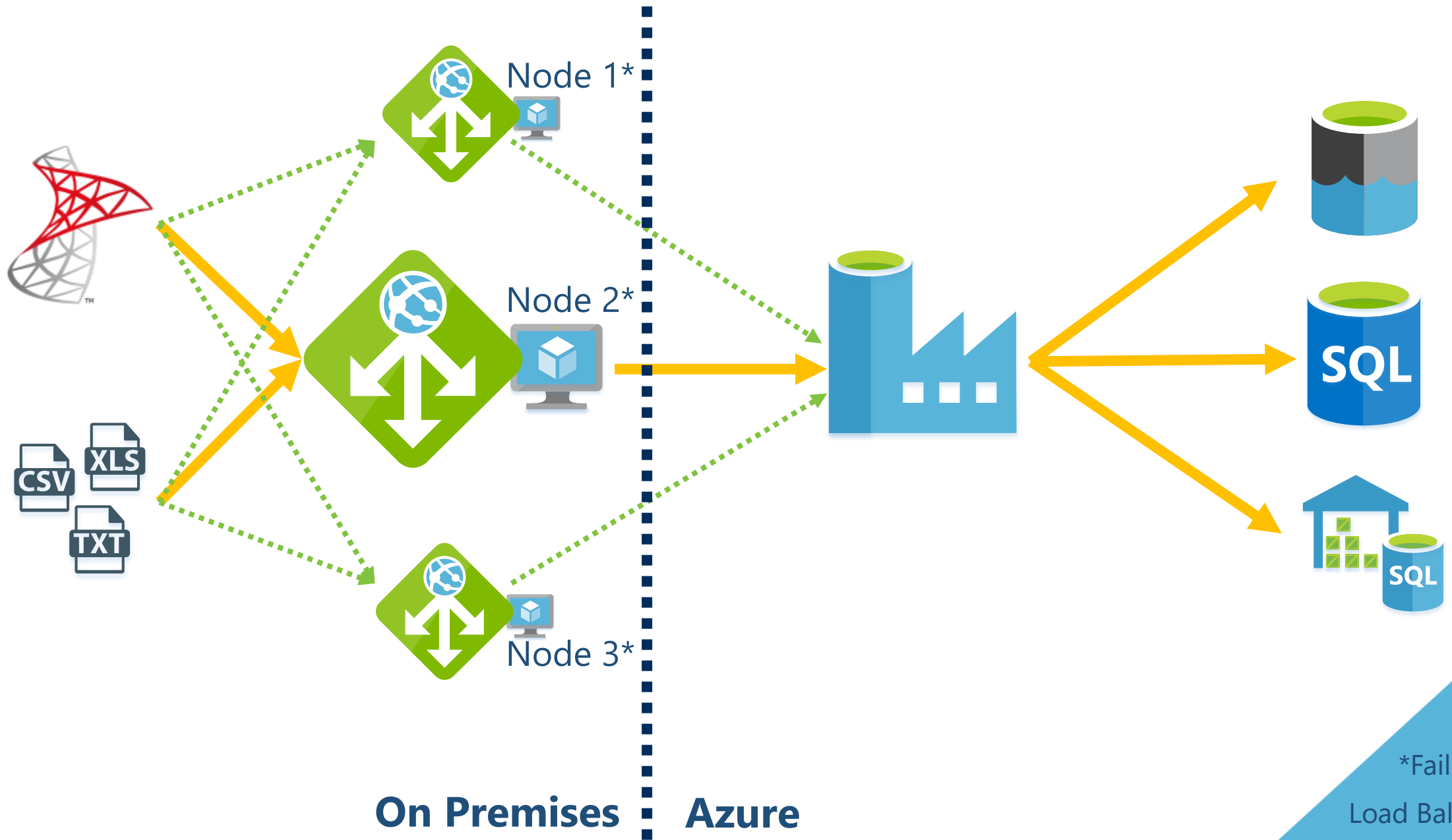
Virtual Machine



# The Hosted Integration Runtime – Basic Pattern

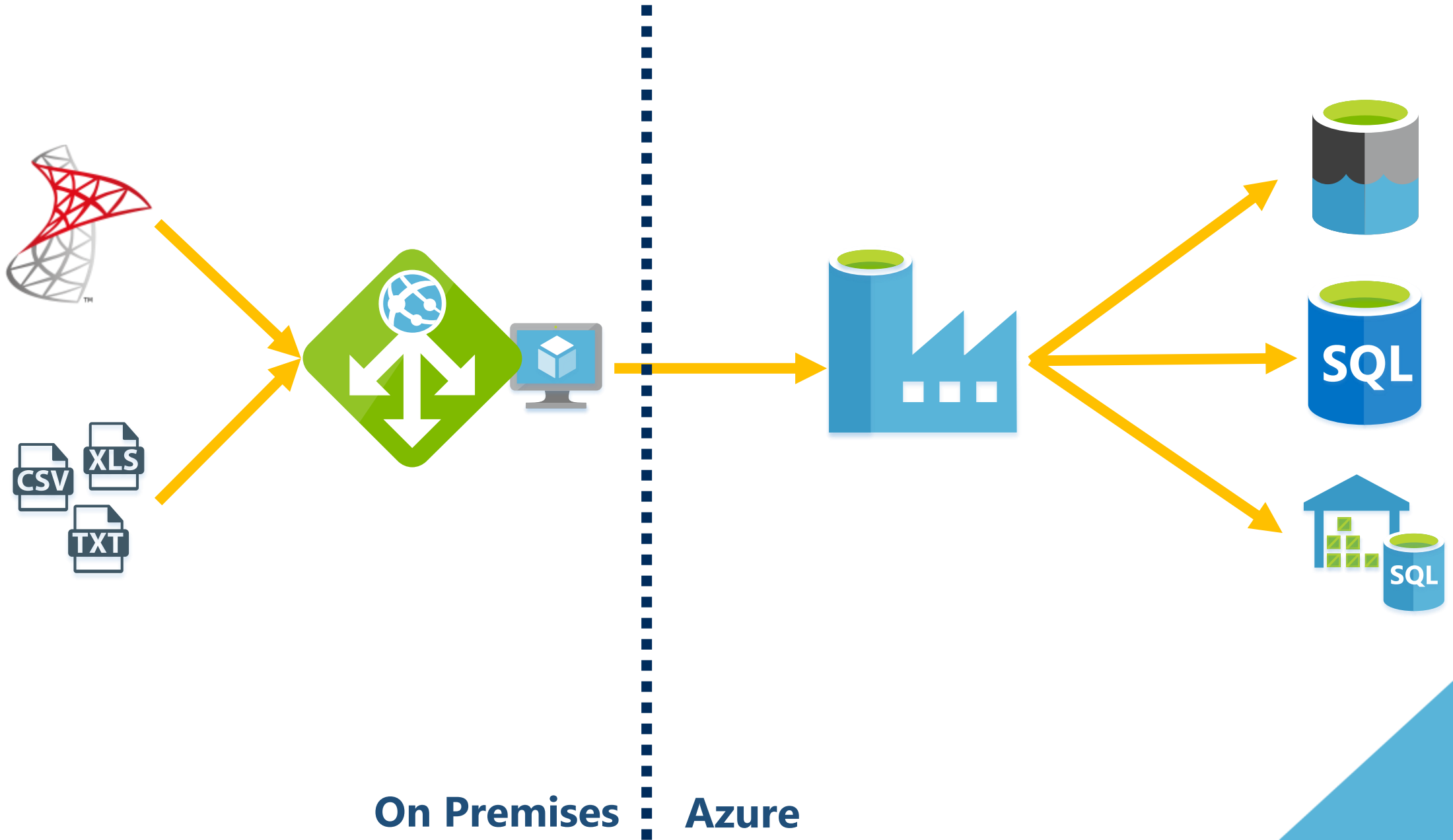


# The Hosted Integration Runtime - Multiple Gateways

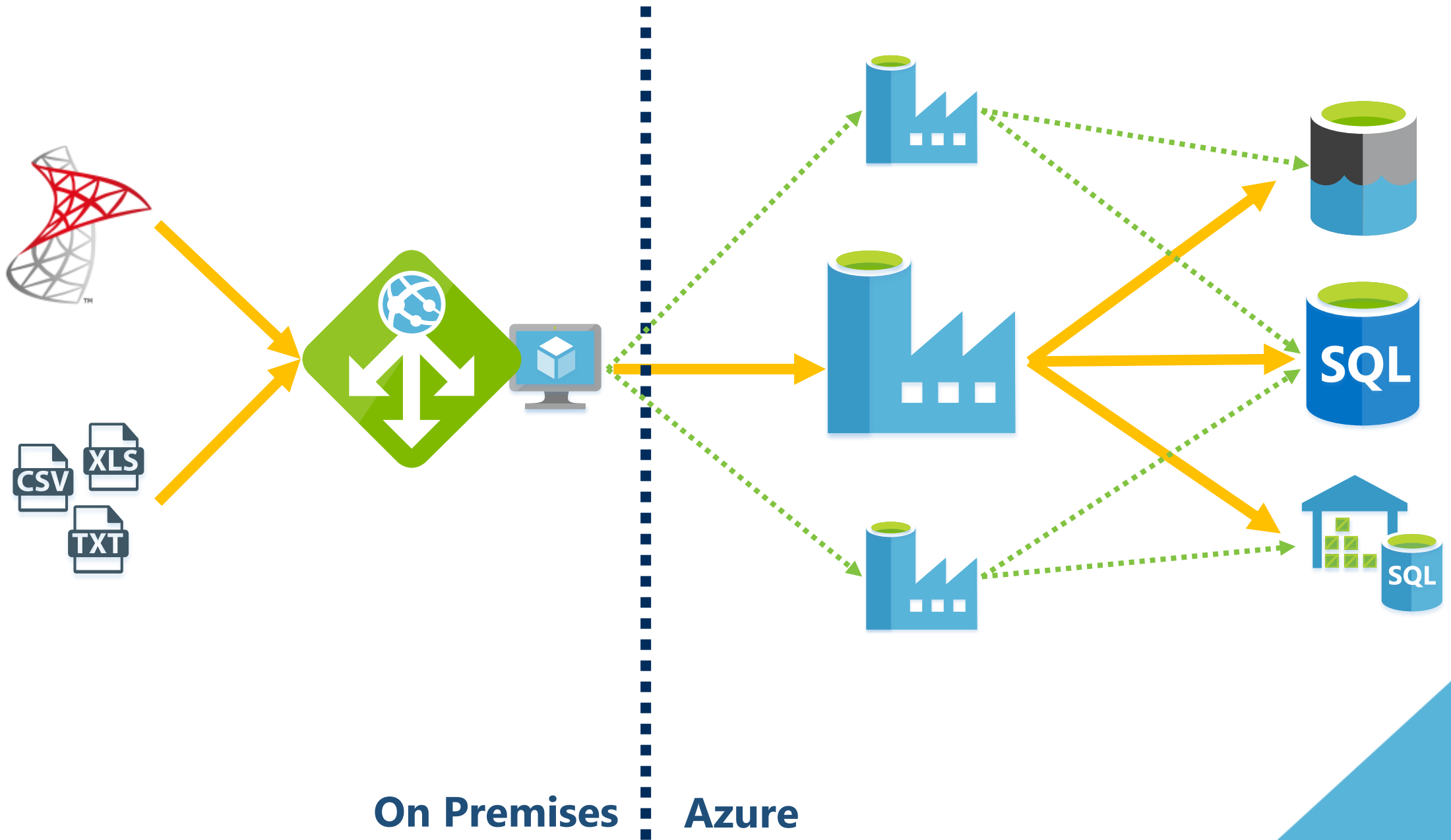




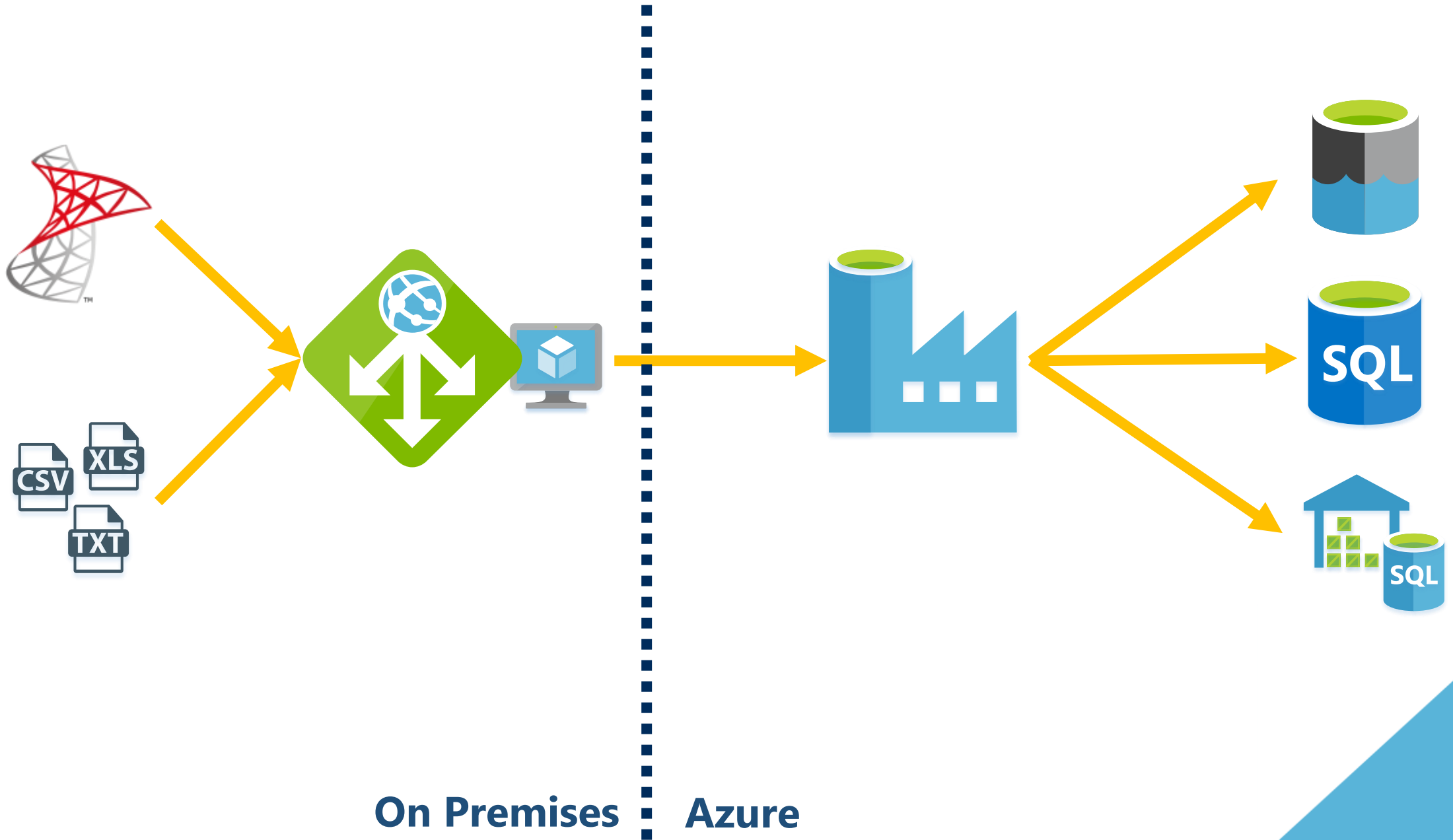
# The Hosted Integration Runtime



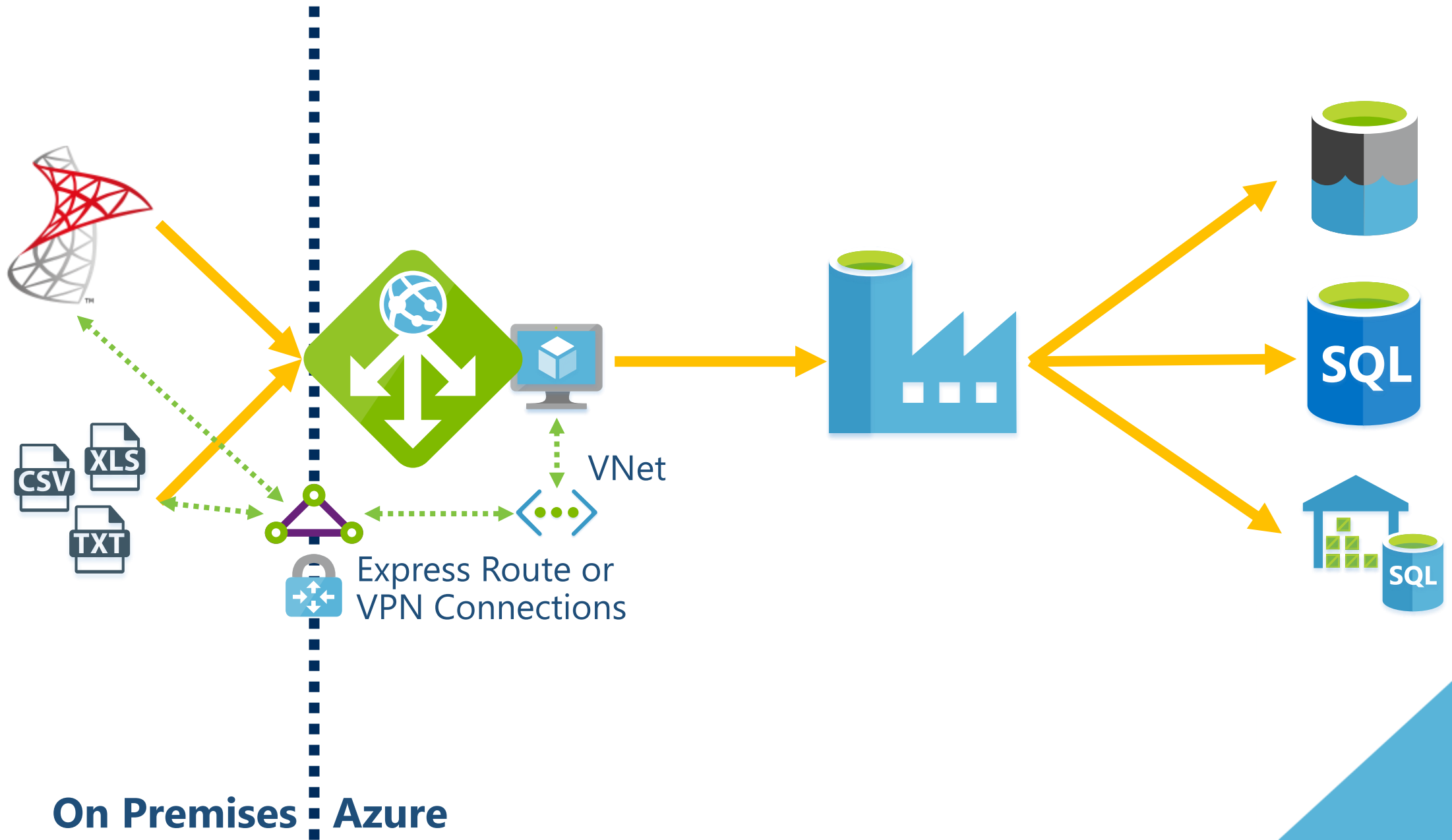
# The Hosted Integration Runtime – Shared Gateways



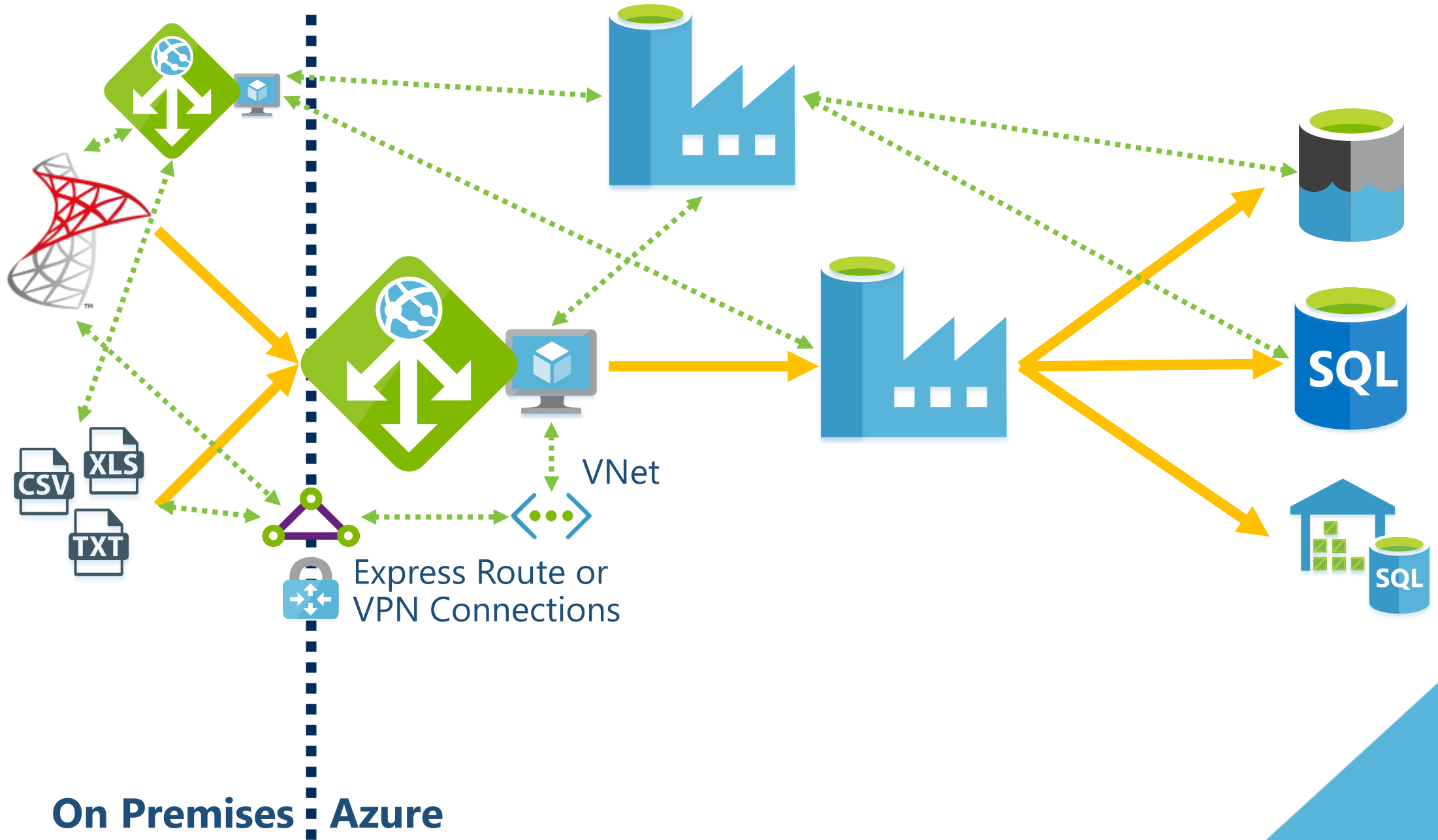
# The Hosted Integration Runtime



# The Hosted Integration Runtime – With Express Route



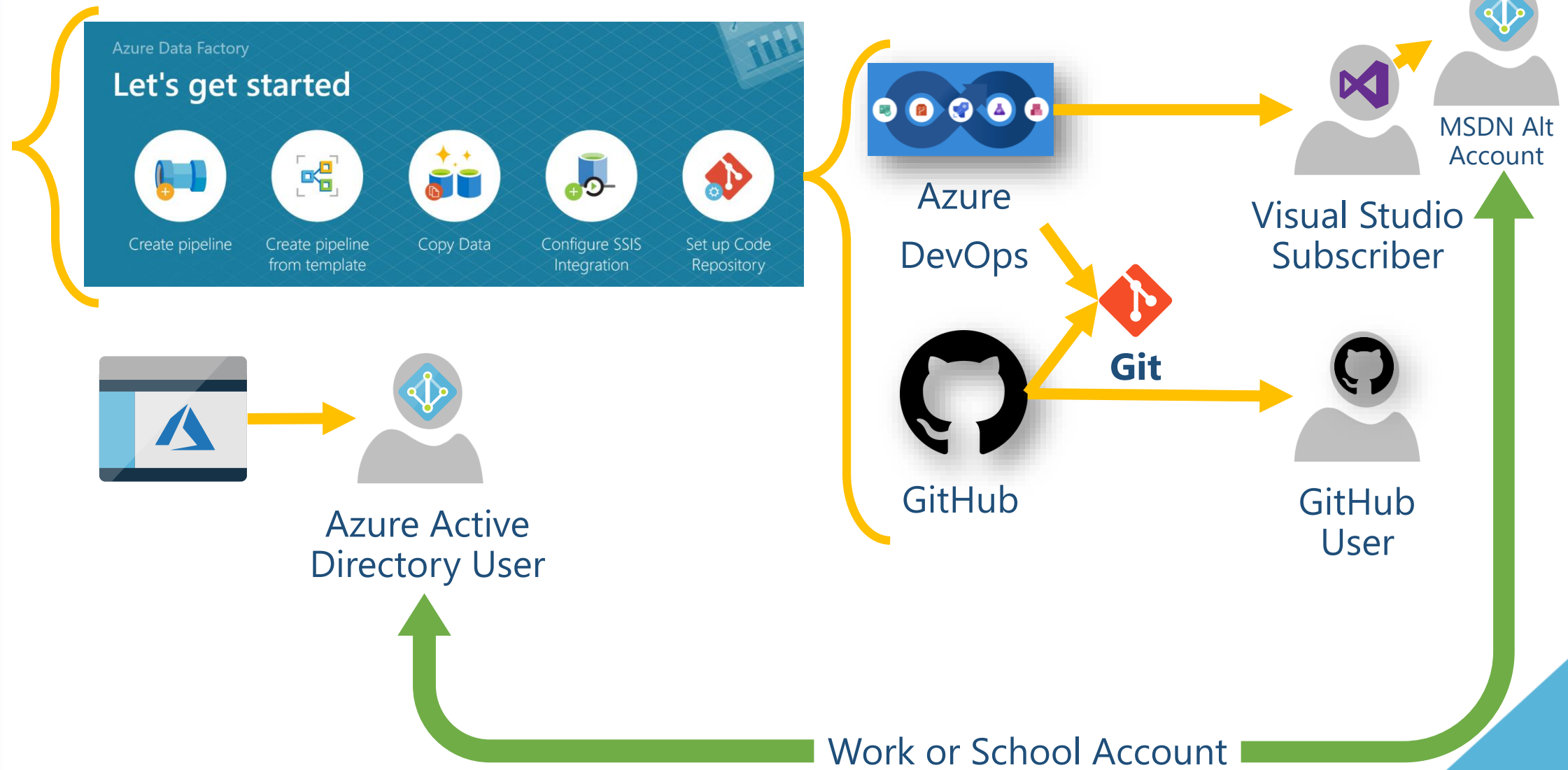
# The Hosted Integration Runtime – Hybrid Gateways



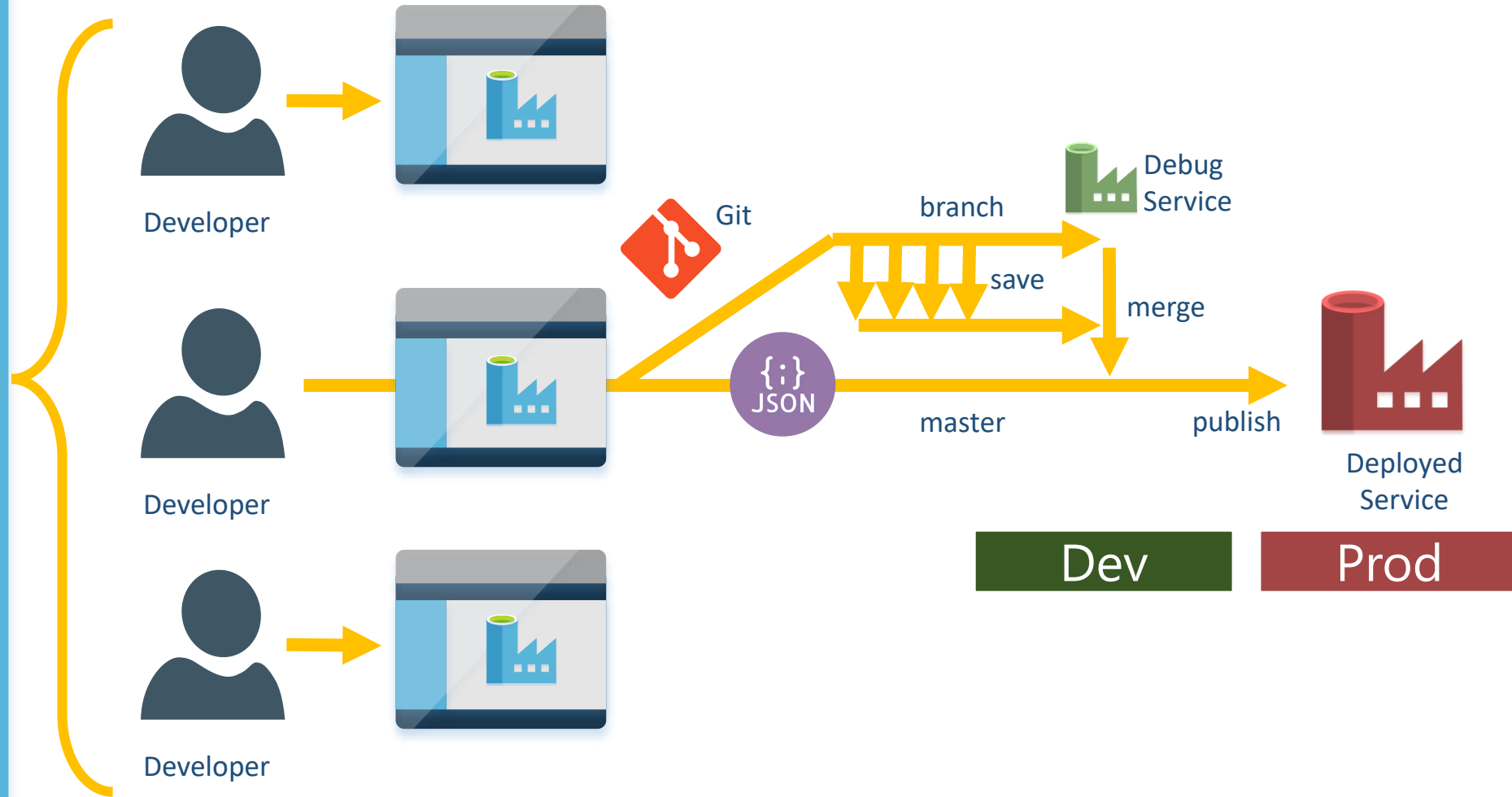
Source Control

A large, solid blue triangle is positioned in the bottom right corner of the slide, pointing towards the top right.

# Configuring Source Control

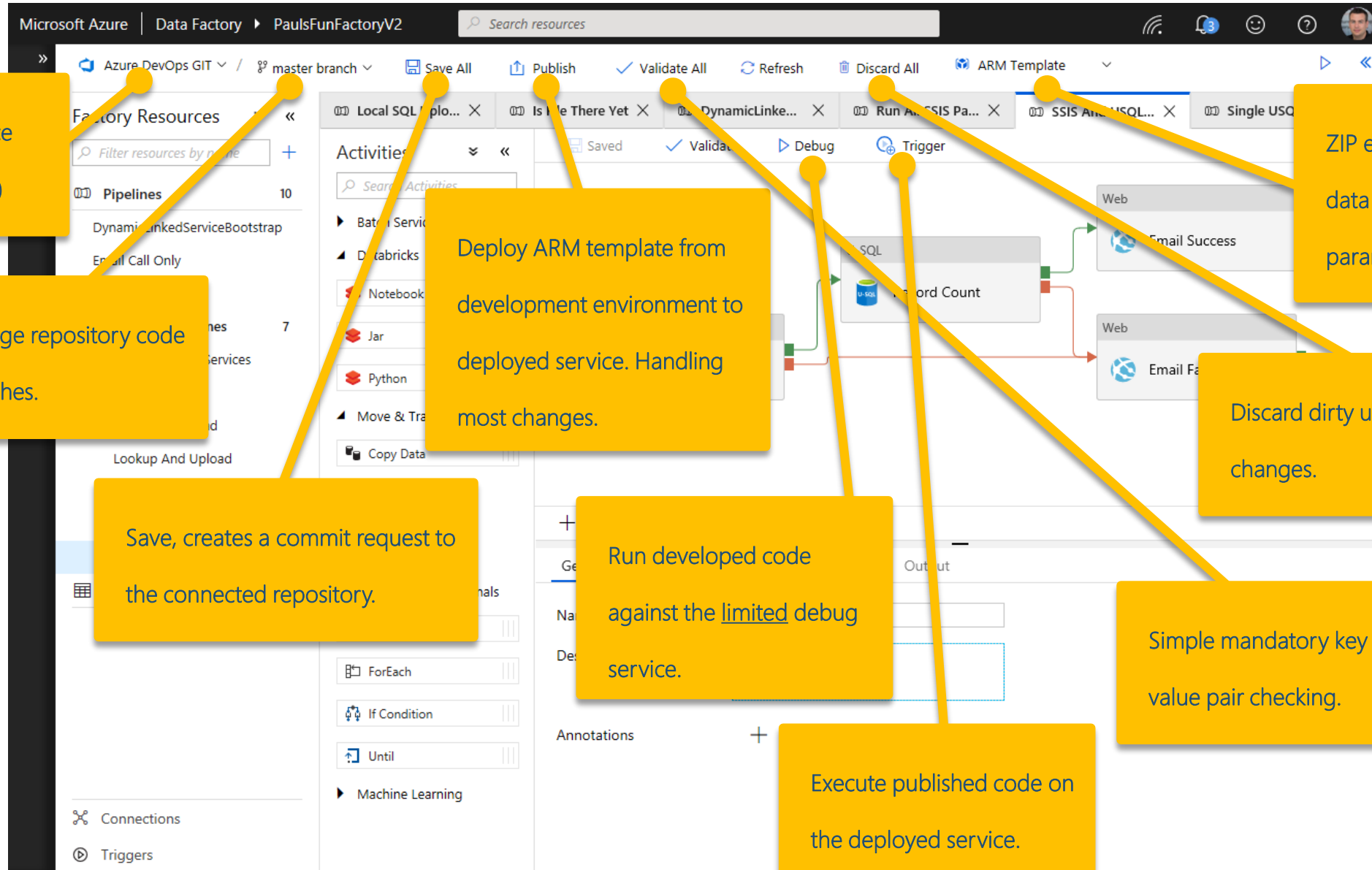


# Data Factory Continuous Integration





# Data Factory Continuous Integration



# Module 1 - What and why Azure Data Factory?

- Service Components
  - Linked Services
  - Datasets
  - Activities
  - Pipelines
- Integration Runtimes
  - Azure
  - SSIS
  - Hosted
- Source Control
  - GitHub
  - Azure DevOps (Git)