



## Creating External Table for Parque File

1. In this lab, we will create the External Table for Parque-based file inside our Azure Synapse Workspace.
2. First, we need to create a new SQL Script using the Develop tab as we did in the previous lab, using the properties section from the right pane change the name of the SQL Script.

### Properties

General      Related (0)

Name \*

02-ExternalTableParque

3. In the new SQL Script, we need to change the database to our appdb. Then we will run the commands to drop the external table and external data source.

```
02-ExternalTablePar... ●
Run Undo Publish Query plan Connect to Built-in Use database appdb
1 DROP EXTERNAL TABLE ActivityLog;
2 DROP EXTERNAL DATA SOURCE srcActivityLog;
```

4. Then we have to create a Master key. So, use the command given to you from the scripts that you have.

```
02-ExternalTablePar... ×
Run Undo Publish Query plan Connect to Built-in
1 DROP EXTERNAL TABLE ActivityLog;
2 DROP EXTERNAL DATA SOURCE srcActivityLog;
3
4 CREATE MASTER KEY ENCRYPTION BY PASSWORD = 'P@ssword@123'
```

5. After that we have to create Scoped credentials using the Shared Access Signature from our Data Lake or Storage Account.
6. Go to the Storage account and look for the Shared access signature then choose the same options as shown below in the snapshot. Scroll down to the bottom and click on Generate SAS.

The screenshot shows the 'Shared access signature' configuration page for a storage account. The 'Shared access signature' tab is selected. The configuration includes:

- Allowed services:** Blob (checked), File, Queue, Table
- Allowed resource types:** Service, Container, Object (checked)
- Allowed permissions:** Read, Write, Delete, List (checked), Add, Create, Update, Process, Immutable storage, Permanent delete
- Blob versioning permissions:** Enables deletion of versions
- Start and expiry date/time:**
  - Start: 12/13/2024, 10:00:45 PM
  - End: 12/14/2024, 6:00:45 AM
- Allowed IP addresses:** For example, 168.1.5.65 or 168.1.5.65-168.1.5.70
- Allowed protocols:** HTTPS only (radio button selected)

7. You will get 3 values here, but we need to use the SAS token Copy this.

The screenshot shows the 'Generate SAS and connection string' section. It includes:

- Connection string:** BlobEndpoint=https://thestorageaccount1201.blob.core.windows.net/?QueueEndpoint=https://thestorageaccount1201.queue.core.windows.net/?FileEndpoint=https://thestorageaccount1201.file.core.windows.net/
- SAS token:** sv=2022-11-02&ss=b&srt=s&sp=rlacyx&se=2024-12-14T00:30:45Z&st=2024-12-13T16:30:45Z&spr=https&sig=o8bpuE7k%2BORa%2BijJT%2BFS%2F2%2FZjaQmVtHa5ekHijwtedU%...
- Blob service SAS URL:** https://thestorageaccount1201.blob.core.windows.net/?sv=2022-11-02&ss=b&srt=s&sp=rlacyx&se=2024-12-14T00:30:45Z&st=2024-12-13T16:30:45Z&spr=https&sig=o8bpuE7k%2...

8. Then use the SAS token and paste it into the Secret value given in the script.

```

02-ExternalTablePar...
Run Undo Publish Query plan Connect to Built-in Use database appdb
1 DROP EXTERNAL TABLE ActivityLog;
2 DROP EXTERNAL DATA SOURCE srcActivityLog;
3
4 CREATE MASTER KEY ENCRYPTION BY PASSWORD = 'P@ssword@123'
5
6 CREATE DATABASE SCOPED CREDENTIAL sasToken
7 WITH IDENTITY = 'SHARED ACCESS SIGNATURE',
8 SECRET='sv=2022-11-02&ss=b&srt=s&sp=rlacyx&se=2024-12-14T00:30:45Z&st=2024-12-13T16:30:45Z&spr=https&sig=
```

9. Now you have to create the External data source and file format for your parquet-based file and create the external table.

```

10  CREATE EXTERNAL DATA SOURCE srcActivityLog
11  WITH
12  (
13      LOCATION='https://datalake50000.blob.core.windows.net/rawdata',
14      CREDENTIAL=sasToken
15  )
16
17  CREATE EXTERNAL FILE FORMAT parquetFileFormat WITH
18  (
19      FORMAT_TYPE=PARQUET,
20      DATA_COMPRESSION='org.apache.hadoop.io.compress.SnappyCodec'
21  )

```

02-ExternalTablePar...

Run Undo Publish Query plan Connect to Built-in

```

16  CREATE EXTERNAL TABLE ActivityLog
17  (
18      [Correlationid] varchar(200),
19      [Operationname] varchar(300),
20      [Status] varchar(100),
21      [Eventcategory] varchar(100),
22      [Level] varchar(100),
23      [Time] varchar(100),
24      [Subscription] varchar(200),
25      [Eventinitiatedby] varchar(1000),
26      [Resourcetype] varchar(300),
27      [Resourcegroup] varchar(1000),
28      [Resource] varchar(2000))
29  WITH (
30      LOCATION='/ActivityLog01.parquet',
31      DATA_SOURCE=srcActivityLog,
32      FILE_FORMAT=parquetFileFormat
33  )

```

10. Now if you run the Select from statement you will be able to see the data inside your table.

44    SELECT \* FROM ActivityLog;

Results    Messages

View    **Table**    Chart    [Export results](#)

Search

Correlationid	Operationname	Status	Eventcategory	Level	Time	Subscription
c5cf5ce4-d26f...	Delete Storage ...	Succeeded	Administrative	Informational	2024-05-30T15:...	387407e5-94af
c5cf5ce4-d26f...	Delete website	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af
c5cf5ce4-d26f...	Delete Namesp...	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af
c5cf5ce4-d26f...	Delete website	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af
c5cf5ce4-d26f...	Delete SQL dat...	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af
c5cf5ce4-d26f...	Delete Storage ...	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af
c5cf5ce4-d26f...	Delete hosting ...	Started	Administrative	Informational	2024-05-30T14:...	387407e5-94af

00:00:08 Query executed successfully.

## 😊 Working with Multiple Parquet Files

- Now we will try to work with multiple parquet-based files. For that first, we need to upload a new parquet file which you can find with the rest of the files you get.

Upload    Add Directory    Refresh    Rename    Delete    Change tier    Acquire lease    Break lease    Give feedback

Authentication method: Access key ([Switch to Microsoft Entra user account](#))  
Location: data

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
ActivityLog01.csv	12/13/2024, 12:26:21...	Hot (Inferred)		Block blob	1.91 MiB	Available
ActivityLog01.parquet	12/13/2024, 10:15:21...	Hot (Inferred)		Block blob	230.51 KiB	Available
ActivityLog02.parquet	12/13/2024, 10:29:03...	Hot (Inferred)		Block blob	294.49 KiB	Available

- Now if we do a counting of rows in our table then we have 5436 rows inside our single parquet file.

```
44  SELECT COUNT(*) FROM ActivityLog;
```

```
45
```

```
--
```

Results Messages

View

Table

Chart

Export results ▾

🔍 Search

(No column name)

5436

3. First, we need to drop our table and then create our table again using the wildcard character so, that both of our files will be used. In the location part, we will define the wildcard character. You will have the script you can use to understand it more accurately.

```
39  WITH (
40    LOCATION='*.parquet',
41    DATA_SOURCE=srcActivityLog,
42    FILE_FORMAT=parquetFileFormat
43 )
```

4. Then after creating our table if we run the counting rows command again we will see that the number of rows has been doubled.

```
45  SELECT COUNT(*) FROM ActivityLog;
```

```
46
```

```
--
```

Results Messages

View

Table

Chart

Export results ▾

🔍 Search

(No column name)

12381

5. Once you are done, click on Publish all to save everything.