

## Lab 2 - Integration Runtimes

---

### Vereisten

Om het lab te kunnen starten is het van belang dat Lab1 is afgerond.

### Doel

De Azure Data Factory maakt gebruik van een *Integration Runtime (IR)*. In feite is dit de plek waar je pipeline "uitgevoerd" wordt: bijvoorbeeld in een specifieke Azure-region of on-premises.

Gedurende dit lab maak je meerdere soorten IRs aan. Ook configureer, installeer en "link" je deze. Volg de opdrachten stap voor stap.

## Opdracht 1 - Azure Integration Runtime aanmaken

Binnen ADF krijg je altijd een Integration Runtime "cadeau": een zogenaamde "Auto Resolve" IR. Dat betekent dat Azure wereldwijd met je meekijkt welke plek handig zou zijn om jouw pipeline uit te gaan voeren en er eentje voor je uitkiest. Dat kan handig zijn om de eerste stappen te maken, maar het is aan te raden om explicieter te zijn waar je data verwerkt wordt. Deels vanuit performance- en kostenoverwegingen, maar ook vanuit security- en compliance-perspectief (weten waar je data verwerkt wordt).

In deze eerste opdracht gaan we daarom een Integration Runtime aanmaken waarin je zelf de locatie aangeeft. In je omgeving heb je twee Data Factories, die later beiden gebruik gaan maken van de ene IR die we aanmaken.

1. Ga naar de [Azure portal](#). Klik de Data Factory aan waarbij **niet** het woord Linked instaat. Een nieuwe pagina zal laden met een knop **Launch Studio** klik deze aan. Je wordt nu naar de pagina van de ADF geleid.
2. Klik links op de **gereedschapskist** (Manage). Klik vervolgens aan de linkerkant op **Integration runtimes**. Je zult zien dat er al een **AutoResolveIntegrationRuntime** bestaat. Dit is de standaard IR binnen Azure waarmee je tussen de verschillende diensten als SQL Database, Azure Data Lake Store e.d. data mee kan verplaatsen.
3. Klik op **New**, klik op **Azure, Self-Hosted** en op **continue** en vervolgens op **Azure** en weer **continue**. Je kunt nu een nieuwe **Azure Integration Runtime** aanmaken, waarbij je ook specifiek de **Region** kunnen aangeven. Hiermee weet je zeker dat de data altijd binnen deze regio blijft.
4. Geef de IR een naam en kies de **Region** waarbinnen de rest van je resources zich bevinden. Qua naamgeving is het handig om te beginnen met **IR-** gevolgd door wat voor een soort IR het is, de - regio en de -omgeving.
  - Praktijkvoorbeeld: **IR-AzureResolve-EastUS-TST**
  - Trainingsvoorbeeld: **IR-AzureResolve-WestEurope-Training**
5. Klik op **Create**

Je IR wordt nu aangemaakt en zou binnen enkele seconden klaar moeten zijn.

## Opdracht 2 - Self-Hosted Integration runtime

Data vanuit een on-premises omgeving of een afgeschermd netwerk is voor Azure Data Factory niet direct benaderbaar. Met een *self-hosted* IR kun je ook deze omgevingen aansluiten: je installeert dan op een VM in je afgeschermd netwerk (bijv. on-premises) een *self-hosted Integration Runtime*. Deze legt vervolgens de verbinding met Azure Data Factory.

1. Ga terug naar je Virtual Machine als je deze open hebt staan, anders je volg de stappen uit Lab1, opdracht 2 weer.
2. Ga naar de Azure portal (portal.azure.com) en login met het account waarmee je de training uitvoert.
3. Ga terug naar de ADF (zonder linked in de naam) via de resourcegroup zoals uitgelegd in opdracht 1.
4. Eenmaal aangekomen op het scherm van de IRs, klik weer op **New**, klik op **Azure, Self-Hosted** en op **continue** en vervolgens op **Self-Hosted** en weer **continue**.

Een Self-Hosted IR draai je meestal op een on-premises Virtual Machine met poort 433 en 1433 open zodat deze bij Azure kan komen voor zowel de Database (1433) als een storage account (433). De minimale vereisten om een Self-Hosted IR goed te draaien zijn:

- 4 cpu cores van 2Ghz of sneller
- 8GB RAM
- 80GB schijfruimte

In de praktijk blijkt dat 16GB RAM en een SSD kunnen helpen om de data beter verplaatsbaar te maken, hierbij kan hooguit de CPU de limiterende factor zijn.

De Integration Runtime is ook direct te downloaden vanaf <https://www.microsoft.com/en-us/download/details.aspx?id=39717>

5. Je hoeft alleen de IR een naam te geven. De regio is afhankelijk van waar de Virtual Machine staat en klik op **Create**.
6. Wanneer de IR is aangemaakt krijg je meteen een scherm met 2 opties voor installatie. Gezien je al op de Virtual Machine zit kies je voor Optie 1 (Express setup). Klik op **Click here to launch the express setup for this computer**. En start vervolgens de installatie door op het gedownloaden bestand te klikken. Dit kan door de popup rechtsbovenin of via **File Explorer** (Het gele mapje) naar **Downloads** te gaan.

Wanneer je niet vanaf de VM op de Azure portal van jouw organisatie kan komen, is **Option 2** een mogelijkheid: je downloadt de IR dan handmatig, en gebruikt één van de weergegeven sleutels tijdens de installatie.

7. Wanneer je de installatie start gebeurt alles voor de rest automatisch, wacht de installatie rustig af.
8. Wanneer de installatie is afgerond, klik op **Close**. In de ADF zou je nu een **Groen bolletje** moeten zien bij de Self-Hosted IR.

Je hebt nu een *self-hosted* IR gemaakt, waardoor Azure Data Factory data van en naar een on-premises omgeving kan verplaatsen.

## Opdracht 3 - Integration runtimes linken

Wanneer je meerdere Data Factories gebruikt, kun je ervoor kiezen een bestaande IR in meerdere Data Factories te "linken". Dit voorkomt dat je een IR meerdere keren moet installeren.

1. Klik de **Self-Hosted IR** aan die je net hebt aangemaakt. Ga naar de tab **Sharing** en kopieer de resourceID door op het **blauwe vakje** te klikken of deze te selecteren en te kopiëren.
2. Klik op het **Blauwe plusje** met de tekst **Grant permissions to another Data Factory or user-assigned managed identity**
3. De adf-linked zou hier automatisch tussen moeten staan, vink het aan en klik vervolgens op **Add**
4. Sluit de tab door op **Apply** te klikken.
5. Rechtsboven in beeld, zie je een rij met iconen. Klik op de 2e van links, het icoon met **de 2 schermen en pijljes** (Switch Data Factory).
6. Er zal een nieuw scherm verschijnen, en het meeste zal al van tevoren zijn ingevuld. Kies bij **Data Factory Name** de adf-linked en klik vervolgens op **OK**.
7. De andere ADF wordt nu geladen. Wanneer dit voltooid is klik je links op de **gereedschapskist** (manage) en vervolgens weer op **Integration runtimes**.
8. Eenmaal aangekomen op het scherm van de IRs, klik weer op **New**, klik op **Azure, Self-Hosted** en op **continue** en vervolgens op **Linked Self-Hosted** en weer **continue**.
9. Plak de gekopieerde resourceID alvast in het veld van de **Resource ID** en geeft de IR een correcte naam.
10. Klik op **Create** en de gelinkede Self-Hosted IR zou moeten verschijnen met een **Groen bolletje**.

Hoewel het koppelen van self-hosted IR's erg handig is, is het niet altijd wenselijk dat bijvoorbeeld een ontwikkel-omgeving via dezelfde IR data kan ontsluiten als een productie-omgeving. Het komt nog wel eens voor dat een Ontwikkel en Test omgeving dezelfde Self-Hosted IR gebruiken, net zoals voor Acceptatie en Productie mits hier een specifieke case voor is.

Het linken voor een Self-Hosted IR wordt vaak gebruikt als er meerdere teams zijn die dezelfde databron gebruiken. Doordat je maar 1 IR nodig hebt (en dus Virtual Machine) bespaar je kosten en resources.

## Inhoudsopgave

1. [De Azure omgeving prepareren](#)
2. [Integration Runtimes](#)
3. [Linked Services](#)
4. [Datasets](#)
5. [Pipelines](#)
6. [Triggers](#)
7. [Global Parameters](#)
8. [Activities](#)
9. [Batching en DIUs](#)