Lab 3 - Linked Services

Vereisten

Om het lab te kunnen starten is het van belang dat Lab2 is afgerond.

Doel

Om data over de zojuist aangemaakte IRs te laten verlopen moeten er connecties met de betreffende diensten gemaakt worden. Gedurende het lab leg je meerdere connecties, met o.a.:

- een SQL database (bijv. een bronsysteem of Data Warehouse)
- een Storage account (bijv. zoals een Data Lake)
- een File system (bijv. een share)

Sommige van deze bronnen kun je benaderen met behulp van *managed identity*: in dat geval worden binnen AAD rechten uitgedeeld aan de Data Factory. Andere bronnen zul je moeten benaderen met een *secret*, bijvoorbeeld een certificaat of een gebruikersnaam/wachtwoord. Deze *secrets* sla je in Azure centraal op in de Key Vault. Vanuit daar kun je dan eenvoudig bepalen welke diensten welke *secrets* mogen bekijken.

Opdracht 1 - Azure Key Vault

Azure Data Factory is eenvoudig te koppelen met Azure Key Vault, waarin we wachtwoorden en connection strings opslaan. We kunnen een verbinding naar een bron dan laten vullen door een *secret* uit de *Key Vault*. Op het moment dat ADF verbinding maakt met die bron, zal ADF eerst de *secret* ophalen uit de Key Vault.

Voordat we echter *secrets* uit de Key Vault kunnen benaderen, zullen we de Key Vault eerst moeten aankoppelen als *Linked Service*.

1. Ga terug naar de **niet** linked ADF. Klik vervolgens weer op Manage. Ga naar **Linked Services**.
2. klik op **New**, en zoek naar **Key vault**. Klik de **Azure Key vault** aan.
3. Geef de Linked services een duidelijke naam. Het aangeraden format is om te beginnen met LS_, de naam van de dienst in je resourcegroup en eindigend met _omgeving.
    - Praktijkvoorbeeld: `LS_KV_Dataplatform_PRD`
    - Trainingsvoorbeeld: `LS_KV_rcc4bh5724jim_Training`
      In de naamgeving is een minteken (`-`) niet toegestaan. Een *underscore* (_) is wel mogelijk.
4. Kies de **Azure Subscription** die je in de training gebruikt
5. Kies bij **Azure Key vault Name** de key vault uit jouw Key Vault (deze start met `kv_`).
6. Klik op de knop **Test Connection** om te valideren dat de verbinding tot stand gebracht kan worden. Gaat dit fout, laat het weten aan de trainer.
7. Als test klaar is en een **Groen bolletje** geeft, kan de Linked Service aangemaakt worden door op **Create** te klikken.
8. De Linked Service naar de Azure Key Vault is nu aangemaakt, maar deze is nog niet gepubliseerd. Klik op de **Blauwe knop** met de tekst **Publish all** en vervolgens op de knop **Publish**. Door te publishen komen de aanpassingen live te staan, en kan de Key Vault gebruikt worden.

Opdracht 2 - Databases

Met de Key Vault aangesloten is het mogelijk om wachtwoorden op te halen om een beveiligde verbinding op te zetten met bijvoorbeeld de databases.

1. Klik op **New**, en zoek naar **SQL**. Dubbelklik de **Azure SQL Databases** aan.
2. Geef de Linked services een duidelijke naam.
3. Kies bij **Connect via integration runtime** de eigen gemaakte **Azure IR**.
4. Kies bij de **Server Name** de Server naam in zoals deze in je resourcegroup staat.
5. Kies bij de **Database Name** de source Database naam in zoals deze in je resourcegroup staat. De source database begint met **sqldb-source-** als naam.
6. Vul bij de **User Name** het SQL admin account in genaamd: **sqladmin**.
7. Bij de optie tussen **Password** en **Azure Key Vault**, kies de Key vault.
8. Kies bij **AKV linked service** de eerder aangemaakte Key Vault Linked Service.
9. Kies bij **Secret Name** de optie **sqladmin**
10. Klik op de knop **Test Connection** om te valideren dat de verbinding tot stand gebracht kan worden. Gaat dit fout, laat het weten aan de trainer.
11. Als test klaar is en een **Groen bolletje** geeft, kan de Linked Service aangemaakt worden door op **Create** te klikken.
12. Doe Opdracht 2 nogmaals, maar nu voor de **sqldb-target** Database.

Je hebt nu twee Linked Services aangemaakt. Dit maakt het voor ADF mogelijk om verbinding te maken met de twee databases.

Opdracht 3 - Storage Account

De tweede bron die we toevoegen is een Storage Account. Deze kunnen we bijvoorbeeld gebruiken als *landing zone* voor de data, of als Data Lake.

1. klik op **New**, en zoek naar **storage**. Klik de **Azure Blob Storage** aan.
2. Geef de Linked services een duidelijke naam.
3. Kies bij **Connect via integration runtime** de eigen gemaakte **Azure IR**.
4. Kies bij **Storage account name** het storage account zoals deze in je resourcegroup staat.
5. Klik op de knop **Test Connection** om te valideren dat de verbinding tot stand gebracht kan worden. Gaat dit fout, laat het weten aan de trainer.
6. Als test klaar is en een **Groen bolletje** geeft, kan de Linked Service aangemaakt worden door op **Create** te klikken.

De rechten op het Storage Account zijn uitgedeeld via Azure AD. Hier heb je dus geen *secret* voor hoeven gebruiken.

Opdracht 4 - File system

De derde bron die we toevoegen is een on-premises filesystem. Omdat het filesystem on-premises staat, moeten we hier de juiste Integration Runtime moeten gebruiken! Ook is deze VM niet in ons domein, waardoor we moeten aangeven met welke username / password we zullen inloggen

1. klik op **New**, en zoek naar **file**. Klik de **File system** aan.
2. Geef de Linked services een duidelijke naam.
3. Kies bij **Connect via integration runtime** de **Self-Hosted IR**
4. Vul bij **Host** het volgende in **D:**

5. Vul bij de **User Name** het SQL admin account in genaamd: **sqladmin**.
6. Bij de optie tussen **Password** en **Azure Key Vault**, kies de Key vault.
7. Kies bij **AKV linked service** de eerder aangemaakte Key Vault Linked Service.
8. Kies bij **Secret Name** de optie **sqladmin**
9. Klik op de knop **Test Connection** om te valideren dat de verbinding tot stand gebracht kan worden. Dit gaat fout, maar dit lossen we zo meteen op!
10. Als test klaar is en een **Groen bolletje** geeft, kan de Linked Service aangemaakt worden door op **Create** te klikken.
11. Klik op de **Blauwe knop** met de tekst **Publish all** en vervolgens op de knop **Publish**.

> ## Waarom gaat de verbinding met het File System fout
>
> Sinds versie 5.22 heeft de Self-Hosted Integration Runtime strengere beveiligingsmaatregelen. Eén van die maatregelen is dat een self-hosted IR niet zomaar bij lokale bestanden mag.
>
> In de training willen we deze bestanden echter juist benaderen om te simuleren dat er een on-premises bron aanwezig is. Daarom zullen we deze beveiligingsmaatregel moeten uitzetten. Dit doe je als volgt:
>
> 1. Maak verbinding met de VM
> 2. Open het startmenu en typ **Powershell**
> 3. Kies **Run as administrator**
>
> In Powershell voer je nu het volgende in:
>
> 4. `cd 'C:\Program Files\Microsoft Integration Runtime\5.0\Shared\'`
> 5. `.\dmgcmd -DisableLocalFolderPathValidation`
>
> De instelling wordt nu aangepast, en de IR herstart. De Linked Service zou nu moeten werken **Test Connection** naar `D:\` moeten werken.

Inhoudsopgave