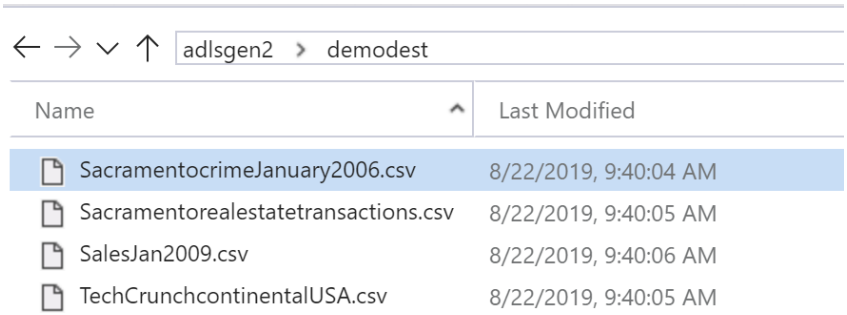
Lab 3: Data Flow Aggregation

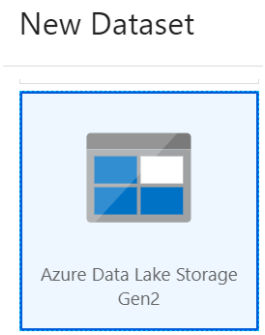
In this lab, we take a file from our ADLSgen2, perform data transformation, then put the result to an Azure SQL DB Table.

Step 1: Build a dataset for source CSV

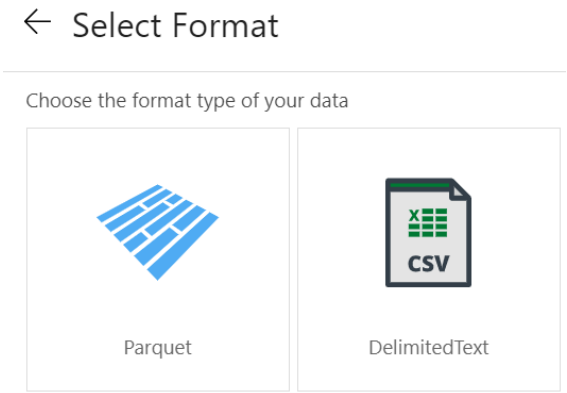
Confirm you have the file in ADLSgen2 after Lab2,



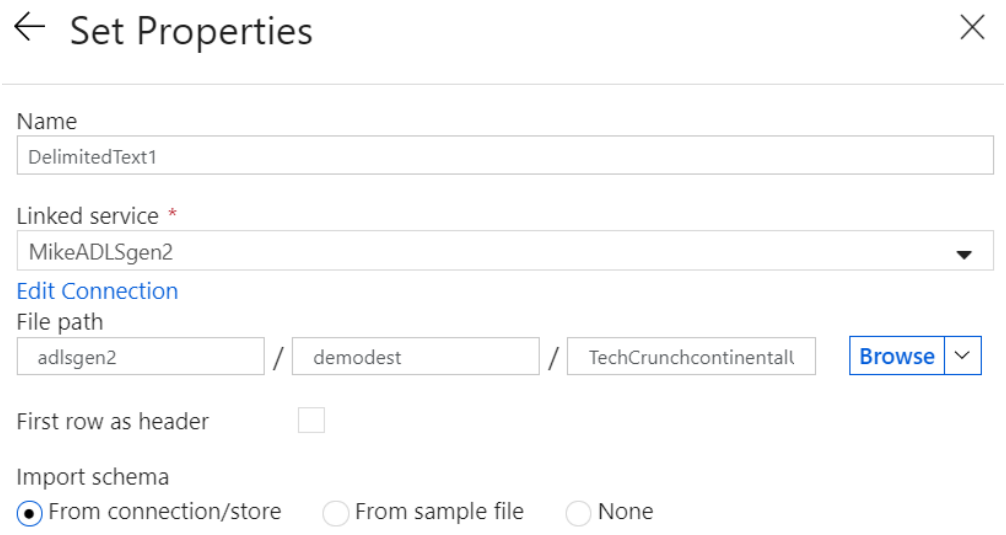
In ADF portal, create a dataset, reuse the ADLSGen2 Linked Services, define a data set:



Select DelimitedText in the box,

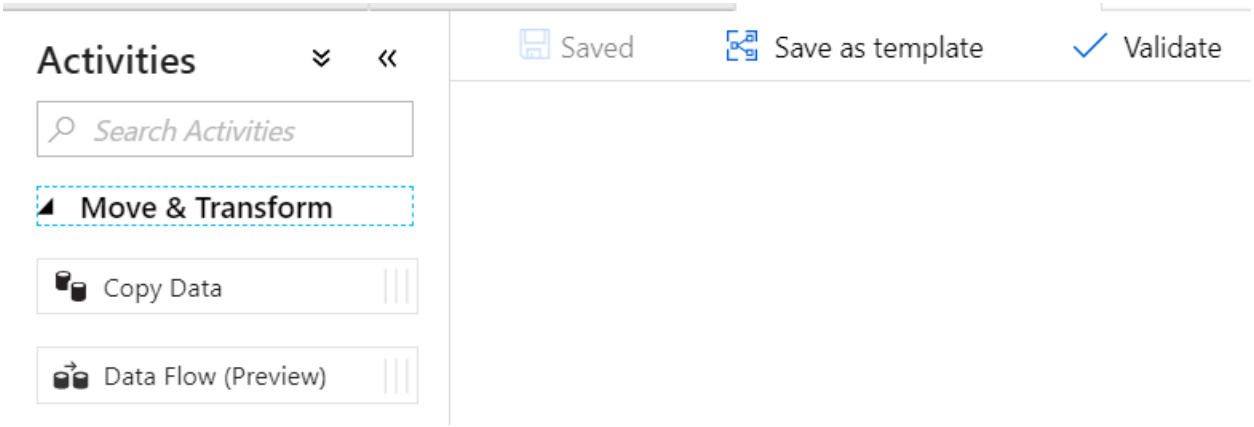


Define below properties with your own ADLS, clicking the First row as header will make your life easier.

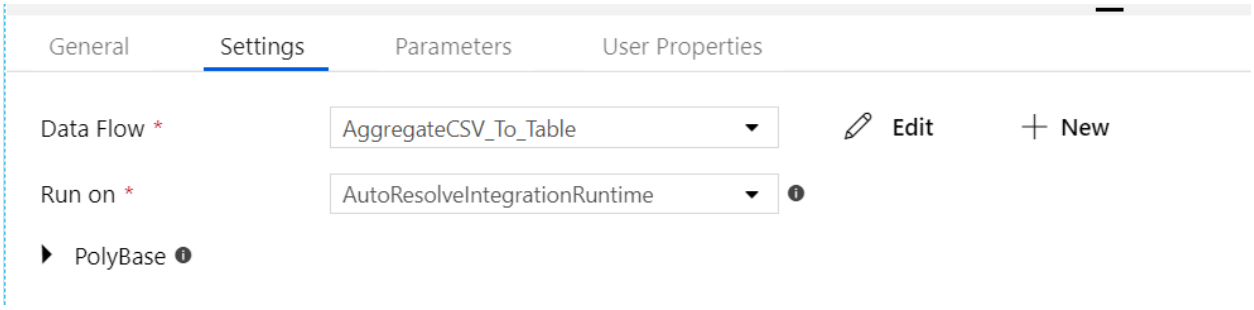


Step 2: Define the required DataFlow transformation

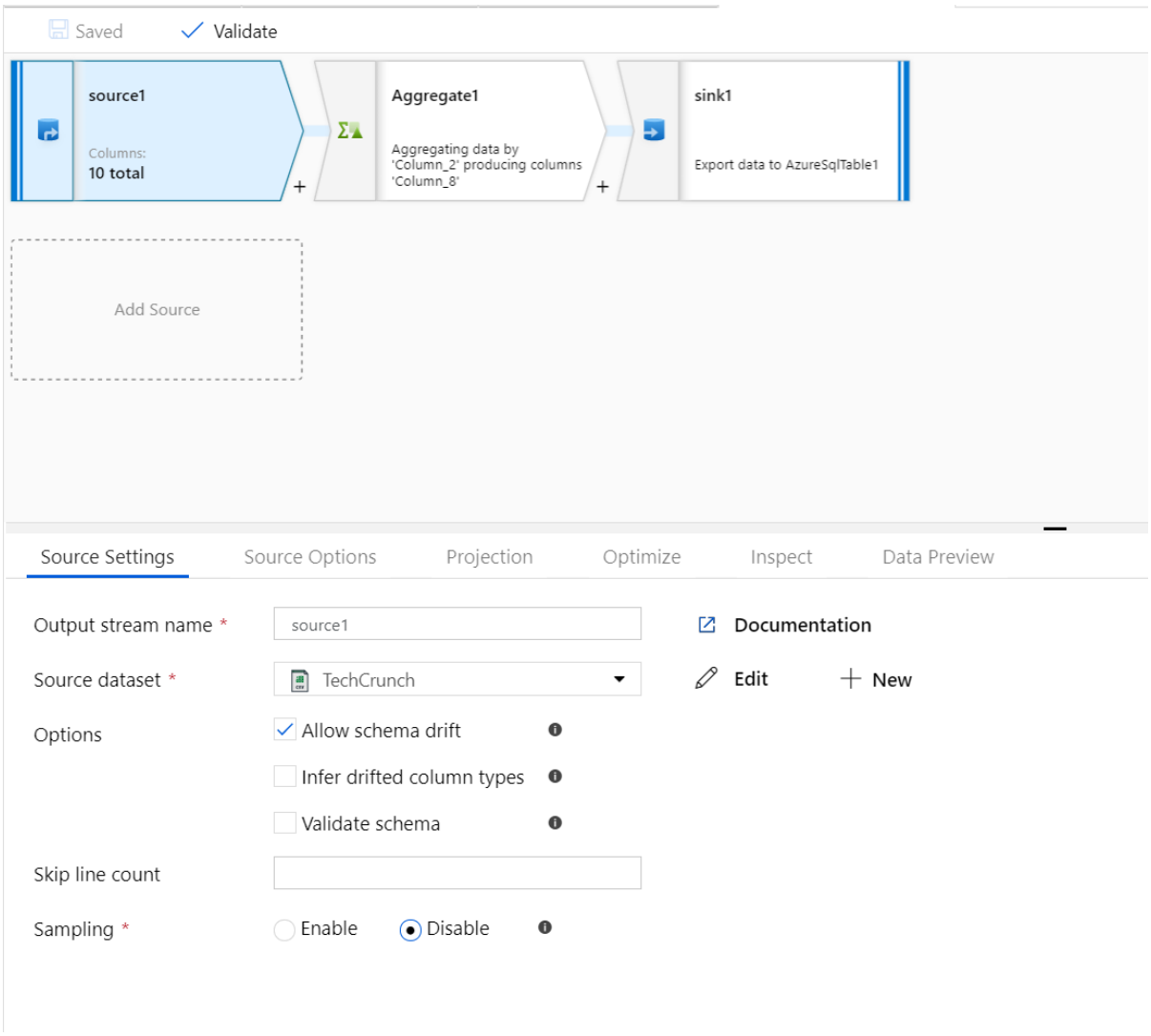
Drag an activity [Data Flow (Preview)] to use



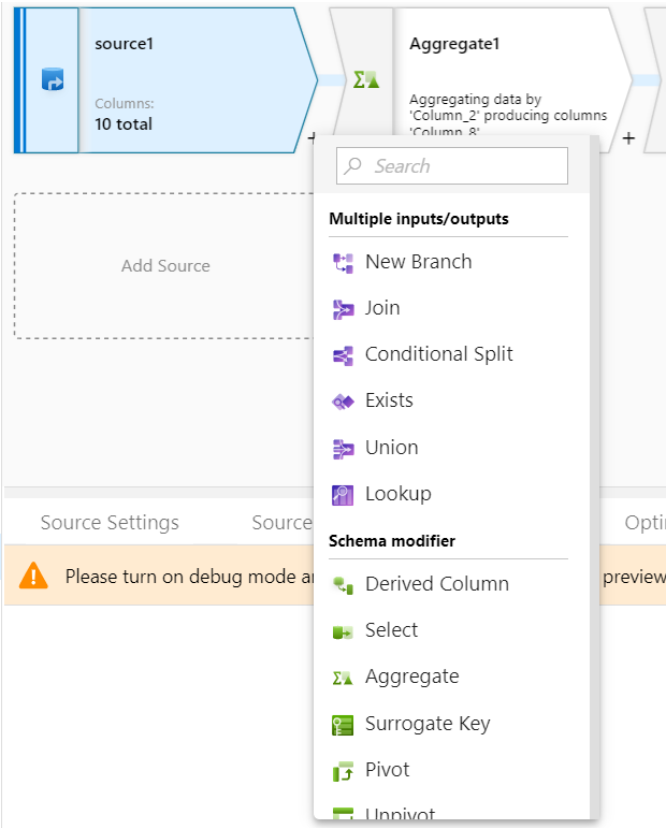
In Settings, click [+New]



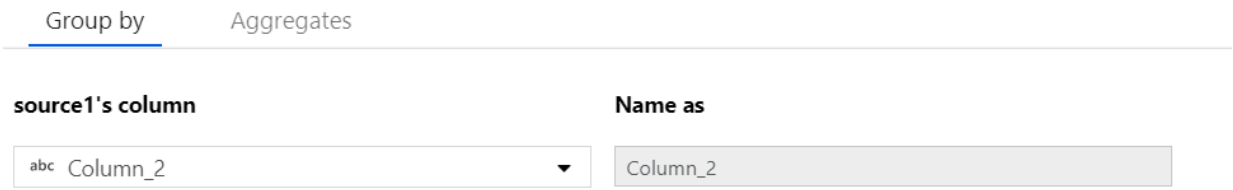
Once in the DataFlow panel click on the first slot as source. Remember you defined a CSV on ADLSgen2, now we use it here in Source dataset.



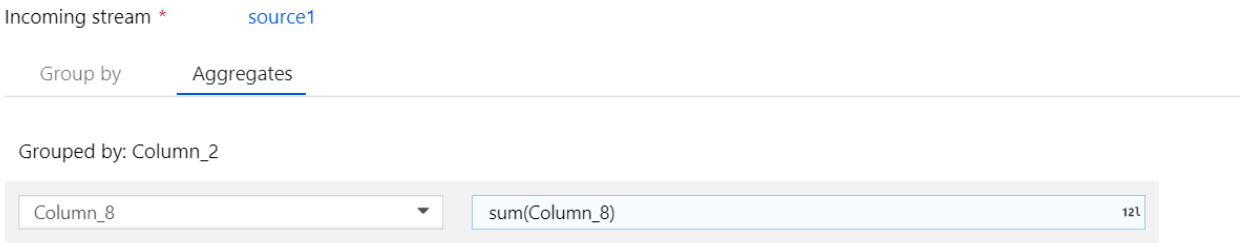
Then go to the small [+] sign underneath,



Choose Aggregate, In Aggregate Settings [Group by] tab, define source1's column as "Column_2"

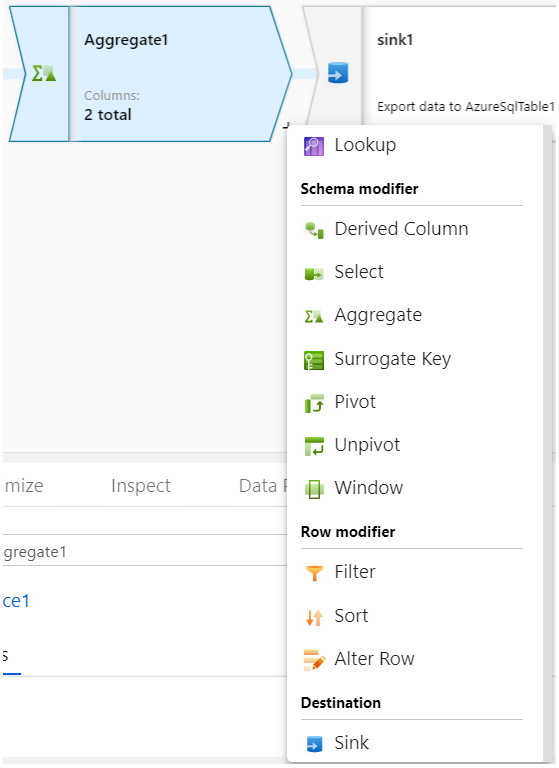


In Aggregates Table, select column 8 and sum(column_8) per below:

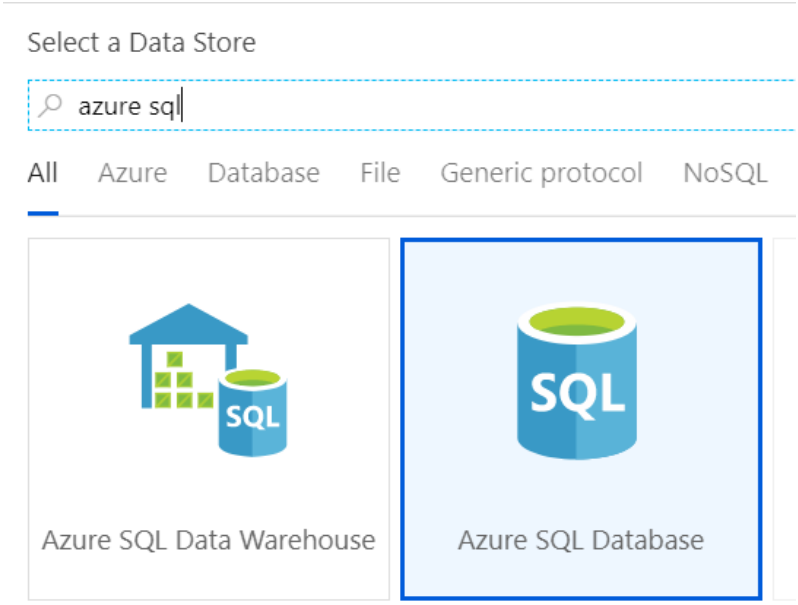


Step 3: Define output table

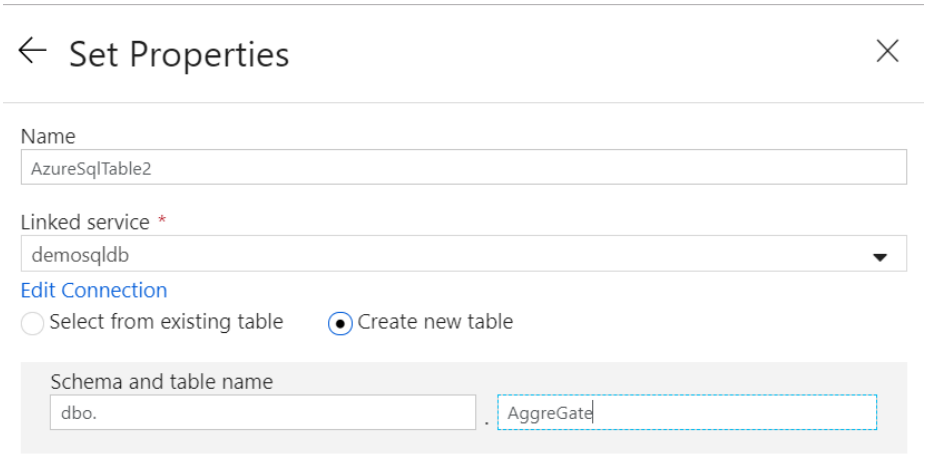
Once the Aggregation is read, click on small [+] again and scroll down to sink



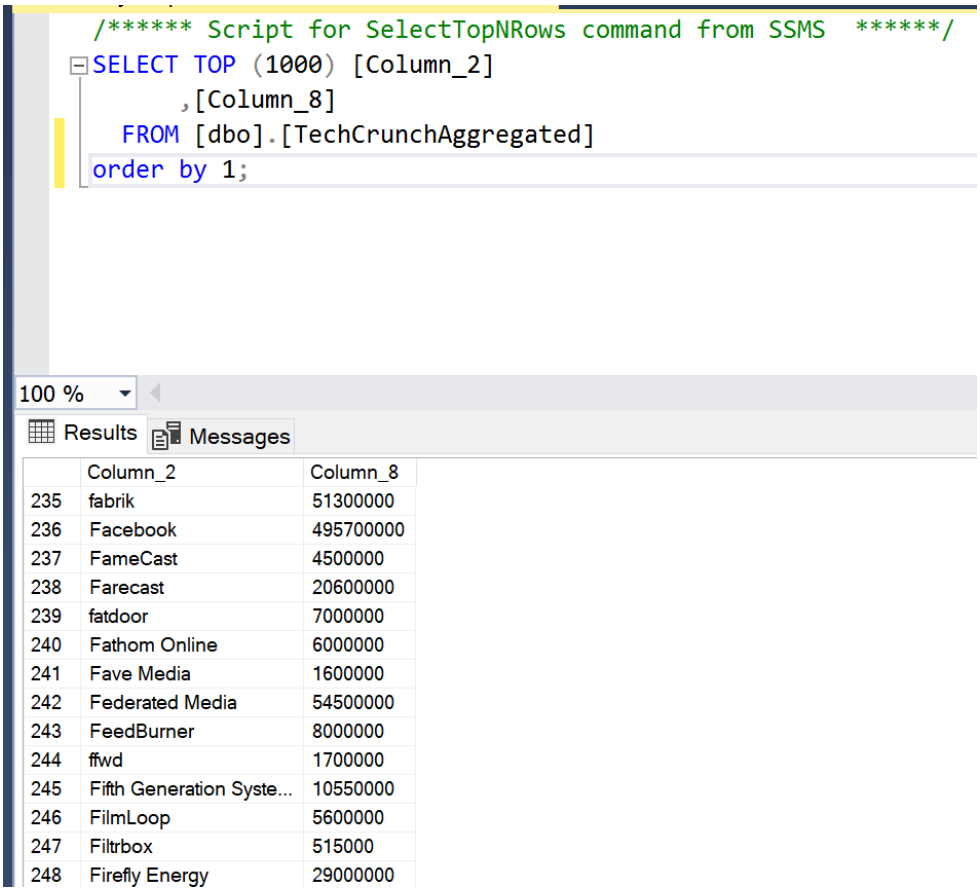
In the Sink tab, new a dataset and select [Azure SQL Database]



In the Dataset properties, setup per below



Your pipeline is ready to be executed and below is the execution result:



```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [Column_2]
      ,[Column_8]
  FROM [dbo].[TechCrunchAggregated]
order by 1;
```

100 %

Results | Messages

	Column_2	Column_8
235	fabrik	51300000
236	Facebook	495700000
237	FameCast	4500000
238	Farecast	20600000
239	fatdoor	7000000
240	Fathom Online	6000000
241	Fave Media	1600000
242	Federated Media	54500000
243	FeedBurner	8000000
244	ffwd	1700000
245	Fifth Generation Syste...	10550000
246	FilmLoop	5600000
247	Filtrbox	515000
248	Firefly Energy	29000000