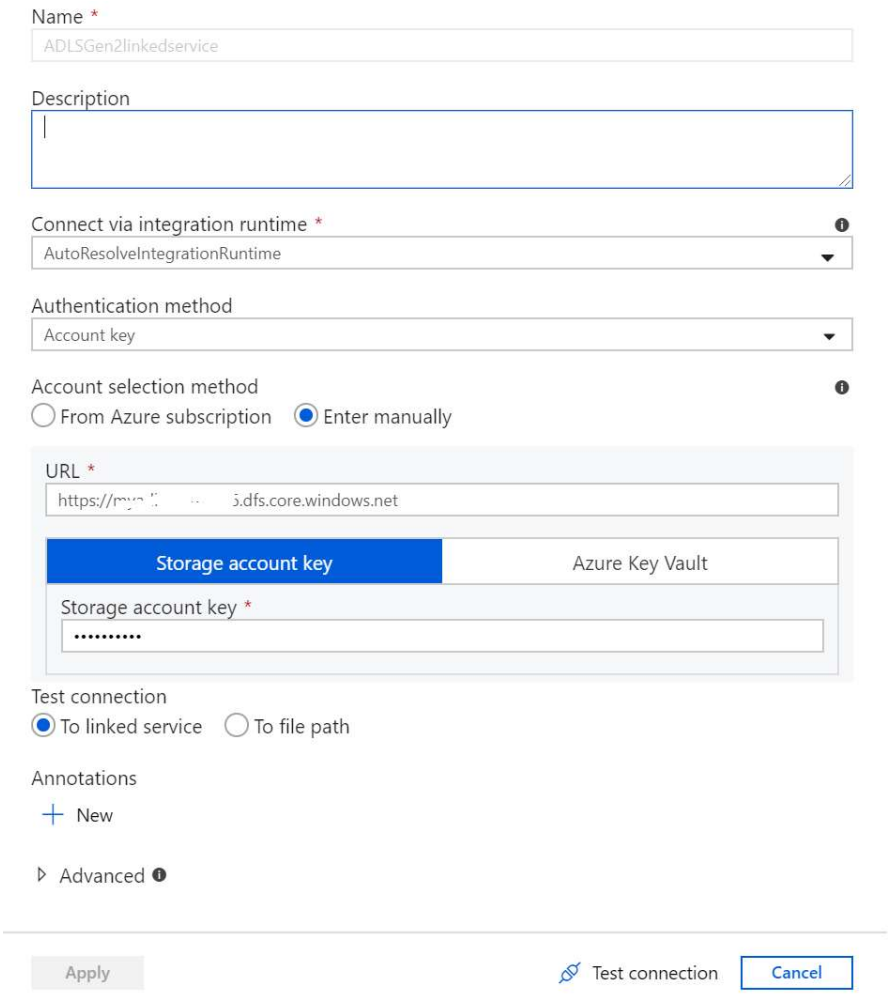
As we know Parquet Apache Parquet is designed for columnar storage, which is more efficiency to compare with CSV files during to query data.

Let's use this article to introduce how to Azure Data Factory to convert CSV file to Parquet.

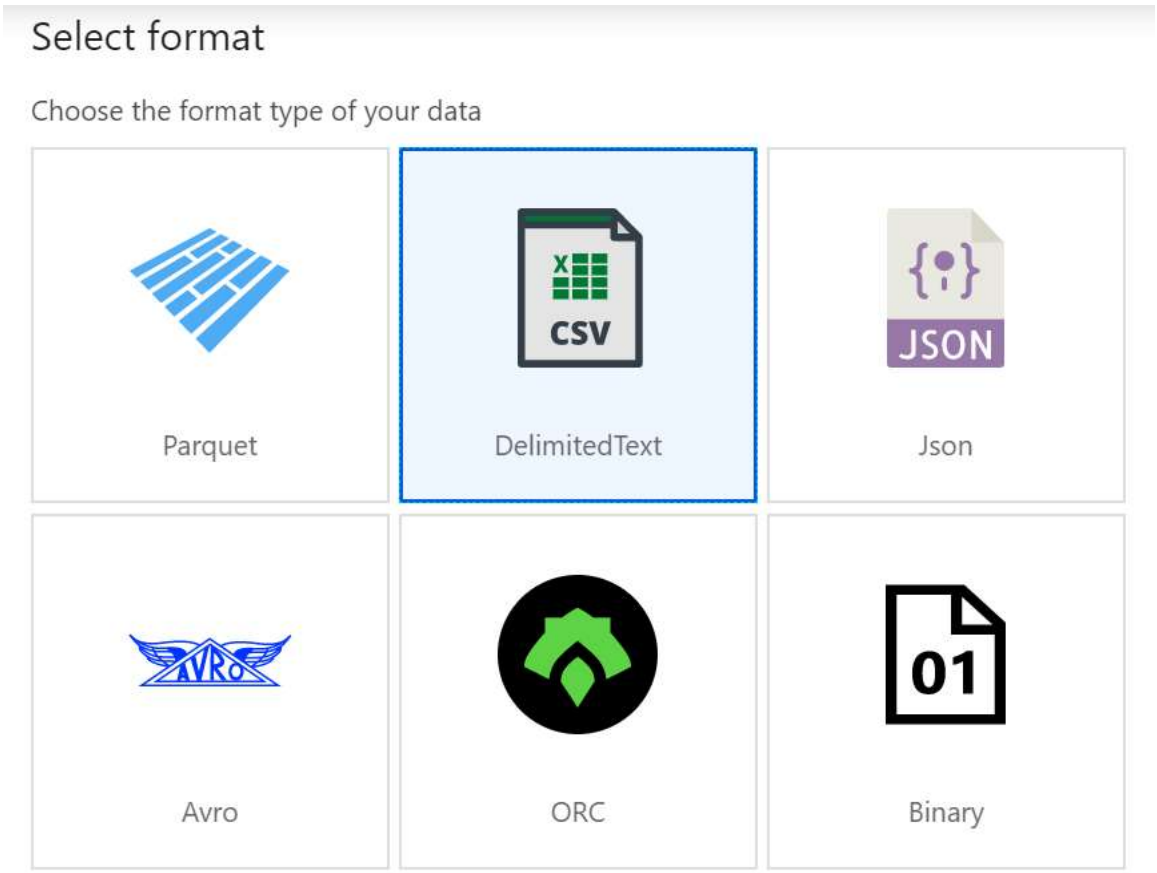
Step1: Create linked service for CSV and Parquet files from Azure Data Lake Gen2.

In this sample, linked service is shared since the both CSV and Parquet files will store one Azure Data Lake Gen 2 account.

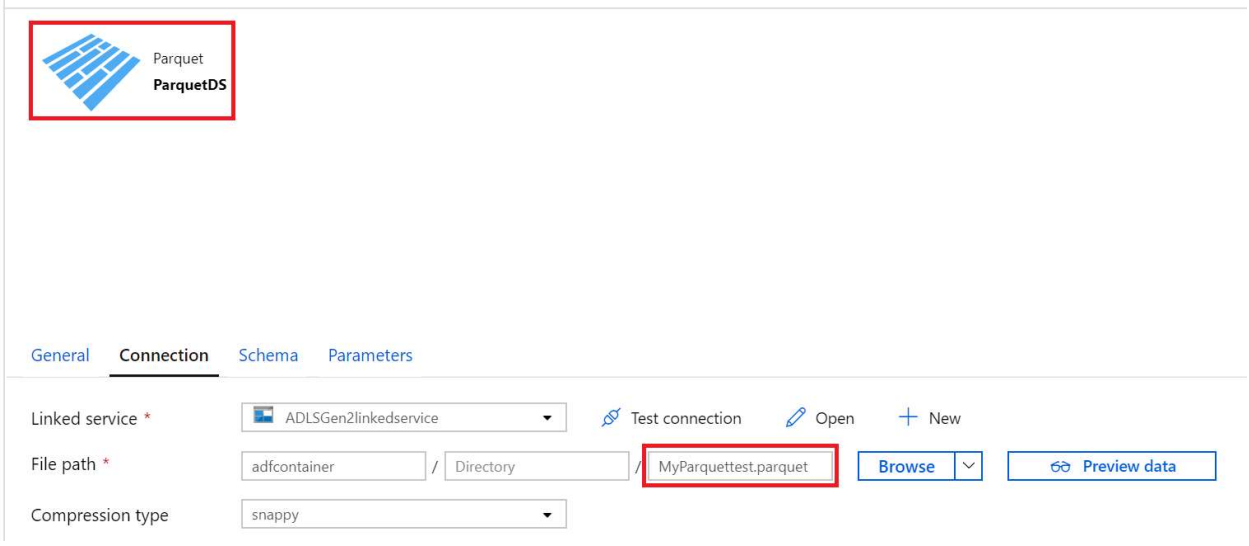


Step 2: Create Datasets for both CSV file and Parquet file.

Please note to create one dataset for CSV file, and another dataset for Parquet file, both datasets will be used in next step.

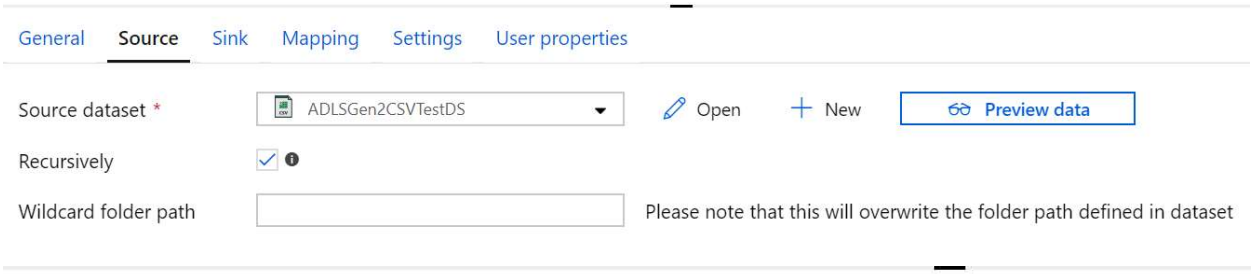


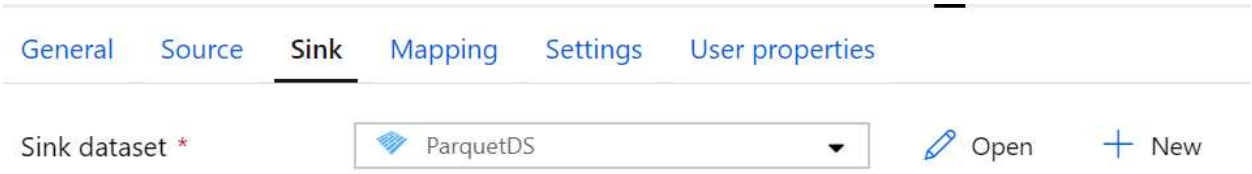
Choose CSV format in dataset (Parquet is similar)



View the Parquet dataset, it will show data format is Parquet

Step 3: Create a copy activity in pipeline, then set source to CSV dataset, and sink to Parquet dataset.





Step 4: Of course, the last one step is to run the pipeline, Cheers!

Post step: No need other tools, it's simple to use ADF to view Parquet data format output.

Preview data

Linked service: ADLSGen2linkedservice
Object: MyParquettest.parquet

Col1	Val1
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
9	10