



Lab2 - Integration Runtimes

Vereisten

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

Doel

De Azure Data Factory maakt gebruik van Integration Runtimes (IR), welke dient als rekeninfrastructuur om de verschillende vormen van data verplaatsingen uit te kunnen voeren. Gedurende de lab gaan we meerdere soorten IRs aanmaken, configureren, installeren en linken. Volg de opdrachten stap voor stap.

Opdracht 1 - Auto Resolve Integration runtime

1. Ga naar de Azure portal (portal.azure.com). Klik de Data Factory aan waarbij **niet** het woord Linked instaat. Een nieuwe pagina zal laden met een knop **Open Azure Data Factory Studio** (of de equivalent in een andere taal) 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 data mee kan verplaatsen.
3. Klik op **New**, klik op **Azure, Self-Hosted** en op **continue** en vervolgens op **Azure** en weer **continue**. We kunnen nu een nieuwe **AutoResolveIntegrationRuntime** aanmaken, waarbij we ook specifiek de **Region** kunnen aangeven. Hiermee weten we zeker dat onze Data altijd binnen deze regio blijft.
4. Geef de IR een naam en kies voor West Europe als **Region**. Het aangeraden format is om te beginnen met IR- gevolgt door wat voor een soort IR het is, de -regio en de -omgeving.

Praktijkvoorbeeld: IR-AutoResolve-WestEU-TST Trainingsvoorbeeld: IR-AutoResolve-WestEU-Training

5. Klik op **Create**, Hoera! je IR wordt nu aangemaakt en zou binnen enkele seconden klaar moeten zijn.

Opdracht 2 - Self-Hosted Integration runtime

1. Ga terug naar je Virtual Machine als je deze open hebt staan, anders je volg de stappen uit Lab1, opdracht 2 weer.
2. Open de **Edge browser** en ga naar de Azure portal (portal.azure.com) en login met het juiste account waar de Azure pass op staat.
3. Ga terug naar de ADF (zonder linked in de naam) via de resourcegroup zoals uitlegd 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 normaal op een On-premise 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 is een Virtual machine met 4 cpu cores van 2Ghz of hoger, 8GB RAM en 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.

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.
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.

Opdracht 3 - Integration runtimes linken

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 **Cancel** te klikken.
5. Rechtboven 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 te voren 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**. Het linken van Self-Hosted integration runtimes zie je vaak bij omgevingen die bij dezelfde data moeten komen. Het is dan ook gebruikelijk om meerdere Virtual Machines te hebben met meerdere Self-Hosted IRs, dit zodat er een goed onderscheid gemaakt kan worden tussen Ontwikkel, Test, Acceptatie en Productie. 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 ook gebruikt als er meerdere teams zijn die bij 1 bepaalde databron moeten zijn, hiervoor heb je dan meer 1 IR nodig (en dus Virtual Machine) wat weer kosten en resources bespaard.

Einde Lab2

/