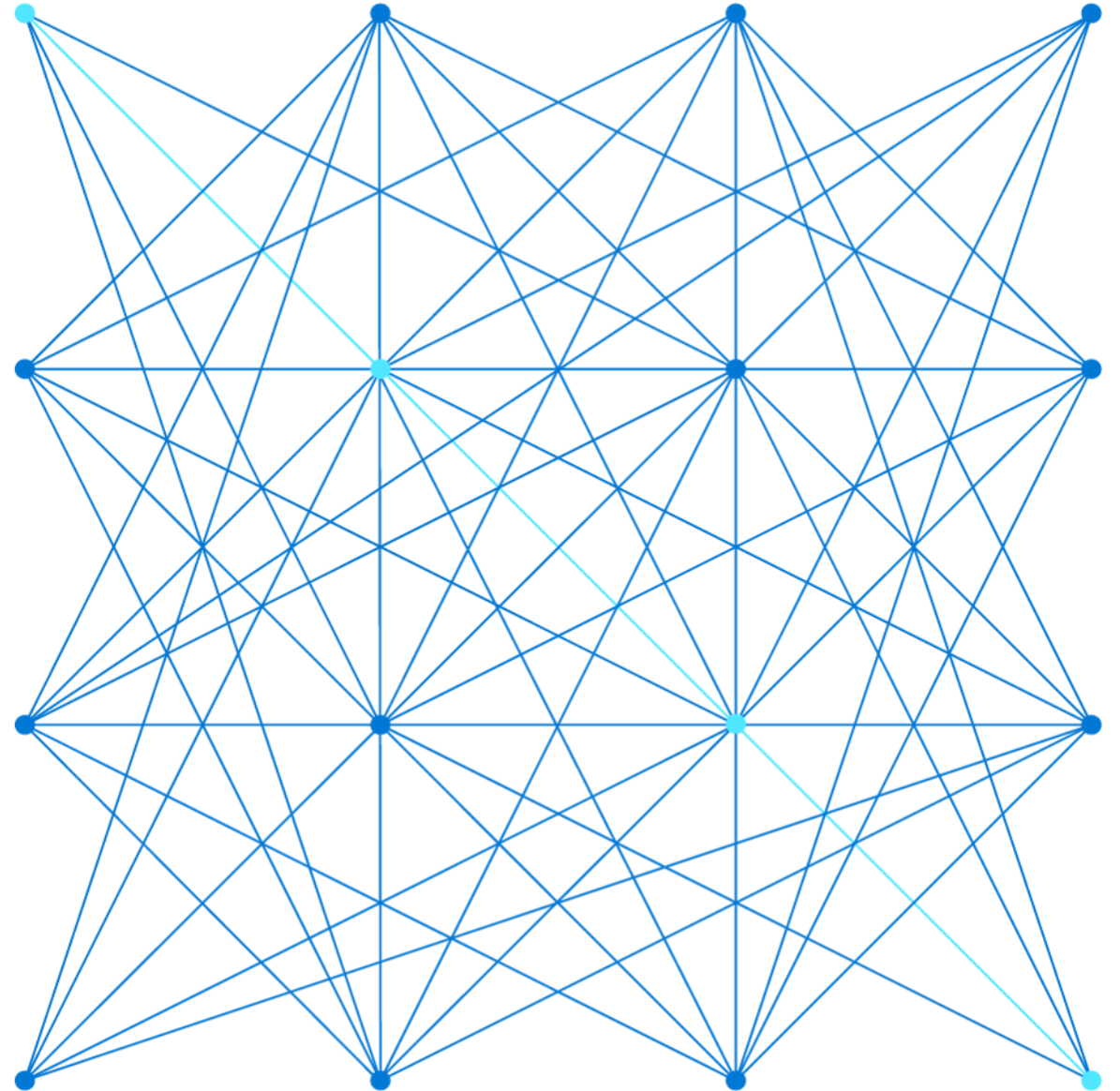
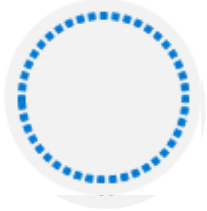


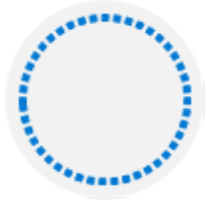
DP-203T00: Transform data with Azure Data Factory or Azure Synapse Pipelines



Agenda



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

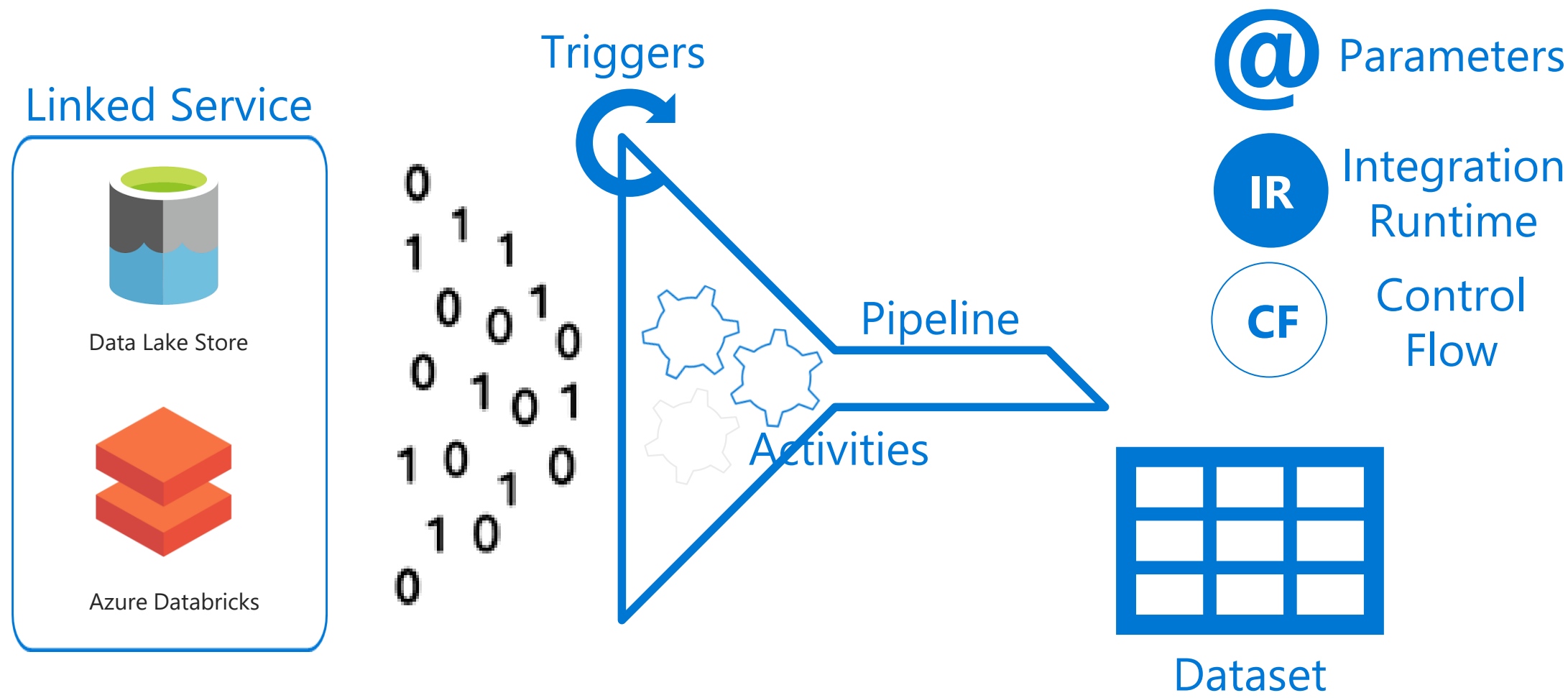


Lesson 02 – Code-free transformation at scale with Azure Data Factory or Azure Synapse Pipelines

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



Data integration with Azure Data Factory or Azure Synapse Pipelines



Linked Services

Microsoft Azure | Synapse Analytics | asaworkspaceinaday84

Analytics pools

- SQL pools
- Apache Spark pools
- External connections
- Linked services**
- Orchestration
- Triggers
- Integration runtimes
- Security
- Access control
- Managed private endpoints

Linked services

Linked services are much like connection strings, which define the connection information needed to connect to external data sources.

+ New

Showing 1 - 10 of 10 items

Name ↑↓	Type ↑↓
asdatalakeinaday84	Azure Data Lake Storage Gen2
asakeyvaultinaday84	Azure Key Vault
asastoreinaday84	Azure Blob Storage
asaworkspaceinaday84-WorkspaceDefaultSqlServer	Azure Synapse Analytics (for SQL)
asaworkspaceinaday84-WorkspaceDefaultStorage	Azure Data Lake Storage Gen2
sqlpool01	Azure Synapse Analytics (for SQL)
sqlpool01_highperf	Azure Synapse Analytics (for SQL)
sqlpool01_import01	Azure Synapse Analytics (for SQL)
sqlpool01_workload01	Azure Synapse Analytics (for SQL)
sqlpool01_workload02	Azure Synapse Analytics (for SQL)

New linked service

Build interactive reports
Learn more

Connect to Power BI

Search

All Azure Compute

Apache Impala

Azure Cosmos DB (SQL API)

Azure Data Lake Storage Gen2

Continue

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
1 ..

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


AllAzureDatabaseFileGeneric protocolNoSQLServices and apps




Azure Cosmos DB (SQL API)




Azure Data Lake Storage Gen2




Azure Database for MariaDB




Azure Blob Storage




Azure Data Explorer (Kusto)




Azure Database for MySQL



Azure Cosmos DB (MongoDB API)



Azure Data Lake Storage Gen1



Azure Data Explorer (Kusto)

Continue

Cancel

Set properties

i 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

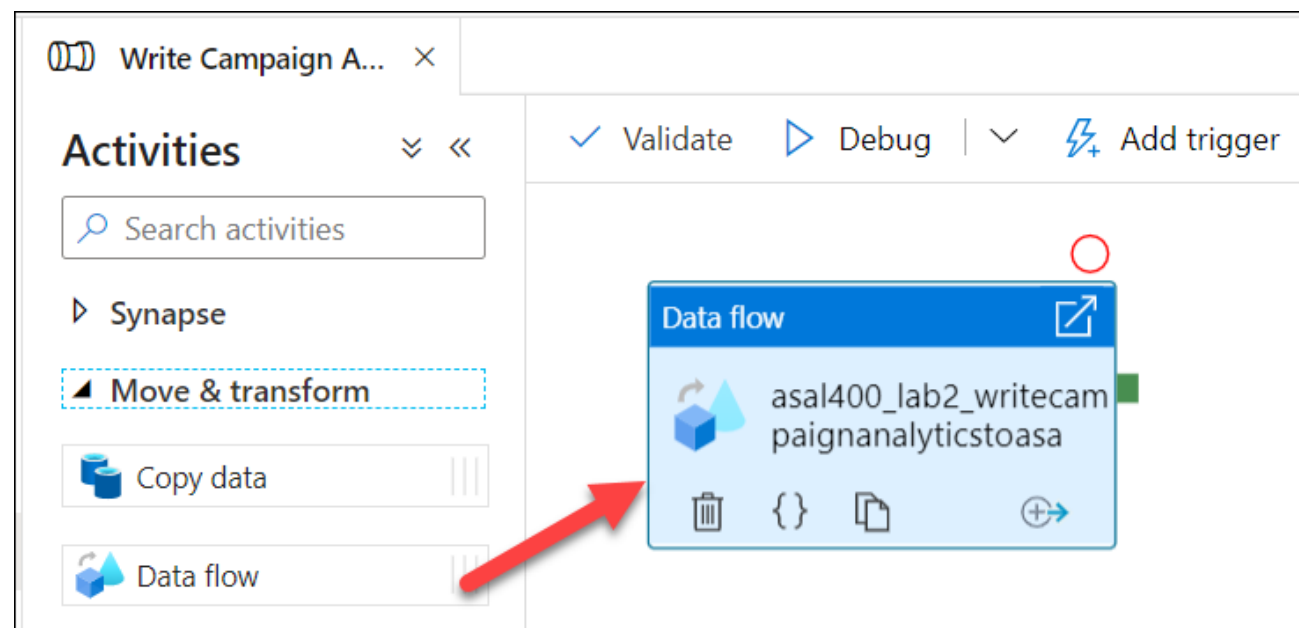
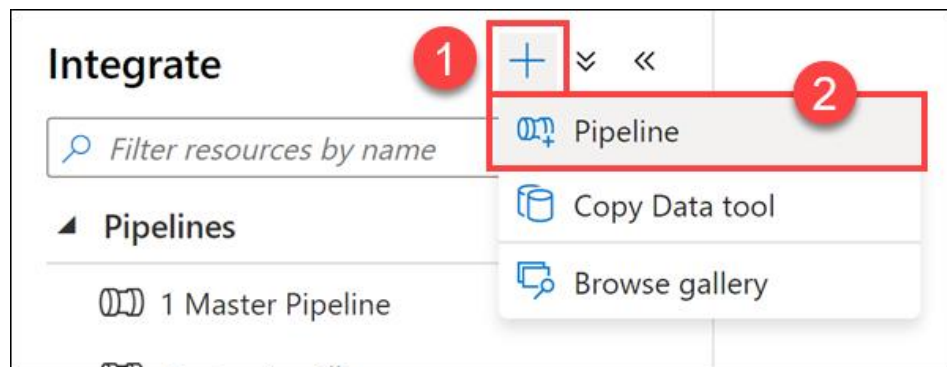
☐ None

OK

Back

Cancel

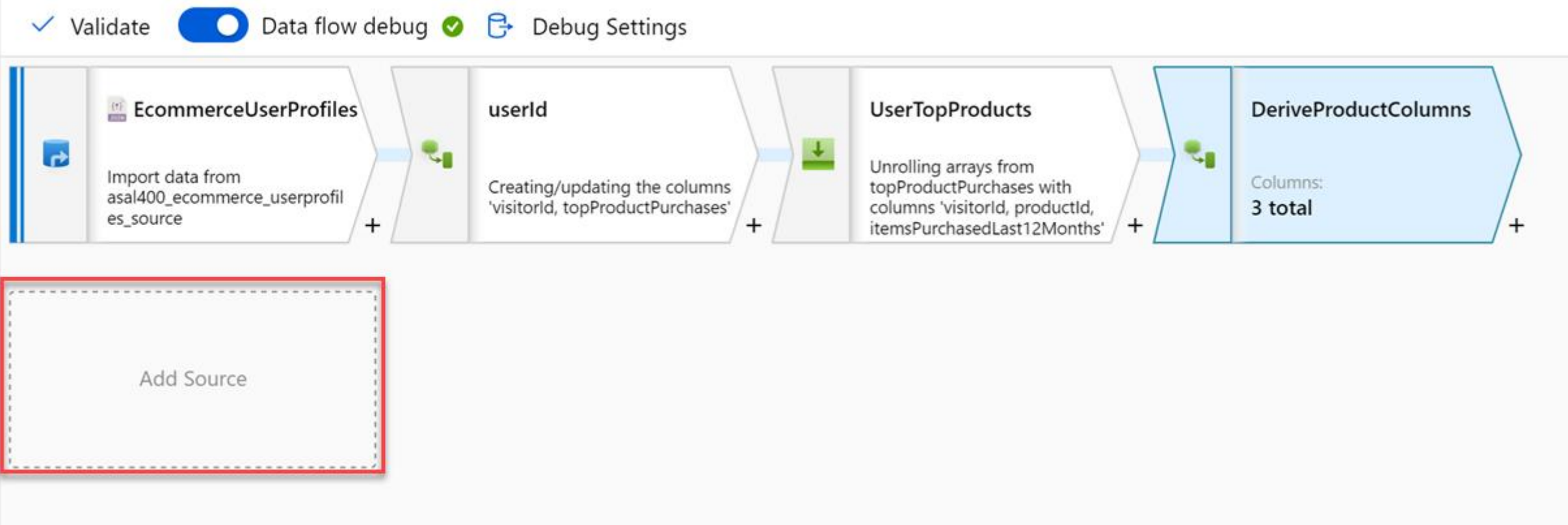
Activities and pipelines



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... e ^x	+ 📄 🗑
<input type="checkbox"/>	RevenueTarget	toDecimal(replace(concat(toString(RevenueTargetP... e ^x	+ 📄 🗑

Formatter transformations

Flatten settings

Optimize

Inspect

Data preview

Description

Output stream name *

UserTopProducts

?

Help

Learn more

Incoming stream *

userId

Unroll by * ⓘ

[] topProductPurchases

Unroll root ⓘ

Options

☐ Skip duplicate input columns ⓘ

☐ Skip duplicate output columns ⓘ

Input columns *

Reset

+ Add mapping

Delete

3 mappings: All inputs mapped

<input type="checkbox"/>	userId's column		Name as	
<input type="checkbox"/>	123 visitorId	→	visitorId	+ <input type="checkbox"/>
<input type="checkbox"/>	abc topProductPurchases.productId	→	productId	+ <input type="checkbox"/>
<input type="checkbox"/>	abc topProductPurchases.itemsPurchasedLast...	→	itemsPurchasedLast12Months	+ <input type="checkbox"/>

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

The screenshot displays a data transformation tool interface. On the left, a pipeline diagram shows a transformation named "DerivedColumnsForMer..." with a green icon and a note "Columns: 7 total". A red box highlights the "Filter settings" dialog box, which has tabs for "Filter settings", "Optimize", "Inspect", and "Data preview". The "Filter settings" tab is active. It contains the following fields:

- Output stream name ***: A text input field containing "Filter1". To its right is a "Learn more" link with an external icon.
- Incoming stream ***: A dropdown menu showing "DerivedColumnsForMerge" with a downward arrow.
- Filter on ***: A text area containing the expression `!isNull(productId)`. The text area is highlighted with a red border. To its right is a small icon with an 'x' and a checkmark.

Below the dialog box, a "Destination" dropdown menu is visible, showing "Sink" as the selected option.

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

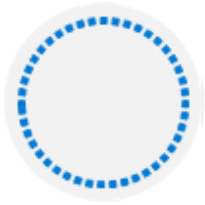
☐ Validate schema

Region

Country

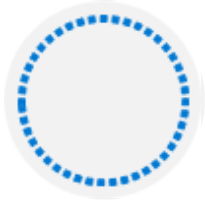
Sink

Review questions



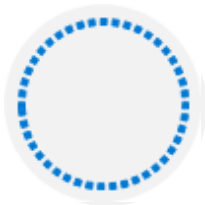
Q01 – You are moving data from an Azure Data Lake Gen2 store to Azure Synapse Analytics. Which Azure integration runtime would be used?

A01 – Azure



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

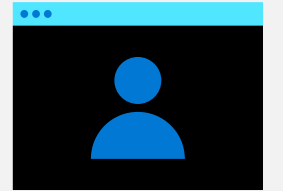
A02 – If Condition activity



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:

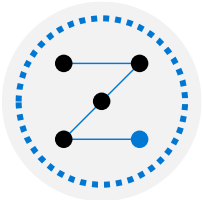
Data integration with Azure Data Factory or Azure Synapse Pipelines

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

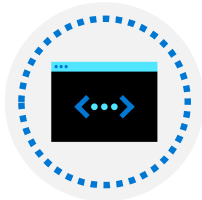
Lab review



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

