Option #2: Use external tables in dedicated SQL pool

2.1 Create a table in dedicated SQL pool to hold metadata information

create table exta.metadata (

Entity nvarchar(50),

Name nvarchar(200),

datatype NVARCHAR(50),

maxLength int,

precision int,

scale int,

 datatypeSQL nvarchar(100)

);

2.2 Create external file format and data source

IF NOT EXISTS (SELECT \* FROM sys.external\_file\_formats WHERE name = 'CSV\_StringDelimiterQuotes')

    CREATE EXTERNAL FILE FORMAT [CSV\_StringDelimiterQuotes]

    WITH ( FORMAT\_TYPE = DELIMITEDTEXT ,

           FORMAT\_OPTIONS (

             FIELD\_TERMINATOR = ',',

             STRING\_DELIMITER = '"',

             USE\_TYPE\_DEFAULT = FALSE

            ))

GO

IF NOT EXISTS (SELECT \* FROM sys.external\_data\_sources WHERE name = 'dataverse-\*\*\*\*-unqc6\*\*\*\*\*\*\*8d9b24e48af72819\_\*\*\*\*\*\_dfs\_core\_windows\_net')

    CREATE EXTERNAL DATA SOURCE [dataverse-\*\*\*\*\*-unqc6\*\*\*\*\*\*\*\*8d9b24e48af72819\_\*\*\*\*\*\*\*\_dfs\_core\_windows\_net]

    WITH (

        LOCATION = 'abfss://dataverse-\*\*\*\*\*-unqc\*\*\*\*\*\*\*\*8d9b24e48af72819@\*\*\*\*\*\*.dfs.core.windows.net',

        TYPE = HADOOP

    )

GO

2.3 Create a synapse pipeline

Graphical user interface, text, application, chat or text message

Description automatically generated

This pipeline will

* Truncate metadata table created in step 2.1 above
* Read metadata stored in dataverse’s model.json
* Iterate through each entity record in model.json
  + Read metadata information from json for specific entity
  + Map data types to relevant SQL data types..eg: guid to nvarchar(36) etc.

*\*\* test this mapping for different tables to make sure mapping works, Otherwise you may need to add new mapping*

* + Insert Entity, column name and data types in metadata table created in 2.1 above
  + Call script activity to execute dynamic script for create table

declare @entity nvarchar(200) = '@{item().name}';

Declare @CreateTableDDL nvarchar(max)

SELECT

@CreateTableDDL  =  'CREATE EXTERNAL TABLE exta.' + @entity +   + '(' + STRING\_AGG(CONVERT(NVARCHAR(max), + '[' + Name + '] ' +  datatypeSQL) , ',') + ')'

+ 'WITH (

    LOCATION = ''cr39c\_njo\_test2/2022-10.csv'',

    DATA\_SOURCE = [dataverse-njotest-unqc6111fa0426b48d9b24e48af72819\_njoadlsext1\_dfs\_core\_windows\_net],

    FILE\_FORMAT = [CSV\_StringDelimiterQuotes]

    )'

FROM exta.metadata

WHERE [Entity] =   'cr39c\_njo\_test2'

execute sp\_executesql  @CreateTableDDL;

*\*\*Replace highlighted portion with appropriate variables or strings. Since this is POC I have hard coded them for one table/entity only.*

\*\* Also, you will see If condition added in foreach activity in the pipeline I have created. You may remove that. Reason I have added that was to test it only for one table. If you don’t want to filter list of tables and apply logic for all tables, please remove “If activity” and place that logic directly under “for each activity”.

**Disclaimer: Sample code provided here is for demo purpose only. This requires modifications as per specific use case and requirements before deploying for production use.**