Dynamic number of columns with ADLS Gen1

Last updated by | Jackie Huang | Jan 4, 2022 at 12:24 AM PST

Achieving dynamic number of columns with ADLS Gen1 as sink

18 de março de 2020
12:37

Issue:

Customer has a source file in ADLS Gen1 that will have multiple rows and each row can have from 19 to 251 columns (for instance), as:
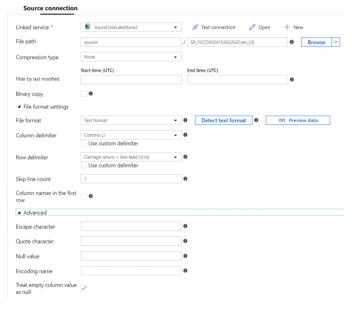

If we don't define schema, it fails complaining that the row 5 has more than 19 columns.
If we force a schema with 251 columns, it fails in the first row saying it has only 19 columns.

The new dataset of ADLS Gen1 doesn't support dynamic number of columns, as the old dataset.

Resolution:

We should manually change the dataset to use the old ADLS Gen1 file type.

1. Open the dataset json payload and change manually to the old dataset from "DelimitedText" to "AzureDataLakeStoreFile"
2. Replace 'schema' by 'structure' in the json payload. Save and close
3. Uncheck binary copy
4. In File Path, select a file with header for the maximum number of columns that you would like to have (for instance, 251). If you don't have such file, just create one in Excel with one row with the header
5. Import the schema
6. Go to connection again and choose the original source file, without any header and with dynamic number of columns
7. Check 'Treat empty column value as null'
8. Execute the copy

Azure Data Lake Storage Gen1
sourceDynamicColumns

	General	Connection	Schema	Parameters

Import schema | Preview data | + New column | ↻ Clear | 🗑 Delete

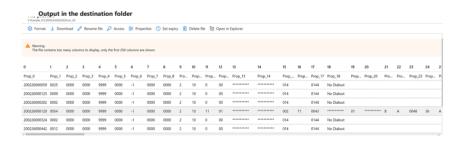|	COLUMN NAME	TYPE	DESCRIPTION
☰	Prop_0	abc String	
☰	Prop_1	abc String	
☰	Prop_2	abc String	
☰	Prop_3	abc String	
☰	Prop_4	abc String	
☰	Prop_5	abc String	
☰	Prop_6	abc String	
☰	Prop_7	abc String	
☰	Prop_8	abc String	
☰	Prop_9	abc String	
☰	Prop_10	abc String	
☰	Prop_11	abc String	
☰	Prop_12	abc String	
☰	Prop_13	abc String	
☰	Prop_14	abc String	
☰	Prop_15	abc String	
☰	Prop_16	abc String	
☰	Prop_17	abc String	
☰	Prop_18	abc String	
☰	Prop_19	abc String	
☰	Prop_20	abc String	
☰	Prop_21	abc String	
☰	Prop_22	abc String	
☰	Prop_23	abc String	

The copy goes through

	General	Parameters	Variables	Output

Pipeline run ID: **b030af21-2b96-46fb-b315-664903ea0a7b**

View debug run con...

NAME	TYPE	RUN START	DURATION	STATUS	INTEGRATION RUNTIME	RUN ID
copyDynamicColumns	Copy	2020-03-06T16:17:47.2694;	00:00:05	✅ Succeeded	DefaultIntegrationRuntime (West Europe)	e695f260-c9a3-4529-9077-6122606a3a19

Output in the destination folder

V1Sample_FCCD0024103022620.en_US

⚙ Format | ⬇ Download | ✏ Rename file | 🔑 Access | ⚙ Properties | 🕐 Set expiry | 🗑 Delete file | ⊞ Open in Explorer

⚠ Warning
The file contains too many columns to display, only the first 250 columns are shown.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	2
Prop_0	Prop_1	Prop_2	Prop_3	Prop_4	Prop_5	Prop_6	Prop_7	Prop_8	Pro...	Prop...	Prop...	Prop...	Prop_13	Prop_14	Prop_...	Prop...	Prop_17	Prop_18	Prop...	Prop_20	Pro...	Pro...	Prop_23	Prop...	P
20026000059	0025	0000	0000	9999	0000	-1	0000	0000	2	10	0	00	**********	**********	014		0144	No Dialout							
20026000125	0009	0000	0000	9999	0000	-1	0000	0000	2	10	0	00	**********	**********	014		0144	No Dialout							
20026000202	0002	0000	0000	9999	0000	-1	0000	0000	2	10	0	00	**********	**********	014		0144	No Dialout							
20026000120	0054	0000	0000	9999	0000	-1	0000	0000	2	10	11	01	**********	**********	002	T1	0043	**********	01	**********	B	A	0048	36	A
20026000324	0002	0000	0000	9999	0000	-1	0000	0000	2	10	0	00	**********	**********	014		0144	No Dialout							
20026000442	0012	0000	0000	9999	0000	-1	0000	0000	2	10	0	00	**********	**********	014		0144	No Dialout							

Additional Information:

- **References:** https://icm.ad.msft.net/imp/v3/incidents/details/178158467/home 🔗
- **Author:** ganespol
- **Reviewer:**
- **Keywords:** Copy, Schema, Column, ADLS

Created with Microsoft OneNote 2016.

How good have you found this content?

🙂 🙁