



# Microsoft Ignite The Tour

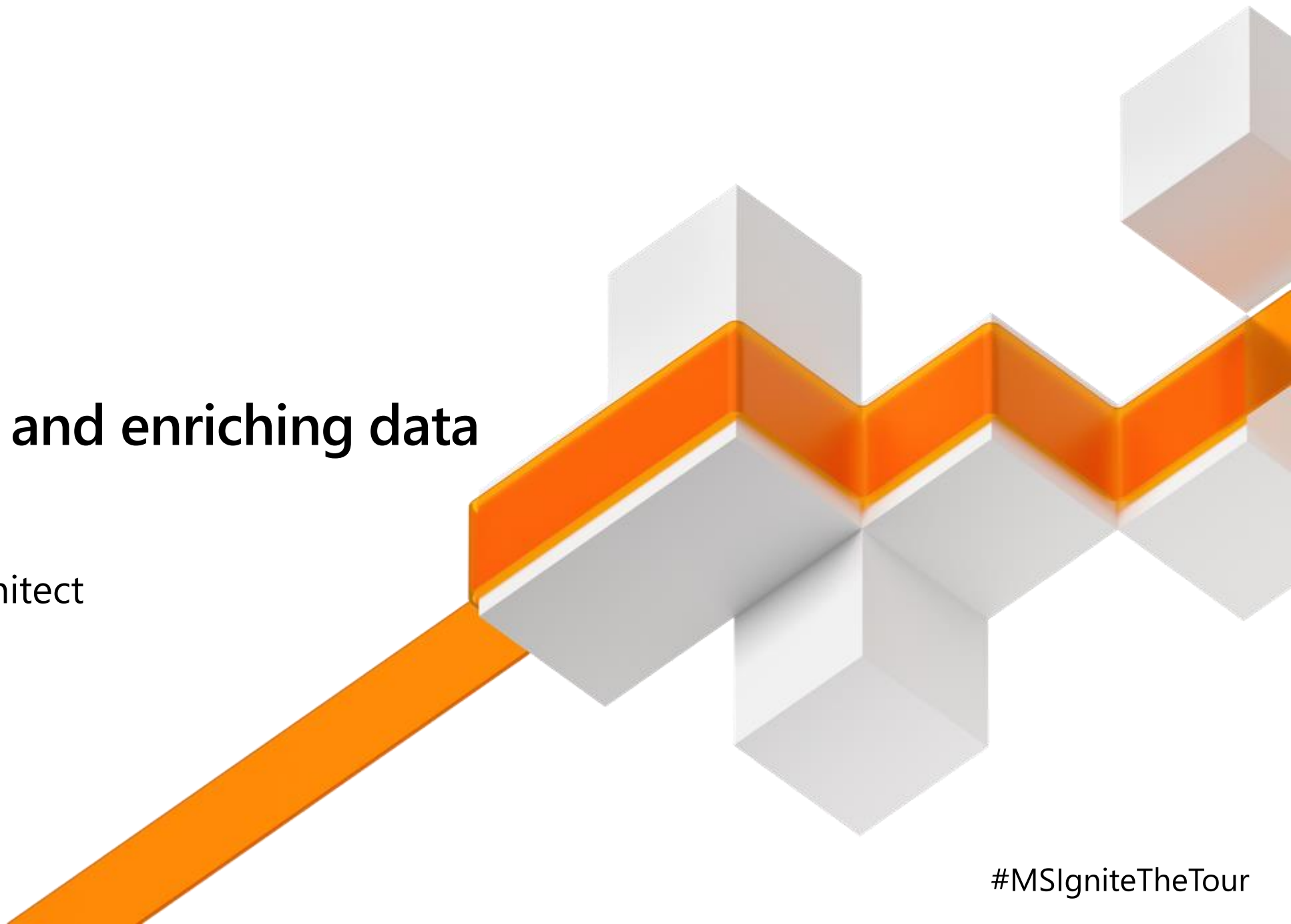
[aka.ms/DATA30](https://aka.ms/DATA30)

#MSIgniteTheTour

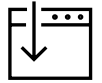


# Transforming and enriching data

René Bremer  
Cloud Solution Architect

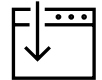


# Resources



Session Resources Hub

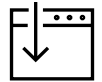
[aka.ms/DATA30](https://aka.ms/DATA30)



Session Code on GitHub

[aka.ms/DATA30Repo](https://aka.ms/DATA30Repo)

[github.com/rebremer/demo30-ignite-mdw-git](https://github.com/rebremer/demo30-ignite-mdw-git)



All Event Session Resources

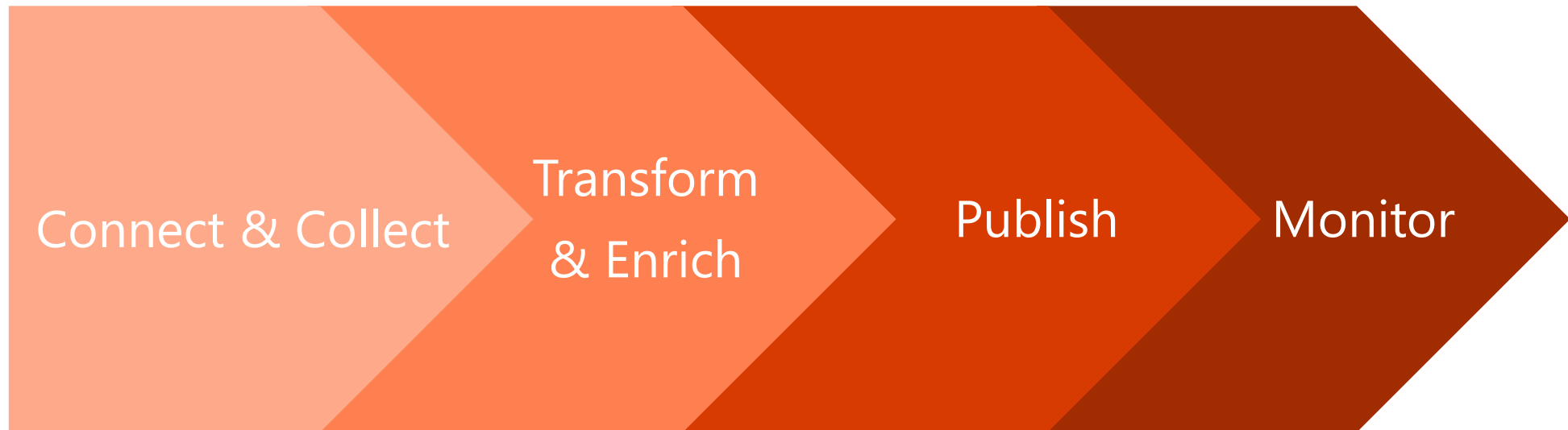
[aka.ms/mysignitethetour](https://aka.ms/mysignitethetour)

# What is Azure Data Factory?

# Azure Data Factory

*A cloud-based data integration service that allows you to orchestrate and automate data movement and data transformation.*

# Azure Data Factory process



# Azure Data Factory Components

## Linked Service

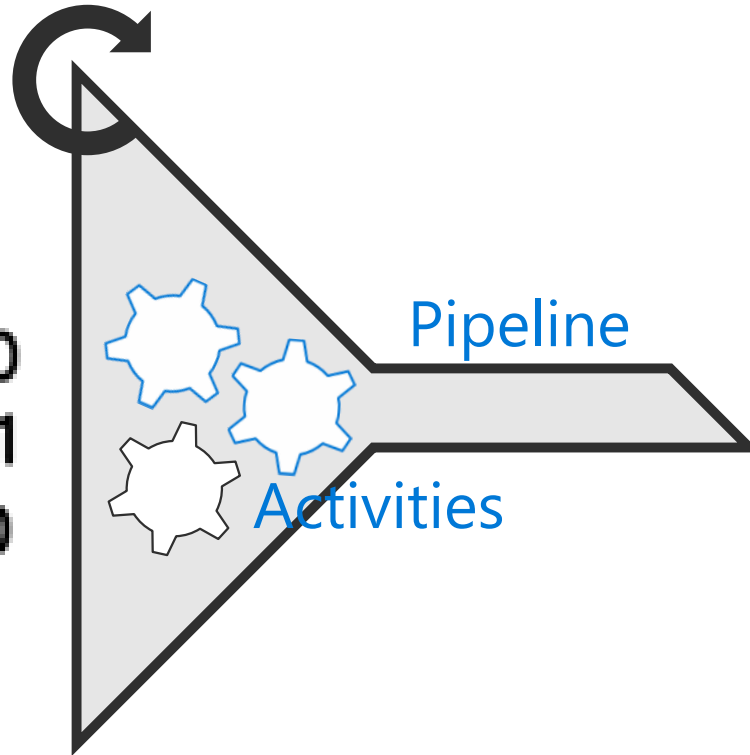
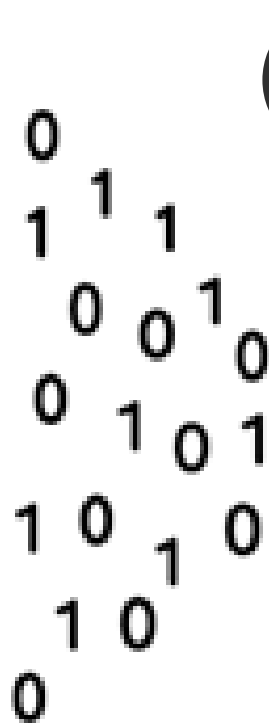


Data Lake Store



Azure Databricks

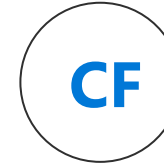
## Triggers



Parameters



Integration Runtime

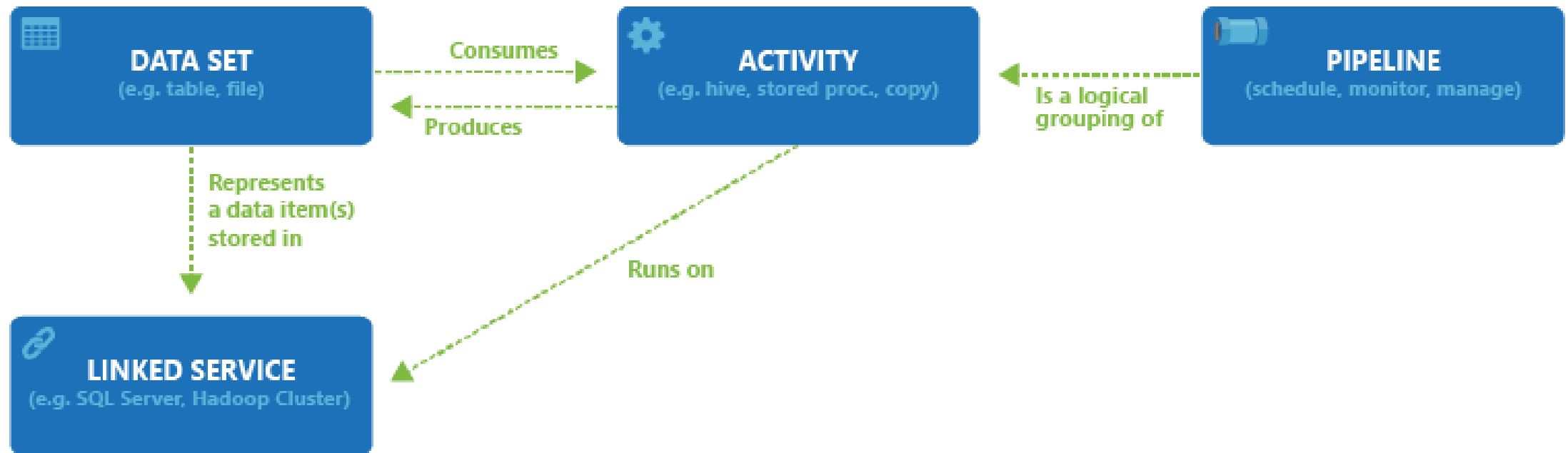


Control Flow



Dataset

# Component dependencies

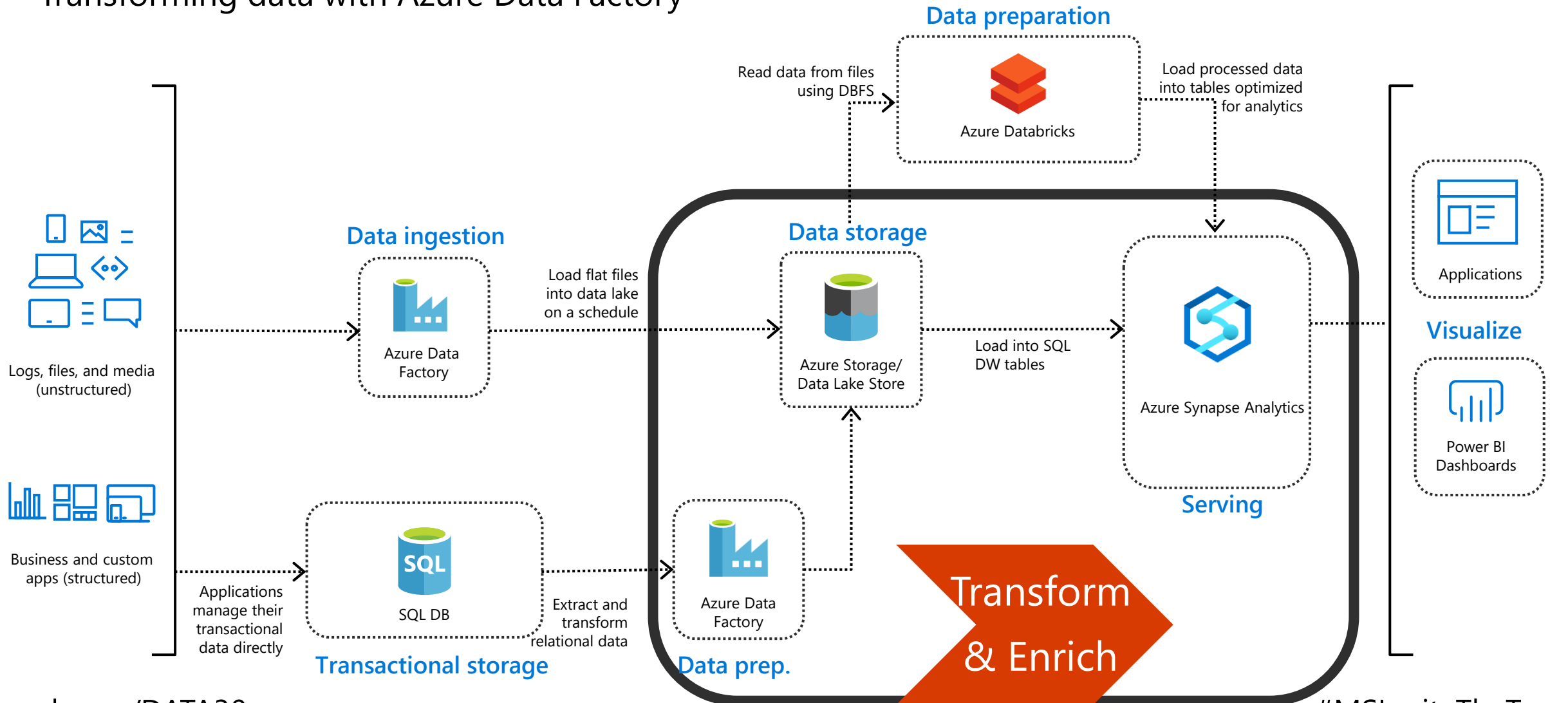




# Transforming data with the ADF Mapping Data Flow

# Data transformation in Azure

## Transforming data with Azure Data Factory



# Methods for transforming in Azure Data Factory

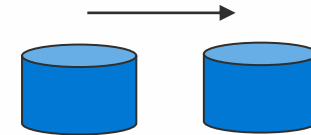
Compute  
resources



SSIS Packages



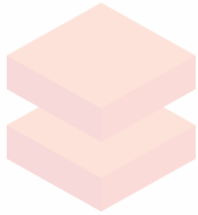
Mapping Data  
Flow



# Methods for transforming data in Azure Data Factory

Code free data transformation at scale

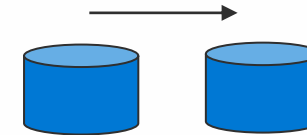
Compute  
resources



SSIS Packages



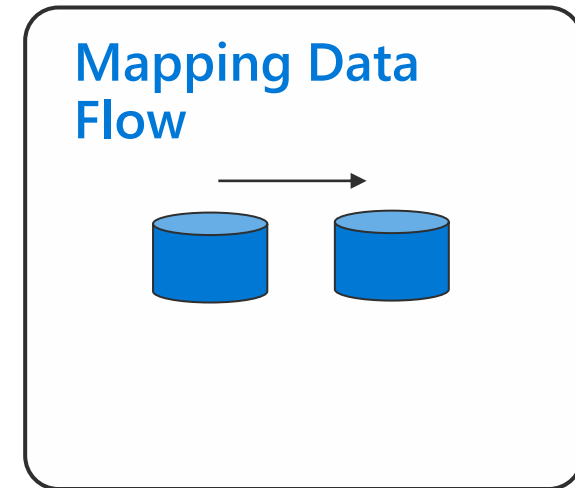
Mapping Data  
Flow



# Benefits of Mapping Data Flow

Code free data transformation at scale

- Perform data cleansing, transformation, aggregations, etc.
- Enables you to build resilient data flows in a code free environment
- Enable you to focus on building business logic and data transformation
- Underlying infrastructure is provisioned automatically with cloud scale via Spark execution



# Using the Mapping Data Flow

## Code free data transformation at scale

The screenshot displays the Azure Data Factory Mapping Data Flow editor. The interface is divided into several sections:

- Left Panel (Factory Resources):** A sidebar showing a hierarchy of resources. Under "Data Flows (Preview)", the flow named "dataflow3" is selected and highlighted in blue.
- Top Bar:** A horizontal bar at the top of the editor containing tabs for "MoviesAnalytics", "DataFlowDemo", "DataFlowDemo", "dataflow2", and "dataflow3". Below the tabs are buttons for "Save", "Validate", and "Debug Settings". A red label "top bar" points to this area.
- Graph:** The central workspace where the data flow is visualized. It contains a source node labeled "source1" with the description "Import data from MoviesDB". Below it is a dashed box labeled "Add Source". A red label "graph" points to this central area.
- Configuration Panel:** A panel at the bottom of the editor with two tabs: "General" and "Parameters". The "General" tab is active, showing fields for "Name" (set to "dataflow3") and "Description". A red label "configuration panel" points to this bottom section.

# Starting the Mapping Data Flow

Code free data transformation at scale

The screenshot shows the 'Adding Data Flow' dialog box in Azure Data Factory. The dialog has a title bar with a close button (X). Below the title bar, there are tabs for 'source1', 'pipeline8', 'pipeline9', and 'DataflowDemo'. The 'source1' tab is active, showing a 'source1' component with 'Columns: 0 total'. Below this, there is a dashed box labeled 'Add Source'. To the right of the 'source1' component, there are buttons for '+', '-', and a search icon. Below the 'source1' component, there are tabs for 'Source Settings', 'Define schema', 'Optimize', 'Inspect', and 'Data Preview'. The 'Source Settings' tab is active, showing the following settings:

- Output stream name \*: source1
- Source Dataset \*: USDOutput (with 'Edit' and '+ New' buttons)
- Options: ☒ Allow schema drift (with an information icon)
- Sampling \*: ☐ Enable ☒ Disable (with an information icon)

At the bottom of the dialog, there are 'Cancel' and 'Finish' buttons.

# Transformation options in the Mapping Data Flow

Unpivot Union Join  
Lookup Window  
Derived Column  
Sink Alter Row New Branch  
aggregate Pivot Filter  
Conditional Split Sort  
Exists Select  
Surrogate Key Source





# ADF Enterprise Ready Features

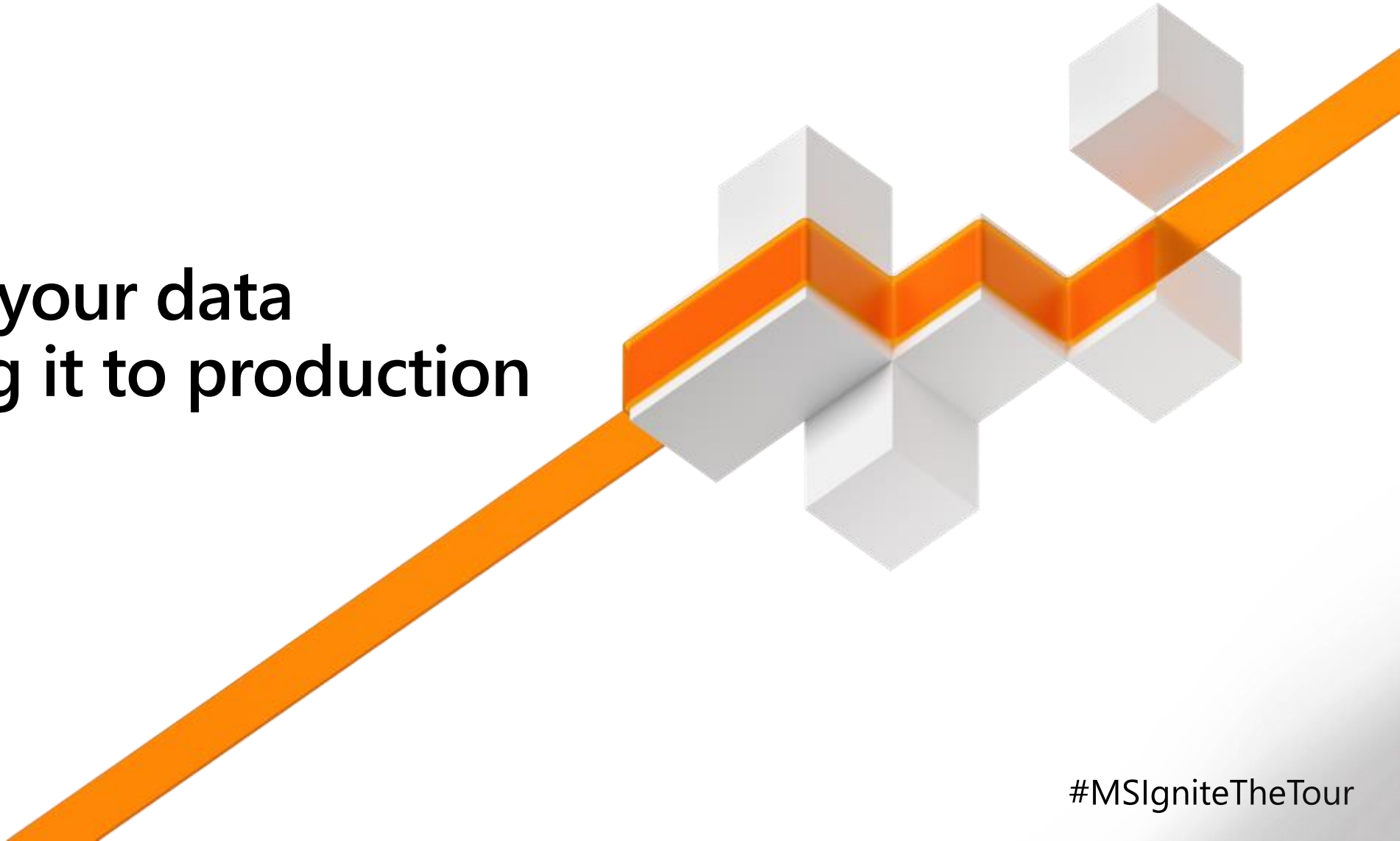
# ADF Enterprise Ready Features

- Full integration with GIT and Azure DevOps
- Self-hosted IR for on-prem connectivity
- Monitoring plane, integrated with Azure Monitor
- Usage of Managed Identities, no need for keys

=> Essential to bring your ADF pipeline to production

# Demo: Transforming your data in ADF & bring it to production

René Bremer



A background image showing two women in a meeting room. One woman is pointing at a whiteboard, and the other is looking at it. The whiteboard has some papers and diagrams on it.

# In Summary:

## Transforming Data with Azure Data Factory

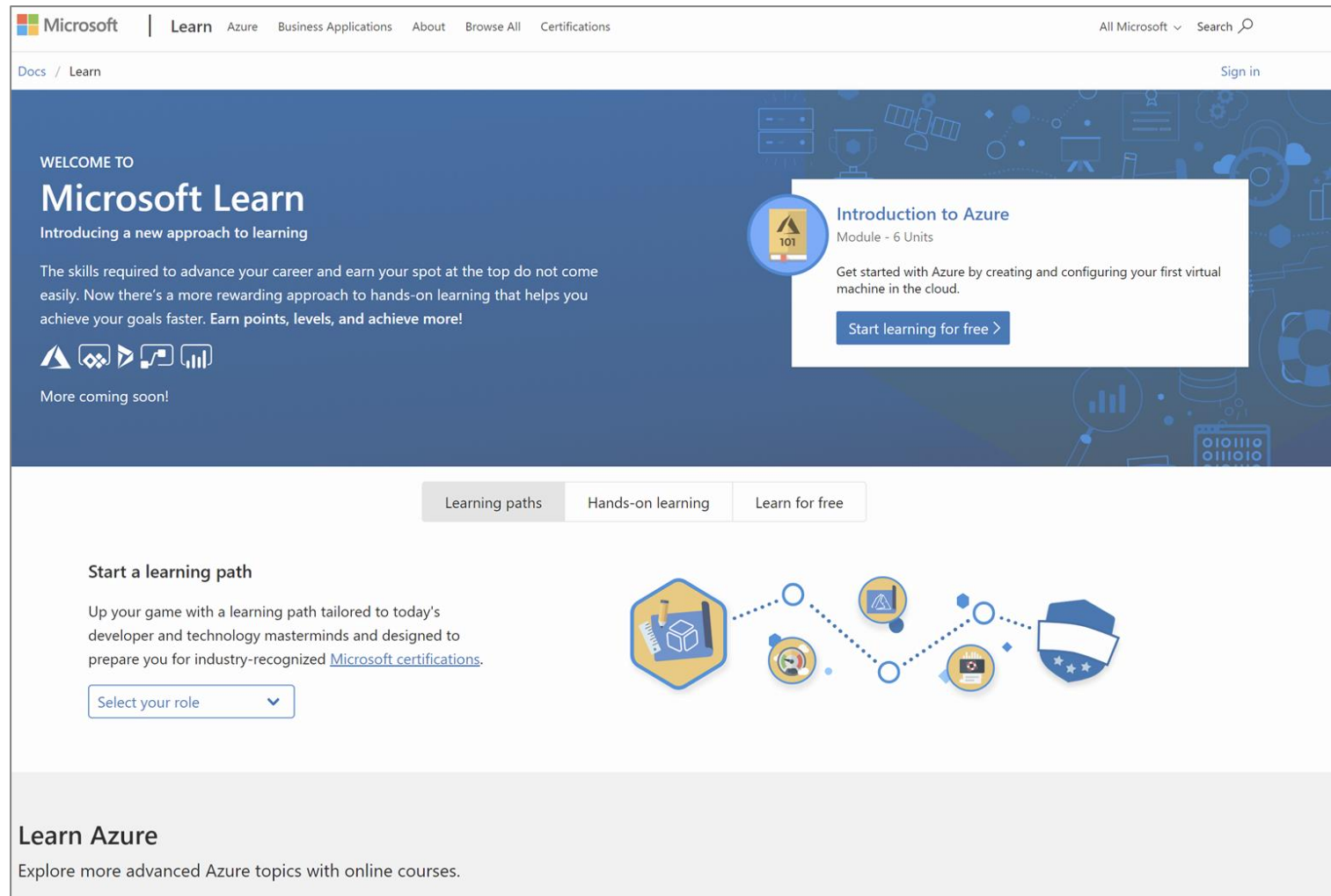
- Azure Data Factory (ADF) is a cloud-based data integration service that allows you to orchestrate and automate data movement and data transformation.
- Transforming data can be performed in ADF by orchestrating a compute resource, calling an SSIS package or using the Mapping Data Flow feature
- The Mapping Data Flow feature enables code free data transformation at scale
- Enable you to focus on building business logic and data transformation
- Enterprise Ready Features such as full Git and Azure DevOps integration, self-hosted integration runtime, Azure Monitor and Managed Identities

#MSIgniteTheTour

# /MS Learn alert

Complete interactive learning exercises, watch videos, and practice and apply your new skills.

[aka.ms/DATA30MSLearnCollection](https://aka.ms/DATA30MSLearnCollection)



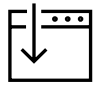
The screenshot shows the Microsoft Learn homepage. At the top is the Microsoft logo and navigation links: Learn, Azure, Business Applications, About, Browse All, and Certifications. On the right, there are links for 'All Microsoft' and a search icon, and a 'Sign in' link. Below the navigation bar, the main header area has a blue background with the text 'WELCOME TO Microsoft Learn' and 'Introducing a new approach to learning'. It includes a paragraph about the skills required to advance one's career and earn a spot at the top, followed by icons representing various learning methods and the text 'More coming soon!'. On the right side of the header, there is a white box titled 'Introduction to Azure' with a 'Module - 6 Units' and a description: 'Get started with Azure by creating and configuring your first virtual machine in the cloud.' Below this is a 'Start learning for free >' button. Below the header, there are three tabs: 'Learning paths', 'Hands-on learning', and 'Learn for free'. Under the 'Learning paths' tab, there is a section titled 'Start a learning path' with a description: 'Up your game with a learning path tailored to today's developer and technology masterminds and designed to prepare you for industry-recognized [Microsoft certifications](#).' Below this is a 'Select your role' dropdown menu. To the right of the text is a diagram showing a sequence of icons connected by dotted lines, representing a learning path. At the bottom, there is a section titled 'Learn Azure' with the text 'Explore more advanced Azure topics with online courses.'

[aka.ms/DATA30](https://aka.ms/DATA30)

#MSIgniteTheTour

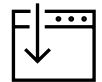


# Resources



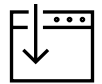
## Session Resources

[aka.ms/DATA30](https://aka.ms/DATA30)



## Session Code on GitHub

[aka.ms/DATA30repo](https://aka.ms/DATA30repo)  
[github.com/rebremer/demo30-ignite-mdw-git](https://github.com/rebremer/demo30-ignite-mdw-git)

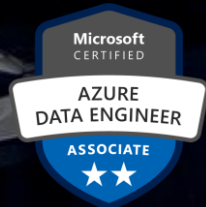


## All Event Resources

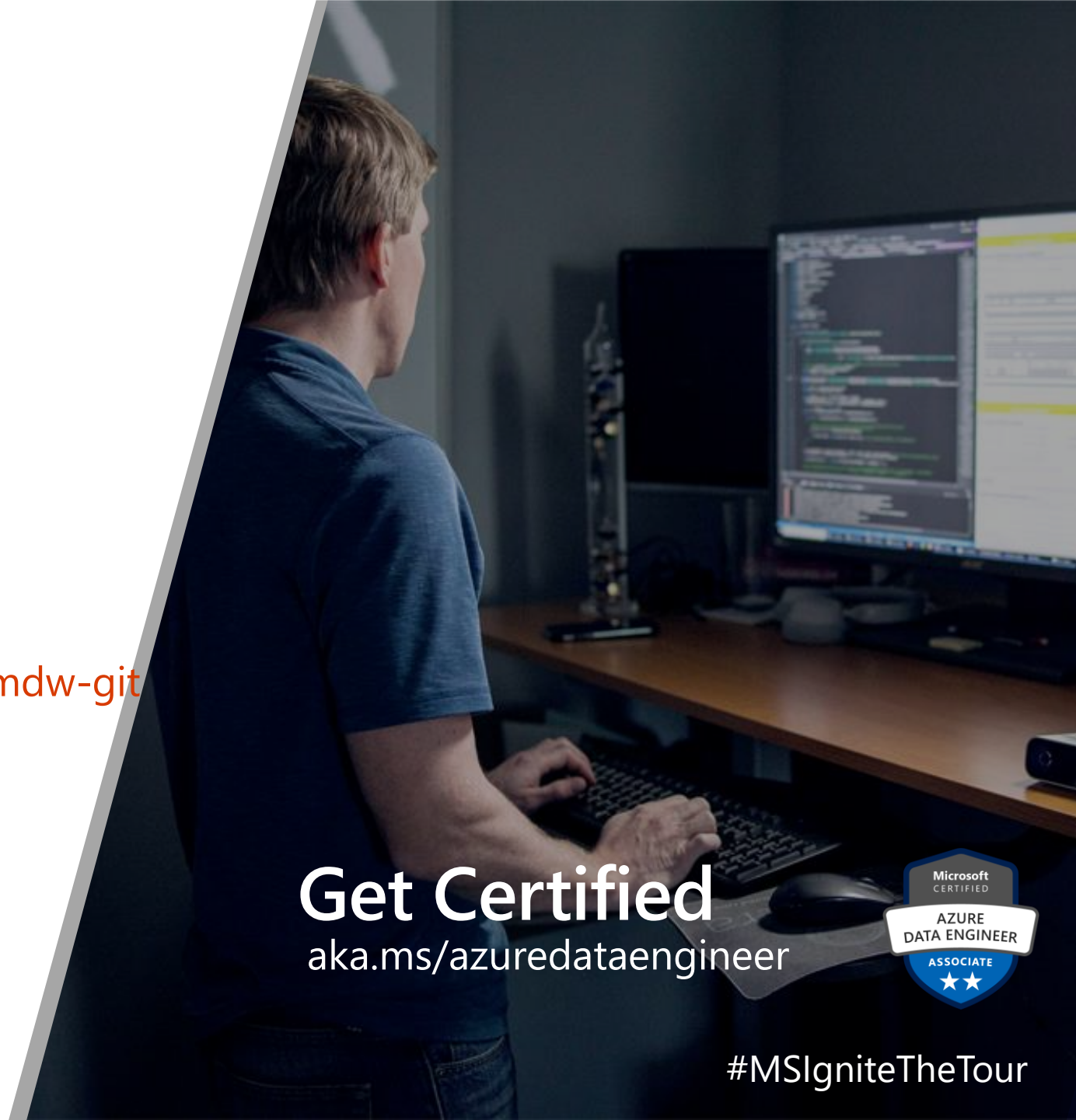
[aka.ms/mysignitethetour](https://aka.ms/mysignitethetour)

[aka.ms/DATA30](https://aka.ms/DATA30)

**Get Certified**  
[aka.ms/azuredataengineer](https://aka.ms/azuredataengineer)



#MSIgniteTheTour





# Transforming and enriching data

René Bremer  
Cloud Solution Architect

[aka.ms/DATA30](https://aka.ms/DATA30)

#MSIgniteTheTour

