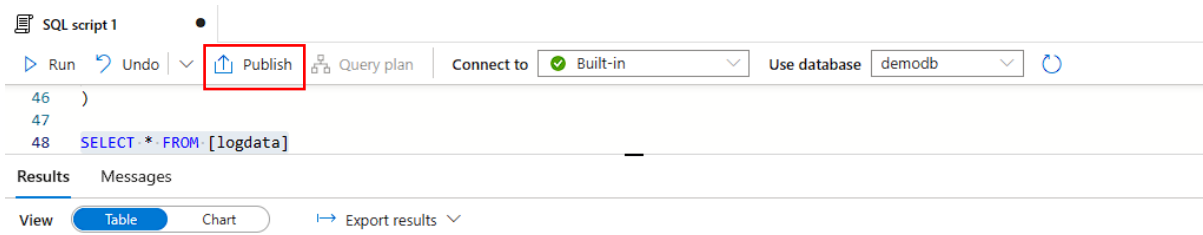


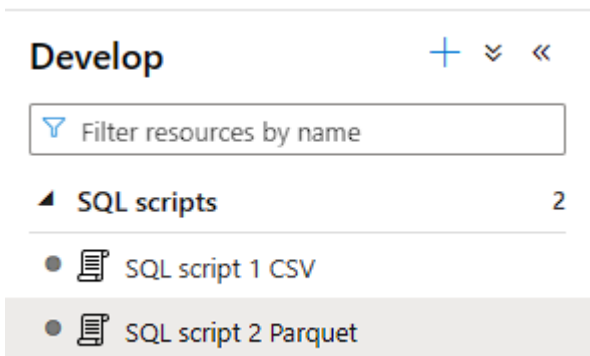


## Using External Tables for Parquet

1. In the previous lab we created an external table for our CSV file but this time we will be creating an external table for our Parquet file.
2. So, in the parquet file the information is the same, the only difference is that parquet stores the data in binary format. So, we are going to make use of an external table to be able to read the data in this file.
3. Now in your Azure portal go to your Synapse Workspace.
4. Now in our workspace we have an external table for CSV. So, you have to click on Publish.

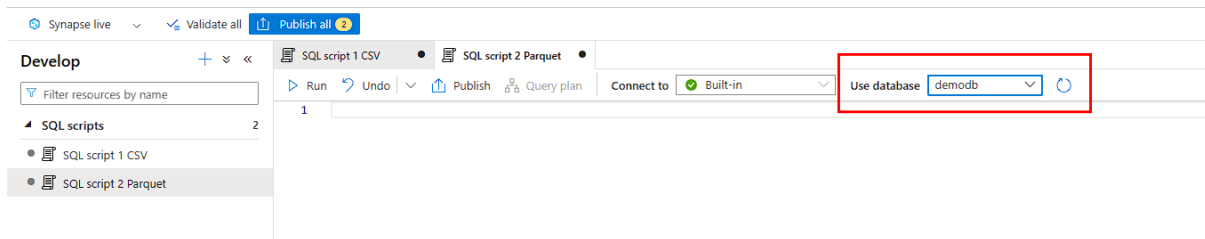


5. Once it is published again you are going to create a new SQL Script. In the Develop section click on the Plus icon and you will get the option to create a new script click on it.
6. Below you can see that you have now a second SQL script on which you are going to work.

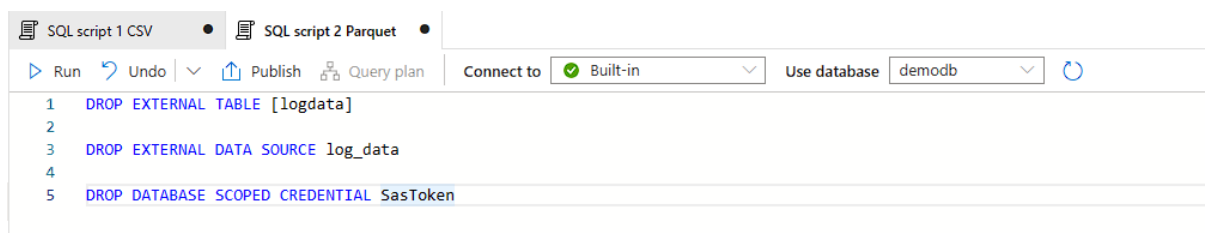


7. We can now copy some of the aspects that we used in our previous script file. Now, do we need to create the database again? No, we can create now our new external table within the same database. So, we don't need to execute the CREATE DATABASE command. Can we make use of the same database scope credential? So yes. If you are working on the same day, definitely you can use the same shared access signature. Since I'm still working on the same day, I will use the same shared access signature and I can still use that same master encryption key.
8. In case you come the next day and your shared access signature is not working because it is already expired, you can go ahead and create the SAS token, the shared access signature again, and then specify a database scope credential.

9. Now just in case if you are working the next day and your SAS token has expired for that first you have to run the command to Drop your external table and data then you have to Drop the Database Scoped Credentials that you have mentioned.
10. For that first you have to use your database from last SQL script then run these commands shown below.



11. Then you have to run these commands one by one shown below.



12. After that you go back to your storage account and create a new SAS token then you have to paste them in the command again and run the command.
13. Then you are going to issue the command to create an external data source.
14. After that you have to issue the command to create an external file format.
15. **You can get the code file from GitHub.**

```

CREATE DATABASE SCOPED CREDENTIAL SasToken
WITH IDENTITY='SHARED ACCESS SIGNATURE'
, SECRET = 'sv=2022-11-02&ss=b&srt=sco&sp=rwd&se=2024-03-31T13:12:23Z&st=2024-03-31T05:12:23Z&spr=https&sig=ybtncr1N185FSP%2FheD8i

CREATE EXTERNAL DATA SOURCE log_data_parquet
WITH (
    LOCATION = 'https://sqlstorage1010.blob.core.windows.net/parquet',
    CREDENTIAL = SasToken
)

CREATE EXTERNAL FILE FORMAT parquetfile
WITH (
    FORMAT_TYPE = PARQUET,
    DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
);

```

16. Then just create your external table.

```

22 DROP EXTERNAL TABLE [log_data_parquet]
23
24 CREATE EXTERNAL TABLE [log_data_parquet]
25 (
26     [Correlationid] [varchar](200) NULL,
27     [Operationname] [varchar](200) NULL,
28     [Status] [varchar](100) NULL,
29     [Eventcategory] [varchar](100) NULL,
30     [Level] [varchar](100) NULL,
31     [Time] [varchar](500) NULL,
32     [Subscription] [varchar](200) NULL,
33     [Event initiated by] [varchar](1000) NULL,
34     [Resourcetype] [varchar](1000) NULL,
35     [Resourcegroup] [varchar](1000) NULL,
36     [Resource] [varchar](2000) NULL)
37 WITH (
38     LOCATION = '/log.parquet',
39     DATA_SOURCE = log_data_parquet,
40     FILE_FORMAT = parquetfile
41 )

```

17. Now use the Select statement to view your table.

18. And below you can view your table.

SQL script 1 CSV

SQL script 2 Parquet

Run

Undo

Publish

Query plan

Connect to

Built-in

Use database

demodb

42

43 SELECT \* FROM [log\_data\_parquet]

Results

Messages

View

Table

Chart

Export results

Search

Correlationid	Operationname	Status	Eventcategory	Level	Time	Subscription	Event initiated ...	Resour
99fe9c3a-e36e-...	Update SQL da...	Succeeded	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	Create Deploy...	Started	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	Create Deploy...	Accepted	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	Registers the M...	Started	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	(NULL)
99fe9c3a-e36e-...	Registers the M...	Succeeded	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	(NULL)
99fe9c3a-e36e-...	Update SQL ser...	Started	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	'audit' Policy ac...	Succeeded	Policy	Warning	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	'auditifNotExist...	Started	Policy	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	Update SQL ser...	Accepted	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	Update SQL ser...	Started	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	Update SQL ser...	Started	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	Update Server ...	Started	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	Update SQL da...	Started	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	Update Server ...	Succeeded	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	Update SQL ser...	Succeeded	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros
99fe9c3a-e36e-...	'auditifNotExist...	Started	Policy	Informational	2023-04-25T03:...	6912d7a0-bc28...	(NULL)	Micros

00:00:06 Query executed successfully.

19. Also, when it comes to now working with external tables. So, even though your data is in the Data Lake Gen 2 storage account because you have now defined an external table with a predefined schema, you can use your SQL-based commands to work with the data. So, for example, let's say you want to select the operation name, the count of the operation name from the log data, and the parquet log data table. So, now we can use SQL-based statements against our data using the external table.

```
SELECT [Operationname] , COUNT([Operationname]) as [Operation Count]
FROM [logdata]
GROUP BY [Operationname]
ORDER BY [Operation Count]
```

20. Below you can see that you are getting the data accordingly.

The screenshot shows a SQL query execution interface. At the top, there are tabs for 'SQL script 1 CSV' and 'SQL script 2 Parquet'. Below the tabs, there are buttons for 'Run', 'Undo', 'Publish', and 'Query plan'. The 'Connect to' dropdown is set to 'Built-in', and the 'Use database' dropdown is set to 'demodb'. The SQL query is displayed in a text area, and the results are shown in a table below. The table has two columns: 'Operationname' and 'Operation Count'. The results are as follows:

Operationname	Operation Count
Add or Update Certificate	1
Smart Detection - Failure Anoma...	1
Update Web App Source Control...	2
Create an Azure Automation job	2
Create or Update Microsoft Polic...	2
Register with the Provider	2
Validate Tagged Traffic Consumer	2
Create or update custom role def...	2
Create or Update Tagged Traffic ...	2
ValidateStorageConfiguration	2
Delete Virtual Machine Extension	2
Register the subscription to Azur...	2
Delete custom role definition	2
Registers the Microsoft SQL Data...	2
Delete record set of type A	2

At the bottom, there is a status bar indicating '00:00:03 Query executed successfully.'

21. Now click on publish to save your SQL Script.