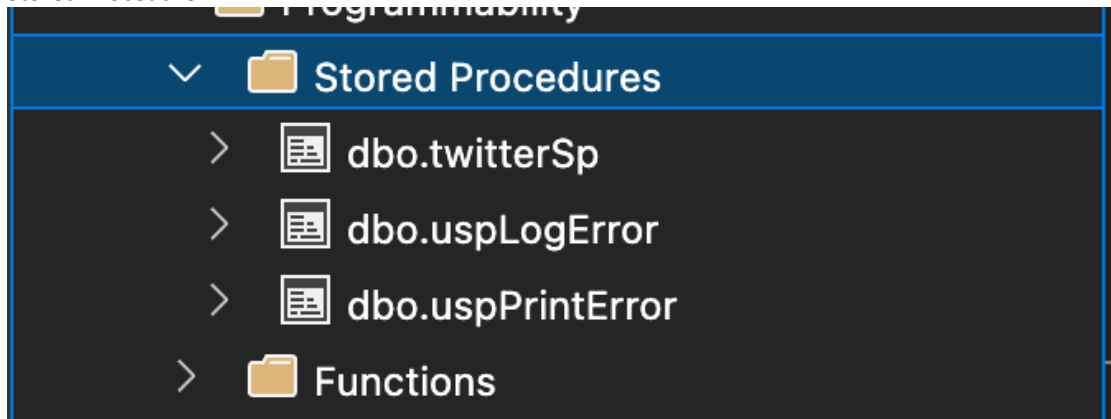


## P4: Work with the Database

### Functions Performed:

- We created a stored procedure to archive data that is already in the database tables.

### Stored Procedure



### Archive Tables

```
select * from sys.tables where name like '%archive%';
```

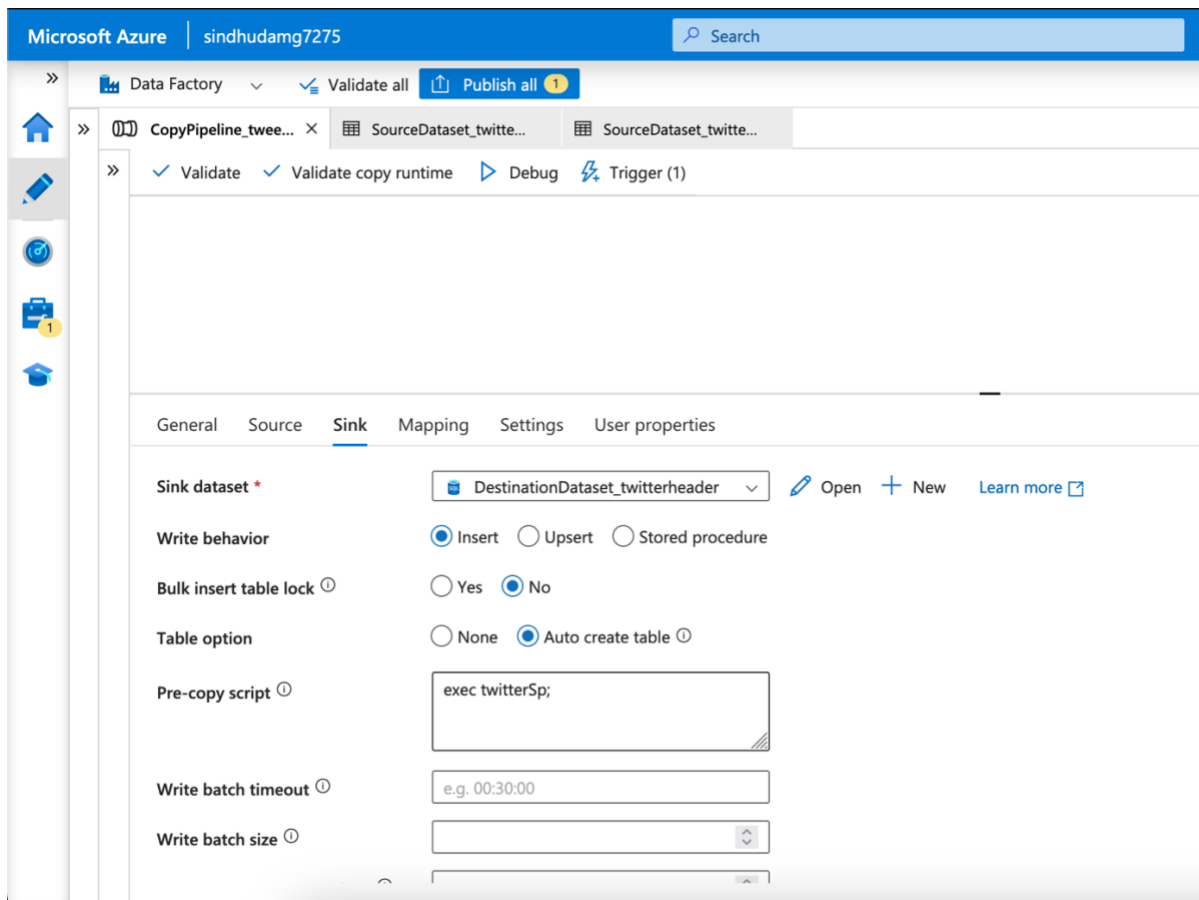
	name	object_id	principal_id	schema_id	parent_object_id	type	type_desc	create_date
1	edgeArchive	1458104235	NULL	1	0	U	USER_TABLE	2023-04-11 19:
2	nodeArchive	1474104292	NULL	1	0	U	USER_TABLE	2023-04-11 19:
3	referencedTweetArchive	1426104121	NULL	1	0	U	USER_TABLE	2023-04-11 19:
4	tweetUrlArchive	1394104007	NULL	1	0	U	USER_TABLE	2023-04-11 19:
5	twitterHeaderArchive	1378103950	NULL	1	0	U	USER_TABLE	2023-04-11 18:
6	twitterUserArchive	1410104064	NULL	1	0	U	USER_TABLE	2023-04-11 19:

### Count of Data in Archive Tables

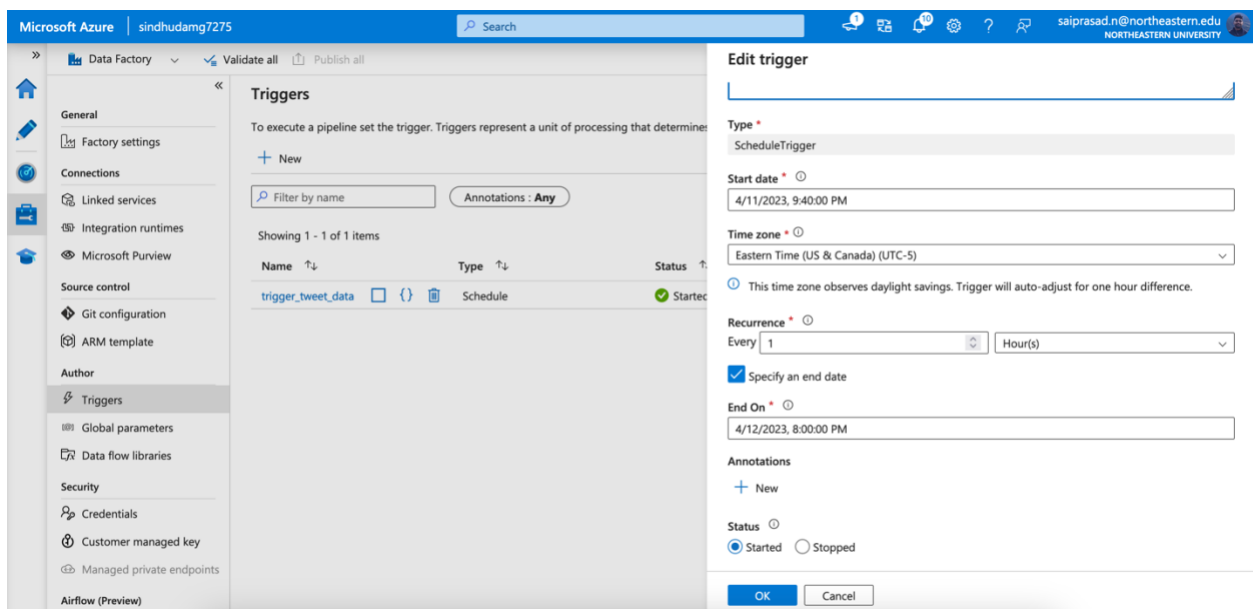
```
select count(tweet_id) as cnt_records from tweetUrlArchive;
```

	cnt_records
1	276

The first step of the pipeline run ensures that the stored procedure above runs:



- We scheduled the ADF pipeline to run every hour so that we can run visualization tools over this.



## Pipeline Runs:

The screenshot shows the Microsoft Azure portal interface. The left sidebar contains navigation options: Dashboards, Runs, Pipeline runs, Trigger runs, Change Data Capture (previ..., Runtimes & sessions, Integration runtimes, Data flow debug, Notifications, and Alerts & metrics. The main area is titled 'Trigger runs' and displays a table of pipeline runs. The table has columns for Trigger name, Trigger type, Trigger time, Status, Pipelines, Run, Message, Properties, and Run ID. Two runs are shown, both for the trigger 'trigger\_tweet\_data' and status 'Succeeded'.

Trigger name	Trigger type	Trigger time	Status	Pipelines	Run	Message	Properties	Run ID
trigger_tweet_data	Schedule trigger	4/11/2023, 6:05:00 P	Succeeded	1	Original			0858
trigger_tweet_data	Schedule trigger	4/11/2023, 5:05:00 P	Succeeded	1	Original			0858

- We also clean up the Azure blob storage after each run so that it does not clutter the container.

The screenshot shows the Azure Data Factory pipeline editor. The left sidebar lists activities: Move & transform, Synapse, Azure Data Explorer, Azure Function, Batch Service, Databricks, Data Lake Analytics, General, HDInsight, Iteration & conditionals, Machine Learning, and Power Query. The main area displays a pipeline with a 'Delete' activity. The activity is named 'Delete1' and is configured with the following settings:

- Dataset: SourceDataset\_twitterheader
- File path type: File path in dataset
- Start time (UTC): [Empty field]
- End time (UTC): [Empty field]
- Filter by last modified: [Empty field]
- Recursively: [Checked]
- Max concurrent connections: [Empty field]