

# Data Flow pipelines and Load

## Stage Data Loading

### 1] title.crew dataset

The screenshot shows a SQL query execution interface. The query is as follows:

```
1 select * from title_crew where tconst = 'tt10930678';
2 select * from title_crew_flatten where tconst = 'tt10930678';
3
4
5 select * from title_basics;
```

The results are displayed in a table with columns: TCONST, DIRECTORS, and WRITERS. The data is as follows:

	TCONST	DIRECTORS	WRITERS
1	tt9042444	nm0059223,nm0083290,nm0123364,nm021	nm0894640
2	tt9042446	nm10145312	nm10145312
3	tt9042448	nm0702142	nm5107308
4	tt9042450	nm0442687	NA
5	tt9042452	nm0776255	nm1075143,nm2360862,nm2917457,nm101
6	tt9042454	nm8412331	nm8412331
7	tt9042456	nm4458749	nm4458749

Query Details:

- Query duration: 3.1s
- Rows: 11.5M
- Query ID: 01baf11f-0001-a4af-005...

Ask Copilot button is visible.

As the data in writers and directors column has values in the form of string, the data is flattened using adf pipeline

The screenshot shows the Microsoft Azure Data Factory interface. The pipeline is named "pipeline5" and is in the "TitleCrewDF" dataset. The pipeline consists of the following stages:

- source1: Import data from TitleCrewSF
- derivedColumn1: Creating/updating the columns TCONST, DIRECTORS, WRITERS
- flatten1: Columns: 3 total
- sink1: Export data to CrewFlattenSF

The "Flatten settings" tab is selected, showing the following configuration:

- Incoming stream: derivedColumn1
- Unroll by: DIRECTORS, WRITERS
- Unroll root: {}
- Options: Skip duplicate input columns, Skip duplicate output columns

02-16 8:17pm 2025-02-16 8:33pm 2025-02-16 8:59pm 2025-02-20 11:47am Midterm Project - IMDB 2025-03-09 7:22pm

ACCOUNTADMIN COMPUTE\_WH (X-Small) Share

IMDB\_DB.IMDB\_SCHEMA Settings Code Versions

```
1 select * from title_crew where tconst = 'tt1093067';
2 select * from title_crew_flatten where tconst = 'tt1093067';
3
4
5 select * from title_basics;
```

Results Chart

	TCONST	DIRECTORS	WRITERS
1	tt1093067	nm0284159	nm0001468
2	tt1093067	nm0284159	nm0611848
3	tt1093067	nm0284159	nm1669351
4	tt1093067	nm0284159	nm0750523
5	tt1093067	nm0284159	nm2227704
6	tt1093067	nm0284159	nm0996024
7	tt1093067	nm0284159	nm0780980

Query Details

Query duration 5.0s

Rows 27.6M

Query ID 01baf122-0001-a4b0-0...

Show more

Ask Copilot

SSP

2:42 PM 3/11/2025

## 2] title.episode

02-16 8:17pm 2025-02-16 8:33pm 2025-02-16 8:59pm 2025-02-20 11:47am Midterm Project - IMDB 2025-03-09 7:22pm

ACCOUNTADMIN COMPUTE\_WH (X-Small) Share

IMDB\_DB.IMDB\_SCHEMA Settings Code Versions

```
8 select * from title_episode;
9 describe table title_episode;
10
11 select * from title_principals;
12
```

Results Chart

	TCONST	PARENTTCONST	SEASONNUMBI	EPISODENUMBI	DI_JOB_ID
1	tt12250694	tt8859520	0	0	24e7bfaa-a0d1-4f5e-9f36-b51ff096e4a0
2	tt12250696	tt8859520	0	0	24e7bfaa-a0d1-4f5e-9f36-b51ff096e4a0
3	tt12250698	tt8859520	0	0	24e7bfaa-a0d1-4f5e-9f36-b51ff096e4a0
4	tt1225070	tt0106086	1	9	24e7bfaa-a0d1-4f5e-9f36-b51ff096e4a0
5	tt12250700	tt4151188	0	0	24e7bfaa-a0d1-4f5e-9f36-b51ff096e4a0
6	tt12250702	tt8859520	0	0	24e7bfaa-a0d1-4f5e-9f36-b51ff096e4a0

Query Details

Query duration 80ms

Rows 26.5M

Query ID 01baf125-0001-a521-00...

Show more

Ask Copilot

SSP

2:45 PM 3/11/2025

## 3] title.principals

https://app.snowflake.com/east-us-2.azure/nua03035/worLh40kjUD#query

02-16 8:17pm 2025-02-16 8:33pm 2025-02-16 8:59pm 2025-02-20 11:47am Midterm Project - IMDB 2025-03-09 7:22pm

ACCOUNTADMIN COMPUTE\_WH (X-Small) Share

IMDB\_DB.IMDB\_SCHEMA Settings

```

11 | select * from title_principals;
12 |
13 | select * from title_akas;
14 | select * from title_akas_flatten where attributes not like 'NA';
15 |

```

Results Chart

	T	CATEGORY	JOB	CHARACTERS	DI_JOB_ID	DI_LOAD_DT
1	3	director	NA	NA	e9b40ce0-fbb4-45e1-8...	2025-03-10
2	0	self	NA	["Self - Host"]	e9b40ce0-fbb4-45e1-8...	2025-03-10
3	0	actress	NA	NA	e9b40ce0-fbb4-45e1-8...	2025-03-10
4	9	self	NA	["Presenter"]	e9b40ce0-fbb4-45e1-8...	2025-03-10
5	3	actor	NA	["DueÀo Casa"]	e9b40ce0-fbb4-45e1-8...	2025-03-10
6	3	director	director	NA	e9b40ce0-fbb4-45e1-8...	2025-03-10

Query Details

Query duration 48ms

Rows 91.3M

Query ID 01baf125-0001-a51f-00...

Show more

Ask Copilot

#### 4) title.akas

https://app.snowflake.com/east-us-2.azure/nua03035/worLh40kjUD#query

02-16 8:17pm 2025-02-16 8:33pm 2025-02-16 8:59pm 2025-02-20 11:47am Midterm Project - IMDB 2025-03-09 7:22pm

ACCOUNTADMIN COMPUTE\_WH (X-Small) Share

IMDB\_DB.IMDB\_SCHEMA Settings

```

12 |
13 | select * from title_akas;
14 | select * from title_akas_flatten where attributes not like 'NA';
15 |
16 | select * from date_dim;

```

Results Chart

	LOAN	LANGUAGE	TYPE	ATTRIBUTE	IS ORIGINAL	DI_JOB_ID	DI_LOAD_DT
1		ja	NA	NA	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10
2		es	NA	NA	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10
3	n	Unknown	original	NA	TRUE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10
4		pt	NA	NA	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10
5		hi	NA	NA	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10
6		fr	NA	NA	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10

Query Details

Query duration 15s

Rows 51.5M

Query ID 01baf127-0001-a4b0-00...

Show more

Ask Copilot

Here the attributes column is flattened and separated into different columns

Microsoft Azure | Data Factory | midterm-project-team2 | Search factory and documentation | pekamwar.s@northeastern.edu | NORTH EASTERN UNIVERSITY

Factory Resources

- Pipelines (9)
  - CrewParquetToSF
  - DateAndRegion
  - DimTitle
  - Language
  - LoadPersonnelDimDF
  - pipeline2
  - pipeline3
  - pipeline4
  - pipeline5
- Change Data Capture (preview) 0
- Datasets (47)
  - CrewFlattenSF
  - DateDimSF

owflakeTable6 • sFTitleEpisode SF\_Fact\_Movie pipeline5 TitleAkasDF

Flatten settings

Incoming stream: derivedColumn1

Unroll by: ATTRIBUTES

Options

- ☐ Skip duplicate input columns
- ☐ Skip duplicate output columns

Input columns

2:54 PM 3/11/2025

https://app.snowflake.com/east-us-2.azure/nua03035/worLh40kjUD#query

Midterm Project - IMDB

ACCOUNTADMIN • COMPUTE\_WH (X-Small) Share

```
select * from title_akas;
select * from title_akas_flatten where attributes not like 'NA';
select * from date_dim;
```

Results

	LANGUAGE	TYPE	ATTRIBUTES	ISORIGINALTITLE	DI_JOB_ID	DI_LOAD_DT
1	Unknown	NA	alternative transliteration	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10
2	Unknown	NA	promotional title	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10
3	Unknown	NA	transliterated title	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10
4	en	NA	informal English title	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10
5	cmn	NA	literal title	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10
6	cmn	NA	literal title	FALSE	33bea6c4-b9bc-420a-b5fe-3ecbe0920ca2	2025-03-10

Query Details

- Query duration: 45ms
- Rows: 296.4K
- Query ID: 01baf12b-0001-a4af-00...

Ask Copilot

2:52 PM 3/11/2025



## 5] title.ratings

The screenshot shows the Snowflake web interface. The left sidebar displays the database structure: **IMDB\_DB** > **IMDB\_SCHEMA** > **Tables** > **TITLE\_RATINGS**. The main editor shows a SQL query:

```
--TABLES FROM DBEAVER
SELECT * FROM TITLE_RATINGS;
SELECT * FROM NAME_BASICS;
SELECT * FROM NAME_BASICS_FLATTEN;
```

The **Results** tab is active, displaying a table with 8 rows and 3 columns:

	TCONST	AVERAGERATING	NUMVOTES
1	TT0580671	8	9
2	TT2342059	7.9	26
3	TT11213434	9.7	31
4	TT0454230	6.1	60
5	TT6464426	7.6	111
6	TT1524275	8.2	6
7	TT8488538	7.2	78
8	TT1563143	7.2	19

Query Details on the right: Query duration 43ms, Rows 4.6M, Query ID 01baf16d-0001-a51f-00...

## 6] name.basics

The screenshot shows the Snowflake web interface. The left sidebar displays the database structure: **IMDB\_DB** > **IMDB\_SCHEMA** > **Tables** > **NAME\_BASICS**. The main editor shows a SQL query:

```
GRANT ROLE IMDB_RoLe TO ROLE ACCOUNTADMIN;
--TABLES FROM DBEAVER
SELECT * FROM TITLE_RATINGS;
```

The **Results** tab is active, displaying a table with 8 rows and 7 columns:

	NCONST	PRIMARYNAME	BIRTHYEAE	DEATHYEAE	PRIMARYPROFESSIC	KNOWNFORTITLES
1	nm10667873	Reine Fassi	0	0	NA	tt4964258
2	nm10667874	Roger Emmenecker	0	0	NA	tt4964258
3	nm10667875	Alain Saustier	0	0	NA	tt4964258
4	nm10667876	Jacques-Alain de Sedouy	1935	2018	NA	tt4964258
5	nm10667877	Lucien Zinger	0	0	NA	tt4964258
6	nm10667878	Manon Pignot	0	0	NA	tt4964258
7	nm10667879	Shelton Salie	0	0	actor	tt4467154
8	nm1066788	Jan Verbraecken	1923	1983	actor,soundtrack	tt3459472,tt0948548,tt

Query Details on the right: Query duration 6.4s, Rows 14.2M, Query ID 01baf175-0001-a51f-00...

Microsoft Azure | Data Factory | Midterm-Project-Team2 | Search factory and documentation

Factory Resources

- dataflow4
- date\_dimDL
- LanguageDimDF
- LoadFactMovie
- LoadGenreDim
- LoadPersonnelDim
- NameBasicsFlatten
- RegionDF
- title\_ratingsDL
- TitleAkasDF
- TitleBasicsDF
- TitleCrewDF
- TitleDimDF
- Power Query: 0
- Templates: 0

Bridge\_Movie | SnowflakeTable13 | pipeline5 | NameBasicsFlatten

Source1: Import data from SnowflakeTable1

derivedColumn1: Creating/updating the columns: 'NCONST, PRIMARYNAME, BIRTHYEAR, DEATHYEAR, PRIMARYPROFESSION, ...'

flatten1: Columns: 6 total

Sink1: Export data to NameBasicsFlattenTable

Flatten settings

Output stream name: flatten1

Description: Unrolling arrays from to with columns: 'NCONST, PRIMARYNAME, BIRTHYEAR, DEATHYEAR, PRIMARYPROFESSION, ...'

Incoming stream: derivedColumn1

Unroll by: PRIMARYPROFESSION, KNOWNFORTITLES

Unroll root: {}

Options: Skip duplicate input columns

app.snowflake.com/east-us-2.azure/aza59135/w2AKgaEUw5k3#query

Sample queries on TPC-H ... | 2025-02-06 12:12pm | 2025-02-20 4:26pm | IMDB Script

ACCOUNTADMIN | COMPUTE\_WH (X-Small) | Share

IMDB\_DB.IMDB\_SCHEMA | Settings

```
38 SELECT * FROM NAME_BASICS;
39 SELECT * FROM NAME_BASICS_FLATTEN;
40 SELECT * FROM DATE_DIM;
41 SELECT * FROM LANGUAGE_DIM;
42 SELECT * FROM BRIDGE_MOVIE;
```

Results

	NCONST	PRIMARYNAME	BIRTHYEAR	DEATHYEAR	PRIMARYPROFESSION	KNOWNFORTITLES
1	nm13598923	Katarina Fiallos	0	0	actor	tt28437533
2	nm13598923	Katarina Fiallos	0	0	actress	tt27155819
3	nm13598923	Katarina Fiallos	0	0	actress	tt28437533
4	nm13598924	Jonah Brown	0	0	NA	NA
5	nm13598925	Ratul Ahmed	0	0	NA	NA
6	nm13598926	Hannah	0	0	actress	tt21524284
7	nm13598927	Bob Jutson	0	0	actor	tt1260682
8	nm13598928	Sierra Segalowitz	0	0	actress	tt18183178

Query Details

Query duration: 38ms

Rows: 38.7M

Query ID: 01baf180-0001-a51f-00...

Show more

Ask Copilot

## 7] title.basics

The screenshot shows a SQL query editor with a query window containing the following SQL code:

```
4  
5 | select * from title_basics;  
6 | select * from title_basics_flatten;  
7  
8 | select * from title_episode;
```

The results pane displays a table with the following columns: TCONST, TITELTYPE, PRIMARYTITLE, ORIGINALTITLE, ISADULT, and STAR. The table contains 6 rows of data.

	TCONST	TITELTYPE	PRIMARYTITLE	ORIGINALTITLE	ISADULT	STAR
1	tt17592272	tvEpisode	Episode #1.29	Episode #1.29	FALSE	
2	tt7079292	tvMovie	Helping Hands	Helping Hands	FALSE	
3	tt17277626	tvEpisode	My Dad's Got Fans?!	My Dad's Got Fans?!	FALSE	
4	tt13563072	tvEpisode	The Sandys	The Sandys	FALSE	
5	tt30789616	tvEpisode	The Agony And Ecstasy Of 2023	The Agony And Ecstasy Of 2023	FALSE	
6	tt13178382	tvEpisode	Armed and Delicious	Armed and Delicious	FALSE	

Query Details: Query duration 6.5s, Rows 11.5M, Query ID 01baf132-0001-a521-00...

Here the flattening of columns on Genres, OriginalTitle and PrimaryTitle into a new table

The screenshot shows a SQL query editor with a query window containing the following SQL code:

```
4  
5 | select * from title_basics;  
6 | select * from title_basics_flatten;  
7  
8 | select * from title_episode;
```

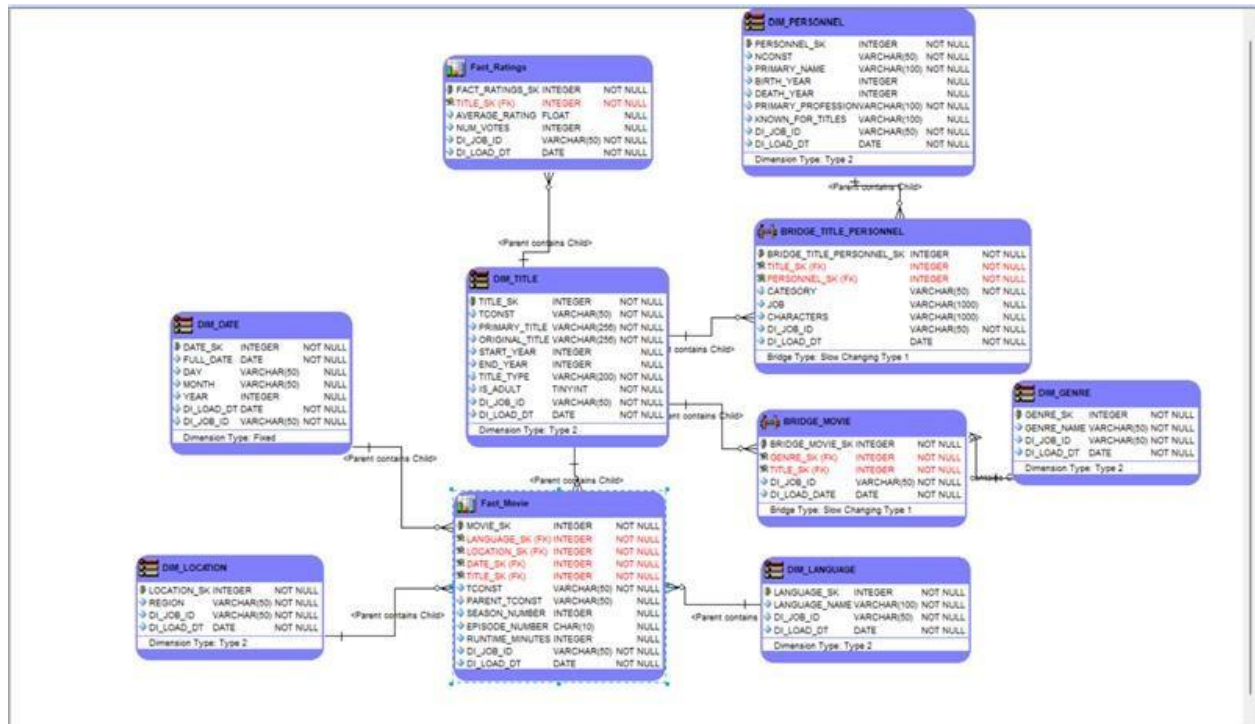
The results pane displays a table with the following columns: TCONST, TITELTYPE, PRIMARYTITLE, ORIGINALTITLE, ISADULT, and STAR. The table contains 6 rows of data.

	TCONST	TITELTYPE	PRIMARYTITLE	ORIGINALTITLE	ISADULT	STAR
1	tt4833930	tvEpisode	True Cheatin' Whore	True Cheatin' Whore	TRUE	
2	tt31242955	tvEpisode	Episode #1.8	Episode #1.8	FALSE	
3	tt14042862	tvEpisode	Episode #1.50	Episode #1.50	FALSE	
4	tt29931681	tvEpisode	Skye's Bathroom Video	Skye's Bathroom Video	TRUE	
5	tt1986887	movie	Rain or Shine	La pluie et le beau temps	FALSE	
6	tt6459118	tvEpisode	Capones Vag	Capones Vag	TRUE	

Query Details: Query duration 7.3s, Rows 18.3M, Query ID 01baf138-0001-a521-00...

**Loading Dims and Facts Tables :**

## ER Studio Diagram :



## 1] Date\_Dim



https://app.snowflake.com/east-us-2.azure/nua03035/worLh40kJUD#query

02-16 8:17pm 2025-02-16 8:33pm 2025-02-16 8:59pm 2025-02-20 11:47am Midterm Project - IMDB 2025-03-09 7:22pm

ACCOUNTADMIN COMPUTE\_WH (X-Small) Share

IMDB\_DB.IMDB\_SCHEMA Settings Code Versions

```
15
16 | select * from date_dim;
17 | select * from genre_dim;
18
19 DESCRIBE TABLE genre_dim;
```

Results Chart

	DATE	DAY_OF_WEEK	MONTH_NAME	YEAR_NUM	DI_JOB_ID	DI_LOAD_DT
1	2021-01-01	Fri	January	2021	2f755f14-f74b-4612-9a47-c886123b373c	2025-03-11
2	2021-01-02	Sat	January	2021	c690f2ca-a330-4f88-af91-416eae6de57d	2025-03-11
3	2021-01-03	Sun	January	2021	776302a3-0568-4399-89dd-034d9c988631	2025-03-11
4	2021-01-04	Mon	January	2021	77bda2ca-df6c-4c4a-97e4-4df6c4ef8ae9	2025-03-11
5	2021-01-05	Tue	January	2021	2d026e9e-ec98-48d6-9a6e-9e09a6b173a	2025-03-11
6	2021-01-06	Wed	January	2021	4c327f0e-d1bd-4325-ad18-b91a3f5b788b	2025-03-11
7	2021-01-07	Thu	January	2021	4d82a5c5-ed5e-4c92-9815-bdaa7622628c	2025-03-11

Query Details

Query duration 76ms

Rows 1.8K

Query ID 01baf13c-0001-a4b0-0...

Show more

DATE Ask Copilot

## 2] Genre\_Dim

02-16 8:17pm 2025-02-16 8:33pm 2025-02-16 8:59pm 2025-02-20 11:47am Midterm Project - IMDB 2025-03-09 7:22pm

ACCOUNTADMIN COMPUTE\_WH (X-Small) Share

IMDB\_DB.IMDB\_SCHEMA Settings Code Versions

```
15
16 | select * from date_dim;
17 | select * from genre_dim;
18
19 DESCRIBE TABLE genre_dim;
```

Results Chart

	#	GENRE_SK	GENRE	DI_JOB_ID	DI_LOAD_DT
1		8847361	Comedy	cc9265bc-c3ce-4431-95ad-6296f8542e78	2025-03-11
2		8847362	Drama	68b283e5-1505-47e9-86f0-61ac3caa89b3	2025-03-11
3		8847363	Drama	9cf3ada4-869f-4138-b5a9-984ee28b9fb3	2025-03-11
4		8847364	Romance	4d239280-a4f1-4025-a864-1da3fa86820f	2025-03-11
5		8847365	Adult	26b0f967-6fb0-459d-9df0-de1d283df778	2025-03-11
6		8847366	Crime	a833082c-d8a3-4583-8b77-49c736337493	2025-03-11
7		8847367	Drama	538ead83-5030-48d6-a0e9-fa5e741d294e	2025-03-11

Query Details

Query duration 55ms

Rows 18.3M

Query ID 01baf13d-0001-a51f-00...

Show more

Ask Copilot

## 3] Region\_Dim

02-16 8:17pm 2025-02-16 8:33pm 2025-02-16 8:59pm 2025-02-20 11:47am Midterm Project - IMDB 2025-03-09 7:22pm

ACCOUNTADMIN COMPUTE\_WH (X-Small) Share

IMDB\_DB.IMDB\_SCHEMA Settings Code Versions

```
23 GRANT INSERT ON ALL TABLES IN SCHEMA IMDB_SCHEMA TO ROLE IMDB_ROLE;  
24  
25 SELECT * FROM REGION_DIM;  
26  
27 SELECT * FROM GENRE_DIM;  
28
```

Results Chart

	REGION_CODE	REGION_NAME	DI_JOB_ID	DI_LOAD_DT	REGION_DIM_SK
1	AF	Afghanistan	96698ed8-175a-4b70-b562-3ce755840dcc	2025-03-11	1
2	AX	Åland Islands	cbccfbee-8d88-4008-9c3c-20406c41fe06	2025-03-11	2
3	AL	Albania	69c74cae-33f4-44d7-9019-274b5f4a3065	2025-03-11	3
4	DZ	Algeria	eb1cb253-df29-45a6-ad51-61a4b01e51a3	2025-03-11	4
5	AS	American Samoa	d383a1f1-60f2-4b25-8595-66baa93f494c	2025-03-11	5
6	AD	Andorra	bdaee182-47c8-4603-ba95-d2d2b10c4462	2025-03-11	6
7	AO	Angola	fb69649b-376c-477d-b419-33b901ea11c5	2025-03-11	7

Query Details

Query duration 117ms

Rows 228

Query ID 01baf13d-0001-a4b0-0...

Show more

REGION\_CODE Ask Copilot

## 4) Language\_Dim

Midterm-Proj Dataset Analy IMDB Script t.announcem... Commits - Sh... Data Profiling... Data Load,FL... +

app.snowflake.com/east-us-2.azure/aza59135/w2AKgaEUw5k3#query

Gmail Yahoo mail LinkedIn Git KK Ja ja DS CS Dhanvardini Rajend... Home | Microsoft 365 Viva Engage - Off C... Yudi J IMP - JOB INTERN >>

Sample queries on TPC-H... 2025-02-06 12:12pm 2025-02-20 4:26pm IMDB Script

ACCOUNTADMIN COMPUTE\_WH (X-Small) Share

IMDB\_DB.IMDB\_SCHEMA Settings Code Versions

```
38 SELECT * FROM NAME_BASICS;  
39 SELECT * FROM NAME_BASICS_FLATTEN;  
40 SELECT * FROM DATE_DIM;  
41 SELECT * FROM LANGUAGE_DIM;  
42 SELECT * FROM BRIDGE_MOVIE;  
43
```

Results Chart

	LANGUAGE_CODE	LANGUAGE_NAME	DI_JOB_ID	DI_LOAD_DT	LANGUAGE_DIM_SK
1	AF	Afrikaans	8eed8e4d-67d8-44bb-881c-ffc8729243b	2025-03-11	1
2	AX	Åland Islands	ad9c8ab5-d662-4379-8296-5f4d1dafd3ae	2025-03-11	2
3	AL	Albanian	ecfe3a73-a99d-4019-81a4-ae7ede20aa1f	2025-03-11	3
4	DZ	Algerian Arabic	687126d8-f9eb-4220-a47f-09aff7273669	2025-03-11	4
5	AS	American Sign Language	a4db349a-40dc-4aac-9fb5-ce3e5bd69497	2025-03-11	5
6	AD	Andorran Catalan	5c033310-a295-4d7e-ae9d-6361875cb50b	2025-03-11	6
7	AO	Angolan Portuguese	b9064141-4b30-4c3e-88d7-e5b3c070f780	2025-03-11	7
8	AI	Anguillian English	e7b7e37e-f9b8-4001-bc18-f35f0c88369c	2025-03-11	8

Query Details

Query duration 121ms

Rows 86

Query ID 01baf181-0001-a4b0-00...

Show more

LANGUAGE\_CODE 100% filled Ask Copilot

## 5) Personnel\_Dim

02-16 8:17pm 2025-02-16 8:33pm 2025-02-16 8:59pm 2025-02-20 11:47am Midterm Project - IMDB 2025-03-09 7:22pm

ACCOUNTADMIN COMPUTE\_WH (X-Small) Share

IMDB\_DB.IMDB\_SCHEMA Settings Code Versions

```

30
31
32 SELECT * FROM REGION_DIM;
33 SELECT * FROM FACT_MOVIE;
34

```

Results Chart

	PERSONNEL_SK	NCONST	PRIMARY_NAME	BIRTH_YEAR	DEATH_YEAR	PRIM
1	23693553	nm1591987	Amaury Gutiérrez	1963	0	soundtrac
2	23693554	nm1591987	Amaury Gutiérrez	1963	0	soundtrac
3	23693555	nm1591987	Amaury Gutiérrez	1963	0	soundtrac
4	23693556	nm1591987	Amaury Gutiérrez	1963	0	soundtrac
5	23693557	nm15919870	Jari Murto	0	0	NA
6	23693558	nm15919871	Jyri Komulainen	0	0	NA

Query Details

Query duration 18s

Rows 38.7M

Query ID 01baf13e-0001-a4af-00...

Show more

Ask Copilot

SSP

3:10 PM 3/11/2025

## 6] Title\_Dim

midterm x Midterm x damg737 x INFO610 x INFO610 x Shiyu-Pei x Google D x Data Prof x Share res x Power BI x

app.snowflake.com/east-us-2.azure/hoa42944/w64u5ku7GP#query

damgspring25 rypd 2025-02-17 8:57pm 2025-02-13 2:24pm 2025-02-23 4:56pm 2025-02-06 6:18pm midtermproject

ACCOUNTADMIN MIDTERM\_PROJECT (X-Small) Share

IMDB\_DB.IMDB\_SCHEMA Settings Code Versions

```

57 ORDER BY DT_LOAD_DT DESC
58 LIMIT 10;
59
60 DESC TABLE IMDB_SCHEMA.BRIDGE_TITLE_PERSONNEL;
61
62 select * from BRIDGE_TITLE_PERSONNEL;
63
64
65 select * from dim_title;
66

```

Results Chart

#	TITLE_SK	TCONST	PRIMARY_TITLE	ORIGINAL_TITLE	START_YEAR	END_YEAR	TITLE_TYPE	DT_LOAD_DT
1	4209535	tt4092626	Die Laubenpieper	Die Laubenpieper	1963	0	tvSeries	FALSE
2	4209536	tt3570924	Episode #1.11	Episode #1.11	1955	0	tvEpisode	FALSE
3	4209537	tt33086649	Episode #2.8	Episode #2.8	1962	0	tvEpisode	FALSE
4	4209538	tt33086649	Episode #2.8	Episode #2.8	1962	0	tvEpisode	FALSE
5	4209539	tt33086649	Episode #2.8	Episode #2.8	1962	0	tvEpisode	FALSE
6	4209540	tt12920816	Prisoners of the Body	Prisoners of the Body	2020	0	short	FALSE
7	4209541	tt9508736	Episode #1.209	Episode #1.209	1996	0	tvEpisode	FALSE
8	4209542	tt1881991	Episode #1.21	Episode #1.21	2011	0	tvEpisode	FALSE
9	4209543	tt6411566	The Letter	The Letter	2017	0	short	FALSE
10	4209544	tt6411566	The Letter	The Letter	2017	0	short	FALSE

Query Details

Query duration 23ms

Rows 18.3M

Query ID 01baf302-0001-a81e-0...

Show more

Ask Copilot

Results too large to display. Limit rows to under 1M.

KPK

10:43 PM 3/11/2025

**FACTS:**

## 1] Fact\_Movie

The screenshot shows a Snowflake query interface with the following SQL query:

```
31 SELECT * FROM REGION_DIM;
32 SELECT * FROM FACT_MOVIE;
```

The results table displays the following data:

	# MOVIE_ID	# TITLE_SK	TCONST	PARENT_TCONST	# SEASON_NUMB	# EPISODE_NUMB
1	0	13302861	tt0031458	tt32857063	0	0
2	0	13302861	tt0041951	tt0041038	1	9
3	0	13302861	tt0042816	tt0989125	1	17
4	0	13302861	tt0042889	tt0989125	0	0
5	0	13302861	tt0043426	tt0040051	3	42
6	0	13302861	tt0043631	tt0989125	2	16

Query Details: Query duration 1.1s, Rows 574.3K, Query ID 01baf2e7-0001-a61b-0...

## 2] Fact\_Ratings

The screenshot shows a Snowflake query interface with the following SQL query:

```
40 SELECT * FROM REGION_DIM;
41 SELECT * FROM LANGUAGE_DIM;
42 SELECT * FROM BRIDGE_MOVIE;
43 SELECT * FROM DIM_TITLE;
44 SELECT * FROM FACT_RATINGS;
```

The results table displays the following data:

	# FACT_RATINGS_SK	# TITLE_SK	# AVERAGE_RATIN	# NUM_VOTE	DI_LOAD_ID	DI_LOAD_D
1	1	13302845	8	9	e2af572e-1442-49e9-9c12-581be3b1b291	2025-03-12
2	2	13302846	8	9	5a11d6e3-8ad8-42eb-bf23-29e9406e3bc2	2025-03-12
3	3	13302847	8	9	59ac98c6-5f08-426a-b508-368acfb9124d	2025-03-12
4	4	13302848	8	9	9c62f86a-1f12-49d1-b6df-4db9e9e23681	2025-03-12
5	5	13302849	8	9	cab5ed53-56a2-4e1c-9ba7-fe9b9586706e	2025-03-12
6	6	13302850	8	9	1515bc1f-ab3b-4b41-b535-4f97dfa4d9ae	2025-03-12
7	7	13302851	8	9	f02ab2d6-4305-4fcd-b8eb-ca4f649c163f	2025-03-12
8	8	13302852	8	9	7edc77ef-8a97-4f2e-abb1-47edf5f99f4d	2025-03-12

Query Details: Query duration 707ms, Rows 2.5K, Query ID 01baf267-0001-a4b0-0...

BRIDGE TABLES :

## 1] Bridge Movie

Dim\_Title and Dim\_Genre

## 2] Bridge Title Personnel

Dim\_Title and Dim\_Personnel

61  
62 | select \* from BRIDGE\_TITLE\_PERSONNEL;  
63  
64

Results Chart

#	TITLE_SK	# PERSONNEL_SK	CATEGORY	JOB	CHARACTERE	DI_JOB_ID	DI_LOAD_DT
1	12783863	5446486	director	director	NA	21063d81-07cf-41ab-9e00-46c89b768ae1	2025-03-11
2	12783863	5446487	director	director	NA	46bfa9b0-e161-4a8a-84f4-aaca9baa86a9	2025-03-11
3	12783863	5446488	director	director	NA	ef497e44-9494-44b7-becb-ff3e3709af71	2025-03-11
4	12783863	5446489	director	director	NA	99b8586e-d2a3-455c-beb0-5e84f4cb83d5	2025-03-11
5	12783863	5446490	director	director	NA	7ca4d8d3-260d-41ed-9d8d-edd4ae4c5a45	2025-03-11
6	12783864	5446479	director	director	NA	2e2b851c-f19a-435e-b441-67a1f01e065d	2025-03-11
7	12783864	5446480	director	director	NA	d9084583-0655-47e1-9292-2b60165abd87	2025-03-11
8	12783864	5446481	director	director	NA	7d69444b-e61d-4151-9b7a-70ac7a69db0e	2025-03-11
9	12783864	5446482	director	director	NA	d6c667db-80f3-4429-ba39-687860fa566e	2025-03-11
10	12783864	5446483	director	director	NA	4e2a434e-c968-4d0a-8b06-4ab3377f5936	2025-03-11

Query Details

Query duration 189ms

Rows 1.38

Query ID 01ba72d2-0001-a52d-0...

Show more

Ask Copilot

Results too large to display. Limit rows to under 1M.

## Loading Fact\_Movie into Power BI



ert

Get data

Pa

### Navigator

Display Options

nua03035.east-us-2.azure.snowflakecompu...

IMDB\_DB [3]

IMDB\_SCHEMA [19]

DATE\_DIM

DIM\_TITLE

**FACT\_MOVIE**

GENRE\_DIM

LANGUAGE\_DIM

PERSONNEL\_DIM

REGION\_DIM

T\_8241\_249C7B01EED5488692EF3E...

T\_8405\_73E56619155E4BE89CFFED...

T\_8422\_AE076D41C4D14F8391C95...

T\_8830\_150CA712A647446C88BE59...

TITLE\_AKAS

TITLE\_AKAS\_FLATTEN

TITLE\_BASICS

TITLE\_BASICS\_FLATTEN

TITLE\_CREW

Select Related Tables

### FACT\_MOVIE

PARENT_TCONST	SEASON_NUMBER	EPISODE_NUMBER	RUNTIME_MINUTES	DI...
tt0793361	2	3	63	8
tt0793361	2	3	63	8
tt0793361	2	3	63	8
tt0797756	1	6	84	8
tt0797756	1	6	84	8
tt0212699	8	4	30	8
tt0212699	8	4	30	8
tt0069572	1	0	90	8
tt0069572	1	0	90	8
tt0806910	1	158	90	8
tt0806910	1	158	90	8
tt0806910	1	158	90	8
tt0149504	1	0	92	8
tt0167732	1	30	80	8
tt0167732	1	30	80	8
tt0252792	1	2	58	8
tt9141320	1	6	27	8
tt9141320	1	6	27	8
tt9141320	1	6	27	8
tt0115122	3	1	90	8
tt0115122	3	1	90	8
tt0270780	2	5	55	8

Load

Transform Data

Cancel

Share

Copilot

Copilot

Visualizations

Build visual

Data

Values

Add data fields here

Drill through

Cross-report

Keep all filters

Update available (click to download)

ENG IN

10:25 PM

3/11/2025

GIT SCREENSHOTS

COMMITTS:

Midterm x Dataset x IMDB S x Shriya-I x Data Pr x Data IL x IMDb I x Midterm x Data IL x +

github.com/Shriya-Pekamwar/DaDaBi\_MidTerm\_Project

Shriya-Pekamwar / DaDaBi\_MidTerm\_Project

Type / to search

<> Code Issues Pull requests Actions Projects Wiki Security Insights

DaDaBi\_MidTerm\_Project Public

Watch 1 Fork 0 Star 0

main 5 Branches 0 Tags

Go to file Add file Code

Shriya-Pekamwar	Updating pipeline: pipeline5	895c373 · 14 minutes ago	198 Commits
dataflow	Updating dataflow: FactRatings	50 minutes ago	
dataset	Updating SnowflakeV2Table: SnowflakeTable2	18 minutes ago	
factory	Merge branch 'main' into Shriya	yesterday	
linkedService	Updating linkedService: MidtermSnowflake	19 hours ago	
pipeline	Updating pipeline: pipeline5	14 minutes ago	
Readme.md	Create Readme.md	3 days ago	
publish_config.json	Update publish_config.json	3 days ago	

README

About

No description, website, or topics provided.

Readme

Activity

0 stars

1 watching

0 forks

Report repository

Releases

No releases published

Create a new release

Packages

No packages published

Midterm x Dataset x IMDB S x Commits x Data Pr x Data IL x IMDb I x Midterm x Data IL x +

github.com/Shriya-Pekamwar/DaDaBi\_MidTerm\_Project/commits/main/

## Commits

main All users All time

Commits on Mar 11, 2025

Updating pipeline: pipeline5	895c373	<>
Shriya-Pekamwar committed 15 minutes ago		
Updating pipeline: Language	da7b056	<>
Shriya-Pekamwar committed 15 minutes ago		
Updating SnowflakeV2Table: SnowflakeTable2	20d4a42	<>
Shriya-Pekamwar committed 19 minutes ago		
Updating pipeline: DimTitle	fe0c78c	<>
Shriya-Pekamwar committed 22 minutes ago		
Updating SnowflakeV2Table: DIM_TITLE_SF	4e5e5c3	<>
Shriya-Pekamwar committed 30 minutes ago		
Updating dataflow: LanguageDimDF	a5b6e1d	<>
Shriya-Pekamwar committed 33 minutes ago		
Updating dataflow: FactRatings	f83eccf	<>
DhanvardiniRajendran committed 51 minutes ago		
Updating dataflow: LoadFactMovie	f6ca4d0	<>

# Respective Git Contribution

The screenshot shows the GitHub repository 'DaDaBi\_MidTerm\_Project' on the 'main' branch. The repository is owned by 'Shriya-Pekamwar' and is public. It has 198 commits, 0 stars, and 0 forks. The commit history is as follows:

Commit	Author	Message	Time
895c373	Shriya-Pekamwar	Updating pipeline: pipeline5	16 minutes ago
		Updating dataflow: FactRatings	52 minutes ago
		Updating SnowflakeV2Table: SnowflakeTable2	20 minutes ago
		Merge branch 'main' into Shriya	yesterday
		Updating linkedService: MidtermSnowflake	19 hours ago
		Updating pipeline: pipeline5	16 minutes ago
		Create Readme.md	3 days ago
		Update publish_config.json	3 days ago

The repository also includes a README file and a publish\_config.json file.

The screenshot shows the same GitHub repository 'DaDaBi\_MidTerm\_Project' but on the 'Dhanvardini' branch. The repository is still owned by 'Shriya-Pekamwar' and is public. It has 31 commits, 0 stars, and 0 forks. The commit history is as follows:

Commit	Author	Message	Time
c106b68	DhanvardiniRajendran	Adding dataflow: dataflow2	yesterday
		Adding dataflow: dataflow2	yesterday
		Adding factory: Midterm-Project-Team2	3 days ago
		Adding linkedService: IMDBSnowflake	2 days ago
		Adding dataflow: date_dimDL	yesterday
		Create Readme.md	3 days ago
		Update publish_config.json	3 days ago

The repository also includes a README file and a publish\_config.json file. A message indicates that this branch is 167 commits behind the main branch.

<> CodeIssuesPull requestsActionsProjectsWikiSecurityInsightsSettings

DaDaBi\_MidTerm\_ProjectPublic

PinUnwatch1Fork0Star0

Komal5 Branches0 Tags

Go to filet+

<> CodeContribute

This branch is 15 commits ahead of, 190 commits behind main.

Komal1799Updating SnowflakeV2Table: Title\_Akas\_SFb68152e · yesterday16 Commits

dataset	Adding SnowflakeV2Table: Title_Akas_SF	2 days ago
factory	Adding factory: Midterm-Project-Team2	3 days ago
linkedService	Adding linkedService: SnowflakeMidterm	2 days ago
pipeline	Adding pipeline: pipeline1	2 days ago
Readme.md	Create Readme.md	3 days ago
publish_config.json	Create publish_config.json	3 days ago

About

No description, website, or topics provided.

ReadmeActivity0 stars1 watching0 forks

Releases

No releases published  
[Create a new release](#)

Packages

No packages published  
[Publish your first package](#)

10:37 PM  
3/11/2025

<> CodeIssuesPull requestsActionsProjectsWikiSecurityInsightsSettings

DaDaBi\_MidTerm\_ProjectPublic

PinUnwatch1Fork0Star0

Shriya5 Branches0 Tags

Go to filet+

<> CodeContribute

This branch is 1 commit ahead of, 94 commits behind main.

Shriya-PekamwarAdding pipeline: pipeline170fa41c · 20 hours ago98 Commits

dataflow	Merge branch 'main' into Shriya	20 hours ago
dataset	Merge branch 'main' into Shriya	20 hours ago
factory	Merge branch 'main' into Shriya	20 hours ago
linkedService	Merge branch 'main' into Shriya	20 hours ago
pipeline	Adding pipeline: pipeline1	20 hours ago
Readme.md	Create Readme.md	3 days ago
publish_config.json	Update publish_config.json	3 days ago

About

No description, website, or topics provided.

ReadmeActivity0 stars1 watching0 forks

Releases

No releases published  
[Create a new release](#)

Packages

No packages published  
[Publish your first package](#)

10:38 PM  
3/11/2025