

# Building Your First Data Pipeline in Azure Data Factory

---

## UNDERSTANDING AZURE DATA FACTORY AND ITS INTERFACE



**Emilio Melo**

PLURALSIGHT AUTHOR

[www.cloudadvantage.tech](http://www.cloudadvantage.tech)



# Module Overview



**What is Azure Data Factory?**

**Main Data Factory elements**

**Demo: creating a Data Factory**

**Demo: Azure Data Factory environment**



# Why Data Factory?

**Data is the new oil**

**Data insights maintain competitive edge**

**Big data: variety, velocity and volume**

**High demand for data professionals**

**ADF: Microsoft's main data engineering tool**



# Azure Data Factory

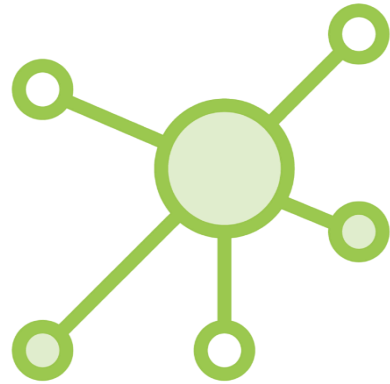
A cloud-based data integration service that allows you to orchestrate and automate data movement and data transformation.



# What Is Data Factory?



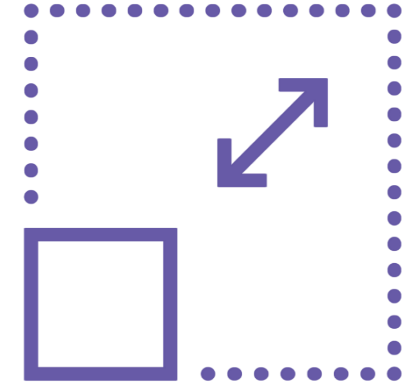
Cloud-based  
PaaS Solution



Data Integration  
Service



Orchestration  
and Automation



Data  
Transformation

# ETL vs. ELT

## ETL

Transform before loading,  
designed for reliability

## ELT

Transform after loading (on  
DW), designed for agility

Azure Data Factory



# Data Factory on Azure Ecosystem

## Migration?

ADF excels on periodic data loads and transformations instead

## Streaming?

ADF can orchestrate, but there are other tools for streaming

## Transformations?

Data Flows for simple ones, external services for complex transforms



# SSIS vs. ADF

## SSIS

More code-free transformations  
On Premises Connectors (e.g. Excel)

## ADF

Much higher scalability  
Cloud and SaaS Connectors  
Event-based Triggers  
Can run SSIS Packages





# Data Factory Considerations

## Two versions

ADF v2 is the current and improved version

## Build options

PowerShell, .NET, Python, REST, ARM

## Highly integrated

DevOps, Key Vault, Monitor, Automation...

## No data storage

Need to persist data by the end

## Security standards

HTTP/TLS whenever possible



# Main Data Factory Elements



## Pipeline

Activity



Activity



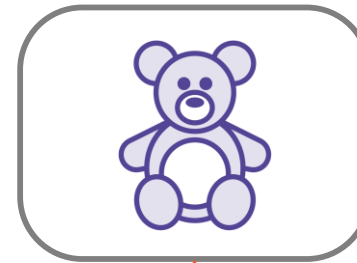
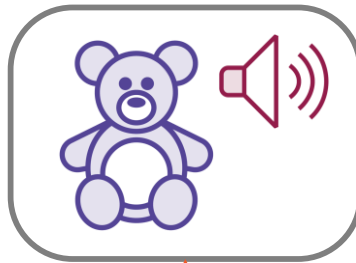
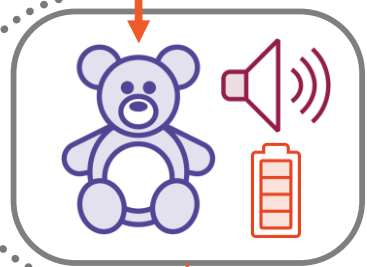
Dataset

Trigger

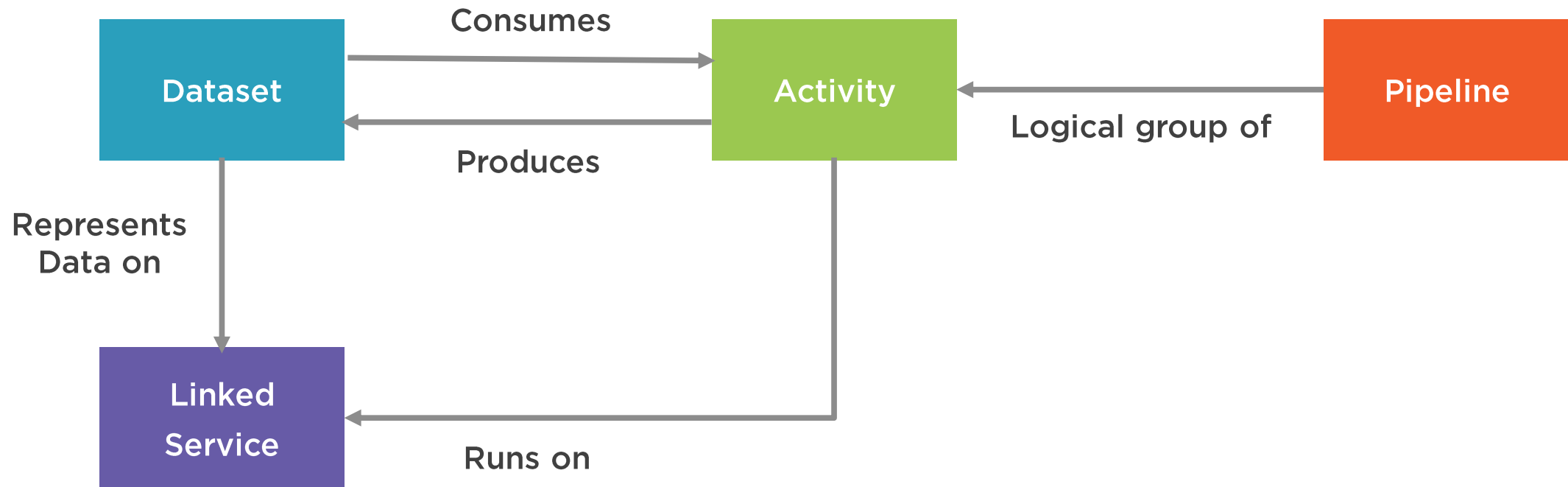


Linked Service

Integration  
Runtime



# Data Factory Elements Relationship



# Microsoft ETL Terminology

## Azure Data Factory

Pipeline

Activity

Linked Services

Datasets

Source/Sink

Json files

## Integration Services

Package

Task

Connection Managers

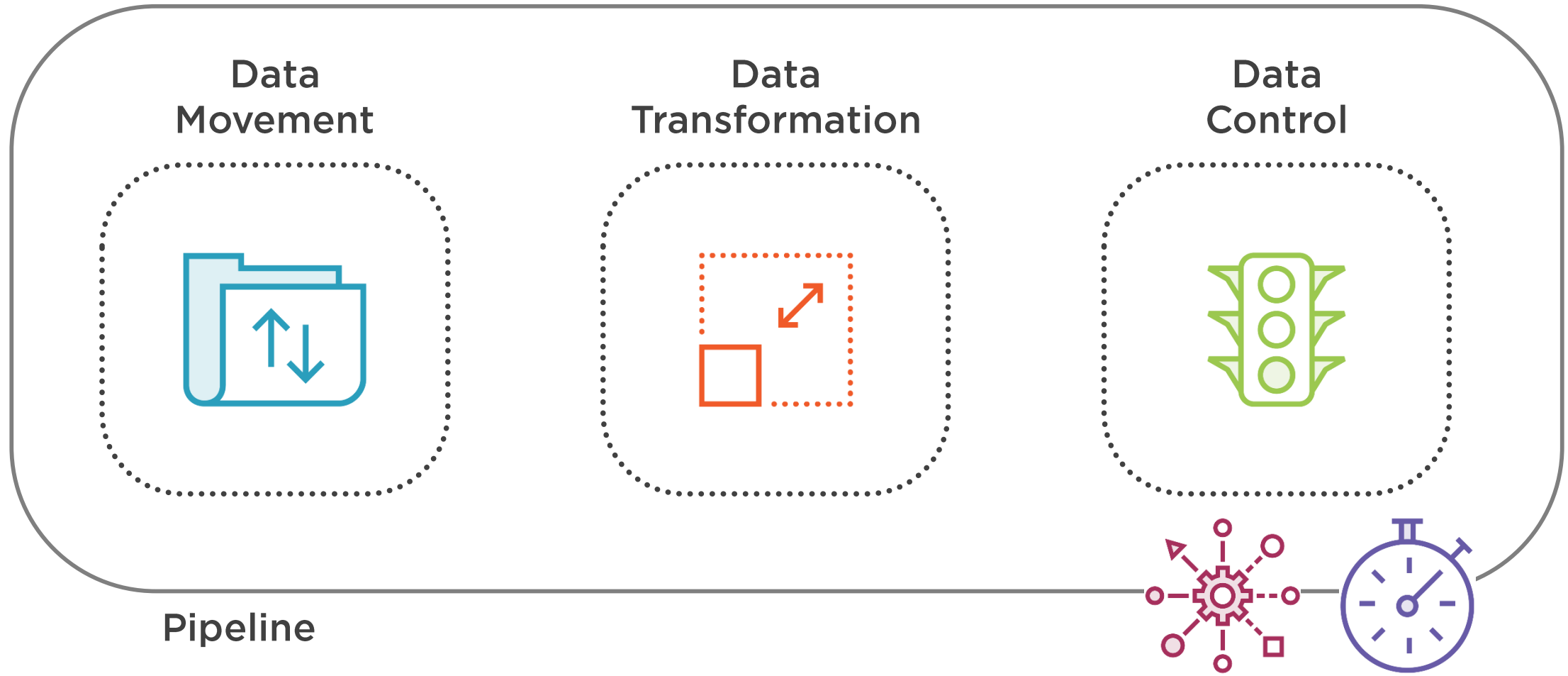
SSIS Sources and Destinations

Source/Destination

DTSX (XML)



# Activities and Pipelines



# Adventure Works Case Study

**Makes bikes and accessories**

**ML experiment on products**

**Based on product:**

- Cost
- Retail price
- Items sold

**SQL to Blob**



# Module Summary



**ADF: Microsoft's Data Engineering Tool**

**Six main elements: Pipelines, Activities, Triggers, IRs, Datasets, and Linked Services**

**ADF resource creation and authoring**

