

Lab 8 - Activities

Vereisten

Om het lab te kunnen starten is het van belang dat Lab4 is afgerond.

Doel

We hebben al enkele activities gebruikt zoals Copy, Web, Wait en Set variable. Er zijn er nog veel meer en sommige zijn heel handig als je deze met elkaar laten samenwerken om zodoende geavanceerde pipelines te creëren. Volg de opdrachten stap voor stap.

Opdracht 1 - Stored Procedure uitvoeren

Stored Procedures zijn opgeslagen programma's op de database. Vaak worden acties op de database (zoals het leegmaken van een tabel, of het starten van een proces binnen de database) in een stored procedure "gevangen". Met ADF kun je deze nu orchestreren.

1. Zorg dat je weer terug bent in ADF. Klik bij Pipelines op **Pipeline Actions** en op **New Pipeline**.
2. Noem de pipeline: **PL_Process_Dates_Training**.
3. Uit de lijst met **Activities**, klik op de optie **General**. Klik en sleep **Stored Procedure** op het canvas.
4. Noem de **Stored Procedure** als volgt: **USP_DL_Dates**. Klik vervolgens op de tab **Settings**.
5. Kies bij **Linked service** voor de **LS_sqldb_target** en bij **Stored Procedure name** de **[Stg].[USP_DL_Dates]**.
6. Klik op **Stored procedure parameters** en vervolgens op **import**. De parameters van de stored procedure worden nu ingeladen.
7. Vul bij **StartYear** het volgende in: **1900**.
8. Vul bij **EndYear** het volgende in: **2099**.
9. Klik op de **Blauwe knop** met de tekst **Publish all** en vervolgens op de knop **Publish**.
10. Klik op **Debug** en wacht tot de pipeline klaar is.

Opdracht 2 - Conditioneel filteren

Je kunt data uit de database ook gebruiken om je orchestratie mee uit te voeren, bijvoorbeeld:

- Een proces dat alleen mag starten als er een bepaalde rij in je instellingen-tabel aanwezig is
- Voor elke klant die er in de Customers-tabel aanwezig is een eigen pipeline starten

Allereerst halen we hier data op uit een SQL-database, en doen een filtering op die data binnen de ADF pipeline.

1. Klik bij Datasets op **Dataset Actions** en op **New Dataset**.

2. Zoek op **SQL** en kies **Azure SQL Database**. Klik vervolgens op **Continue**.
3. Noem de Dataset als volgt: **DS_asql_SalesLT_Customers_Training** en kies als linked service de **LS_sqldb_source**.
4. Kies bij **Table** voor **SalesLT.Customer** en klik op **OK**.
5. Klik bij Pipelines op **Pipeline Actions** en op **New Pipeline**.
6. Noem de pipeline: **PL_Filter_SalesPersonal_Training**.
7. Uit de lijst met **Activities**, klik op de optie **General**. Klik en sleep **Lookup** op het canvas.
8. Noem de lookup: **Lookup_SalesPersonal**.
9. Ga naar de tab **Settings** en kies bij **Source dataset** de **DS_asql_SalesLT_Customers_Training**.
10. Zet het vinkje uit bij **First row only**.
11. Klik bij **Use query** op de optie **query** en type of plak de volgende code:


```
SELECT COUNT(*) AS Registered_Customers, SalesPerson
FROM [SalesLT].[Customer]
GROUP BY SalesPerson
```
12. Uit de lijst met **Activities**, klik op de optie **Iteration & conditionals**. Klik en sleep **filter** op het canvas.
13. Sleep het **groene blokje** van de Lookup naar de filter activity. Zodat deze aan elkaar zijn verbonden.
14. Geef de **filter** de naam **Best seller**.
15. klik op de tab **Settings** en klik bij **Items** op het veld er naast en vervolgens op **Add dynamic content**.
16. Klik onder **Activity outputs** op **Lookup_SalesPersonal value array** en klik op **OK**.
17. Klik bij **Condition** op het veld er naast en vervolgens op **Add dynamic content**.
18. Ga naar **Functions**, en type of plak de volgende code:


```
@greaterOrEquals(item().Registered_Customers,100)
```
19. Klik op de **Blauwe knop** met de tekst **Publish all** en vervolgens op de knop **Publish**.
20. Klik op **Debug** en wacht tot de pipeline klaar is, bekijk de resultaten door op de **Output** van de **Best seller** stap te kijken.

Opdracht 3 - Dynamische pipelines/ Datasets

Tot nu toe hebben we alle tabellen stuk voor stuk ingeladen, met eigen datasets en pipelines. Dat is echter (gelukkig) niet nodig in ADF: je kunt je pipelines en datasets *dynamisch* maken. Dat houdt in:

- Je voegt parameters toe aan je dataset (bijvoorbeeld voor de tabelnaam)
- Je laat het *schema* leeg. ADF doet nu een *schema infer*, wat betekent dat het schema op het moment van uitvoeren bepaald wordt.

- Bij het gebruiken van de dataset geef je de benodigde parameters mee.

Zo kun je bijvoorbeeld een lijst op te halen tabellen uitlezen uit een CSV-bestand of SQL-configuratietabel, waarna je ze met een ForEach-loop één voor één uitleest.

1. Klik bij Datasets op **Dataset Actions** en op **New Dataset**.
2. Zoek op **SQL** en kies **Azure SQL Database**. Klik vervolgens op **Continue**.
3. Noem de Dataset als volgt: **DS_asql_sqldb_SourceTables_Training** en kies als linked service de **LS_sqldb_source**.
4. Laat de **Table name** leeg en klik op **OK**.
5. Herhaal stap 1 t/m 4 voor de **sqldb-target** en noem de Dataset als volgt: **DS_asql_sqldb_TargetTables_training**.
6. Wanneer de Dataset voor de **sqldb-target** is aangemaakt, ga naar de tab **Parameters** en klik op **New**.
7. Geef de parameter de naam **TargetTableName**.
8. Ga naar de tab **Connection** en vink **Edit** aan. Vul in het eerste veld **Stg** in en klik op het 2e veld en vervolgens op **Add dynamic content**.
9. Kies uit de lijst de parameter genaamd: **TargetTableName** en klik vervolgens op **OK**.
10. Klik bij Pipelines op **Pipeline Actions** en op **New Pipeline**.
11. Noem de Pipeline als volgt: **PL_copy_deltaload_Training**.
12. Uit de lijst met **Activities**, klik op de optie **General**. Klik en sleep **Lookup** op het canvas.
13. Geef de **lookup** de naam **Lookup_SourceTables**, klik vervolgens op de **Settings** tab en kies bij **Source dataset** de **DS_aqsl_sqldb_SourceTables_training**.
14. Klik bij **Use query** de optie **Query** aan. Klik vervolgens in het query veld en plak of type de volgende query:

```
SELECT
TABLE_SCHEMA AS Table_Schema,
TABLE_NAME AS Table_Name
FROM INFORMATION_SCHEMA.TABLES
WHERE TABLE_SCHEMA = 'SalesLT' AND TABLE_TYPE = 'BASE TABLE'
```

Indien **First row only** staat aangevinkt, zet deze uit.

15. Uit de lijst met **Activities**, klik op de optie **Iteration & Conditionals**. Klik en sleep **ForEach** op het canvas.
16. Sleep het **Groene blokje** van **Lookup_SourceTables** naar de **ForEach** zodat deze opeenvolgend aan elkaar zijn verbonden.

17. Geef de **ForEach** de naam **ForEachTable** en klik op de tab **Settings**.
18. Klik op het vlak naast **Items** en vervolgens op **Add dynamic content**. Kies voor **Lookup_SourceTables** **value array**.
19. klik op de **plus** in de **ForEachTable**. Kies nu de activity **Copy Data**.
20. Geef de **Copy data** de naam **Copy Tables** en klik vervolgens op de tab **Source** en kies voor de **Source dataset** de **DS_aqsl_sqldb_SourceTables_training**.
21. Klik bij **Use query** de optie **Query** aan. Klik vervolgens in het query veld en vervolgens op **Add dynamic content** en type of plak:

```
SELECT * FROM @{item().Table_Schema}.@{item().Table_Name}
```

22. Klik op de tab **Sink** en kies vervolgens de **DS_aqsl_sqldb_TargetTables_training** linked service, klik daarna op het veld naast **TargetTableName** gevolgd door **Add dynamic content** en plak of type:

```
@item().Table_Name
```

23. Klik op het veld naast **Pre-copy script** en vervolgens op **Add dynamic content** en plak of type:

```
TRUNCATE TABLE Stg.@{item().Table_Name}
```

24. Klik op de **Blauwe knop** met de tekst **Publish all** en vervolgens op de knop **Publish**.
25. Klik op **Debug** en wacht tot de pipeline klaar is.

Einde Lab 8

Inhoudsopgave

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