

Project – Writing Queries Using Synapse SQL Serverless Pool

Overview:

In this project we will be reading CSV, JSON, Parquet format files along with Folder and sub-folders in Synapse SQL Serverless Pool from ADLS Gen 2.

ADLS Storage:

Created a container Project in the storage account, created a directory called ‘Data’ and uploaded the documents with Parquet, CSV, JSON formats.

Home > synapsestorageadls12 | Containers >

project

Container

Search

◊ ◀

Upload

Add Directory

Refresh

Rename

Delete

Change tier

Acquire lease

Break lease

Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: project / Data

Search blobs by prefix (case-sensitive)

Name	Modified	Access tier
<input type="checkbox"/> [.]		
<input type="checkbox"/> business_employment_parquet		
<input type="checkbox"/> business_employment_csv		
<input type="checkbox"/> business_employment_1.parquet	08/09/2024, 17:24:52	Hot (Inferred)
<input type="checkbox"/> business_employment.csv	04/09/2024, 13:41:18	Hot (Inferred)
<input type="checkbox"/> Employees.json	09/09/2024, 16:26:28	Hot (Inferred)
<input type="checkbox"/> Sample.json	06/09/2024, 16:52:43	Hot (Inferred)
<input type="checkbox"/> sample1.parquet	08/09/2024, 11:43:34	Hot (Inferred)

Reading a CSV file after uploading to ADLS account and running auto generated 100 rows:

```
SELECT
TOP 100 *
FROM
    OPENROWSET(
```

```

BULK 'https://synapsestorageadls12.dfs.core.windows.net/project/Data/**',
FORMAT = 'CSV',
PARSER_VERSION = '2.0'
) AS [result]

```

C1	C2	C3	C4	C5	C6
Series_reference	Period	Data_value	Suppressed	STATUS	UNITS
BDCQ.SEA1AA	2011.06	80078	(NULL)	F	Number
BDCQ.SEA1AA	2011.09	78324	(NULL)	F	Number
BDCQ.SEA1AA	2011.12	85850	(NULL)	F	Number
BDCQ.SEA1AA	2012.03	90743	(NULL)	F	Number

To correct the header row in CSV in place of C1, C2, C3...

```

SELECT
TOP 100 *
FROM
OPENROWSET(
BULK
'abfss://project@synapsestorageadls12.dfs.core.windows.net/Data/business_employment.csv',
FORMAT = 'CSV',
PARSER_VERSION = '2.0',
HEADER_ROW = TRUE,
FIELDTERMINATOR = ',',
ROWTERMINATOR = '\n'
) AS [result]

```

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS
BDCQ.SEA1AA	2011.06	80078	(NULL)	F	Number
BDCQ.SEA1AA	2011.09	78324	(NULL)	F	Number
BDCQ.SEA1AA	2011.12	85850	(NULL)	F	Number
BDCQ.SEA1AA	2012.03	90743	(NULL)	F	Number
BDCQ.SEA1AA	2012.06	81780	(NULL)	F	Number
BDCQ.SEA1AA	2012.09	79261	(NULL)	F	Number
BDCQ.SEA1AA	2012.12	87793	(NULL)	F	Number
BDCQ.SEA1AA	2013.03	91571	(NULL)	F	Number

Given Datatypes for all the columns:

```

SELECT *
FROM
OPENROWSET(
BULK
'abfss://project@synapsestorageadls12.dfs.core.windows.net/Data/business_employment.csv',
FORMAT = 'CSV',
PARSER_VERSION = '2.0',
HEADER_ROW = TRUE,
FIELDTERMINATOR = ';',
ROWTERMINATOR = '\n'
)
WITH (
Series_reference NVARCHAR(100),
Period NVARCHAR(50),
Data_value FLOAT,
Suppressed NVARCHAR(10),
STATUS NVARCHAR(50),

```

```

UNITS NVARCHAR(50),
Magnitude FLOAT,
Subject NVARCHAR(100),
[Group] NVARCHAR(100),
Series_title_1 NVARCHAR(100),
Series_title_2 NVARCHAR(100),
Series_title_3 NVARCHAR(100),
Series_title_4 NVARCHAR(100),
Series_title_5 NVARCHAR(100)
)AS [result]

```

Search									
Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Series_title_1
BDCQ.SEA1AA	2011.06	80078	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2011.09	78324	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2011.12	85850	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2012.03	90743	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2012.06	81780	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2012.09	79261	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2012.12	87793	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2013.03	91571	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2013.06	81687	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2013.09	81471	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2013.12	93950	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
BDCQ.SEA1AA	2014.03	97208	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs

Created External Two Data Sources:

CREATE EXTERNAL DATA SOURCE Project

WITH(

LOCATION = 'abfss://project@synapsestorageadls12.dfs.core.windows.net',

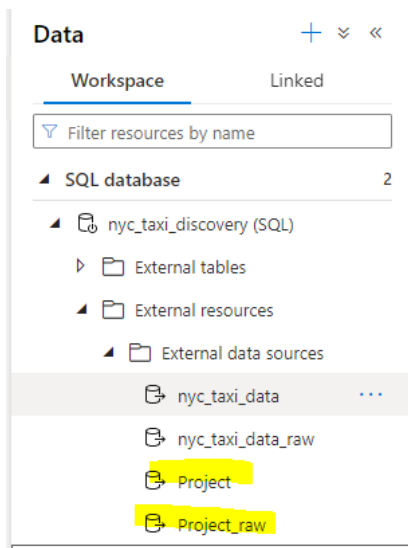
)

CREATE EXTERNAL DATA SOURCE Project_raw

WITH(

LOCATION = 'abfss://project@synapsestorageadls12.dfs.core.windows.net/Data'

)



Reading the CSV now with using the Data Source:

After creating the external data source, we are using the below query

Data (abfss://project@synapsestorageadls12.dfs.core.windows.net/Data) will be picked from data source BULK 'business_employment.csv'- file name will pull from here

SELECT

*

FROM

OPENROWSET(

BULK 'business_employment.csv',

DATA_SOURCE = 'Project_raw',

FORMAT = 'CSV',

PARSER_VERSION = '2.0',

FIRSTROW = 2,

FIELDTERMINATOR = ',',

ROWTERMINATOR = '\n'

)

WITH (

Series_reference NVARCHAR(100),
Period NVARCHAR(50),
Data_value FLOAT,
Suppressed NVARCHAR(10),
STATUS NVARCHAR(50),
UNITS NVARCHAR(50),
Magnitude FLOAT,
Subject NVARCHAR(100),
[Group] NVARCHAR(100),
Series_title_1 NVARCHAR(100),
Series_title_2 NVARCHAR(100),
Series_title_3 NVARCHAR(100),
Series_title_4 NVARCHAR(100),
Series_title_5 NVARCHAR(100)

)AS [result]

ResultsMessages

ViewTableChartExport results

Search

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Si
BDCQ.SEA1AA	2011.06	80078	(NULL)	F	Number	0	Business Data ...	Industry by em...	Fi
BDCQ.SEA1AA	2011.09	78324	(NULL)	F	Number	0	Business Data ...	Industry by em...	Fi
BDCQ.SEA1AA	2011.12	85850	(NULL)	F	Number	0	Business Data ...	Industry by em...	Fi
BDCQ.SEA1AA	2012.03	90743	(NULL)	F	Number	0	Business Data ...	Industry by em...	Fi
BDCQ.SEA1AA	2012.06	81780	(NULL)	F	Number	0	Business Data ...	Industry by em...	Fi
BDCQ.SEA1AA	2012.09	79261	(NULL)	F	Number	0	Business Data ...	Industry by em...	Fi
BDCQ.SEA1AA	2012.12	87793	(NULL)	F	Number	0	Business Data ...	Industry by em...	Fi
BDCQ.SEA1AA	2013.03	91571	(NULL)	F	Number	0	Business Data ...	Industry by em...	Fi
BDCQ.SEA1AA	2013.06	81687	(NULL)	F	Number	0	Business Data ...	Industry by em...	Fi
BDCQ.SEA1AA	2013.09	81471	(NULL)	F	Number	0	Business Data ...	Industry by em...	Fi
-----	-----	-----	-----	-	-----	-	-----	-----	-

Using Collate for UTF 8 in CSV

SELECT

*

FROM

OPENROWSET(

BULK 'business_employment.csv',

DATA_SOURCE = 'Project_raw',

FORMAT = 'CSV',

PARSER_VERSION = '2.0',

FIRSTROW = 2,

FIELDTERMINATOR = ',',

ROWTERMINATOR = '\n'

)

WITH (

Series_reference NVARCHAR(100) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

Period NVARCHAR(50) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

Data_value FLOAT,

Suppressed NVARCHAR(10) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

STATUS NVARCHAR(50) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

UNITS NVARCHAR(50) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

Magnitude FLOAT,

Subject NVARCHAR(100) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

[Group] NVARCHAR(100) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

Series_title_1 NVARCHAR(100) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

Series_title_2 NVARCHAR(100) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

Series_title_3 NVARCHAR(100) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

Series_title_4 NVARCHAR(100) COLLATE Latin1_General_100_CI_AI_SC_UTF8,

Series_title_5 NVARCHAR(100) COLLATE Latin1_General_100_CI_AI_SC_UTF8

)AS [result]

Search											
Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Series_title_1	Series_title_2	Series_title_3
BDCQ.SEA1AA	2011.06	80078	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2011.09	78324	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2011.12	85850	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2012.03	90743	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2012.06	81780	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2012.09	79261	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2012.12	87793	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2013.03	91571	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2013.06	81687	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual

If we do not want to use the Collate function multiple times, we can use the Data base as below and specify it UTF8.

USE nyc_taxi_discovery;

--Since we had the database setup used it other wise, we CREATE DATABASE----

ALTER DATABASE nyc_taxi_discovery COLLATE Latin1_General_100_CI_AI_SC_UTF8;

Removed the "0" values from a column using EscapeChar:

SELECT

*

FROM

OPENROWSET(

BULK 'business_employment.csv',

DATA_SOURCE = 'Project_raw',

FORMAT = 'CSV',

PARSER_VERSION = '2.0',

HEADER_ROW = TRUE,

ESCAPECHAR = '0'

) AS vendor;

Results Messages

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

Search

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Series_title_1	Series_title_2	Series_title
BDCQ,SEA1AA	2011.06	8078	(NULL)	F	Number	,Business Data Collection - BDC	Industry by em...	Filled jobs	Agriculture, For...	Actual	(NULL)
BDCQ,SEA1AA	2011.09	78324	(NULL)	F	Number	,Business Data Collection - BDC	Industry by em...	Filled jobs	Agriculture, For...	Actual	(NULL)
BDCQ,SEA1AA	2011.12	8585,	F	Number	,Business Data Colle...	Industry by employment varia...	Filled jobs	Agriculture, For...	Actual	(NULL)	(NULL)
BDCQ,SEA1AA	2012.03	9743	(NULL)	F	Number	,Business Data Collection - BDC	Industry by em...	Filled jobs	Agriculture, For...	Actual	(NULL)
BDCQ,SEA1AA	2012.06	8178,	F	Number	,Business Data Colle...	Industry by employment varia...	Filled jobs	Agriculture, For...	Actual	(NULL)	(NULL)
BDCQ,SEA1AA	2012.09	79261	(NULL)	F	Number	,Business Data Collection - BDC	Industry by em...	Filled jobs	Agriculture, For...	Actual	(NULL)

Reading a JSON:

Using Fieldterminator below is given data to read JSON:

SELECT *

FROM OPENROWSET(

BULK 'Sample.json',

DATA_SOURCE = 'Project_raw',

FORMAT = 'CSV',

PARSER_VERSION = '1.0',

FIELDTERMINATOR = '0x0b',

FIELDQUOTE = '0x0b',

ROWTERMINATOR = '0x0a'

)

WITH

(

jsonDoc NVARCHAR(MAX)

) AS payment_type;

Results	Messages
View	Table Chart
Export results	
Search	
jsonDoc	
{	
"fruit": "Apple",	
"size": "Large",	
"color": "Red"	
}	

Reading a JSON using Cast and arranging values in tabular form:

```
SELECT CAST(JSON_VALUE(jsonDoc, '$.fruit') AS VARCHAR(15)) Apple,
CAST(JSON_VALUE(jsonDoc, '$.size') AS VARCHAR(15)) Large,
CAST(JSON_VALUE(jsonDoc, '$.color') AS VARCHAR(15)) Red
FROM OPENROWSET(
BULK 'Sample.json',
DATA_SOURCE = 'Project_raw',
FORMAT = 'CSV',
PARSER_VERSION = '1.0',
FIELDTERMINATOR = '0x0b',
FIELDQUOTE = '0x0b',
ROWTERMINATOR = '0x0a'
)
WITH
```

```
(
jsonDoc NVARCHAR(MAX)
) AS [result];
```

Apple	Large	Red

Reading another JSON file with Cross Apply Showing the JSON Document and Tabular format side by side:

```
SELECT *
FROM OPENROWSET(
BULK 'Employees.json',
DATA_SOURCE = 'Project_raw',
FORMAT = 'CSV',
PARSER_VERSION = '1.0',
FIELDTERMINATOR = '0x0b',
FIELDQUOTE = '0x0b',
ROWTERMINATOR = '0x0a'
)
WITH
(
jsonDoc NVARCHAR(MAX)
) AS [result]
CROSS APPLY OPENJSON(jsonDoc)
WITH(
name VARCHAR(20) ,
```













email VARCHAR(20) '\$.email'

);

Results Messages		
View	Table	Chart
Export results		
Search		
jsonDoc	name	email
{"name":"Shyam","email":"shyamjaiswal@..."	Shyam	shyamjaiswal@gmail.c
{"name":"Bob","email":"bob32@gmail.com"}	Bob	bob32@gmail.com
{"name":"Jai","email":"jai87@gmail.com"}	Jai	jai87@gmail.com

Reading Folder & Subfolders:

Reading Folder using *:

<input type="checkbox"/>  business_employment_csv		
name	Name	
<input type="checkbox"/>  [.]	<input type="checkbox"/>  [.]	
<input type="checkbox"/>  Folder=1	<input type="checkbox"/>  Week=1	
<input type="checkbox"/>  Folder=2	<input type="checkbox"/>  Week=2	
		<input type="checkbox"/>  business_employment_1.csv
	Name	
	<input type="checkbox"/>  [.]	
	<input type="checkbox"/>  Week=3	
	<input type="checkbox"/>  Week=4	
		<input type="checkbox"/>  business_employment_3.csv

SELECT *

FROM

OPENROWSET(

```

BULK 'business_employment_csv/Folder=1/Week=1/*.csv',
DATA_SOURCE = 'Project_raw',
FORMAT = 'CSV',
PARSER_VERSION = '2.0',
HEADER_ROW = TRUE
) AS [result]

```

Results Messages

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

Search

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Series_title_1	Series_title_2	Series_title_3
BDCQ,SEA1AA	2011.06	80078	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2011.09	78324	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2011.12	85850	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.03	90743	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.06	81780	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.09	79261	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.12	87793	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2013.03	91571	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual

Reading Subfolder by using "***"

```

SELECT *
FROM
OPENROWSET(
BULK 'business_employment_csv/Folder=1/**',
DATA_SOURCE = 'Project_raw',
FORMAT = 'CSV',
PARSER_VERSION = '2.0',
HEADER_ROW = TRUE
) AS [result]

```

Results

Messages

View

Table

Chart

Export results

Search

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Series_title_1	Series_title_2	Series_title_3
BDCQ,SEA1AA	2013.09	81471	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2013.12	93950	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2014.03	97208	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2014.06	85879	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2014.09	84447	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2014.12	95075	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2015.03	98202	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2015.06	87987	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual

Reading subfolder for only Week1:

```
SELECT *
FROM
OPENROWSET(
BULK 'business_employment_csv/Folder=1/Week=1/*.csv',
DATA_SOURCE = 'Project_raw',
FORMAT = 'CSV',
PARSER_VERSION = '2.0',
HEADER_ROW = TRUE
) AS [result]
```

Results

Messages

View

Table

Chart

Export results

Search

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Series_title_1	Series_title_2	Series_title_3
BDCQ.SEA1AA	2011.06	80078	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2011.09	78324	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2011.12	85850	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2012.03	90743	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2012.06	81780	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2012.09	79261	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2012.12	87793	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ.SEA1AA	2013.03	91571	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual

00:00:00

Query executed successfully.

Reading all folders & Subfolders present:

```
SELECT *  
  
FROM  
  
OPENROWSET(  
  
BULK 'business_employment_csv/Folder=*/Week=*/*.csv',  
  
DATA_SOURCE = 'Project_raw',  
  
FORMAT = 'CSV',  
  
PARSER_VERSION = '2.0',  
  
HEADER_ROW = TRUE  
  
) AS [result]
```

Results Messages

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

Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Series_title_1	Series_title_2	Series_title_3
BDCQ,SEA1AA	2011.09	78324	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2011.12	85850	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.03	90743	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.06	81780	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.09	79261	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.12	87793	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2013.03	91571	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2013.06	81687	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual

Reading all folders & subfolders using filename:

```
SELECT  
  
TOP 100  
  
result.filename() AS file_name,  
  
result.*  
  
FROM  
  
OPENROWSET(  
  
BULK 'business_employment_csv/Folder=*/Week=*/*.csv',
```

```

DATA_SOURCE = 'Project_raw',

FORMAT = 'CSV',

PARSER_VERSION = '2.0',

HEADER_ROW = TRUE

) AS [result]

```

Results Messages

View Table Chart Export results

Search

file_name	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Series_title
business_employment_2.csv	BDCQ,SEA1AA	2015.03	98202	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
business_employment_2.csv	BDCQ,SEA1AA	2015.06	87987	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
business_employment_2.csv	BDCQ,SEA1AA	2015.09	84529	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
business_employment_2.csv	BDCQ,SEA1AA	2015.12	96848	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
business_employment_3.csv	BDCQ,SEA1AA	2016.03	99291	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
business_employment_3.csv	BDCQ,SEA1AA	2016.06	88716	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
business_employment_3.csv	BDCQ,SEA1AA	2016.09	85933	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
business_employment_3.csv	BDCQ,SEA1AA	2016.12	96540	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs
business_employment_3.csv	BDCQ,SEA1AA	2017.03	88884	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs

00:00:00 Query executed successfully.

Reading the count of files using group by and order by :

```

SELECT

result.filename() AS file_name,

COUNT(1) AS record_count

FROM

OPENROWSET(

BULK 'business_employment_csv/Folder=*/Week=*/*.csv',

DATA_SOURCE = 'Project_raw',

FORMAT = 'CSV',

PARSER_VERSION = '2.0',

HEADER_ROW = TRUE

) AS [result]

GROUP BY result.filename()

ORDER BY result.filename();

```


Results Messages	
View	Table Chart Export results
Search	
file_name	record_count
business_employment_1.csv	9
business_employment_2.csv	10
business_employment_3.csv	10
business_employment_4.csv	10

Reading all count of files and filtering it with only 1 & 2 folder using 'Where' and 'IN':

```

SELECT
result.filename() AS file_name,
COUNT(1) AS record_count
FROM
OPENROWSET(
BULK 'business_employment_csv/Folder=*/Week=*/*.csv',
DATA_SOURCE = 'Project_raw',
FORMAT = 'CSV',
PARSER_VERSION = '2.0',
HEADER_ROW = TRUE
) AS [result]
WHERE result.filename() IN ('business_employment_1.csv', 'business_employment_2.csv')

```

```
GROUP BY result.filename()  
ORDER BY result.filename();
```

Results Messages	
View	Table Chart
Export results	
Search	
file_name	record_count
business_employment_1.csv	9
business_employment_2.csv	10

Reading folders using filepath to get the count per folder and week:

```
SELECT  
result.filepath(1) AS Folder,  
result.filepath(2) AS Week,  
COUNT(1) AS record_count  
FROM  
OPENROWSET(  
BULK 'business_employment_csv/Folder=*/Week=*/*.csv',  
DATA_SOURCE = 'Project_raw',  
FORMAT = 'CSV',  
PARSER_VERSION = '2.0',  
HEADER_ROW = TRUE  
) AS [result]  
GROUP BY result.filepath(1), result.filepath(2)  
ORDER BY result.filepath(1), result.filepath(2);
```

Results Messages		
View	Table	Chart
Export results		
Search		
Folder	Week	record_count
1	1	9
1	2	10
2	3	10
2	4	10

Reading folders using filepath to get the count per folder and week and filtering for Folder 1 & Week 2 data:

```

SELECT
result.filepath(1) AS Folder,
result.filepath(2) AS Week,
COUNT(1) AS record_count
FROM
OPENROWSET(
BULK 'business_employment_csv/Folder=*/Week=*/*.csv',
DATA_SOURCE = 'Project_raw',
FORMAT = 'CSV',
PARSER_VERSION = '2.0',
HEADER_ROW = TRUE
) AS [result]
WHERE result.filepath(1) = '1'

```

```

AND result.filepath(2) IN ('2')

GROUP BY result.filepath(1), result.filepath(2)

ORDER BY result.filepath(1), result.filepath(2);

```

Results Messages		
View	Table	Chart
Export results		
Search		
Folder	Week	record_count
1	2	10

Reading Parquet file:

```

SELECT
    TOP 100 *
FROM
    OPENROWSET(
        BULK 'https://synapsestorageadls12.dfs.core.windows.net/project/Data/sample1.parquet',
        FORMAT = 'PARQUET'
    ) AS [result]

```

(Or)

```

SELECT
    TOP 100 *
FROM
    OPENROWSET(
        BULK 'sample1.parquet',
        DATA_SOURCE = 'Project_raw',
    )

```

FORMAT = 'PARQUET'

) AS [result]

Results		Messages
View		Table Chart
Export results		
Search		
column0	column1	
first	last	
Jorge	Frank	
Hunter	Moreno	
Esther	Guzman	
Dennis	Stephens	
Nettie	Franklin	
Stanley	Gibson	
Eugenia	Greer	
Jeffery	Delgado	
00:00:01 Query executed successfully.		

Reading Parquet using 'with' for datatypes for columns:

SELECT

TOP 100 *

FROM

OPENROWSET(

BULK 'sample1.parquet',

DATA_SOURCE = 'Project_raw',

FORMAT = 'PARQUET'

)

WITH (

column0 NVARCHAR(15),

column1 NVARCHAR(15)

) AS [result]

View

Table

Chart

↗ Export results ▾

🔍 Search

column0	column1
first	last
Jorge	Frank
Hunter	Moreno
Esther	Guzman
Dennis	Stephens
Nettie	Franklin
Stanley	Gibson

Reading just one column name i.e., first name:

```

SELECT
    TOP 100 *
FROM
    OPENROWSET(
        BULK 'sample1.parquet',
        DATA_SOURCE = 'Project_raw',
        FORMAT = 'PARQUET'
    )
    WITH (
        column0 NVARCHAR(15)
    ) AS [result]

```

Results	Messages
View	Table Chart
Export results	
Search	
column0	
first	
Jorge	
Hunter	
Esther	
Dennis	
Nettie	
Stanley	

Reading Parquet from a specific folder and all files from Folder 1, Week 1:

```

SELECT TOP 100 *
FROM OPENROWSET (
    BULK 'business_employment_parquet/Folder=1/Week=1/*.parquet',
    DATA_SOURCE = 'Project_raw',
    FORMAT = 'PARQUET'
) AS Result

```

Results Messages											
View Table Chart Export results											
Search											
Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Series_title_1	Series_title_2	Series_title_3
BDCQ,SEA1AA	2011.06	80078	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2011.09	78324	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2011.12	85850	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.03	90743	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.06	81780	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual
BDCQ,SEA1AA	2012.09	79261	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	Agriculture, For...	Actual

Reading the parquet file with 'Filename':

SELECT TOP 100

result.filename() AS file_name,

result.*

FROM OPENROWSET (

BULK 'business_employment_parquet/Folder=1/Week=1/*.parquet',

DATA_SOURCE = 'Project_raw',

FORMAT = 'PARQUET'

) AS result

Results Messages											
View Table Chart Export results											
Search											
file_name	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject	Group	Series_title_1	Series_title_2
business_employment_1.parquet	BDCQ,SEA1AA	2011.06	80078	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	
business_employment_1.parquet	BDCQ,SEA1AA	2011.09	78324	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	
business_employment_1.parquet	BDCQ,SEA1AA	2011.12	85850	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	
business_employment_1.parquet	BDCQ,SEA1AA	2012.03	90743	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	
business_employment_1.parquet	BDCQ,SEA1AA	2012.06	81780	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	
business_employment_1.parquet	BDCQ,SEA1AA	2012.09	79261	(NULL)	F	Number	0	Business Data ...	Industry by em...	Filled jobs	

Reading subfolder level data in parquet using **:

SELECT TOP 100

result.filename() AS file_name,

result.*

```
FROM OPENROWSET (  
    BULK 'business_employment_parquet/**',  
    DATA_SOURCE = 'Project_raw',  
    FORMAT = 'PARQUET'  
) AS result
```

Results Messages

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

Search

file_name	Series_reference	Period	Data_value	Suppressed	STATUS	UNITS	Magnitude	Subject
business_employment_1.parquet	BDCQ.SEA1AA	2011.06	80078	(NULL)	F	Number	0	Business Data ...
business_employment_1.parquet	BDCQ.SEA1AA	2011.09	78324	(NULL)	F	Number	0	Business Data ...
business_employment_1.parquet	BDCQ.SEA1AA	2011.12	85850	(NULL)	F	Number	0	Business Data ...
business_employment_1.parquet	BDCQ.SEA1AA	2012.03	90743	(NULL)	F	Number	0	Business Data ...
business_employment_1.parquet	BDCQ.SEA1AA	2012.06	81780	(NULL)	F	Number	0	Business Data ...
business_employment_1.parquet	BDCQ.SEA1AA	2012.09	79261	(NULL)	F	Number	0	Business Data ...

Reading the count of Parquet file with Filepath:

```
SELECT result.filepath(1) AS Folder,  
    result.filepath(2) AS Week,  
    COUNT(1) AS record_count  
FROM OPENROWSET (  
    BULK 'business_employment_parquet/Folder=*/Week=*/*.parquet',  
    DATA_SOURCE = 'Project_raw',  
    FORMAT = 'PARQUET'  
) AS result
```

```
GROUP BY result.filepath(1), result.filepath(2)  
ORDER BY result.filepath(1), result.filepath(2);
```

Results Messages		
View	Table	Chart
↗ Search		
Folder	Week	record_count
1	1	9
1	2	10
2	3	10
2	4	10

Reading the count of Parquet file with Filepath for a specific folder and subfolder using Where and IN:

```

SELECT result.filepath(1) AS Folder,
       result.filepath(2) AS Week,
       COUNT(1) AS record_count
FROM OPENROWSET (
    BULK 'business_employment_parquet/Folder=*/Week=*/*.parquet',
    DATA_SOURCE = 'Project_raw',
    FORMAT = 'PARQUET'
) AS result
WHERE result.filepath(1) = '1'
     AND result.filepath(2) IN ('2')
GROUP BY result.filepath(1), result.filepath(2)
ORDER BY result.filepath(1), result.filepath(2);

```

Results Messages		
View	Table Chart	Export results
Search		
Folder	Week	record_count
1	2	10

Creating External File Format (using parser version 2.0):

```
IF NOT EXISTS (SELECT * FROM sys.external_file_formats WHERE name = 'csv_file_format')
CREATE EXTERNAL FILE FORMAT csv_file_format
WITH (
    FORMAT_TYPE = DELIMITEDTEXT, FORMAT_OPTIONS (
        FIELD_TERMINATOR = ','
        , STRING_DELIMITER = '"' , First_Row = 2 , USE_TYPE_DEFAULT = FALSE , Encoding = 'UTF8' ,
        PARSER_VERSION = '2.0' )
);
```

Creating External File Format (using parser version 1.0):

```
IF NOT EXISTS (SELECT * FROM sys.external_file_formats WHERE name
='csv_file_format_pv1')
CREATE EXTERNAL FILE FORMAT csv_file_format_pv1
WITH (
    FORMAT_TYPE = DELIMITEDTEXT,
    FORMAT_OPTIONS (
        FIELD_TERMINATOR = ','
        , STRING_DELIMITER = '"'
        , First_Row = 2
        , USE_TYPE_DEFAULT = FALSE
```

```
, Encoding = 'UTF8'  
, PARSER_VERSION = '1.0' )  
);
```

Creating External File Format parquet:

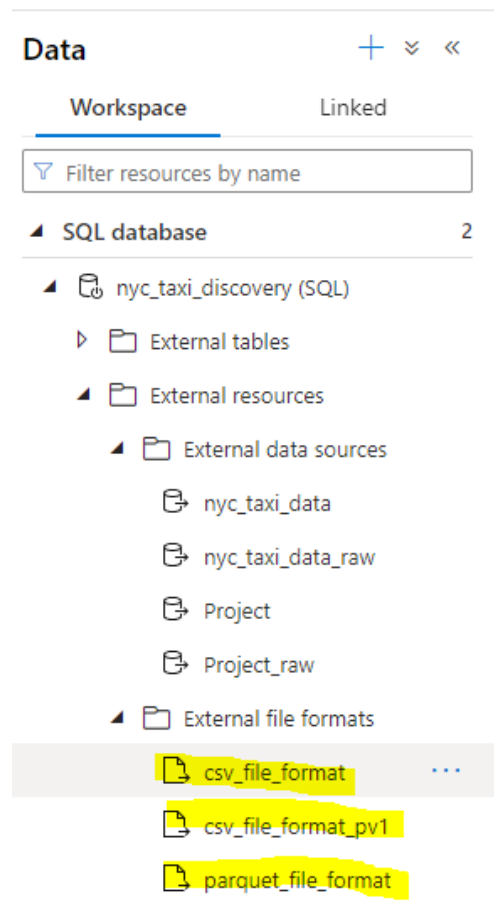
```
IF NOT EXISTS (SELECT * FROM sys.external_file_formats WHERE name = 'parquet_file_format')  
CREATE EXTERNAL FILE FORMAT parquet_file_format
```

```
WITH (
```

```
FORMAT_TYPE = PARQUET,
```

```
DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
```

```
);
```



Error Log:

1. While querying JSON changed Row terminator from 0x0a & 0x0b values.
2. Made sure the keys and values are in quotations for JSON file, removed extra spaces from the file, made necessary corrections and uploaded the file to storage account to read.
3. Tried a couple of variation with the JSON file and executed automated query.