



**SIGMA**DATA  
Learning

# Azure data Factory

Sigma Data Academy 2021

2021



# Introductie

- Wie ben je?
- Wat is je achtergrond/ervaring ELT?
- Wat verwacht je van de training?



## Training tot nu tot?

- Breakout rooms? werkt dat goed voor jullie?
- Zijn er nog vast momenten dat iemand even weg moet?



# Module 1

- Voorbereiding
- ETL
- Wat is Azure Data Factory (ADF)
- ADF onderdelen



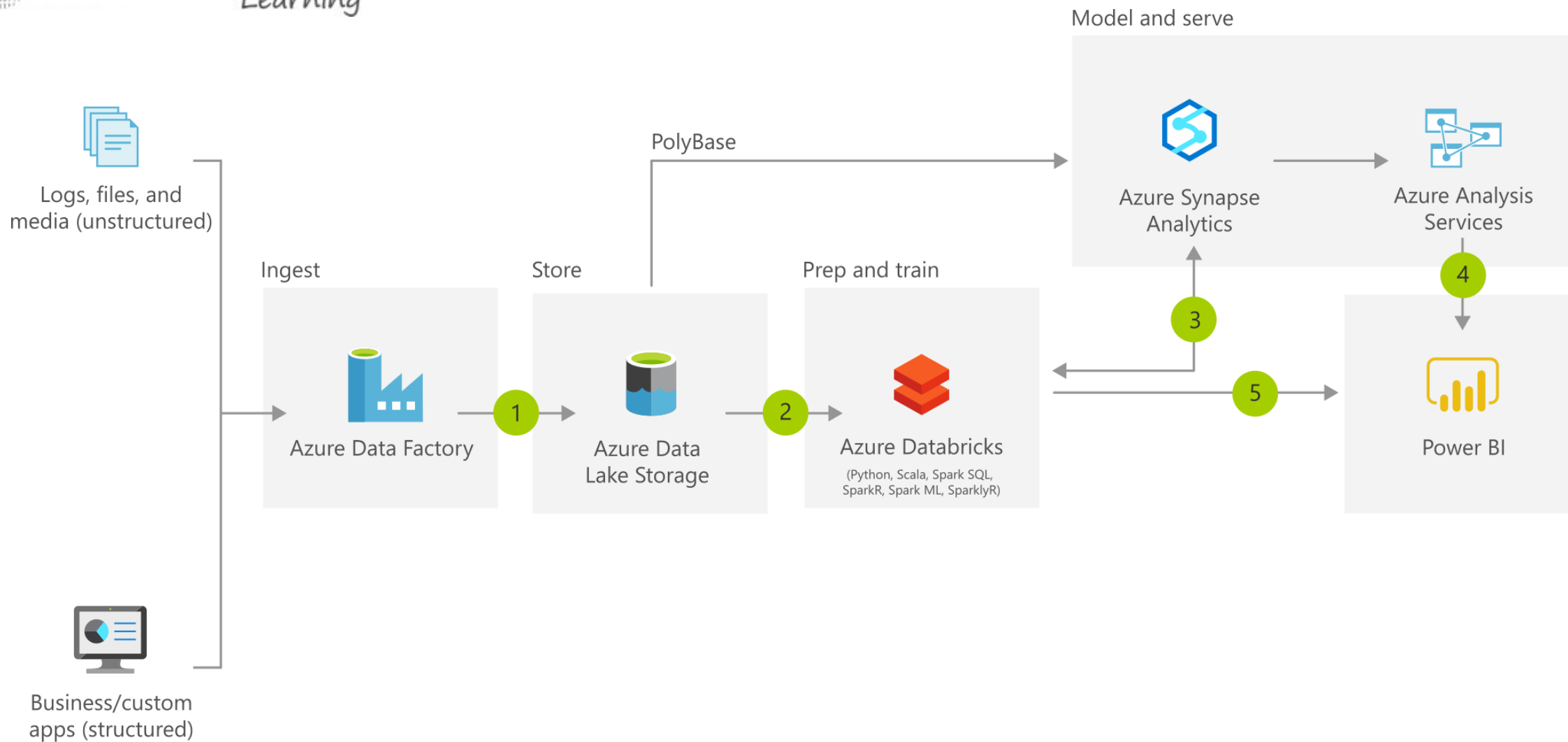
# Voorbereiding

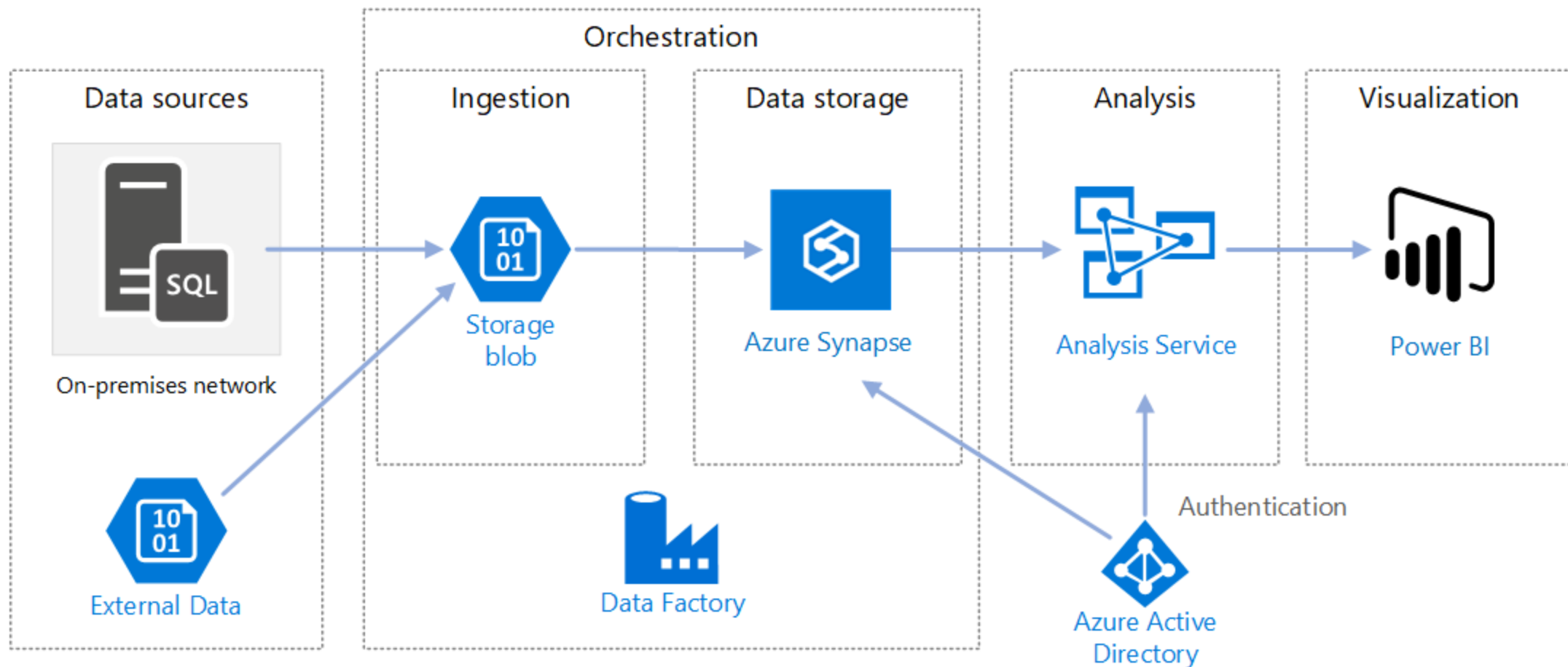
- Azure Portal
- Github of Azure Devops repository
- SQL Server Firewall
  - Azure Services
  - Home IPAdres
- KeyVault - Access policies ADF / User



# Data van A naar B

Wat doen we eigenlijk met ETL  
Waar in de architectuur zijn we bezig.









## ETL vs ELT

- Wat is ETL vs ELT?
- Waarom ETL?



# ETL vs ELT

- Waarom ETL?
  - Organisatie breed
  - Beheerbaarheid
  - Performance
  - Automatiseren van de dataflow



# Azure Dafa Factory (ADF)

- ADF
- Interface
  - | URL: <https://adf.azure.com/>
- Documentatie
  - | Docs: <https://docs.microsoft.com/en-us/azure/data-factory>



## Link Services

- Type verbinding
- Authentication
- Authorization



# Get started with ADF Project

- Git

Opdracht: Configuratie



# Project Links Services

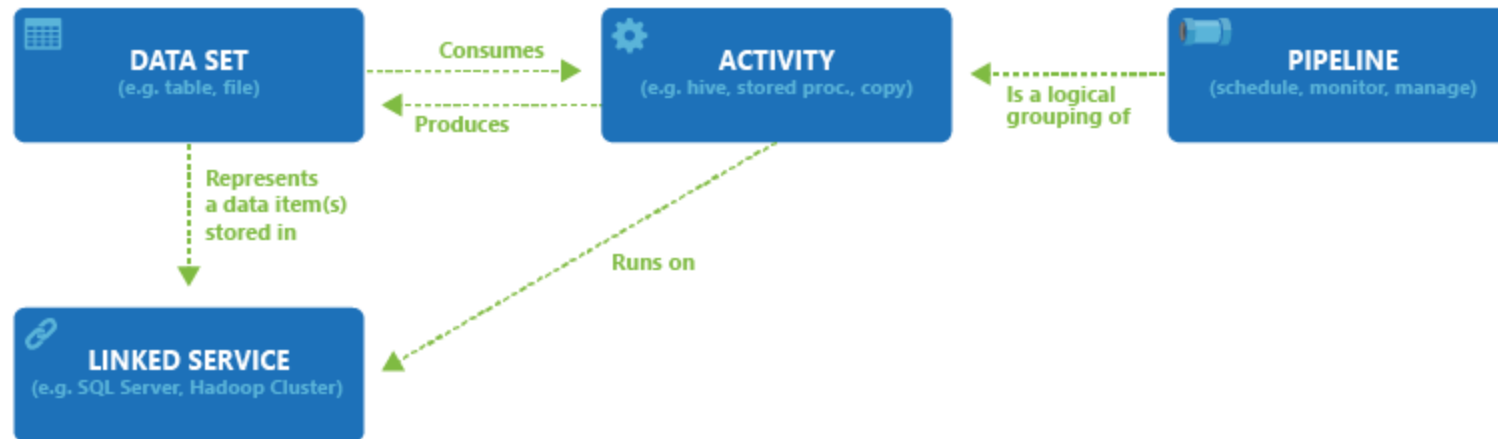
Opdracht: Maken van een Linked Services naar:

- Key Vault
- Storage Containter
- Database

[https://github.com/pcpronk/SDL\\_ADF\\_2021/blob/main/Opdrachten/linked\\_services.md](https://github.com/pcpronk/SDL_ADF_2021/blob/main/Opdrachten/linked_services.md)



# Pipelines





# Pipelines

- Activiteiten
  - data movement activities
  - data transformation activities
  - control activities
- Control flow
  - Status van een activiteit
- Parameters & Variables





# Data Set

- Wat is een data set
- Hoe op te zetten
- Kolommen & Data types
- Maken van een Data set
- Parameters



# Opdrachten

Bron > Datalake (CopyDate)

We gaan verder door een tabel Customer uit een database AdventureWorksLite (awlt) op te halen en te kopiëren naar ADLS. Het bestands type in ADLS is parquet.



# Integration runtime

- Waar draait het dan
  - Best effort voor Locatie
- Drie spaken
  - Azure
  - Self-hosted
  - Azure-SSIS



# Execution & Triggers

- Onces
- Schedule trigger
- Tumbling window trigger
- Event-based trigger

**Docs:** <https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers>



# Monitoring/ Logging



# Recap Module 1

- Wat is er allemaal besproken.
- Zijn er nog vragen?



# Traning

Zijn er nog ding dat je graag anders zou willen zien in deze training?

Gaat het snel of langzaam?

Genoeg pauze?