Task 1: Export MySQL Data to CSV, Parquet, and Avro

1.) Set up Microsoft Azure

2.) Create an azure data factory and created linked services using ngrok (mysql_local_ngrok) to create a tunnel using ngrok which enables us to send the data from our local MYSQL Workbench to Microsoft Aure

3.) Created an azure storage account, used it to make another linked service azure_blob_storage. Used a contained called output-data for storing the files.


4.) Created a source dataset called MYSQL_Source and used the customers table in MYSQL Workbench to feed data into it using ngrok

5.) Then created a pipeline named CopyToAllFormats that did the conversion of MYSQL data to CSV, Parquet and Avro respectively when the trigger command was sent.

6.) Created Sink Datasets
  - CSV Output
    - Created a dataset: AzureBlob_CSV
    - Format: DelimitedText
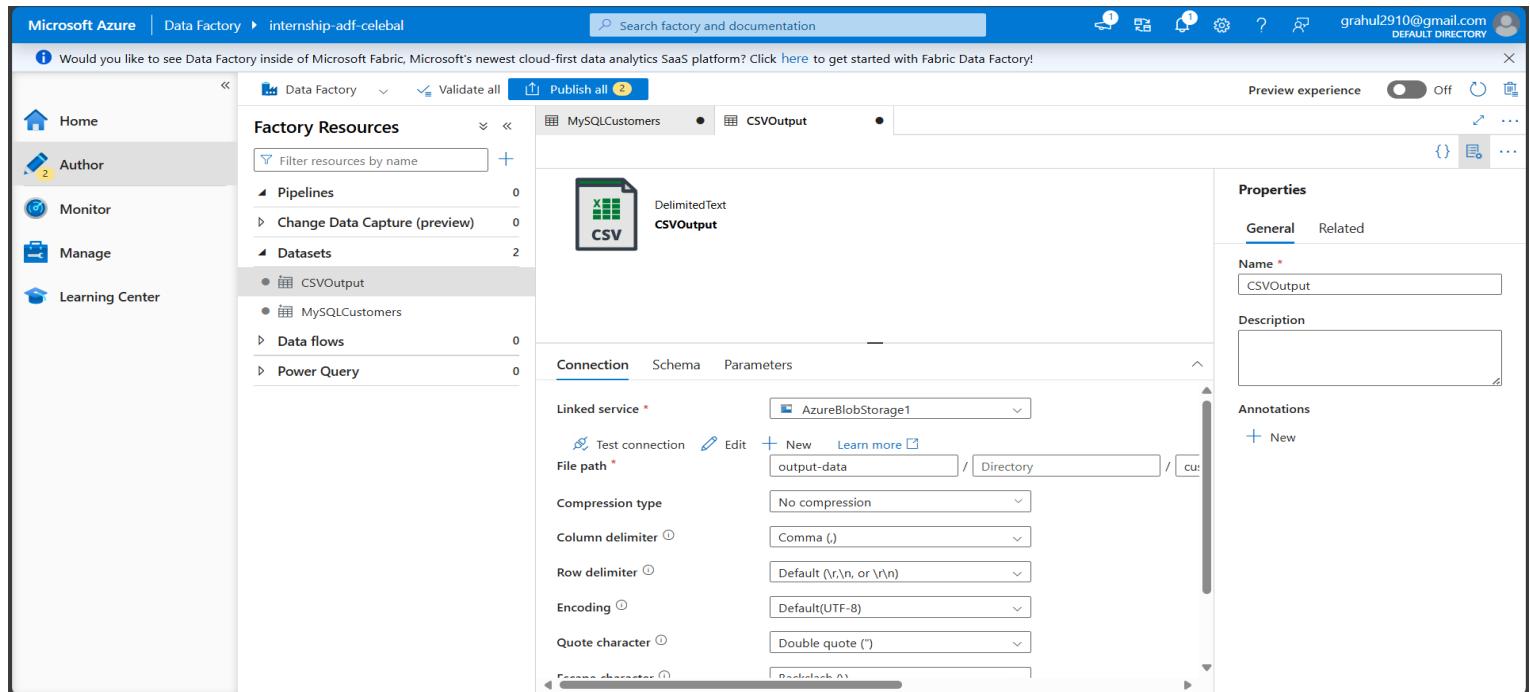
  - Parquet Output
    - Dataset: AzureBlob_Parquet
    - Format: Parquet
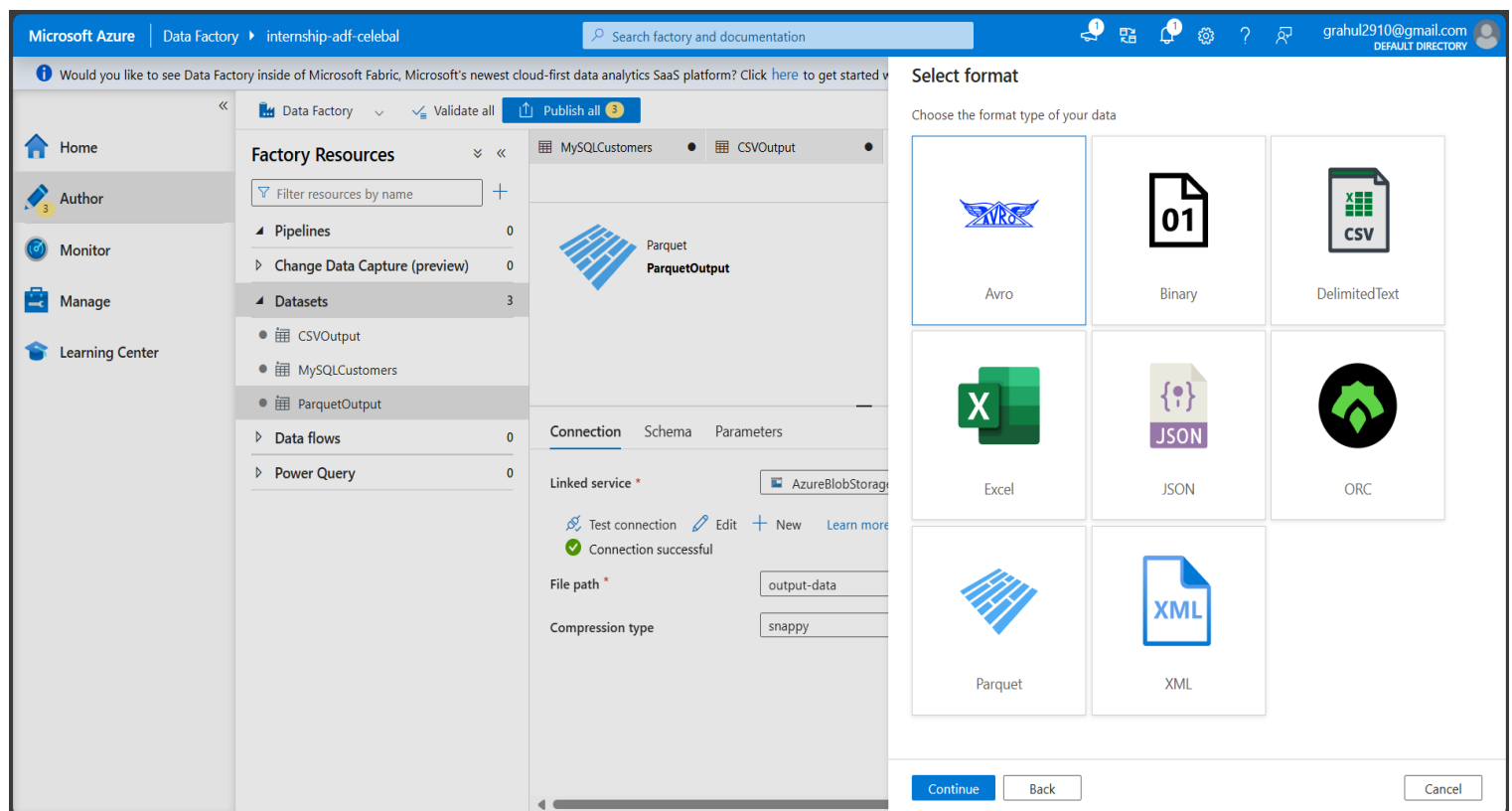
o Avro Output
   • Dataset: AzureBlob_Avro
   • Format: Avro

Output- The output-data file is the blob file that stores all the data that is converted from MYSQL to CSV, Avro and parguet.

Notes

- MySQL connection was done via Ngrok
- Output files are stored in Azure Blob Storage container 'output-data'