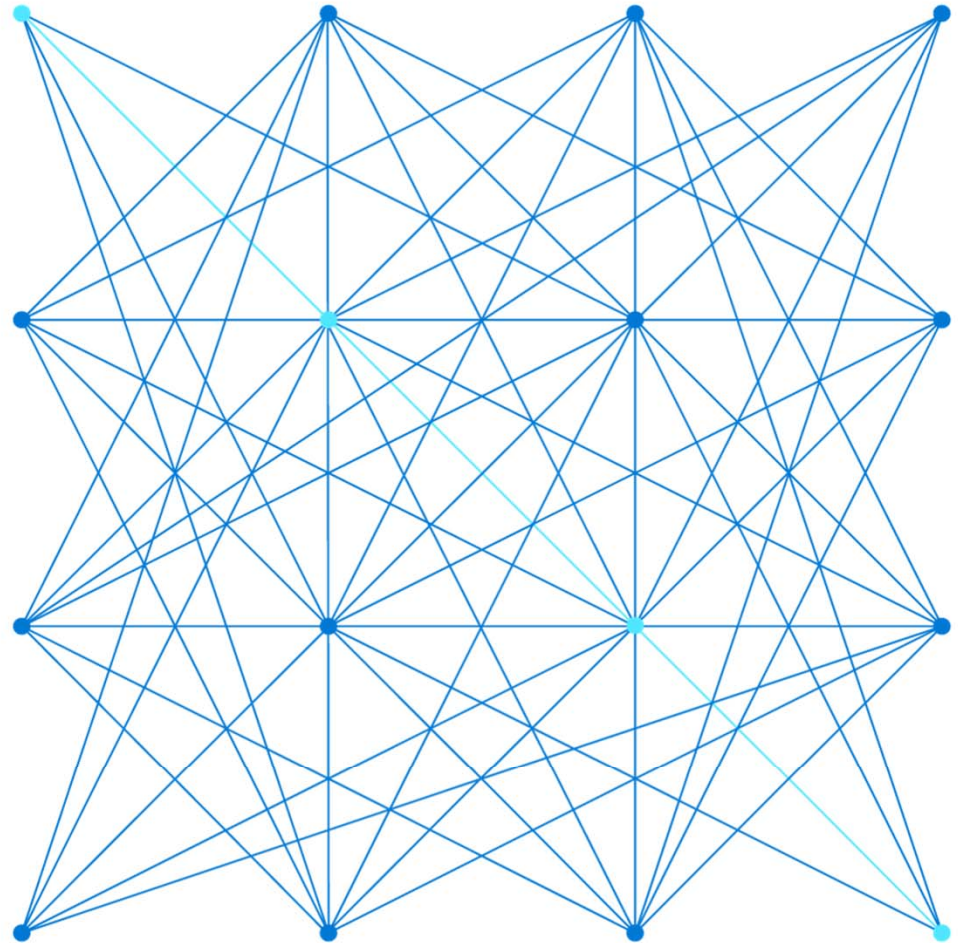


DP-203T00: Transform data with Azure Data Factory or Azure Synapse Pipelines





Agenda



Lesson 01 – Data integration with Azure Data Factory or Azure Synapse Pipelines

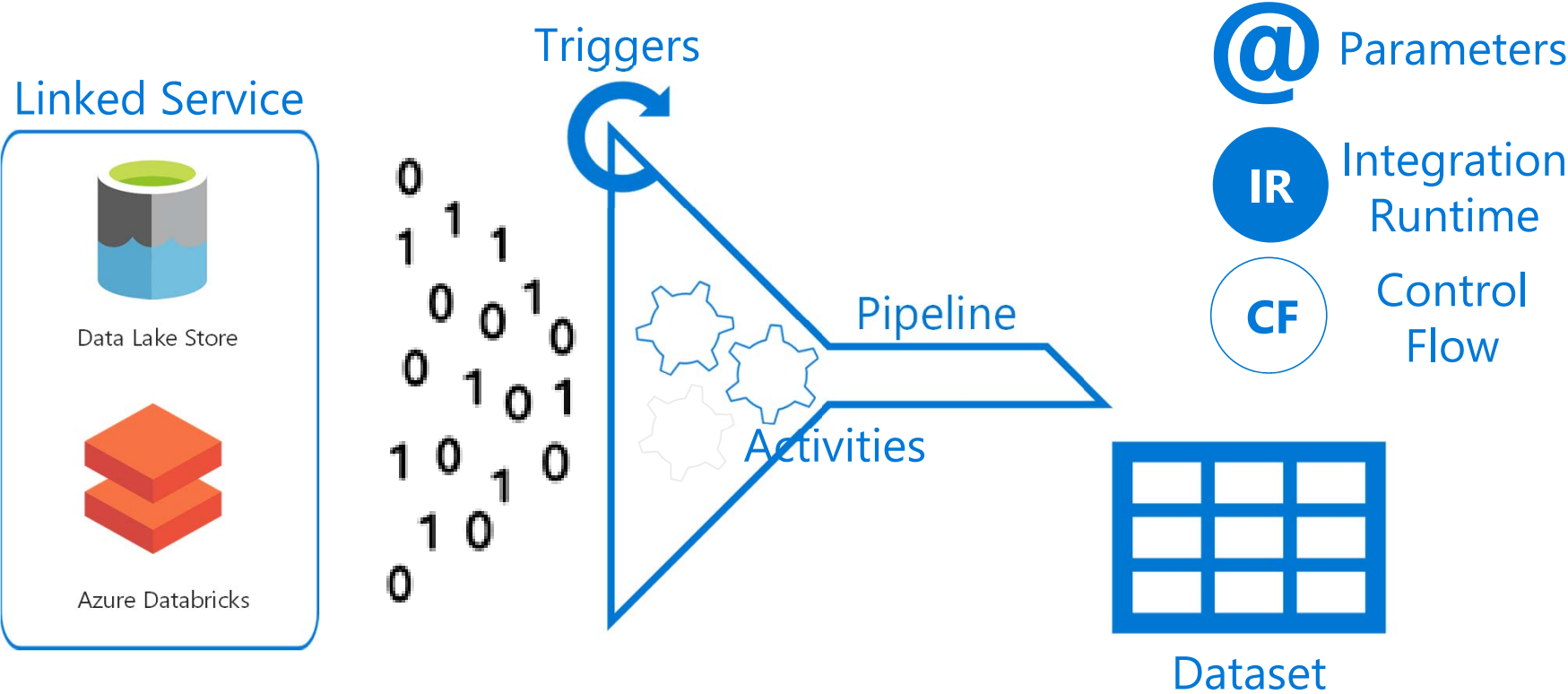


Lesson 02 – Code-free transformation at scale with Azure Data Factory or Azure Synapse Pipelines

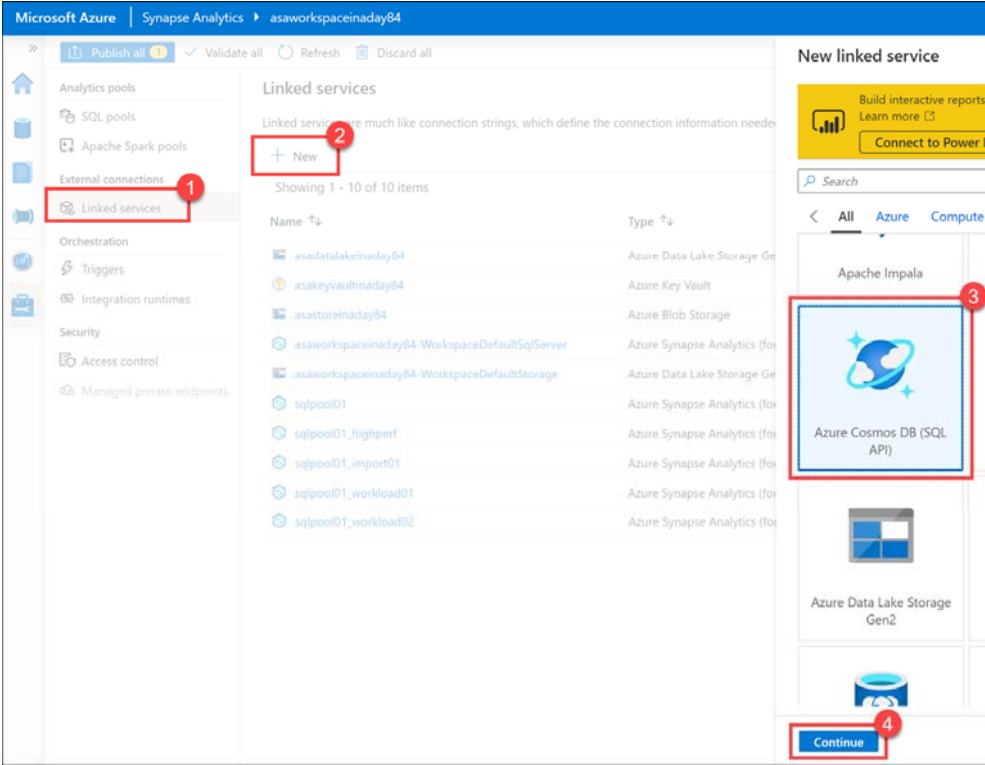
Lesson 01: Data integration with Azure Data Factory or Azure Synapse Pipelines



Data integration with Azure Data Factory or Azure Synapse Pipelines



Linked Services



New linked service (Azure Cosmos DB (SQL API))

Choose a name for your linked service. This name cannot be updated later.

Name *

asacosmosdb01

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Connection string

Azure Key Vault

Account selection method

From Azure subscription

Enter manually

Azure subscription

Select all

Azure Cosmos DB account name *

asacosmosdbinaday84

Database name *

CustomerProfile

Additional connection properties

+ New

Annotations

Create

Back

Connection successful

Test connection

Cancel

Datasets

New integration dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

Search

All

Azure

Database

File

Generic protocol

NoSQL

Services and apps

Apache Impala

Azure Cosmos DB (SQL API)

Azure Data Lake Storage Gen2

Azure Database for MariaDB

Azure Data Explorer (Kusto)

Azure Data Lake Storage Gen1

Azure Database for MySQL

Azure Blob Storage

Continue

Cancel

Set properties

Choose a name for your dataset. This name can be updated at any time until it is published.

Name

asal400_customerprofile_cosmosdb

Linked service *

asacosmosdb01

Connect via integration runtime *

AutoResolveIntegrationRuntime

Collection

OnlineUserProfile01

Edit

Import schema

From connection/store

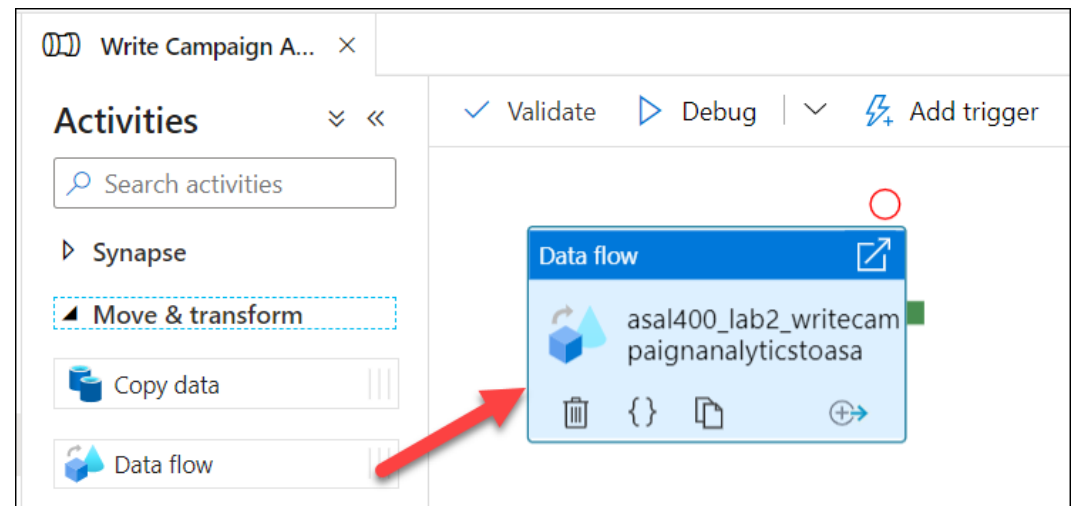
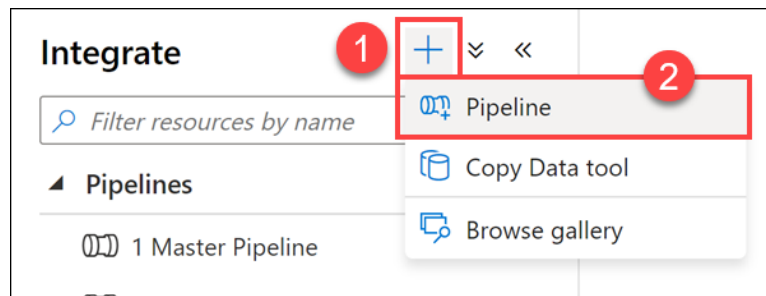
None

OK

Back

Cancel

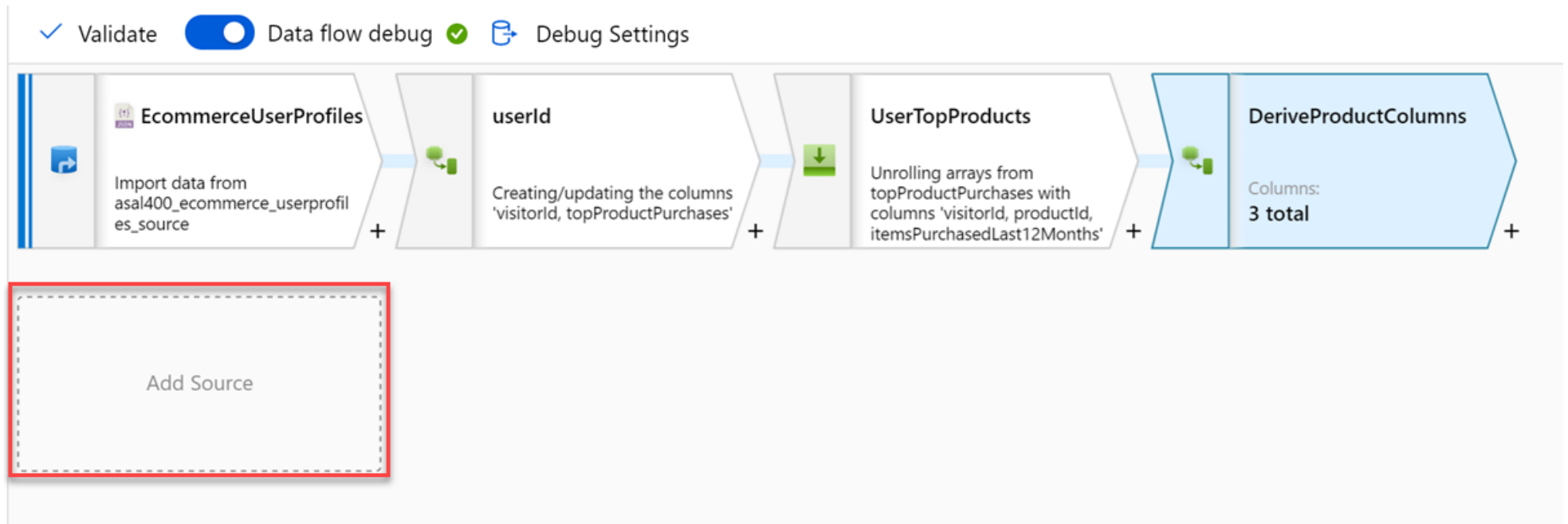
Activities and pipelines



Lesson 01: Code-free transformation at scale with Azure Data Factory or Azure Synapse Pipelines



Code-free transformation at scale with Azure Data Factory or Azure Synapse Pipelines



Adding a source

✓ Validate

Source settings Source options Projection Optimize Inspect Data preview

Output stream name * [Learn more](#)

Source type *

☒ Dataset ☐ Inline

Dataset * [Test connection](#)

Options

☒ Allow schema drift ⓘ

☐ Infer drifted column types ⓘ

☐ Validate schema ⓘ

Sampling * ⓘ ☐ Enable ☒ Disable

Schema modifier transformations

Derived column's settings

Optimize

Inspect

Data preview

Output stream name *

ConvertColumnTypesAndValues


Learn more [↗](#)


Incoming stream *





MapCampaignAnalytics ▼

Columns * ⓘ

+ Add

 Duplicate

 Delete

<input type="checkbox"/>	Column	Expression	
<input type="checkbox"/>	Revenue ▼	toDecimal(replace(concat(toString(RevenuePart1), t... <i>e^x</i>	<div><div>+</div><div></div><div></div></div>
<input type="checkbox"/>	RevenueTarget ▼	toDecimal(replace(concat(toString(RevenueTargetP... <i>e^x</i>	<div><div>+</div><div></div><div></div></div>

Formatter transformations

Flatten settings Optimize Inspect Data preview Description

Output stream name *

Input stream *

Unroll by * ⓘ

Unroll root ⓘ

Options

☐ Skip duplicate input columns ⓘ

☐ Skip duplicate output columns ⓘ

Input columns *

Reset

+ Add mapping

Delete

3 mappings: All inputs mapped

☐

userId's column

123 visitorId

abc topProductPurchases.productId

abc topProductPurchases.itemsPurchasedLast...

Name as

visitorId

productId

itemsPurchasedLast12Months

+

+

+

Multiple inputs/outputs transformations

Join settings

Optimize

Inspect

Data preview ●

Description

Output stream name *

JoinTopProductsWithPreferredProducts

[Learn more](#)


Left stream *


DeriveProductColumns


Right stream *


UserPreferredProducts


Join type *

Full outer

Inner

Left outer

Right outer

Custom (cross)

Join conditions *

Left: DeriveProductColumns's column

123 visitorId

==

Right: UserPreferredProducts's column

123 userId

+

🗑️

Row modifier transformations

Pr...
d

DerivedColumnsForMer...
Columns:
7 total

Filter settings **Optimize** **Inspect** **Data preview** ●

Output stream name * [Learn more](#)

Incoming stream * ▼

Filter on *

✕ ✓

Alter ROW

Destination

Sink

Sink transformation

SelectCampaignAnalyti...

Sink

Settings

Mapping

Optimize

Inspect

Data preview

Output stream name *

CampaignAnalyticsASA

[Learn more](#)

Incoming stream *

SelectCampaignAnalyticsColumns

Sink type *

Integration dataset

Inline

Cache

Dataset *

asal400_wwi_campaign_analytics_asa

Test connection

Open

New

Options

☒ Allow schema drift

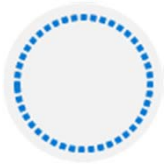
☐ Validate schema

Region

Country

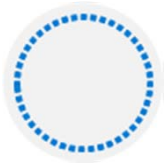
Sink

Review questions



Q01 – You are moving data from an Azure Data Lake Gen2 store to Azure Synapse Analytics. Which Azure integration runtime would be used?

A01 – Azure



Q02 – Which control activity can be used to branch an activity on a condition that evaluates to true or false?

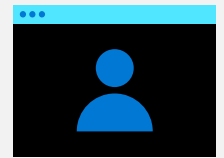
A02 – If Condition activity



Q03 – What feature enables you to interact with the Mapping Data Flow transformations that you create?

A03 – Debug

Lab: Transform data with Azure Data Factory or Azure Synapse Pipelines



Lab overview

This lab teaches students how to build data integration pipelines to ingest data from multiple data sources, transform data using mapping data flows and notebooks, and perform data movement into one or more data sinks.

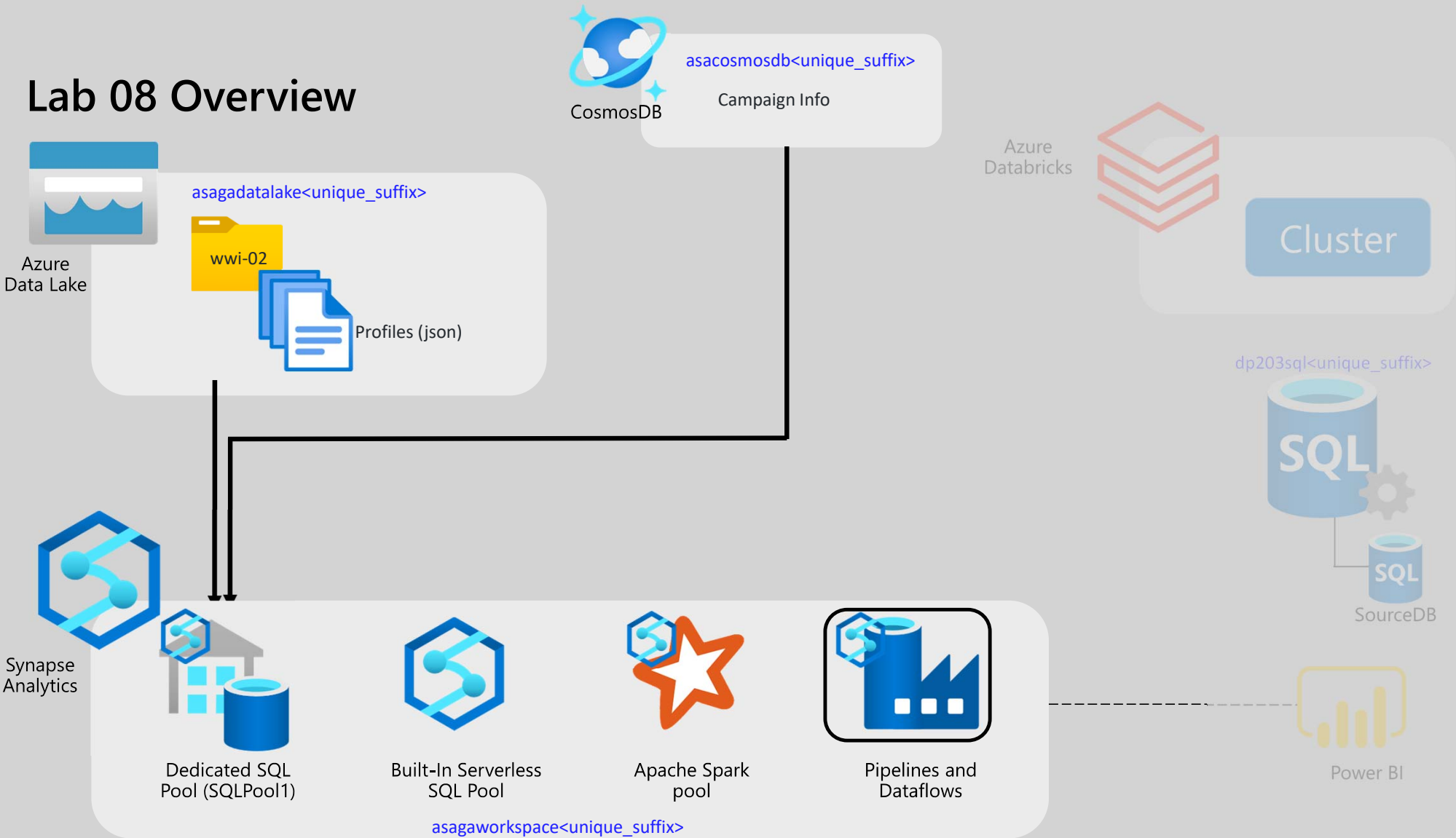
Lab objectives

After completing this lab, you will be able to:

Data integration with Azure Data Factory or Azure Synapse Pipelines

Code-free transformation at scale with Azure Data Factory or Azure Synapse Pipelines

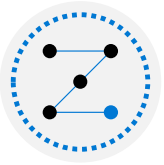
Lab 08 Overview



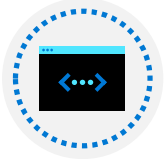
Lab review



Question 1 – What are the first two components do you create in Azure Data Factory/Azure Synapse Pipelines to connect to data in a data source?



Question 2 – Which mapping data flow transformation enables you to remove records during the execution of a data flow?



Question 3 – How can you save and persist the work that you create in Azure Data Factory/Azure Synapse Pipelines?



Question 4 – How can you view detailed information about a data flow?

Module summary

In this module, you have learned about:

Data integration with Azure Data Factory or Azure Synapse Pipelines

Code-free transformation at scale with Azure Data Factory or Azure Synapse Pipelines

Next steps

After the course, consider visiting [[the Microsoft Customer Case Study site](#)]. Use the search bar to search by an industry such as healthcare or retail, or by a technology such as Azure Synapse Analytics or Azure Data Factory. Read through some of the customers stories

