



Draaiboek Azure Data Factory (ADF) Fundamentals Dag 1

2 daagse / 22 & 23 november en 24 & 25 november 2021 / max 11 deelnemers.

9:00 - Sheet 1 t/m 4

Inleiding:

Welkom heten bij de training.

De leerdoelen voor deze training zijn: Na de training heeft de deelnemer inzicht in:

- Wat is data orchestratie.
- Wat is ADF en wat kan ik ermee.
- Welke basis onderdelen bestaat de ADF uit.
- Hoe maak je een basis pipeline.

Na de training kan de deelnemer:

- Aangeven wat data orchestratie is.
- Uitleggen wat de ADF is en waarvoor je het inzet.
- Uitleggen wat de basis onderdelen zijn en hoe deze zich tot elkaar verhouden
- Een eigen pipeline maken voor data verplaatsing.

De Agenda voor deze training is:

- Inleiding
- Kennismaking
- Hoofdstuk 1: Data orchestratie
- Korte pauze
- Hoofdstuk 2: De Azure Data Factory
- Lab1, Azure environment deployen
- Hoofdstuk 3: Integration Runtimes
- Lunch
- Korte recap
- Lab2, Integration Runtimes
- Hoofdstuk 4: Linked Services
- Lab3, Linked Services
- Hoofdstuk 5: Datasets
- Lab4, Datasets
- Korte pauze
- Hoofdstuk 6: Pipelines
- Lab5, Pipelines

- Hoofdstuk 7: Triggers
 - Lab6, Triggers
 - Recap training
 - Vragen, advies en feedback
-

9:15 - Sheet 5 voor virtueel. Voor klassikaal, schrijf de vragen op een whitebord/ flapover.

Kennismaking:

Stel eerst jezelf kort voor. Het doel van de kennismaking is: Van elkaar weten wat je komt leren zodat je elkaar kunt helpen. Vragen bij de kennismaking:

- Naam / leeftijd.
 - Wat is je functie en/of achtergrond.
 - Wat wil je in deze training leren.
 - Wat weet je al van ADF.
-

9:45 - Sheet 6 t/m 9

Hoofdstuk 1: Data orchestratie

Doel:

Na deze les weet de deelnemer wat de volgende zaken zijn:

- Wat data orchestratie is.
- Waarvoor data orchestratie gebruikt wordt.
- Welke onderdelen binnen data orchestratie bestaan.
- Welke uitdagingen data orchestratie met zich meebrengt.

Structuur:

- Wat is data orchestratie?
- Met welk doeleinde wordt het gebruikt?
- Uit welke onderdelen bestaat het (Ingest, Store, Prep and train, Model and serve).
- Welke uitdagingen brengt het met zich mee?

Theorie:

- Data orchestratie uitleggen.
- Uitleggen dat data orchestratie gebruikt wordt voor het samenbrengen, samenvoegen en prepareren van data voor analyse doeleide en t.b.v. besluitvorming.
- Vraag: Uit welke onderdelen denk jij dat data orchestratie bestaat?
- Uitleggen welke onderdelen er zijn en wat bij elke onderdeel plaats vind.
- Uitleggen welke uitdagingen data orchestratie met zich mee brengt.
- Vraag: Wie heeft er wel eens eerder één van deze problemen mee gemaakt?
- Vraag: Hoe heb jij dit opgelost?

10:15

Koffiepauze

10:30 - Sheet 10 t/m 14, adf.azure.com

Hoofdstuk 2: Azure Data Factory

Doel:

Na deze les weet de deelnemer:

- Wat de Azure Data Factory is.
- Wat je ermee kan doen.
- Hoe het in de theorie werkt.
- Uit welke componenten het bestaat.

Structuur:

- Wat is de Azure Data Factory?
- Waar gebruik je het voor?
- Hoe werkt het? (Demo de kant en klare omgeving)
- Welke belangrijke componenten bevat de ADF (Pipelines, Activities, Datasets, Linked services, Integration Runtimes)

Theorie:

- Wat is de Azure Data Factory (ADF).
- Welke rol speelt de ADF binnen Azure/ Dataplatform.
- Vraag: Waar zou jij de ADF voor willen gebruiken?
- Hoe gaat de ADF te werk.
- Welke componenten zijn nodig voor een complete orchestratie.
- Vraag: Wie kan aangeven welke componenten er allemaal zijn?
- Vraag: In welke volgorde denk jij dat de componenten horen?

Reflectie:

- Vraag: Hoe staan jullie hier tegenover?
-

11:00 - Sheet 15, [Portal.azure.com](https://portal.azure.com), GitHub voor virtueel. Handout voor klassikaal

Lab1, Azure environment deployen

Doel:

De trainings omgeving uit te rollen zodat elke deelnemer voor zich aan de slag kan in Azure.

Instructie:

Jullie kunnen aan de hand van Github/ handout de uitleg voor Lab1 volgen. Deze stap voor stap uitvoeren om de trainings omgeving klaar te zetten in Azure.

Reflectie:

Met deze omgeving kunnen we verder in de training aan de slag met de bijhorende labs.

- Vraag: Hoe vonden jullie het inrichten van de omgeving?
-

11:45 - Sheet 16 t/m 18, adf.azure.com

Hoofdstuk 3: Integration Runtimes

Doel:

Na deze les weet de deelnemer welke verschillende soorten Integration Runtimes er zijn en wat ze doen.

Structuur:

- Wat is een Integration Runtime?
- Waar gebruik je het voor?
- Welke soorten zijn er?

Theorie:

- Wat is een Integration Runtime (IR).
 - Waar gebruik je het voor.
 - Vraag: Waar moet je op letten bij een IR?
 - Welke soorten IR's zijn er.
 - Vraag: Stel, ik werk in een ander team, Dit team heeft een eigen ADF voor zijn/ haar data doeleinde. Het team heeft een bron nodig van on-premise die al **eerder** is ontsloten, welke soort IR heb ik nodig?
-

12:00

Lunch

13:00 - Sheet 19, Sheet 6 t/m 18

Recap ochtend

Doel:

- Kennis stimuleren.

Structuur:

- Vraag: Wat is data orchestratie?
- Vraag: Welke 4 onderdelen waren er?
- Vraag: Noem minstens 2 uitdagingen waar je tegen aan kan lopen?
- Vraag: Wat is de ADF?

- Vraag: Noem minstens 2 componenten van de ADF?
- Vraag: Welk component hebben we als laatste voor de lunch behandeld?
- Vraag: Welke soorten IR's waren er?

Reflectie:

- Hoe vonden jullie de ochtend?
 - Hebben jullie nog vragen over iets uit het ochtend gedeelte?
-

13:30 - Sheet 20, adf.azure.com, GitHub voor virtueel. Handout voor klassikaal.

Lab2, Integration Runtimes

Doel:

Na deze oefening weet de deelnemer hoe hij/zij een Integration Runtime (IR) moet aanmaken, configureren en installeren binnen ADF en Azure.

Instructie:

Jullie kunnen aan de hand van Github/ handout de uitleg voor Lab2 volgen. Deze stap voor stap uitvoeren om jullie eerste IR's aan te maken.

Reflectie:

- Vraag: Hoe vonden jullie het inrichten van de IR?
 - Vraag: Kan je hier iets mee in de praktijk?
-

14:00 - Sheet 21 t/m 23, adf.azure.com

Hoofdstuk 4: Linked Services

Doel:

Na deze les weet de deelnemer wat een Linked Service is en waarvoor het wordt gebruikt.

Structuur:

- Wat is een Linked Service?
- Wat doet een Linked Service?
- Welke soorten connectors zijn er allemaal?

Theorie:

- Wat is een Linked Service.
 - Wat dient een Linked Service voor.
 - Vraag: Weet iemand welke een mogelijke connector van een Linked Service?
 - Soorten (90+) connectoren zijn er?
 - Vraag: Welke connectoren verwacht jij te gaan gebruiken in de praktijk?
-

14:15 - Sheets 24, adf.azure.com, GitHub voor virtueel. Handout voor klassikaal.

**Lab3, Linked Services*

Doel:

Na deze oefening weet de deelnemer hoe hij/zij een Linked Service moet aanmaken binnen ADF.

Instructie:

Jullie kunnen aan de hand van Github/ handout de uitleg voor Lab3 volgen. Deze stap voor stap uitvoeren om jullie eerste Linked Services aan te maken.

Reflectie:

- Vraag: Hoe vonden jullie het aanmaken van de Linked Services?
 - Vraag: Kan je hier iets mee in de praktijk?
-

14:30 - Sheets 25 t/m 27, adf.azure.com

Hoofdstuk 5: Datasets

Doel: Na deze les weet de deelnemer wat een Dataset is en waarvoor het wordt gebruikt.

Structuur:

- Wat is een Dataset?
- Wat doet een Dataset?
- Welke mogelijkheden heeft een Dataset?

Theorie:

- Wat is een Dataset.
 - Waar dient een Dataset voor.
 - Vraag: Weet iemand een voorbeeld van een mogelijke Dataset?
-

14:45 - Sheets 28, adf.azure.com, GitHub voor virtueel. Handout voor klassikaal.

Lab4, Datasets

Doel:

Na deze oefening weet de deelnemer hoe hij/zij een Dataset moet aanmaken binnen ADF.

Instructie:

Jullie kunnen aan de hand van Github/ handout de uitleg voor Lab4 volgen. Deze stap voor stap uitvoeren om jullie eerste Datasets aan te maken.

Reflectie:

- Vraag: Hoe vonden jullie het aanmaken van de Dataset?

- Vraag: Kan je hier iets mee in de praktijk?
-

15:00

Koffiepauze

15:15 - Sheet 29 t/m 31, adf.azure.com

Hoofdstuk 6: Pipelines

Doel: Na deze les weet de deelnemer wat een Pipeline is en waarvoor het wordt gebruikt.

Structuur:

- Wat is een Pipeline?
- Wat doet een Pipeline?
- Welke mogelijkheden heeft een Pipeline?

Theorie:

- Wat is een Pipeline.
 - Waar dient een Pipeline voor.
 - Wat kan je allemaal met een Pipeline doen.
 - Vraag: Wie kan een voorbeeld geven van een Pipeline?
-

15:30 - Sheet 32, adf.azure.com, GitHub voor virtueel. Handout voor klassikaal.

Lab5, Pipelines

Doel:

Na deze oefening weet de deelnemer hoe hij/zij een Pipeline moet aanmaken binnen ADF.

Instructie:

Jullie kunnen aan de hand van Github/ handout de uitleg voor Lab5 volgen. Deze stap voor stap uitvoeren om jullie eerste Pipeline aan te maken.

Reflectie:

- Vraag: Hoe vonden jullie het aanmaken van de Pipeline?
 - Vraag: Kan je hier iets mee in de praktijk?
-

15:45 - Sheet 33 t/m 35, adf.azure.com

Hoofdstuk 7: Triggers

Doel:

Na deze les weet de deelnemer wat een Triggers is en waarvoor het wordt gebruikt.

Structuur:

- Wat is een Trigger?
- Wat doet een Trigger?
- Welke soorten Triggers zijn er?

Theorie:

- Wat is een Trigger.
 - Waar dient een Trigger voor.
 - Welke soorten Triggers zijn er.
 - Vraag: Als ik een Pipeline op een bepaald moment wil aan laten gaan, welke soort trigger heb ik nodig?
 - Vraag: Als team zijn wij afhankelijk van een ander team die elke dag een bestand dient up te loaden op ons storage account. Nadat dit geupload is kunnen wij pas onze Pipelines gaan draaien. Welke soort trigger dienen wij in te stellen?
-

16:00 - Sheet 36, adf.azure.com, GitHub voor virtueel. Handout voor klassikaal.

Lab6, Triggers

Doel:

Na deze oefening weet de deelnemer hoe hij/zij een Trigger moet aanmaken binnen ADF.

Instructie:

Jullie kunnen aan de hand van Github/ handout de uitleg voor Lab6 volgen. Deze stap voor stap uitvoeren om jullie eerste Trigger aan te maken.

Reflectie:

- Vraag: Hoe vonden jullie het aanmaken van de Trigger?
 - Vraag: Kan je hier iets mee in de praktijk?
-

16:30 - Sheet 37

Leerdoelen aanhalen

Doel:

Dag 1 doornemen, wat hebben we allemaal behandeld ten opzichte van de gestelde leerdoelen voor vandaag.

16:45 - Sheet 38

Vragen, adviezen en/of feedback

Afsluiting

Iedereen hartelijk bedanken voor de inzet en het delen van kennis en inzichten. Goede reis naar huis.

17:00 - Einde dag 1.

Draaiboek Azure Data Factory (ADF) Fundamentals Dag 2

2 daagse / 22 & 23 november en 24 & 25 november 2021 / max 11 deelnemers.

9:00 - Sheet 1 t/m 4

Inleiding:

Welkom heten bij de training.

De leerdoelen voor deze training zijn: Na de training heeft de deelnemer inzicht in:

- Hoe maak je geavanceerde pipelines.
- Hoe je dynamische content toepast binnen de ADF.
- Hoe versiebeheer werkt via GIT.
- Hoe je via DevOps (CI/CD) de ADF kan deployen.

Na de training kan de deelnemer:

- Eigen pipelines maken via de Best Practises.
- Dynamische parameters toepassen.
- Uitleggen hoe GIT werkt i.c.m. ADF.
- Uitleggen hoe DevOps (CI/CD) toegepast kan worden.

De Agenda voor deze training is:

- Inleiding
- Recap dag 1
- Hoofdstuk 8: Global Parameters
- Lab7, Global Parameters
- Korte pauze
- Hoofdstuk 9: Activities
- Lunch
- Korte recap
- Lab8, Activities
- Hoofdstuk 10: Batches en DIUs
- Lab9, Batches en DIUs
- Korte pauze
- Hoofdstuk 11: GIT
- Hoofdstuk 12: DevOps
- Recap training
- Vragen, advies en feedback

9:15 - Sheet 5, Wissel naar de presentatie van dag 1.

Recap dag 1

Doel:

- Kennis stimuleren
- Opfrisser

Structuur:

- Wat hebben wij gisteren allemaal behandeld?
- Wat is data orkestratie?
- Wat is de ADF?
- Wat was een IR?
- Wat was een Linked Service?
- Wat was een Dataset?
- Wat was een Pipeline?
- Wat was een Trigger?

Theorie:

- Wat hebben wij gisteren behandeld.
- Vraag: Wat is data orkestratie?
- Vraag: Welke 4 onderdelen waren er?
- Slides data orkestratie doorlopen.
- Vraag: Wat is de ADF?
- Vraag: Noem de componenten uit de ADF?
- Slides ADF doorlopen.
- Vraag: Wat is een IR?
- Vraag: Welke soort IR's zijn er?
- Slides IR doorlopen.
- Vraag: Wat is een Linked Services?
- Vraag: Noem minstens 5 connectoren?
- Slides van de Linked Services doorlopen.
- Vraag: Wat is een Dataset?
- Vraag: Waar kan je doen met een Dataset?
- Slides van de Datasets doorlopen.
- Vraag: Wat is een Pipeline?
- Vraag: Wat kan je met een Pipeline?
- Slides van de Pipelines doorlopen.
- Vraag: Wat is een Trigger?
- Vraag: Welke soorten Triggers zijn er?
- Slides van de Triggers doornemen.

10:00

Koffiepauze

10:15 - Sheet 6 t/m 8, adf.azure.com

Hoofdstuk 8: Global Parameters

Doel:

Na deze les weet de deelnemer wat een global parameters zijn en waarvoor het wordt gebruikt.

Structuur:

- Wat is een global parameter?
- Wat kan je met een global parameter?

Theorie:

- Wat is een global parameter.
 - Waar dient een global parameter voor.
 - Vraag: Waar zou je een global parameter wellicht nog meer voor kunne gebruiken?
-

10:30 - Sheet 9, adf.azure.com, GitHub voor virtueel. Handout voor klassikaal.

Lab7, Global Parameters

Doel:

Na deze oefening weet de deelnemer hoe hij/zij een global parameter moet aanmaken binnen ADF.

Instructie:

Jullie kunnen aan de hand van Github/ handout de uitleg voor Lab7 volgen. Deze stap voor stap uitvoeren om jullie eerste global parameter aan te maken en te verwerken in een pipeline.

Reflectie:

- Vraag: Hoe vonden jullie het toepassen van globale parameters?
 - Vraag: Kan je hier iets mee in de praktijk?
-

11:00 - Sheet 10 t/m 14

Hoofdstuk 9: Activities

Doel:

Na deze les weet de deelnemer wat activities zijn, hoe deze samenhangen en gebruikt kunnen worden, wat de best practises zijn en hoe je pipeline content dynamisch tot stand kan brengen.

Structuur:

- Wat is een activity?

- Welke activiteiten zijn er allemaal?
- Welke activiteiten gebruik je voor welk scenario?
- Hoe kan je activiteiten samen laten werken?
- Welke best practises zijn er allemaal?
- Hoe maak je pipeline content dynamisch?

Theorie:

- Wat is een activity.
 - Welke soort activiteiten zijn er allemaal.
 - Welke activity gebruik je wanneer.
 - Vraag: Welke activiteiten kan je gebruiken als je iets conditioneel wilt maken?
 - Best practises voor activity pipelines
 - Vraag: Denk je dat er bepaalde best practises zijn die jullie kunnen gebruiken in de praktijk?
 - Wat kan je allemaal dynamisch maken.
-

12:00

Lunch

13:00 - Sheet 15, Sheet 6 t/m 14

Recap ochtend

Doel:

- Kennis stimuleren.

Structuur:

- Vraag: Wat is een global parameter?
- Vraag: Wat kan ik met een global parameter?
- Vraag: Wat is een activity?
- Vraag: Waar gebruik ik een ForEach activity voor?
- Vraag: Wat doet een LookUp activity?
- Vraag: Welke andere activiteiten ken je nog meer?
- Vraag: Hoe ziet het proces eruit van dynamisch data overhalen?
- Vraag: Wat moet er gedaan worden om in een stored procedure te inserten?

Reflectie:

- Hoe vonden jullie de ochtend?
 - Hebben jullie nog vragen over iets uit het ochtend gedeelte?
-

13:30 - Sheet 16, adf.azure.com, GitHub voor virtueel. Handout voor klassikaal.

Lab8, Activities

Doel:

Na deze oefening weet de deelnemer hoe hij/zij een verschillende activiteiten moet aanmaken, deze moet laten samenwerken en via best practises pipelines kan aanmaken met de juiste activiteiten op dynamisch manier toe te passen.

Instructie:

Jullie kunnen aan de hand van Github/ handout de uitleg voor Lab8 volgen. Deze stap voor stap uitvoeren om met de verschillende activiteiten en best practises aan de slag te gaan.

Reflectie:

- Vraag: Hoe vonden jullie het maken van geavanceerde pipelines?
- Vraag: Kan je hier iets mee in de praktijk?

14:30 - Sheet 17 t/m 19, adf.azure.com

Hoofdstuk 10: Batches en DIUs

Doel:

Na deze les weet de deelnemer hoe hij/zij data overdacht kan versnellen door data integration units en data kan partitioneren in batches.

Structuur:

- Welke versnellende stappen kan ik maken in de ADF?
- Wat is een DIU?
- Wat kan ik met DIUs?
- Wat is Batching?
- Wat kan ik met Batching?

Theorie:

- Hoe versnel ik mijn pipelines.
- Wat zijn DIUs.
- Wat kan ik met DIUs.
- Vraag: Betekent meer DIUs ook gegarandeerd betere performance?
- Wat is Batching?
- Hoe werkt Batching?
- Vraag: Hebben jullie workloads waarbij je batching verwacht nodig te hebben?
- Vraag: Wie weet een best practice op te noemen die behandeld is, waar batching wellicht handig voor is?

14:45 - Sheet 20, adf.azure.com, GitHub voor virtueel. Handout voor klassikaal.

Lab9, Batches en DIUs

Doel:

Na deze oefening weet de deelnemer hoe hij/zij Batching en DIUs moet toepassen op pipelines binnen de ADF.

Instructie:

Jullie kunnen aan de hand van Github/ handout de uitleg voor Lab8 volgen. Deze stap voor stap uitvoeren om jullie pipelines te voorzien van batching en extra DIUs.

Reflectie:

- Vraag: Hoe vonden jullie het toepassen van batches en DIUs?
 - Vraag: Kan je hier iets mee in de praktijk?
-

15:00

Koffiepauze

15:15 - Sheet 21 t/m 23, adf.azure.com, dev.azure.com

Hoofdstuk 11: GIT

Doel: Na deze les weet de deelnemer hoe hij/zij GIT moet toepassen op de ADF.

Structuur:

- Wat is GIT?
- Hoe verhoudt GIT zich ten opzichten van TFS?
- Hoe pas je GIT toe in ADF? (DEMO)

Theorie:

- Wat is GIT.
 - GIT vs TFS als versiebeheer.
 - Vraag: Wie heeft er wel eens met GIT gewerkt?
 - Vraag: Hoe heb je dat ervaren?
 - Welke soorten GIT zijn er voor ADF.
 - Hoe pas je GIT toe in ADF.
-

15:45 - Sheets 24 t/m 27, adf.azure.com, dev.azure.com

Hoofdstuk 12: DevOps

Doel:

Na deze les weet de deelnemer hoe hij/zij DevOps kan toepassen op het CI/CD deployen van de ADF.

Structuur:

- Wat is DevOps?
- Wat is CI/CD?
- Hoe werkt ADF met CI/CD? (DEMO)

Theorie:

- Wat is DevOps.
 - Vraag: Wie heeft er wel eens eerder DevOps gewerkt?
 - Wat is CI/CD.
 - Vraag: Wie heeft er wel via CI/CD iets gedeployed?
 - Hoe werkt ADF met CI/CD.
-

16:30 - Sheet 28

Leerdoelen aanhalen

Doel:

Dag 2 doornemen, wat hebben we allemaal behandeld ten opzichte van de gestelde leerdoelen voor vandaag.

16:45 - Sheet 29

Vragen, adviezen en/of feedback

Afsluiting

Iedereen hartelijk bedanken voor de inzet en het delen van kennis en inzichten. Goede reis naar huis.

17:00 - Einde dag 2.

/