3. Hands on practice

Activity -1 – Source activity

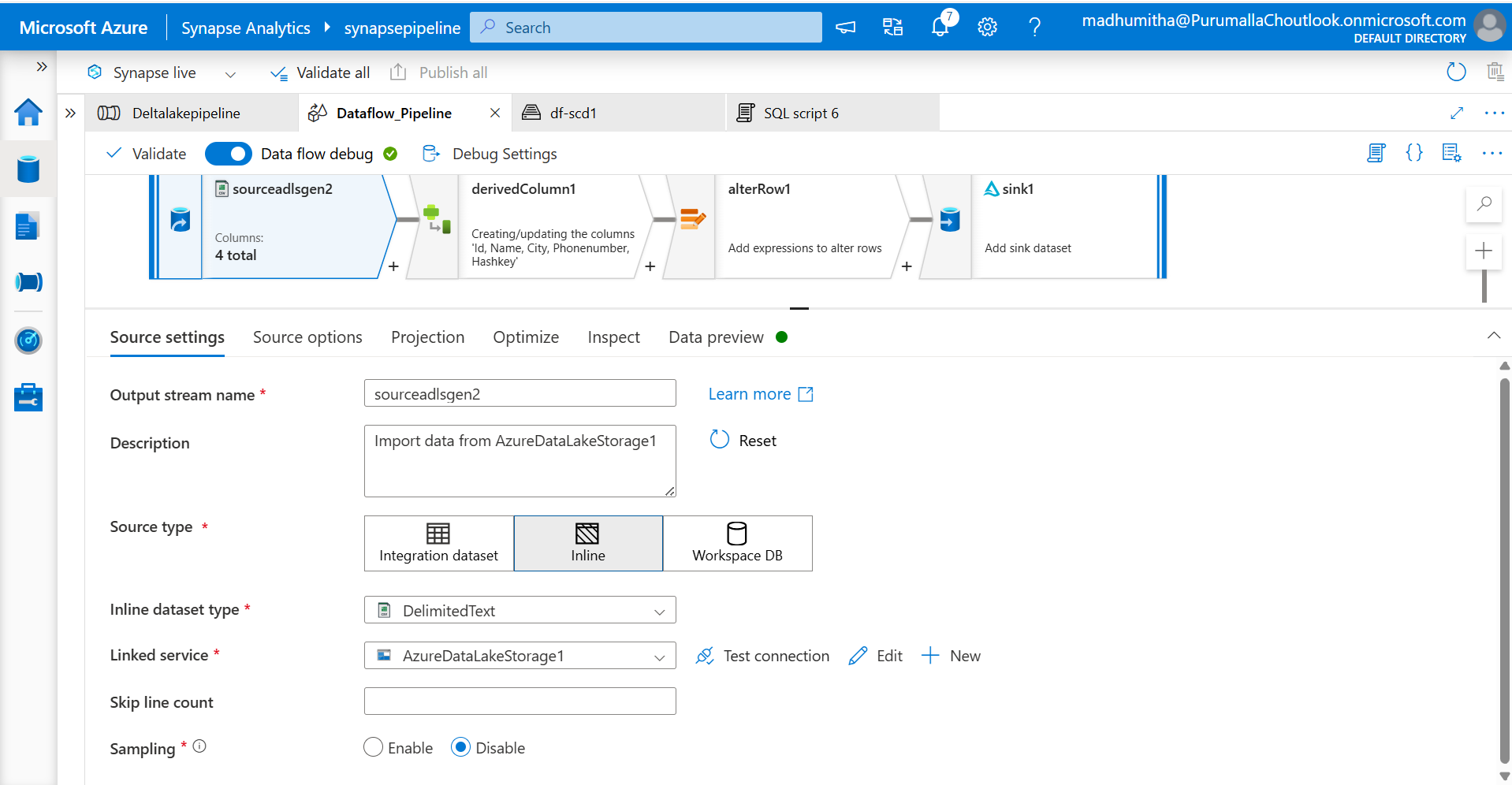
Step 1: Under source settings, name sourceadlsgen2 in O.P stream name.

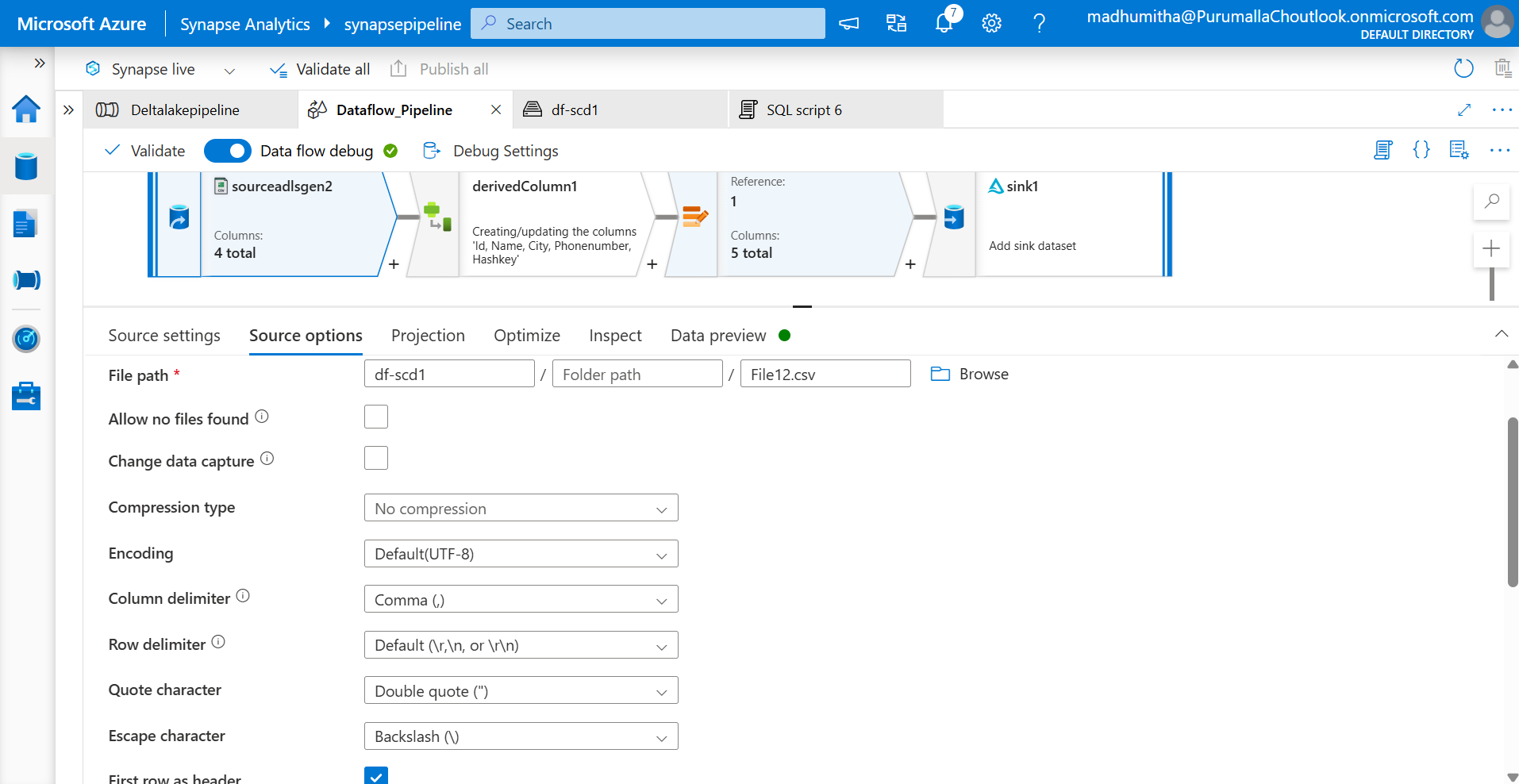
Step 2: Select the source type as Inline, Delimited text in inline dataset type and select existing linked service.

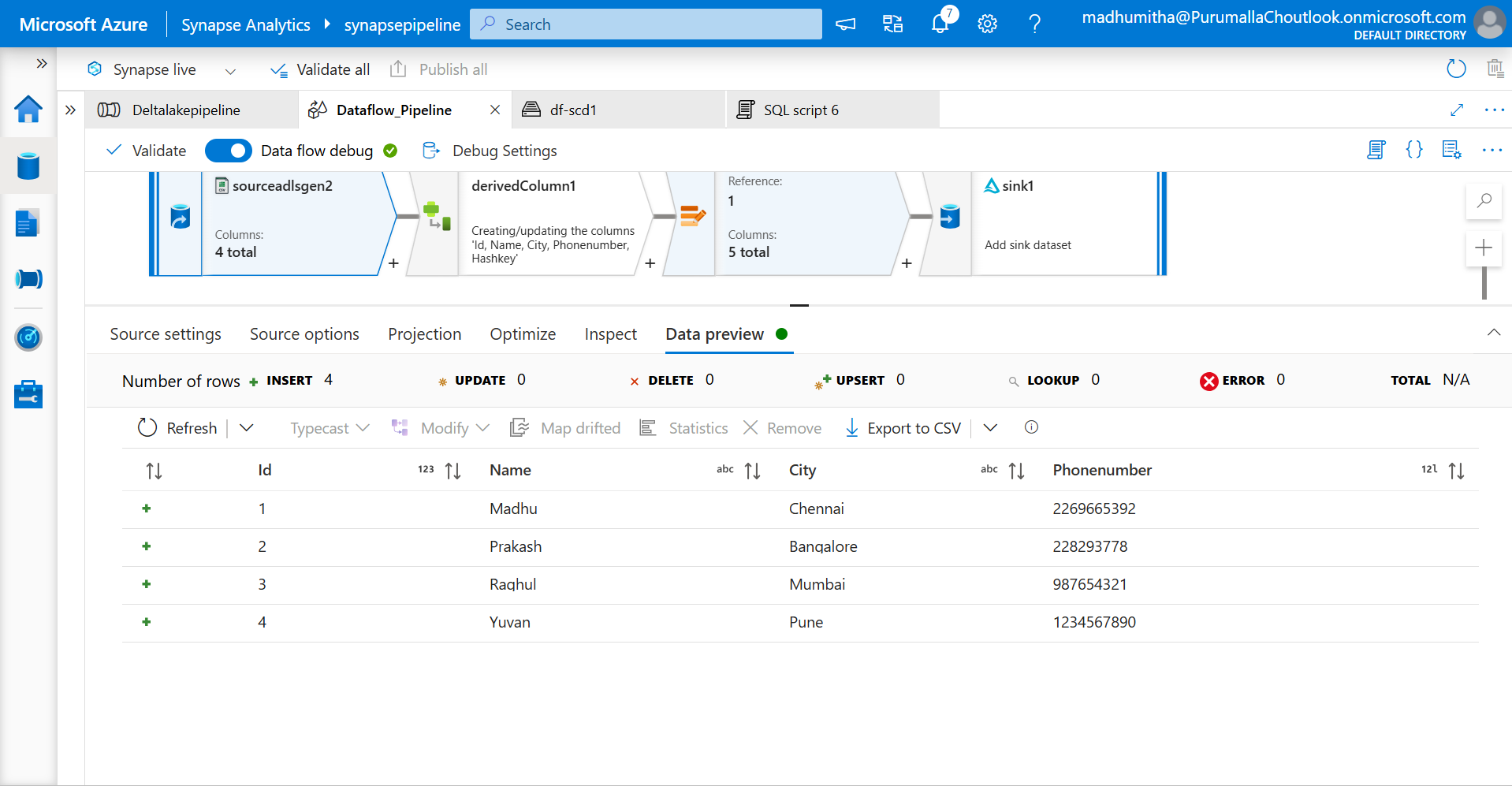
Step 3: Select the file path and select the check box as first row as header.

Step 4: Import the schema as int under projection tab.

Step 5: Navigate to data preview and give refresh.





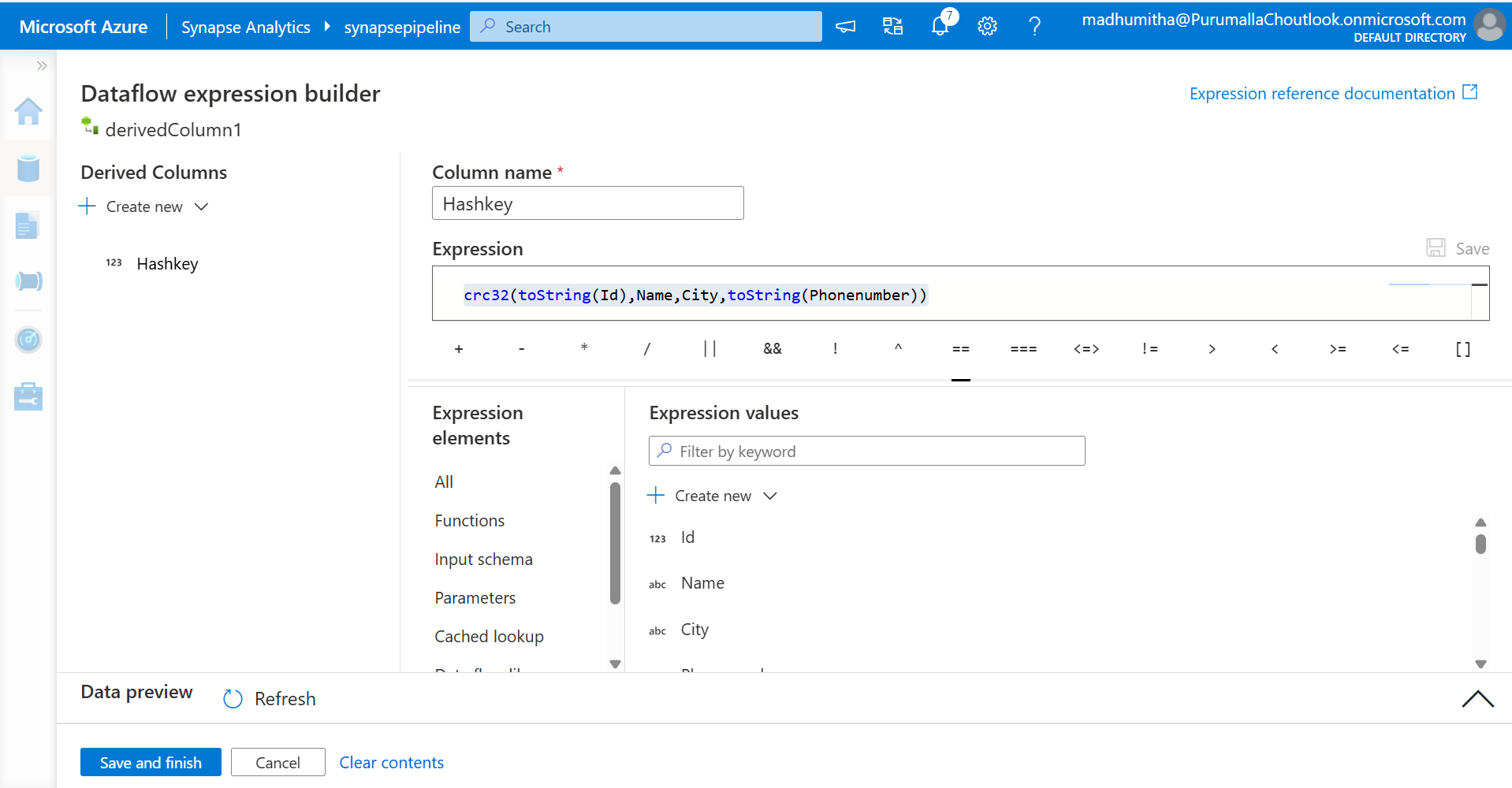


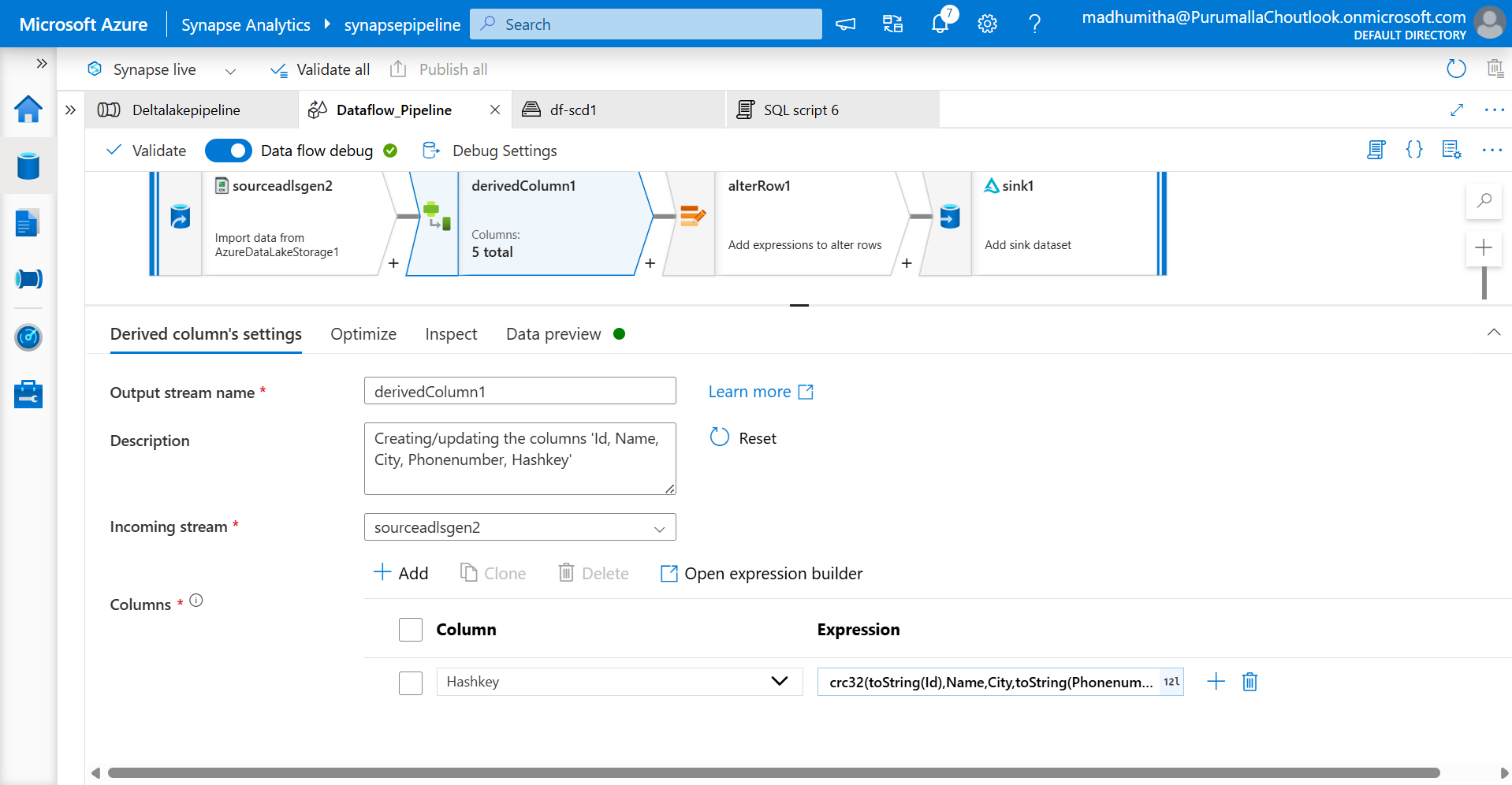
Activity 2 – Derived column

Derived column is used to add new column in the table

Step 1: Name the o/p stream as derivedColumn1

Step2: Add column as Haskey and expression is crc32(toString(Id),Name,City,toString(Phonenumber))





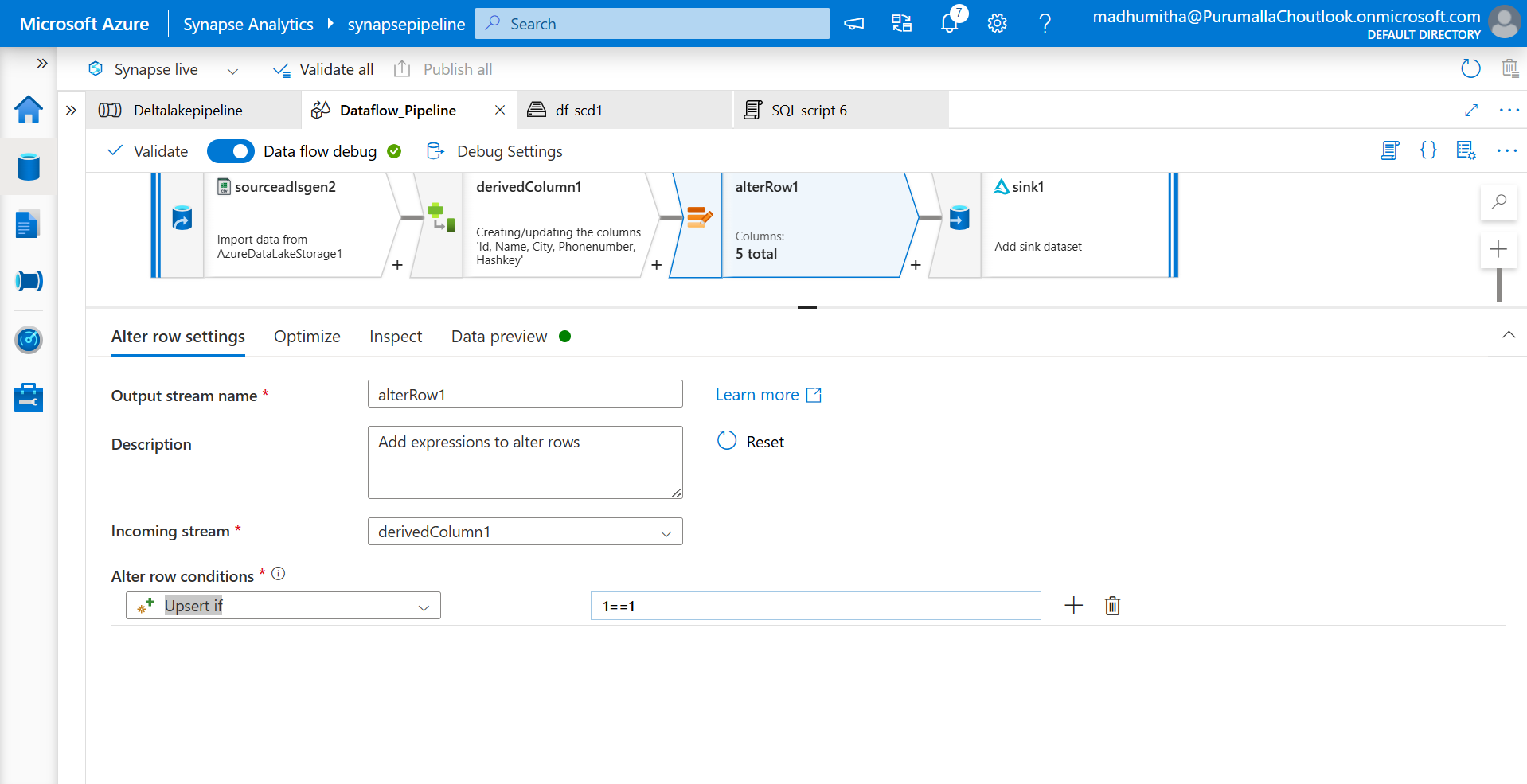
Activity 3- Alter row

Use Alterrow for modification of table

Step1: Name the o/p stream as alterRow1

Step2: Select incoming stream as derivedColumn1

Step3: Under Alter row condition, select Upsert if , 1==1



Activity 4 – Sink

Step 1: Name o/p stream as sink1

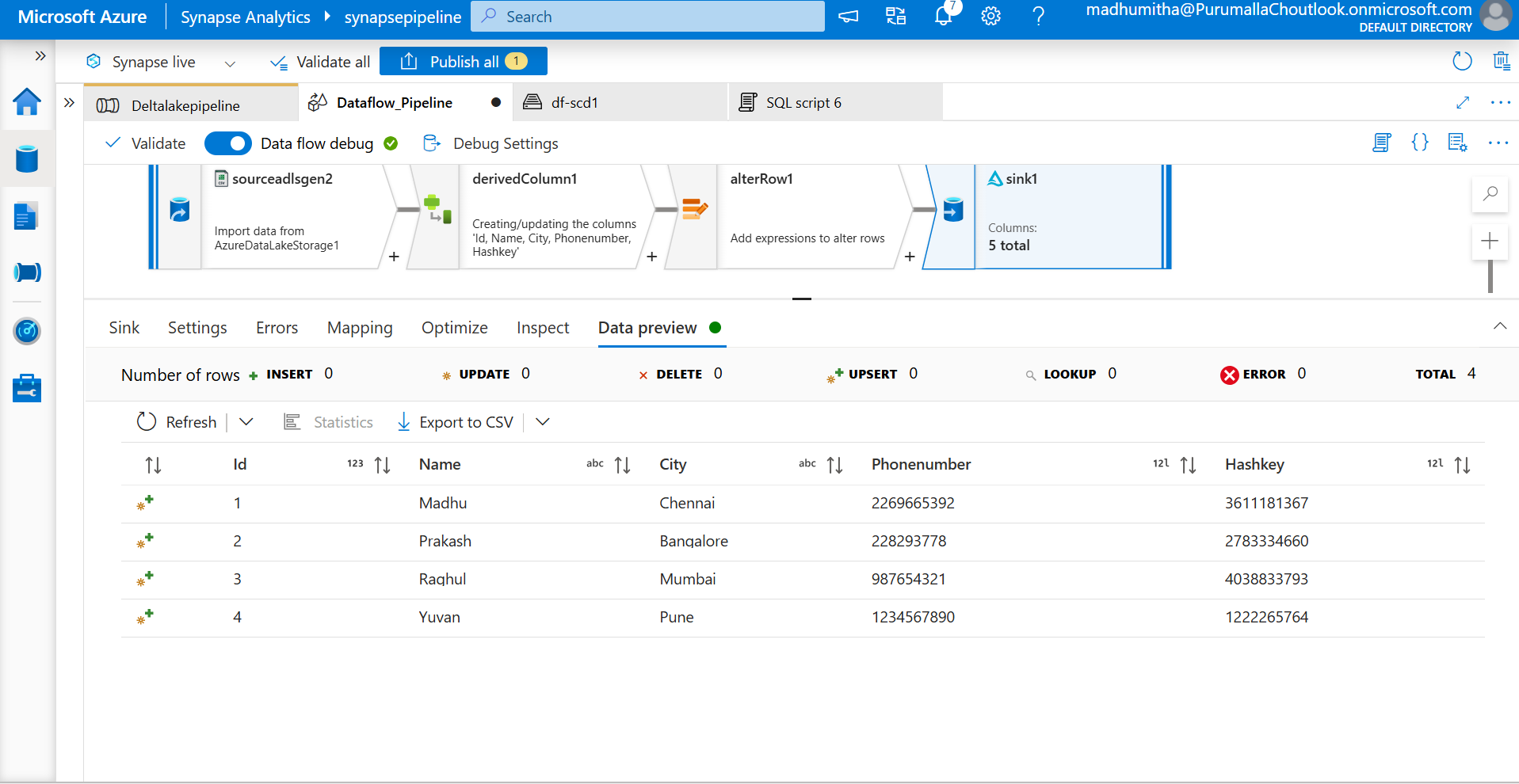
Step 2: Select Inline under sink type, under Inline dataset type, click delta format and select existing ADLSGEN2 account.

Step 3: Navigate to settings, Select the folder path and click the existing container and create new folder.

Step 4: Select allow upsert

Step 5: Click Id in list of column for key columns

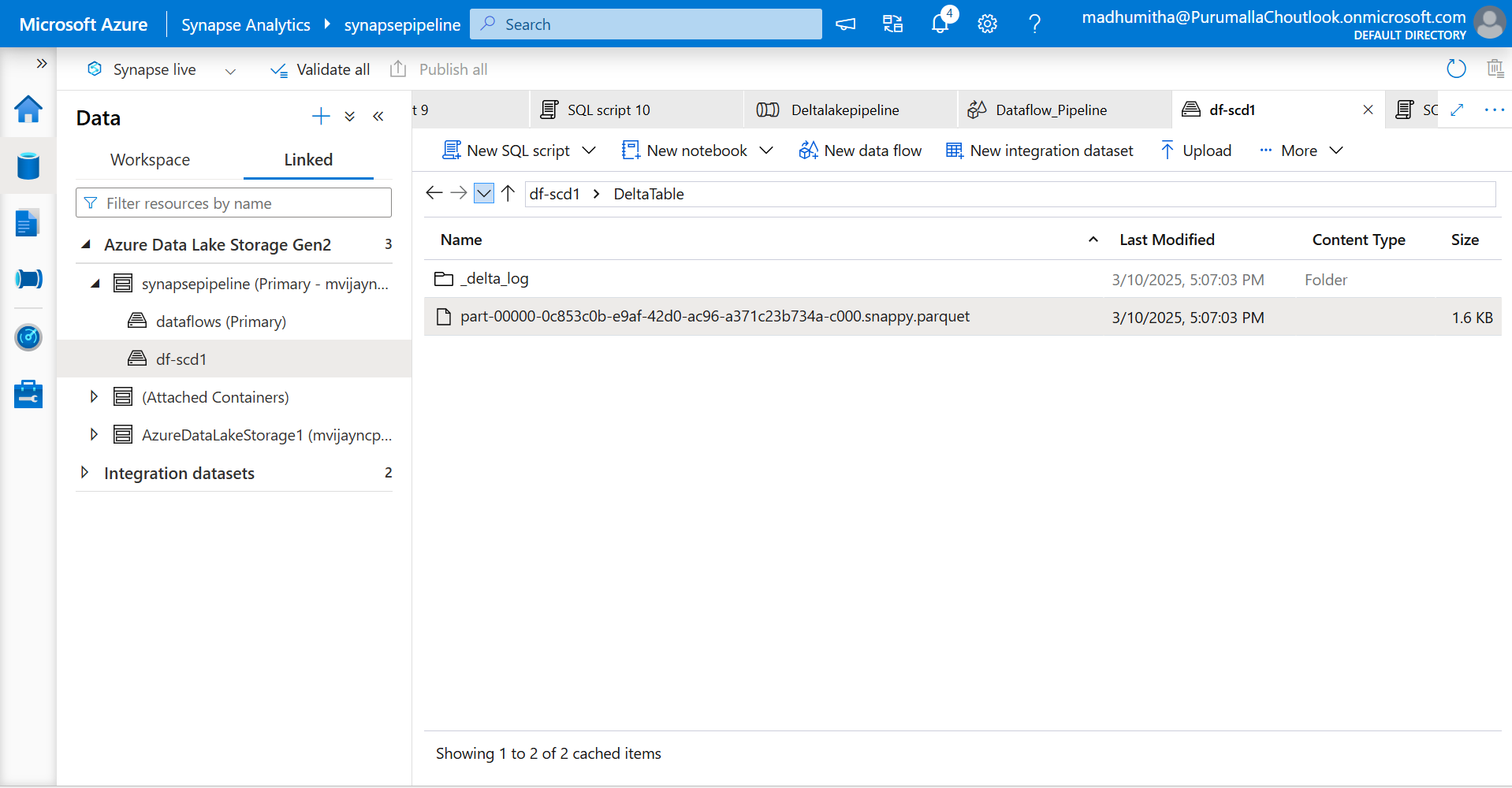
Step4: Navigate to Data preview, click refresh

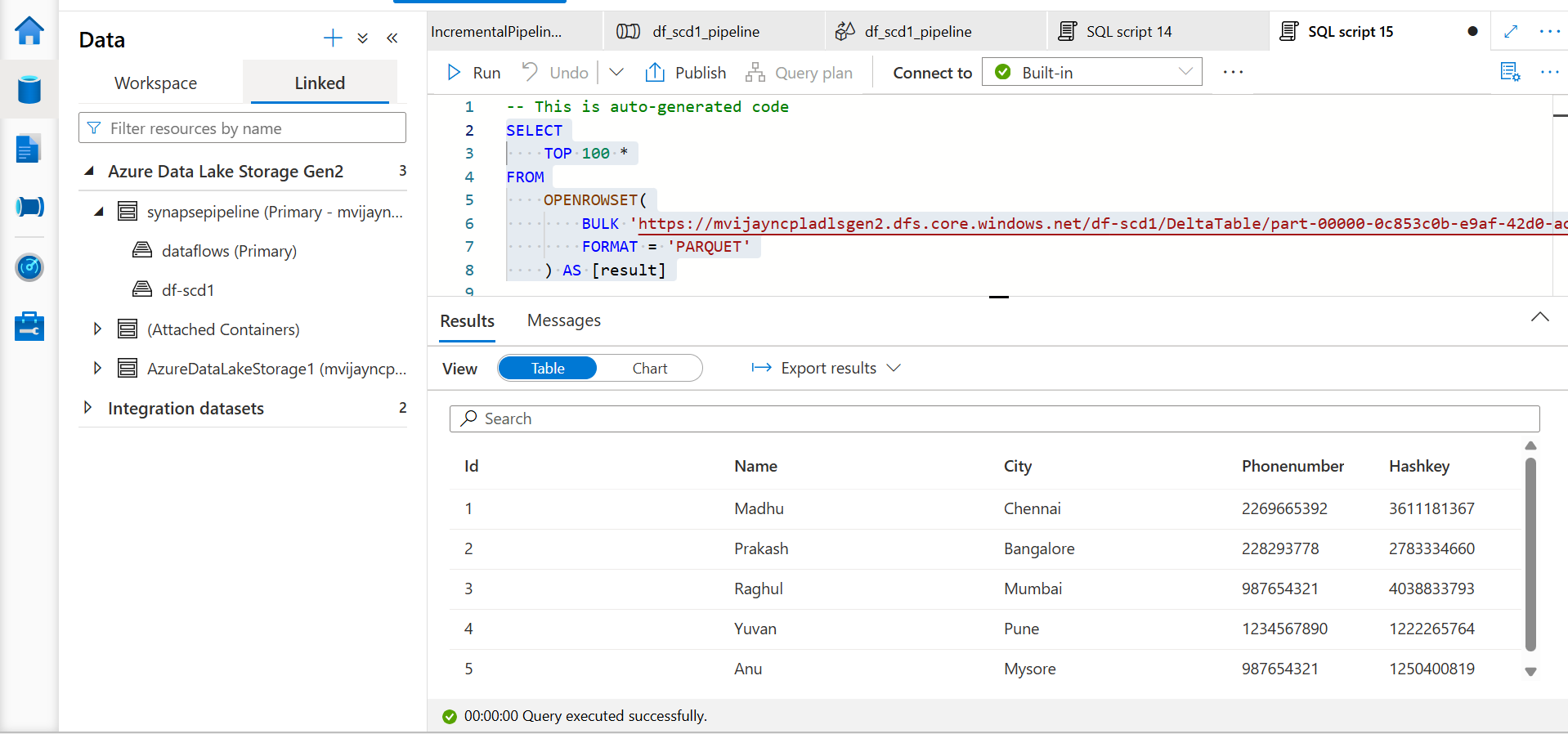


Output

Navigate to th data tab, click the linked and select the container.

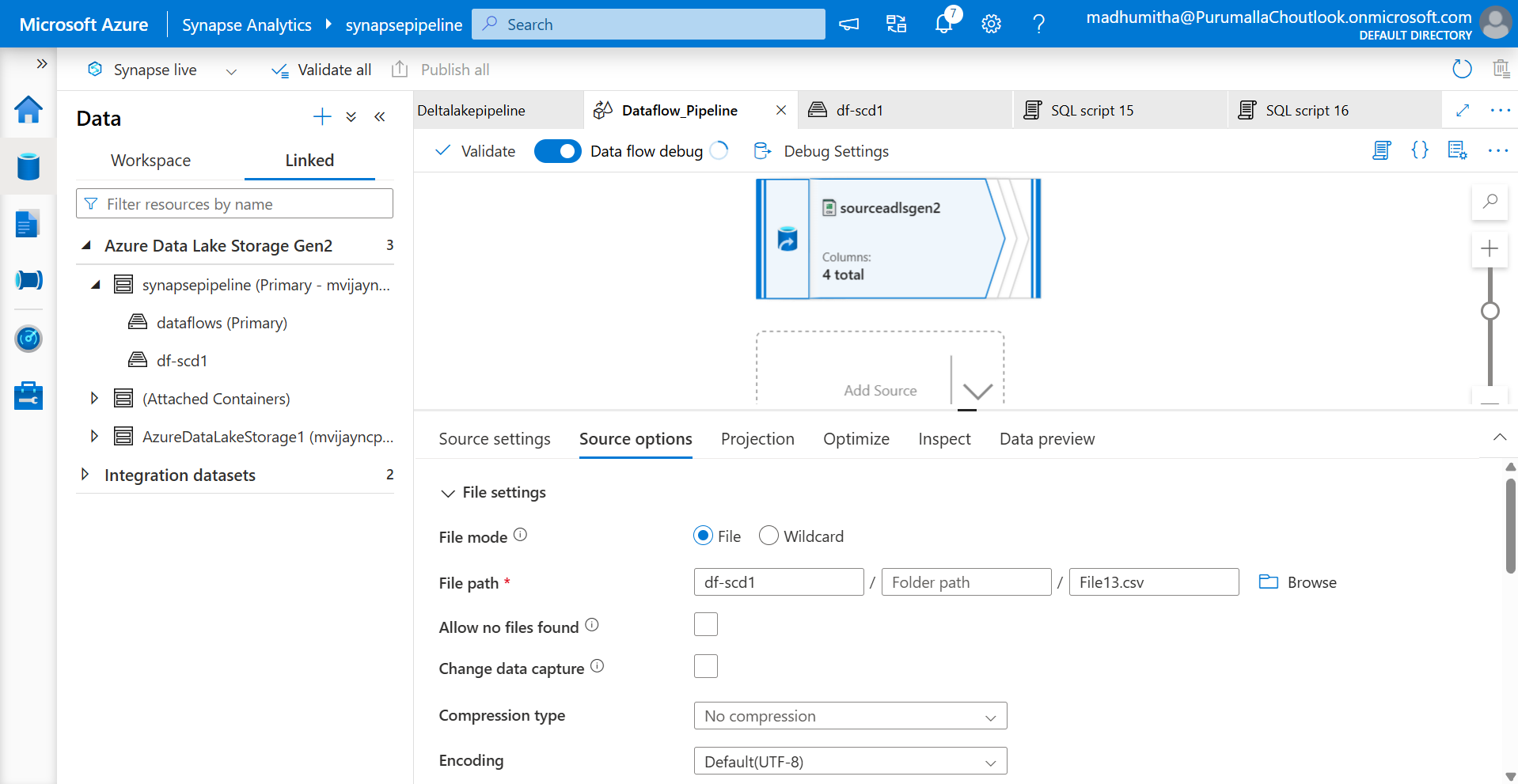
Under df-scd1 container, 1 file will be created with file extension .parquet and log files will be generated.





For insert

Step1 : Change the file path in source; add the File13.csv from the same container



Step 2: Run the pipeline

Two .parquet files and one log file will be generated in ADLSGEN2 container. 1st file will have the File12.csv data and 2nd file will have File13.csv data. Log file will show the insert operation has done.

