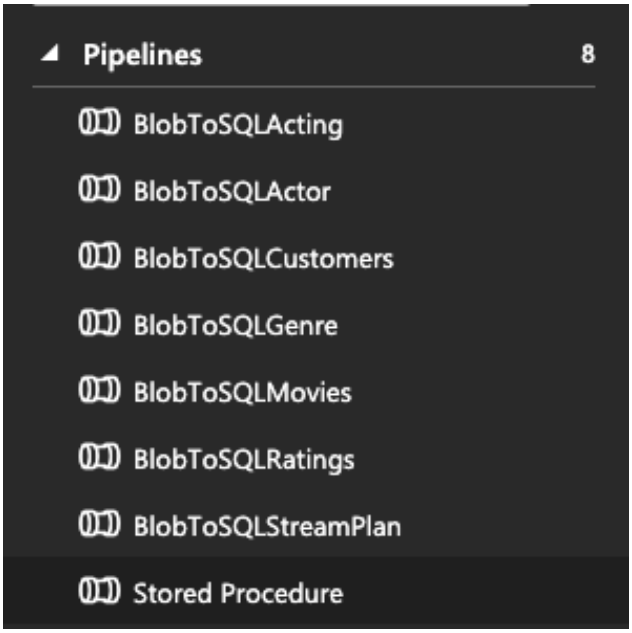
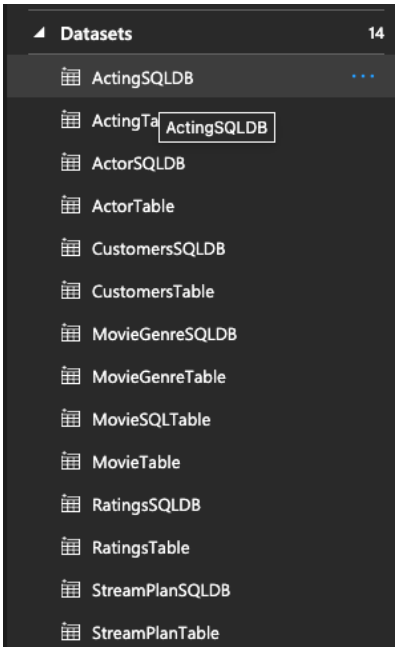
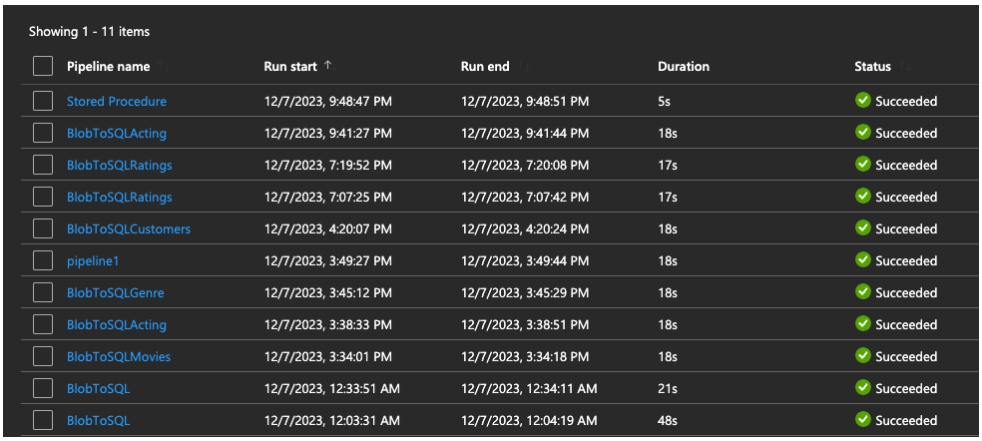
P4- Implementation
Streaming Service Analysis
Team 6
Abhishek Tikam Ramchandani – 002743745
Jurreyah Firdaws Mohammed – 002747514
Atharva Uplenchwar – 002990536
Samhitha Mereddy – 002796140

In this phase of our project implementation, we have utilized linked services in Azure Data Factory to connect and manage data across different storage and database services. As we now have cleaned datasets that we transformed using Alteryx and Python in Phase-3, we established linked services to Azure Blob Storage and Azure SQL Database, which acted as the sources and destinations for our data. Blob Storage was primarily used for handling large volumes of structured + unstructured data, whereas SQL Database served as our structured data's destination storage system.

The core of our Azure Data Factory implementation involved setting up Data Pipelines and Scheduled Triggers. These pipelines were designed to efficiently transfer and transform data between the source and sink database, "xxxTable" & "xxxSQLDB", respectively.







Showing 1 - 11 items

Pipeline name	Run start	Run end	Duration	Status
Stored Procedure	12/7/2023, 9:48:47 PM	12/7/2023, 9:48:51 PM	5s	Succeeded
BlobToSQLActing	12/7/2023, 9:41:27 PM	12/7/2023, 9:41:44 PM	18s	Succeeded
BlobToSQLRatings	12/7/2023, 7:19:52 PM	12/7/2023, 7:20:08 PM	17s	Succeeded
BlobToSQLRatings	12/7/2023, 7:07:25 PM	12/7/2023, 7:07:42 PM	17s	Succeeded
BlobToSQLCustomers	12/7/2023, 4:20:07 PM	12/7/2023, 4:20:24 PM	18s	Succeeded
pipeline1	12/7/2023, 3:49:27 PM	12/7/2023, 3:49:44 PM	18s	Succeeded
BlobToSQLGenre	12/7/2023, 3:45:12 PM	12/7/2023, 3:45:29 PM	18s	Succeeded
BlobToSQLActing	12/7/2023, 3:38:33 PM	12/7/2023, 3:38:51 PM	18s	Succeeded
BlobToSQLMovies	12/7/2023, 3:34:01 PM	12/7/2023, 3:34:18 PM	18s	Succeeded
BlobToSQL	12/7/2023, 12:33:51 AM	12/7/2023, 12:34:11 AM	21s	Succeeded
BlobToSQL	12/7/2023, 12:03:31 AM	12/7/2023, 12:04:19 AM	48s	Succeeded

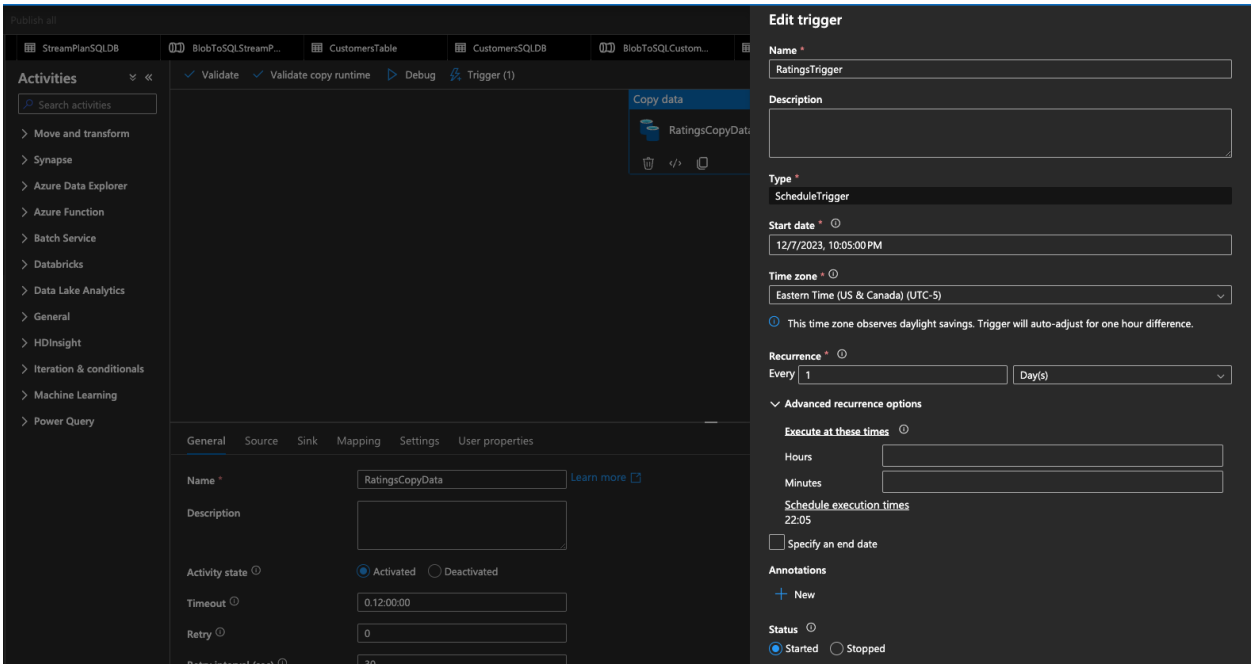
We focused on creating a seamless flow for specific datasets, particularly our document model (Customers, Stream-Plan & Movie-Ratings) as it. This document represents each Customer's info, the stream plan they are on and the Movies they have rated on IMDB. To do the same, we wrote a Stored Procedure (code below) and scheduled an **Ongoing Data Refresh** that executes this SP on a 24-hour basis. This was also done for our Customers table.

```
CREATE PROCEDURE GetCustomerDetails
AS
BEGIN
    SELECT
        c.CustID as id,
        c.Cust_FName as firstName,
        c.Cust_LName as lastName,
        c.Email as email,
        sp.Description as streamPlanDescription,
        sp.Monthly_Cost as streamPlanMonthlyCost,
        (SELECT
             r.MovieID as movieID,
             r.IMDB_Rating as rating
         FROM Ratings r
         WHERE r.CustID = c.CustID
         FOR JSON PATH) as ratedMovies
    FROM Customers c
    INNER JOIN Stream_Plan sp ON sp.Stream_Plan_ID = c.Stream_Plan_ID
    FOR JSON PATH;
END;
```

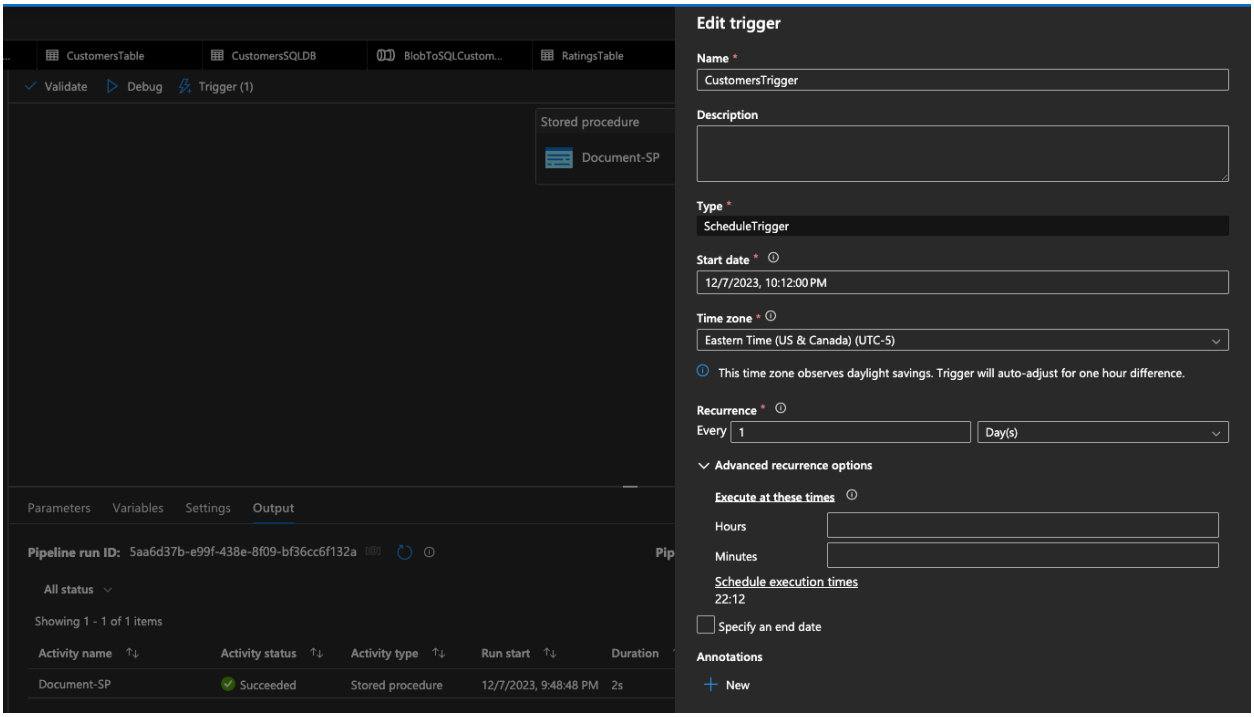
Generated JSON Output (Customers, Stream-Plan & Movie-Ratings):

To manage and automate the entire data flow, we set up triggers in ADF. These triggers were configured to initiate pipeline runs based on our requirements, ensuring that data processing was timely and consistent. This setup not only streamlined our data management tasks but also provided a scalable and flexible solution for handling our data processing needs in ADF.

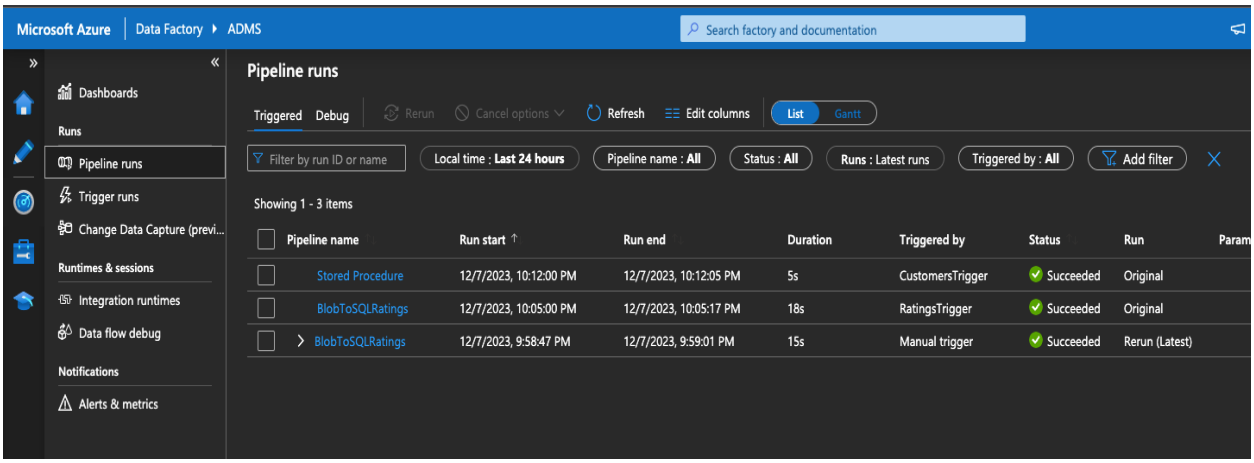
Ratings Table will need updating regularly as many people are more likely to rate many movies, so this trigger updates the same every 24 hours.



Similarly, for Customers table as new customers can join the platform every day:



Successful runs of our Scheduled and Manual Triggers:



Here are some reference screenshots to show that the tables update & populated successfully:

```
18  SELECT * FROM Movies
19  --done pipelining
20  CREATE TABLE Acting(
```

Results | Messages

	MovieID	Title	Year	Movie_Rating	IMDB_Rating
1	tm1000037	Je suis Karl	2021	R	5.4
2	tm1000147	Zone 414	2021	R	4.9
3	tm100015	Takers	2010	PG-13	6.2
4	tm1000166	Wave of Cinema: Surat Dar…	2020	PG-13	7.5
5	tm1000185	Squared Love	2021	PG-13	5.1
6	tm100027	Alibaba Aur 40 Chor	1979	PG-13	6.2
7	tm1000296	New Gods: Nezha Reborn	2021	PG-13	6.8
8	tm1000551	Namaste Wahala	2020	PG-13	5.1
9	tm1000599	The Last Forest	2021	PG-13	7.3
10	tm1000619	Radhe Shyam	2022	PG-13	5.3
11	tm1000797	Ride or Die	2021	R	5.6
12	tm100106	My Amnesia Girl	2010	PG-13	6.7
13	tm1001095	Roohi	2021	PG-13	4.3
14	tm1001097	Beauty	2022	R	3.9
15	tm1001108	Biggie: I Got a Story to …	2021	R	6.8

Row Count validation:

```
9  SELECT count(*) FROM dbo.Movies;
```

Results | Messages

	(No column name)
1	5489

```
6  SELECT count(*) FROM dbo.Ratings;
```

Results | Messages

	(No column name)
1	1000