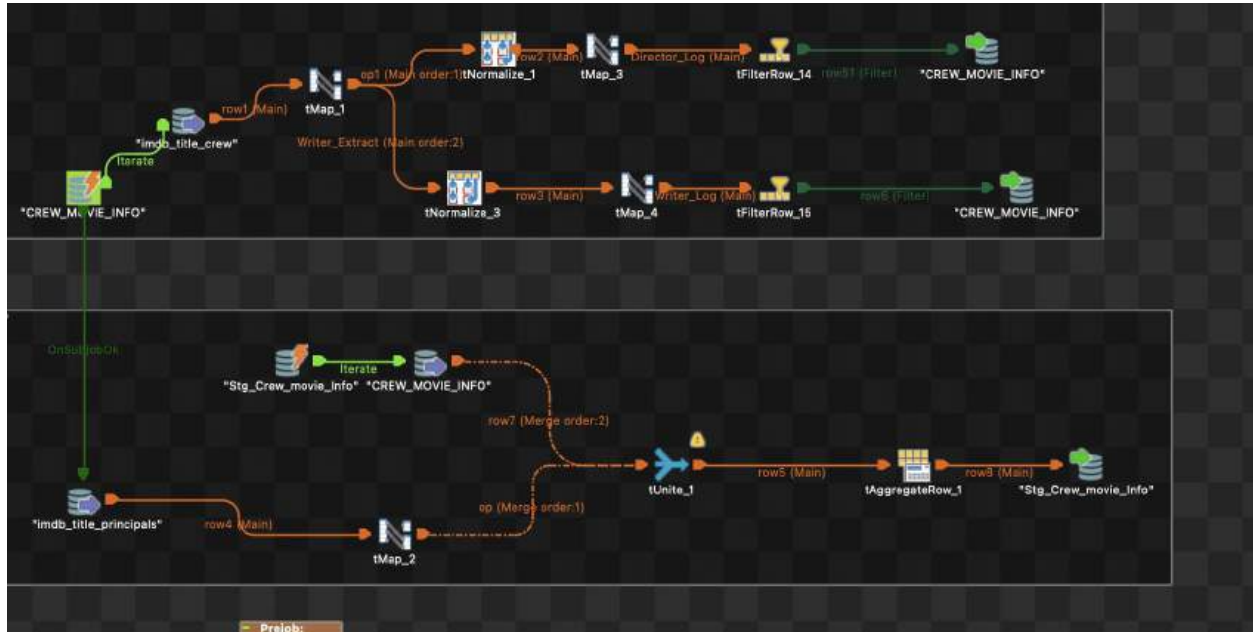


IMDB STG_CREATE

Crew_movie EXTRACT



```
1 • SELECT count(*) FROM imdb_src_movie.Stg_Crew_movie_Info;
```

100% 16:1

Result Grid Filter Rows: Search Export:

count(*)
4662016

Revenue Loading

The workflow diagram shows the following components and connections:

- tFileList_1** connects to **tFileInputDelimited_1** via an **Iterate** connector.
- tFileInputDelimited_1** connects to **tUnite_1** via a **row1 (Merge order:1)** connector.
- tUnite_1** connects to **tJavaRow_1** via a **row5 (Main)** connector.
- tJavaRow_1** connects to **tMap_1** via a **row2 (Main)** connector.
- tMap_1** connects to **tConvertType_1** via an **op (Main)** connector.
- tConvertType_1** connects to **tLogRow_1** via a **row3 (Main)** connector.
- tLogRow_1** connects to a database icon labeled **"imdb_revenue"** via a **row4 (Main)** connector.

Below the workflow, there is a **Prejob** section with **tPrejob_1** and a **Postjob** section with **Pre_Job**. A green arrow labeled **OnComponentOk** connects **tPrejob_1** to **Pre_Job**.

The **Designer** tab shows the following SQL query:

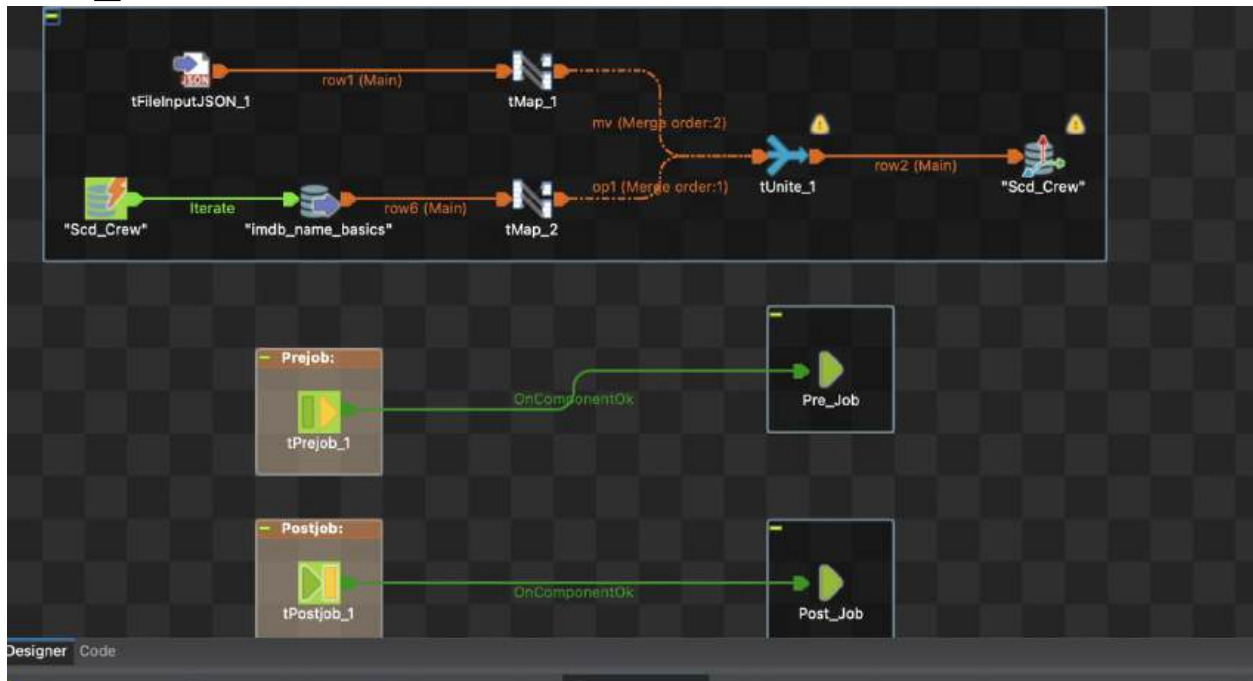
```
1 • SELECT * FROM imdb_src_movie.imdb_revenue;
```

The **Result Grid** shows the following data:

tconst	title	Date	Rank	Gross	YD	LW	Theaters	Per_Theaters	Total
tt0499549	Avatar	2009-12-18 00:00:00	1	26752099	0	0	3452	7750	26752
tt0499549	Avatar	2009-12-19 00:00:00	1	25529036	-5	0	3452	7395	52281
tt0499549	Avatar	2009-12-20 00:00:00	1	24744346	-3	0	3452	7168	77021
tt0499549	Avatar	2009-12-21 00:00:00	1	16385820	-34	0	3452	4747	93411
tt0499549	Avatar	2009-12-22 00:00:00	1	16086461	-2	0	3452	4660	10941
tt0499549	Avatar	2009-12-23 00:00:00	2	16445291	2.0	0	3452	4784	12591
tt0499549	Avatar	2009-12-24 00:00:00	1	11150596	-32	0	3452	3230	13701
tt0499549	Avatar	2009-12-25 00:00:00	2	23065046	107.00001	-14	3456	6683	16011
tt0499549	Avatar	2009-12-26 00:00:00	1	28274406	22.0	11.0	3456	8181	18841
tt0499549	Avatar	2009-12-27 00:00:00	1	24247681	-14	-2	3456	7016	21271
tt0499549	Avatar	2009-12-28 00:00:00	1	19418139	-20	19.0	3456	5619	25211
tt0499549	Avatar	2009-12-29 00:00:00	1	18290628	-6	14.0	3456	5292	25041
tt0499549	Avatar	2009-12-30 00:00:00	1	18466123	1.0	12.0	3456	5343	26881
tt0499549	Avatar	2009-12-31 00:00:00	1	14738136	-20	32.0	3456	4265	28361
tt0499549	Avatar	2010-01-01 00:00:00	1	25274008	71.0	9.0	3461	7303	30881
tt0499549	Avatar	2010-01-02 00:00:00	1	25835551	2.0	-9	3461	7465	33471
tt0499549	Avatar	2010-01-03 00:00:00	1	17381129	-33	-28	3461	5022	35211
tt0499549	Avatar	2010-01-04 00:00:00	1	8094554	-53	-58	3461	2339	36021
tt0499549	Avatar	2010-01-05 00:00:00	1	7327233	-9	-60	3461	2117	36751
tt0499549	Avatar	2010-01-06 00:00:00	1	5000167	6	63	3461	1465	37441

SCD Load

SCD_CREW



```
1 SELECT * FROM imdb_src_movie.Scd_Crew
2 where person_id in (Select distinct(person_id) from imdb_src_movie.Scd_Crew where scd_active=0)
```

100% 96:2

Result Grid Filter Rows: Search Edit: Export/Import:

Scd_crew_id	perso...	primary_name	birth_year	death_year	scd_start	scd_end	scd_version	scd_active
1343739	nm0000375	Robert Downey Jr.	1965	W	2022-04-20	2023-12-02	1	0
2797968	nm0000375	Robert Downey Jr. (aka IRON MAN)	1965	W	2023-12-02	HULL	2	1
1437536	nm0262635	Chris Evans	1981	W	2022-04-20	2023-12-02	1	0
2797969	nm0262635	Chris Evans (aka Captain America)	1981	W	2023-12-02	HULL	2	1
1489138	nm0424060	Scarlett Johansson	1984	W	2022-04-20	2023-12-02	1	0
2797972	nm0424060	Scarlett Johansson (aka Black Wid...	1984	W	2023-12-02	HULL	2	1
1591941	nm0749263	Mark Ruffalo	1967	W	2022-04-20	2023-12-02	1	0
2797970	nm0749263	Mark Ruffalo (aka Hulk)	1967	W	2023-12-02	HULL	2	1
1801647	nm1165110	Chris Hemsworth	1983	W	2022-04-20	2023-12-02	1	0
2797971	nm1165110	Chris Hemsworth (aka Thor)	1983	W	2023-12-02	HULL	2	1
1839166	nm1212722	Benedict Cumberbatch	1976	W	2022-04-20	2023-12-02	1	0
2797973	nm1212722	Benedict Cumberbatch (aka Dr. Str...	1976	W	2023-12-02	HULL	2	1
2305406	nm4043618	Tom Holland	1996	W	2022-04-20	2023-12-02	1	0
2797974	nm4043618	Tom Holland (aka Spiderman)	1996	W	2023-12-02	HULL	2	1
HULL	HULL	HULL	HULL	HULL	HULL	HULL	HULL	HULL

SCD_Movie

The screenshot displays a data pipeline workflow on a dark background. The main workflow consists of the following components and connections:

- tFileInputJSON_3** (Main) connects to **tLogRow_3** (Main).
- tLogRow_3** (Main) connects to **tMap_dop** (Merge order:2).
- "Scd_movie"** (Main) connects to **Iterate**, which then connects to **"imdb_title_basics"** (Main).
- "imdb_title_basics"** (Main) connects to **tMap_2** (Main).
- tMap_dop** (Merge order:2) and **tMap_2** (Main) both connect to **op** (Merge order:1).
- op** (Merge order:1) connects to **tUnite_3** (Main).
- tUnite_3** (Main) connects to **"Scd_movie"** (Main).

Below the main workflow, there are two sub-workflows:

- Prejob:** **tPrejob_2** connects to **Pre_Job** via **OnComponentOk**.
- Postjob:** (An empty box with a placeholder icon).

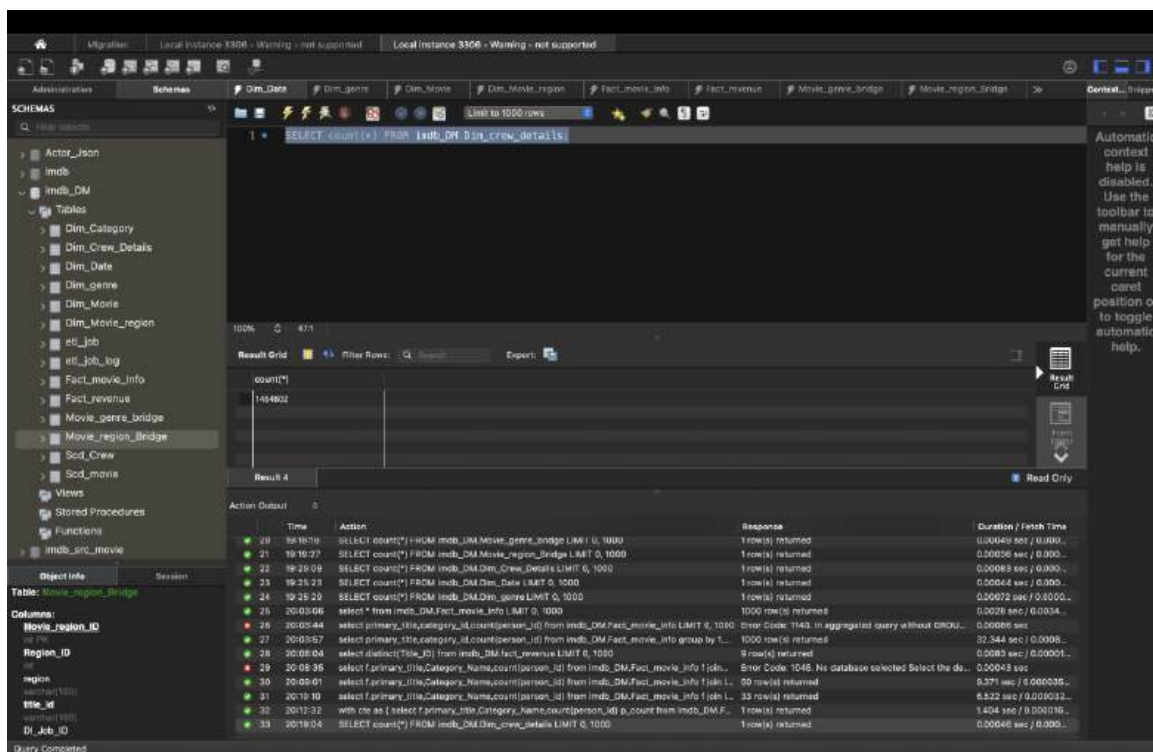
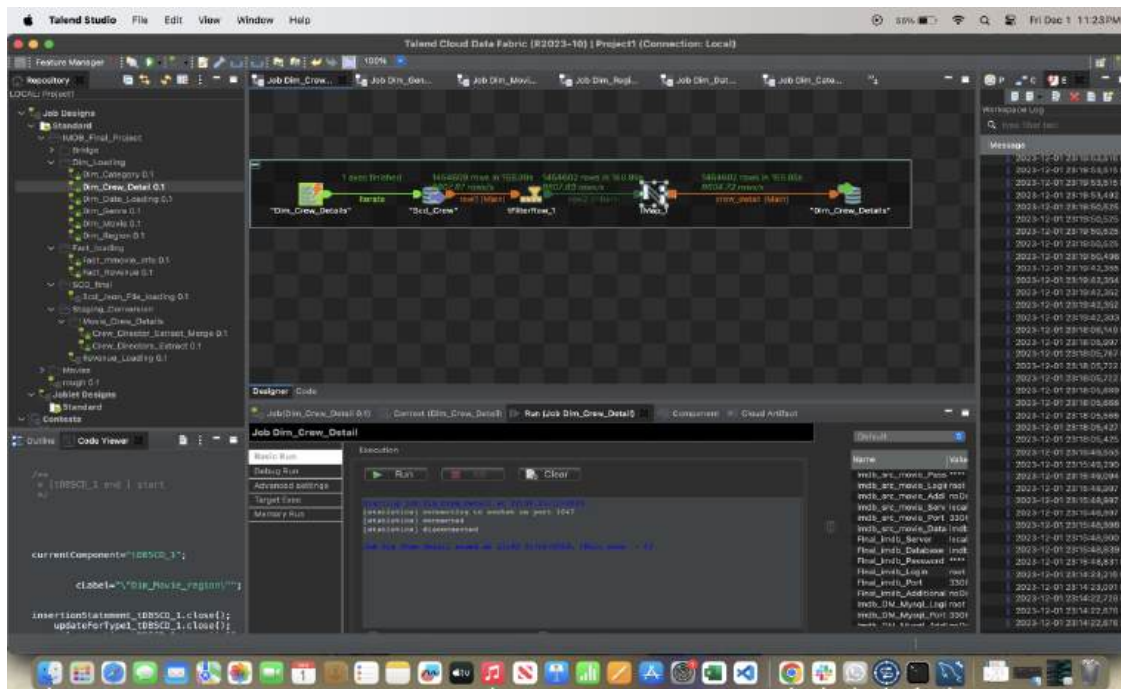
At the bottom, a SQL query is shown:

```
2 SELECT * FROM imdb_DM.Scd_movie
3 where title_id in (Select distinct(title_id) from imdb_DM.Scd_movie where scd_active=0)
```

Below the query is a **Result Grid** showing a table of movie data:

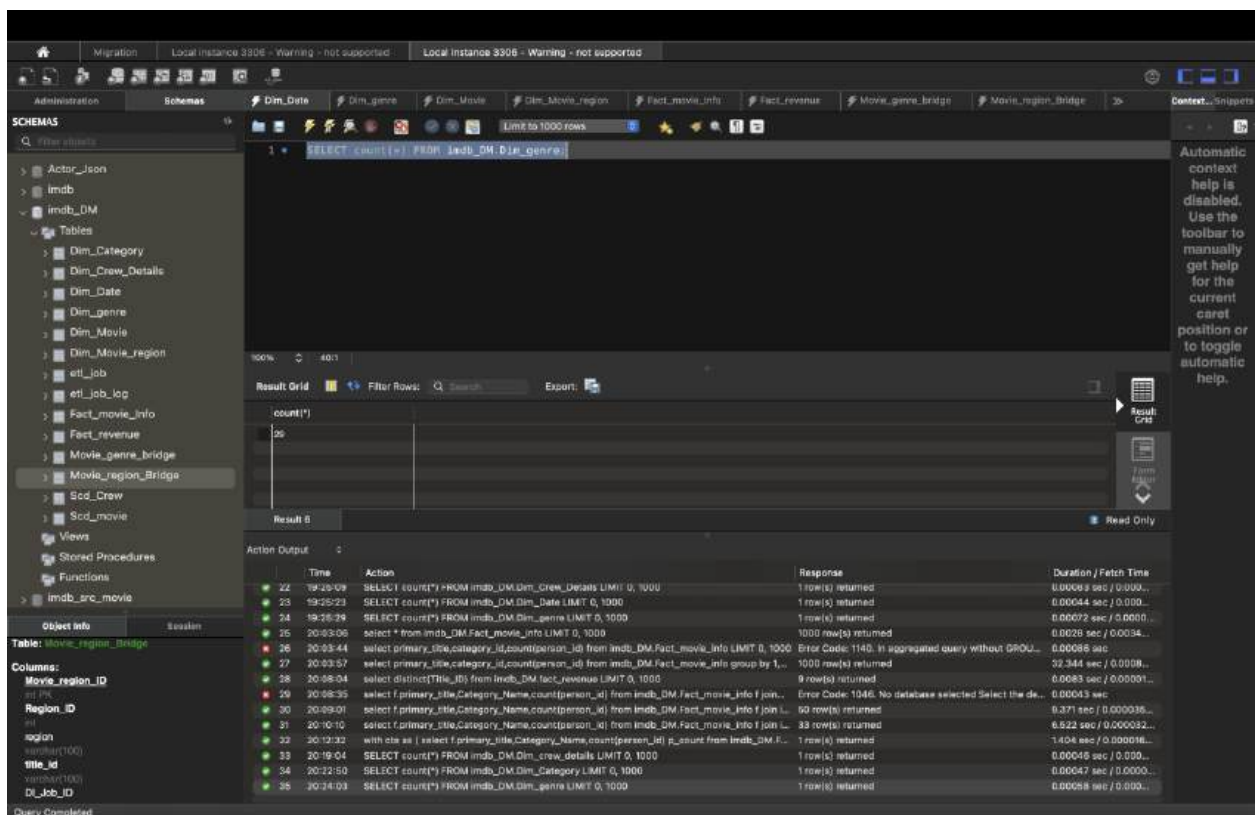
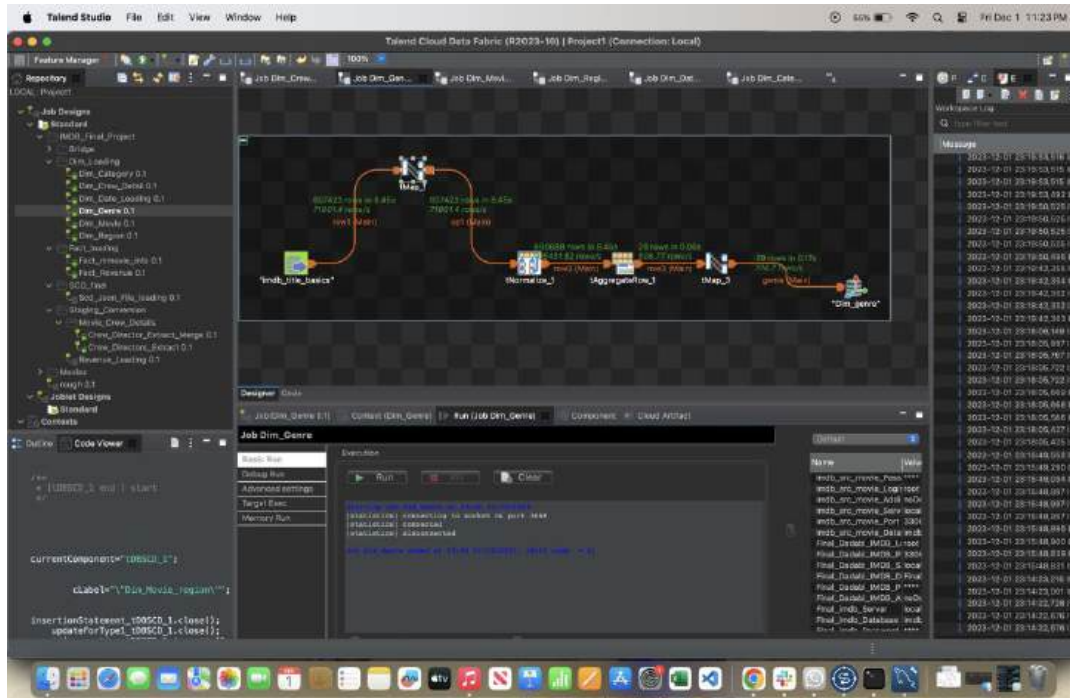
SCD_Movie_ID	title_id	primary_title	isAdult	startYear	runtime	scd_start	scd_end	scd_version	scd_active
82447	tt0120338	Titanic	0	1997	194	2022-04-20	2023-12-02	19	0
607424	tt0120338	Titanic (1997)	0	1997	194	2023-12-02	2023-12-02	20	0
607433	tt0120338	Titanic (1997) remake	0	1997	194	2023-12-02	2023-12-02	21	1
214938	tt0499549	Avatar	0	2009	162	2022-04-20	2023-12-02	13	0
607425	tt0499549	Avatar (2009)	0	2009	162	2023-12-02	2023-12-02	14	1
220622	tt0848228	The Avengers	0	2012	143	2022-04-20	2023-12-02	13	0
607426	tt0848228	The Avengers (2012)	0	2012	143	2023-12-02	2023-12-02	14	1
248001	tt10872600	Spider-Man: No Way Home	0	2021	148	2022-04-20	2023-12-02	13	0
607427	tt10872600	Spider-Man: No Way Home (...)	0	2021	148	2023-12-02	2023-12-02	14	1
380972	tt1825683	Black Panther	0	2018	134	2022-04-20	2023-12-02	13	0
607428	tt1825683	Black Panther (2018)	0	2018	134	2023-12-02	2023-12-02	14	1
418657	tt2395427	Avengers: Age of Ultron	0	2015	141	2022-04-20	2023-12-02	13	0
607429	tt2395427	Avengers: Age of Ultron (2015)	0	2015	141	2023-12-02	2023-12-02	14	1
422201	tt2488496	Star Wars: Episode VII - The...	0	2015	138	2022-04-20	2023-12-02	13	0
607430	tt2488496	Star Wars: Episode VII - The...	0	2015	138	2023-12-02	2023-12-02	14	1
473346	tt4154756	Avengers: Infinity War	0	2018	149	2022-04-20	2023-12-02	13	0
607431	tt4154756	Avengers: Infinity War (2018)	0	2018	149	2023-12-02	2023-12-02	14	1
473347	tt4154796	Avengers: Endgame	0	2019	181	2022-04-20	2023-12-02	13	0
607432	tt4154796	Avengers: Endgame (2019)	0	2019	181	2023-12-02	2023-12-02	14	1

Dim_Crew_details

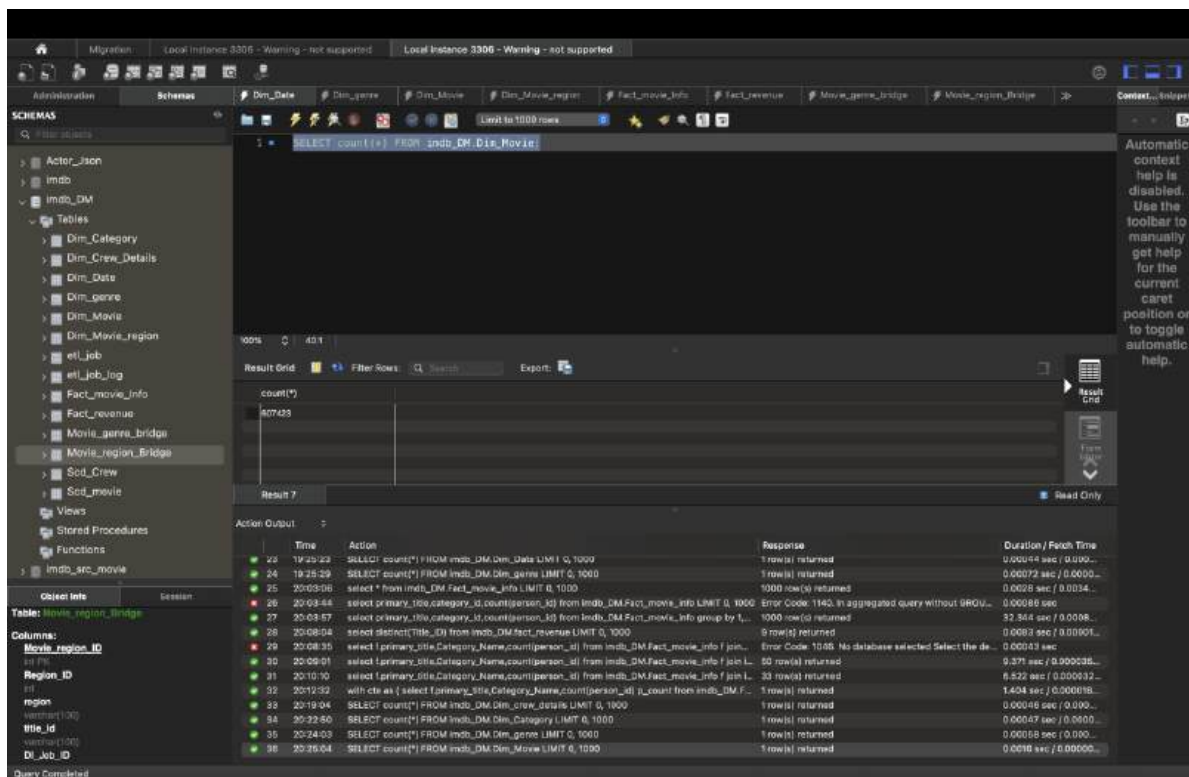
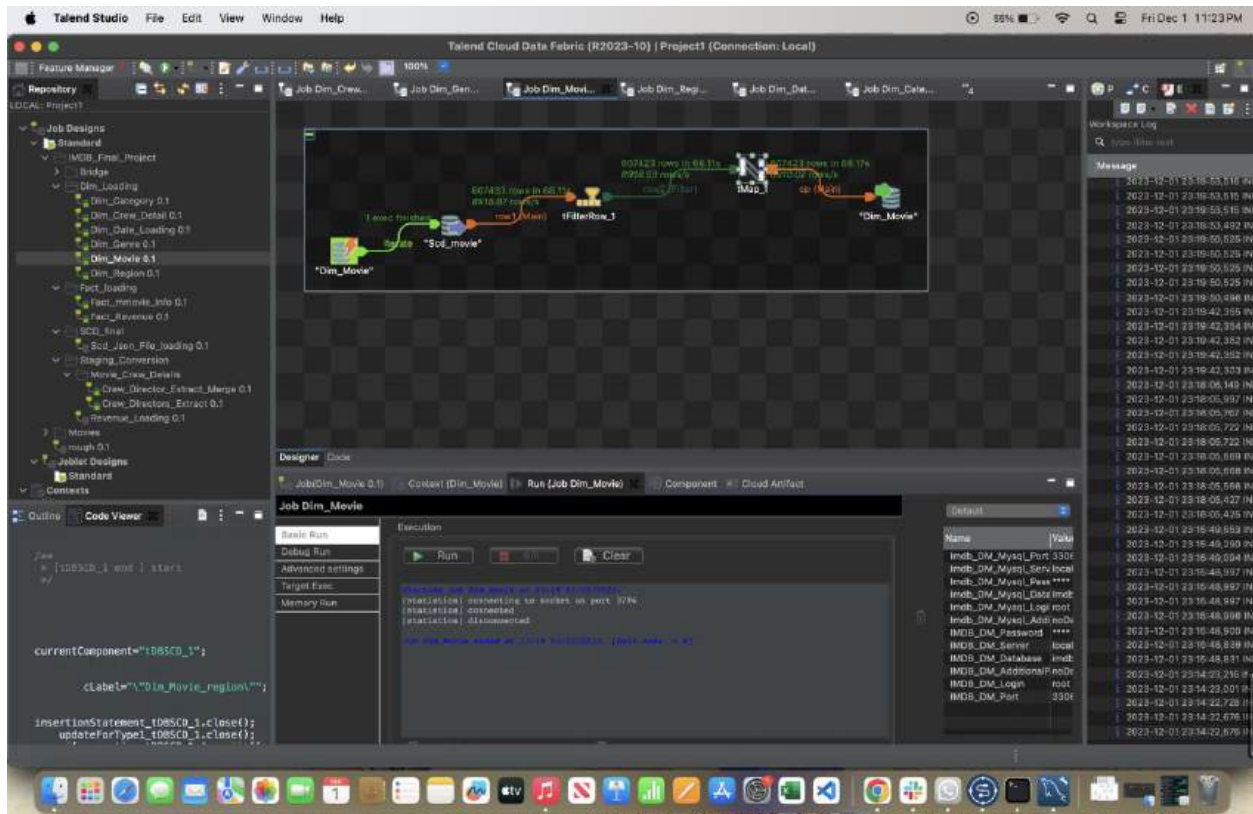


Dim_Category

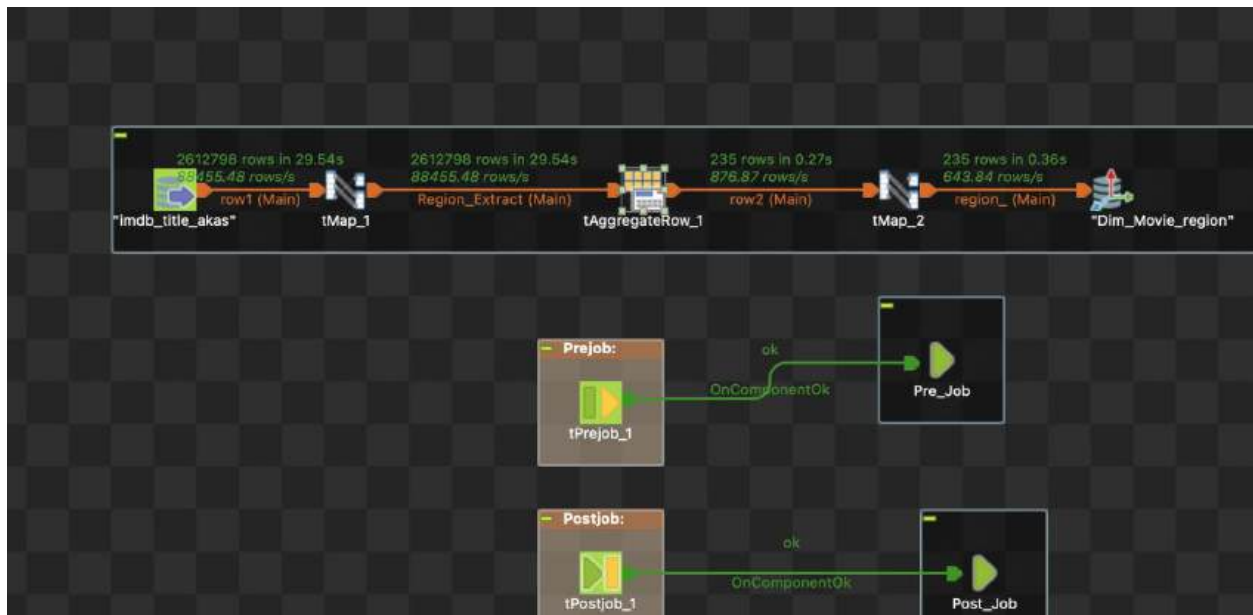
Dim_Genre



DIM_Movie



DIM_MOVIE_REGION



Administration Local Instance 3906 - Warning - not supported Local Instance 3906 - Warning - not supported

Schemas Administration Schemas Dim_Date Dim_genre Dim_Movie Dim_Movie_region Fact_movie_info Fact_revenue Movie_genre_bridge Movie_region_bridge Context Snippets

Filter objects

Actor_json imdb imdb_DM

Tables

- Dim_Category
- Dim_Crew_Details
- Dim_Date
- Dim_genre
- Dim_Movie
- Dim_Movie_region
- etl_job
- etl_job_log
- Fact_movie_info
- Fact_revenue
- Movie_genre_bridge
- Movie_region_bridge
- Scd_Crew
- Scd_movie

Views

Stored Procedures

Functions

imdb_src_movie

Object info Location

Table: Movie_region_Bridge

Columns:

- Movie_region_ID
- Region_ID
- Region
- title_id
- year(100)
- title_id
- year(100)
- title_id

Query: SELECT count(*) FROM imdb_DM.Dim_Movie_region

Result Grid

count(*)
235

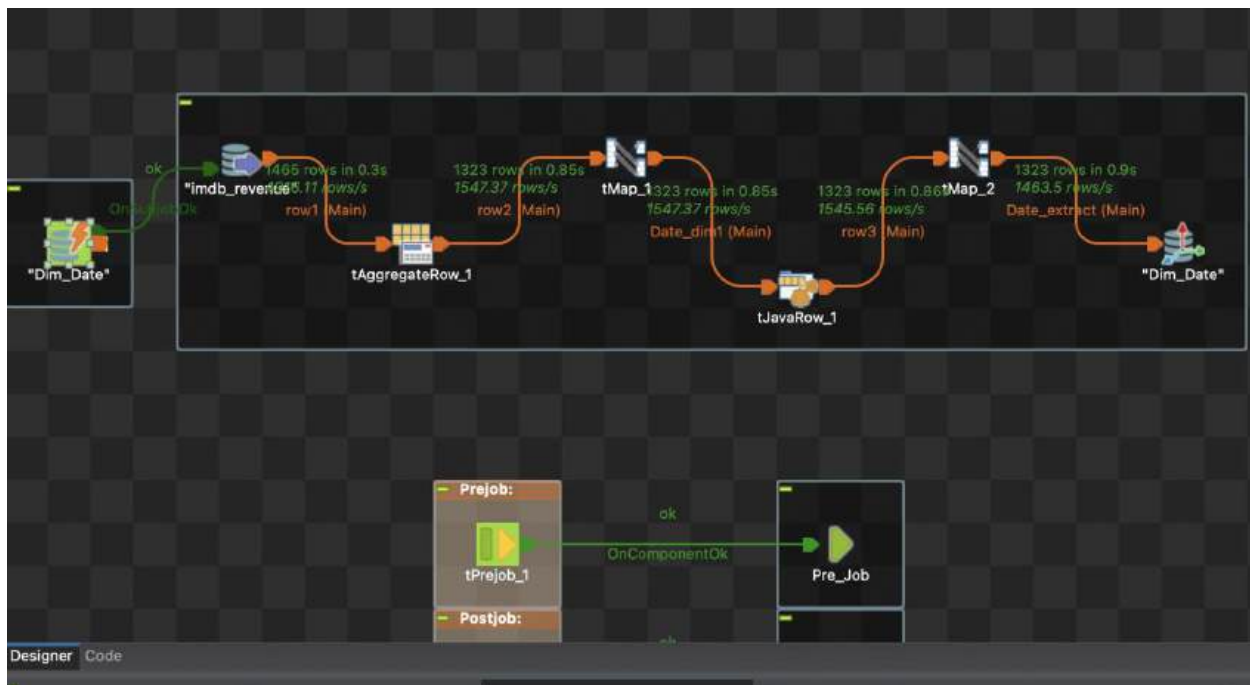
Action Output

Time	Action	Response	Duration / Fetch Time
24 19:29:29	SELECT count(*) FROM imdb_DM.Dim_genre LIMIT 0, 1000	1 row(s) returned	0.0002 sec / 0.0000...
25 20:03:06	select * from imdb_DM.Fact_movie_info LIMIT 0, 1000	1000 row(s) returned	0.0029 sec / 0.0034...
26 20:03:44	select primary_title,category_id,count(person_id) from imdb_DM.Fact_movie_info LIMIT 0, 1000	Error Code: 1145. In aggregated query without GROU...	0.00086 sec
27 20:03:57	select primary_title,category_id,count(person_id) from imdb_DM.Fact_movie_info group by 1,...	1000 row(s) returned	32.344 sec / 0.0008...
28 20:08:04	select distinct(Title_ID) from imdb_DM.Fact_revenue LIMIT 0, 1000	9 row(s) returned	0.0083 sec / 0.00001...
29 20:08:35	select f.primary_title,Category_Name,count(person_id) from imdb_DM.Fact_movie_info f join...	Error Code: 1046. No database selected Select the de...	0.00043 sec
30 20:09:01	select f.primary_title,Category_Name,count(person_id) from imdb_DM.Fact_movie_info f join...	50 row(s) returned	6.371 sec / 0.000038...
31 20:10:10	select f.primary_title,Category_Name,count(person_id) from imdb_DM.Fact_movie_info f join...	33 row(s) returned	6.522 sec / 0.000032...
32 20:10:32	with one as (select f.primary_title,Category_Name,count(person_id) p_count from imdb_DM.F...	1 row(s) returned	1.204 sec / 0.000016...
33 20:10:04	SELECT count(*) FROM imdb_DM.Dim_crew_details LIMIT 0, 1000	1 row(s) returned	0.00046 sec / 0.000...
34 20:12:10	SELECT count(*) FROM imdb_DM.Dim_Category LIMIT 0, 1000	1 row(s) returned	0.00017 sec / 0.0000...
35 20:24:03	SELECT count(*) FROM imdb_DM.Dim_genre LIMIT 0, 1000	1 row(s) returned	0.00088 sec / 0.000...
36 20:25:04	SELECT count(*) FROM imdb_DM.Dim_Movie LIMIT 0, 1000	1 row(s) returned	0.0016 sec / 0.00000...
37 20:26:04	SELECT count(*) FROM imdb_DM.Dim_Movie_region LIMIT 0, 1000	1 row(s) returned	0.00045 sec / 0.000...

Query Completed

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

DIM_DATE



Migration Local Instance 3306 - Warning - not supported Local Instance 3306 - Warning - not supported

Administration Schemas

Schemas

- Actor_json
- Imdb
- Imdb_DM
 - Tables
 - Dim_Category
 - Dim_Crew_Details
 - Dim_Date
 - Dim_Genre
 - Dim_Movie
 - Dim_Movie_region
 - Est_Job
 - Est_Job_Log
 - Fact_Movie_Info
 - Fact_Revenue
 - Movie_Genre_Bridge
 - Movie_region_Bridge
 - Scd_Crew
 - Scd_Movie
 - Views
 - Stored Procedures
 - Functions
 - Imdb_src_Movie

Object info

Table: Movie_region_Bridge

Columns:

- Movie_region_ID
- Region_ID
- title_id
- DI_Job_ID

Query Completed

1 * SELECT count(*) FROM Imdb_DM.Dim_date

Result Grid

count(*)
1323

Action Output

Time	Action	Response	Duration / Fetch Time
20:03:09	select * from imdb_DM.Fact_movie_info LIMIT 0, 1000	1000 row(s) returned	0.0028 sec / 0.00094...
20:03:44	select primary_title_category_id, count(person_id) from imdb_DM.Fact_movie_info LIMIT 0, 1000	Error Code: 1140. In aggregated query without GROUP...	0.00086 sec
20:03:57	select primary_title_category_id, count(person_id) from imdb_DM.Fact_movie_info group by 1,...	1000 row(s) returned	32.344 sec / 0.0008...
20:04:04	select distinct(TITLE_ID) from imdb_DM.Fact_revenue LIMIT 0, 1000	9 row(s) returned	0.0083 sec / 0.000091...
20:04:35	select f.primary_title, Category_Name, count(person_id) from imdb_DM.Fact_movie_info f join L...	Error Code: 1046. No database selected Select the de...	0.00043 sec
20:09:01	select f.primary_title, Category_Name, count(person_id) from imdb_DM.Fact_movie_info f join L...	50 row(s) returned	9.371 sec / 0.000035...
20:10:10	select f.primary_title, Category_Name, count(person_id) from imdb_DM.Fact_movie_info f join L...	33 row(s) returned	6.522 sec / 0.000032...
20:12:32	with cte as (select f.primary_title, Category_Name, count(person_id) p_count from imdb_DM.F...	1 row(s) returned	1.404 sec / 0.000018...
20:16:04	SELECT count(*) FROM imdb_DM.Dim_crew_details LIMIT 0, 1000	1 row(s) returned	0.00046 sec / 0.000...
20:22:50	SELECT count(*) FROM imdb_DM.Dim_Category LIMIT 0, 1000	1 row(s) returned	0.00047 sec / 0.0000...
20:24:09	SELECT count(*) FROM imdb_DM.Dim_Genre LIMIT 0, 1000	1 row(s) returned	0.00058 sec / 0.000...
20:26:04	SELECT count(*) FROM imdb_DM.Dim_Movie LIMIT 0, 1000	1 row(s) returned	0.0016 sec / 0.00000...
20:26:04	SELECT count(*) FROM imdb_DM.Dim_Movie_region LIMIT 0, 1000	1 row(s) returned	0.00045 sec / 0.000...
20:27:16	SELECT count(*) FROM imdb_DM.Dim_date LIMIT 0, 1000	1 row(s) returned	0.0012 sec / 0.00000...

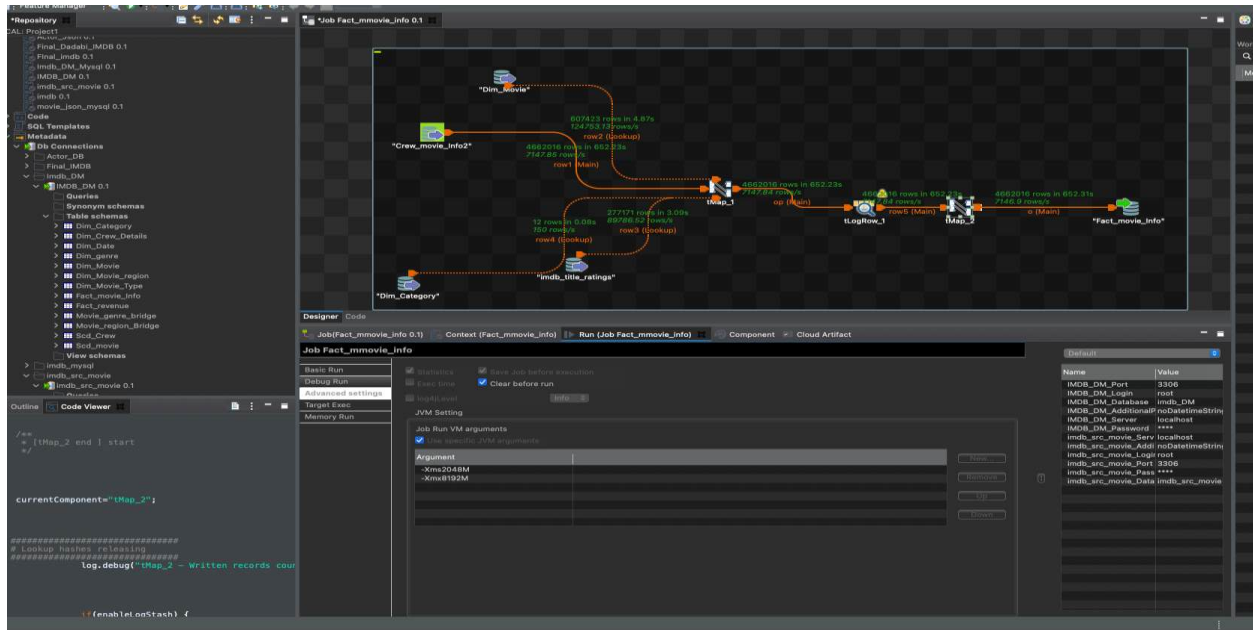
MOVIE_GENRE_BRIDGE

The screenshot displays the SAP Data Services Designer interface with a data flow diagram. The diagram includes the following components and data:

- Movie_region_Bridge**: A component that initiates the process with "1 exec finished". It connects to a database icon labeled **"imdb_title_akas"**.
- imdb_title_akas**: A database icon that feeds into a **tmap_1** component. Performance metrics for this step are: "2612798 rows in 254.26s" and "10276.13 rows/s".
- tmap_1**: A mapping component that connects to another database icon labeled **"Movie_region_Bridge"**. Performance metrics for this step are: "2612798 rows in 254.29s" and "10274.71 rows/s".
- Dim_Movie_region**: A database icon that connects to the **tmap_1** component. Performance metrics for this step are: "235 rows in 0.09s" and "2640.45 rows/s".
- Pre_Job**: A component that receives input from the **tmap_1** component via a green arrow labeled "OnComponentOk".

The diagram also shows a **Postjob** component at the bottom, which is currently empty.

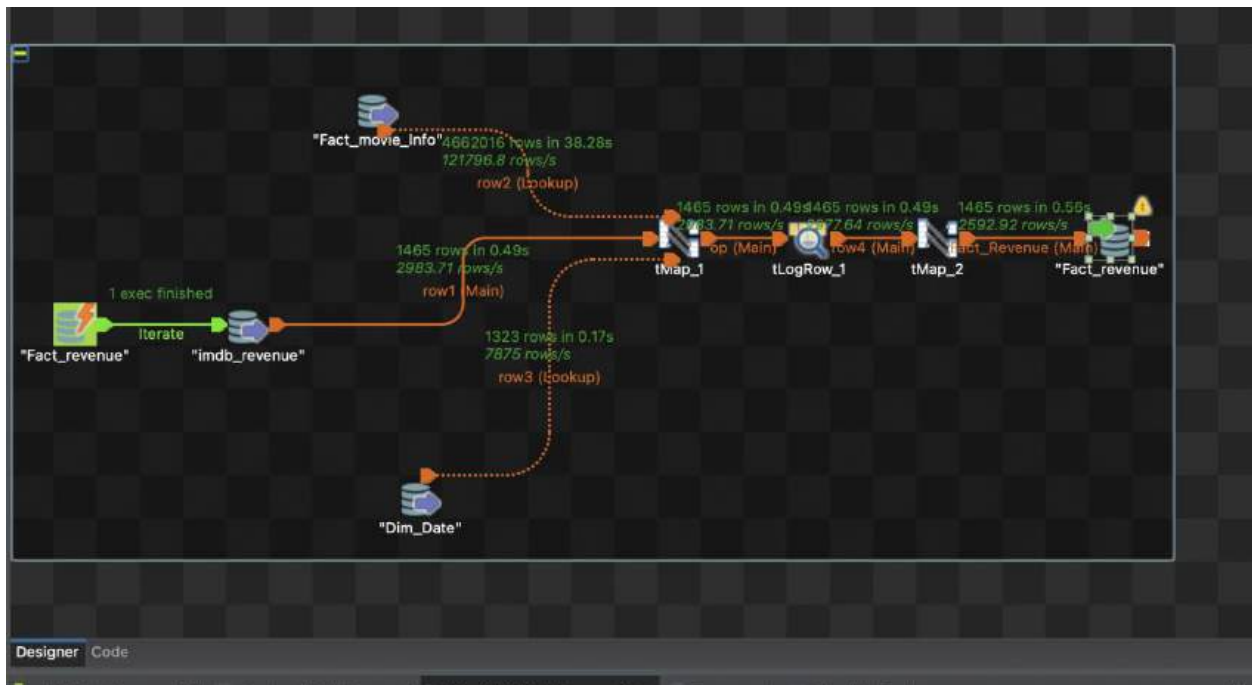
Fact_Movie_Info



The screenshot displays a database management tool interface. The 'Schemas' pane on the left shows a tree view of the database schema. The 'Query' pane shows a SQL query: `SELECT count(*) FROM imdb_DM.Fact_movie_Info`. The 'Result Grid' shows the query result, which is a single row with the value 4662016. The 'Action Log' pane shows the execution of the query, with a table of actions and their responses.

Time	Action	Response	Duration / Fetch Time
20:09:29	select primary_dim,Category_Name,count(person_id) from imdb_DM.Fact_movie_Info join L...	Error Code: 1046. No database selected select the da...	0.00043 sec
20:09:01	select primary_dim,Category_Name,count(person_id) from imdb_DM.Fact_movie_Info join L...	50 row(s) returned	0.00039 sec / 0.00039...
20:09:01	select primary_dim,Category_Name,count(person_id) from imdb_DM.Fact_movie_Info join L...	33 row(s) returned	0.00039 sec / 0.00039...
20:09:02	with cte as (select primary_dim,Category_Name,count(person_id) p_count from imdb_DM.F...	1 row(s) returned	1.024 sec / 0.000016...
20:10:04	SELECT count(*) FROM imdb_DM.dim_crew_details LIMIT 0, 1000	1 row(s) returned	0.00048 sec / 0.00...
20:12:50	SELECT count(*) FROM imdb_DM.dim_Category LIMIT 0, 1000	1 row(s) returned	0.00047 sec / 0.000...
20:14:03	SELECT count(*) FROM imdb_DM.dim_genre LIMIT 0, 1000	1 row(s) returned	0.00058 sec / 0.00...
20:15:04	SELECT count(*) FROM imdb_DM.dim_Movie LIMIT 0, 1000	1 row(s) returned	0.0016 sec / 0.0000...
20:16:04	SELECT count(*) FROM imdb_DM.dim_Movie_region LIMIT 0, 1000	1 row(s) returned	0.00045 sec / 0.00...
20:17:19	SELECT count(*) FROM imdb_DM.dim_data LIMIT 0, 1000	1 row(s) returned	0.0012 sec / 0.0000...
20:18:16	SELECT count(*) FROM imdb_DM.dim_movie_genre_bridge LIMIT 0, 1000	Error Code: 1146. Table 'imdb_dm.dim_movie_genre_...	0.0005 sec
20:18:27	SELECT count(*) FROM imdb_DM.movie_genre_bridge LIMIT 0, 1000	1 row(s) returned	0.00037 sec / 0.000...
20:19:45	SELECT count(*) FROM imdb_DM.movie_region_bridge LIMIT 0, 1000	1 row(s) returned	0.00057 sec / 0.000...
20:20:51	SELECT count(*) FROM imdb_DM.Fact_movie_Info LIMIT 0, 1000	1 row(s) returned	0.00053 sec / 0.000...

FACT_REVENUE



Migration Local Instance 3306 - Warning - not supported Local Instance 3306 - Warning - not supported

Administration Schemas

Filter objects

Actor_person

imdb

imdb_DM

Tables

Dim_Category

Dim_Crew_Details

Dim_Date

Dim_Genre

Dim_Movie

Dim_Movie_region

etl_job

etl_job_log

Fact_movie_Info

Fact_revenue

Movie_genre_bridge

Movie_region_Bridge

Scd_Draw

Scd_Movie

Views

Stored Procedures

Functions

imdb_src_movie

Object info Execution

Table: Movie_region_Bridge

Columns:

Movie_region_ID

PK

Region_ID

FK

region

varchar(100)

title_id

varchar(100)

DI_Job_ID

Query Completed

1 * SELECT count(*) FROM imdb_DM.Fact_revenue

Result Grid

count(*)
1465

Result 13

Action Output

Time	Action	Response	Duration / Fetch Time
20:09:01	select f.primary_title,category_name,count(person_id) from imdb_DM.Fact_movie_Info f join L...	50 row(s) returned	0.000000 sec / 0.000000...
20:10:10	select f.primary_title,Category_Name,count(person_id) from imdb_DM.Fact_movie_Info f join L...	33 row(s) returned	6.522 sec / 0.000032...
20:12:32	with cte as (select f.primary_title,Category_Name,count(person_id) p_count from imdb_DM.F...	1 row(s) returned	1.404 sec / 0.000016...
20:18:04	SELECT count(*) FROM imdb_DM.Dim_crew_Details LIMIT 0, 1000	1 row(s) returned	0.00046 sec / 0.000...
20:22:50	SELECT count(*) FROM imdb_DM.Dim_Category LIMIT 0, 1000	1 row(s) returned	0.00037 sec / 0.0000...
20:24:03	SELECT count(*) FROM imdb_DM.Dim_Genre LIMIT 0, 1000	1 row(s) returned	0.00058 sec / 0.000...
20:25:04	SELECT count(*) FROM imdb_DM.Dim_Movie LIMIT 0, 1000	1 row(s) returned	0.0018 sec / 0.00000...
20:26:04	SELECT count(*) FROM imdb_DM.Dim_Movie_region LIMIT 0, 1000	1 row(s) returned	0.00045 sec / 0.000...
20:27:15	SELECT count(*) FROM imdb_DM.Dim_date LIMIT 0, 1000	1 row(s) returned	0.0012 sec / 0.00000...
20:28:16	SELECT count(*) FROM imdb_DM.Dim_movie_genre_bridge LIMIT 0, 1000	Error Code: 1148. Table 'imdb_dm.dim_movie_genre_...	0.0005 sec
20:28:37	SELECT count(*) FROM imdb_DM.Movie_genre_bridge LIMIT 0, 1000	1 row(s) returned	0.00037 sec / 0.000...
20:29:45	SELECT count(*) FROM imdb_DM.Movie_region_Bridge LIMIT 0, 1000	1 row(s) returned	0.00057 sec / 0.000...
20:30:51	SELECT count(*) FROM imdb_DM.Fact_movie_Info LIMIT 0, 1000	1 row(s) returned	0.00083 sec / 0.000...
20:31:48	SELECT count(*) FROM imdb_DM.Fact_revenue LIMIT 0, 1000	1 row(s) returned	0.00050 sec / 0.000...

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

BI Questions:

Trend Analysis:

1. Average Movie Runtime over year

```
1 SELECT year_release, round(avg(cast(replace(runtime, '\\N', '') as float)), 2) FROM imdb_DM.Fact_movie_Info
2 where year_release < 2024
3 group by 1
4 order by 1 desc;
```

100% 17:4

Result Grid Filter Rows: Search Export:

year_release	round(avg(cast(replace(runtime, '\\N', '') as flo...
2023	14.89
2022	40.03
2021	66.12
2020	72.57
2019	80.11
2018	76.21
2017	76.62
2016	76.06
2015	77.82
2014	77.4
2013	78.12

2. Correlation Between avg rating and year release

```
1 SELECT year_release, round(avg(rating), 2) FROM imdb_DM.Fact_movie_Info
2 where rating < 10
3 group by 1
4 order by 2 desc;
```

00% 17:4

Result Grid Filter Rows: Search Export:

year_release	round(avg(rating)...
2021	6.32
1949	6.32
1929	6.31
1948	6.31
1947	6.28
1950	6.26
1946	6.26
1927	6.25
1955	6.23
1952	6.22
1940	6.21

3. Movie realease per year

```

1 SELECT year_release, round(avg(rating),2) FROM imdb_DM.Fact_movie_Info
2 where rating<10
3 group by 1
4 order by 2 desc;

```

00% 17:4

Result Grid Filter Rows: Search Export:

year_release	round(avg(rating),2)
2021	6.32
1949	6.32
1929	6.31
1948	6.31
1947	6.28
1950	6.26
1946	6.26
1927	6.25
1955	6.23
1952	6.22
1940	6.21

2.Average Rating and year relation

Actor and Actress Film Record :

```

1)with cte as
(SELECT d.primary_name,
       c.category_name,
       count(primary_title),
       rank()over(partition by c.category_name order by count(primary_title) desc) rnk
FROM imdb_DM.Fact_movie_Info f
join dim_crew_details d
on d.person_id = f.person_id
join dim_category c
on c.category_id = f.category_id
where c.category_name in ('actor','actress') and f.Rating between 4 and 7
group by 1,2
order by 2, 3 desc
)
select * from cte where rnk<11

```


Result Grid



Filter Rows:



Search

Export:



	primary_name	category_name	count(primary_tit...	rnk	
	Brahmanandam	actor	356	1	
	Cüneyt Arkin	actor	275	2	
	Mithun Chakraborty	actor	269	3	
	Tom Byron	actor	198	4	
	Dharmendra	actor	192	5	
	Ron Jeremy	actor	191	6	
	Mammooty	actor	184	7	
	Mohanlal	actor	183	8	
	Sadri Alisik	actor	173	9	
	Jeetendra	actor	172	10	
	Jamie Gillis	actor	172	10	

	primary_name	category_name	count(primary_tit...	rnk	
	Jamie Gillis	actor	172	10	
	Sridevi	actress	186	1	
	Türkan Soray	actress	171	2	
	Aachi Manorama	actress	150	3	
	Rekha	actress	148	4	
	Fatma Girik	actress	140	5	
	Hülya Koçyigit	actress	136	6	
	Aruna Irani	actress	131	7	
	Filiz Akin	actress	109	8	
	Hema Malini	actress	99	9	
	Catherine Deneuve	actress	97	10	

Genre Analysis

Popularity of genre over the years

```
1 SELECT genre_name,count(distinct(b.title_id)) FROM imdb_DM.Movie_genre_br
2 join imdb_DM.Fact_movie_Info i
3 on b.title_id=i.title_id
4 where year_release>year(current_date())-10
5 group by 1
6
```

100% 1:6

Result Grid Filter Rows: Search Export: Result Grid Form Editor

genre_name	count(distinct(b.ti...
Drama	49379
Documentary	45237
Comedy	23703
Thriller	12447
Horror	11498
Action	10710
Romance	8945
Crime	6950
Biography	5837

Top 5 genres as compared to Gross Earning

```
1 SELECT genre_name,sum(gross) FROM imdb_DM.Fact_revenue r
2 join imdb_DM.Movie_genre_bridge g
3 on g.title_id=r.Title_ID
4 group by 1
5 order by 2 desc
6 limit 5
7
```

100% 16:5

Result Grid Filter Rows: Search Export: Fetch rows:

genre_name	sum(gross)
Action	6090283475
Adventure	6090283475
Sci-Fi	3557064362
Fantasy	1614846113
Drama	1562451638

Seasonal_Analysis

```
1 SELECT Season, sum(gross) FROM imdb_DM.Fact_revenue r
2 join imdb_DM.Dim_Date d on
3 d.Date=r.date
4 group by 1
```

100% 1:1

Result Grid Filter Rows: Search Export:

	Season	sum(gro...	
	Spring	3033650624	
	Summer	1962604141	
	Fall	1738107348	

2. top 3 movie based on season

```
with cte as
(SELECT season, primary_title, sum(gross),
row_number() over (partition by season order by sum(gross) desc)
rw FROM imdb_DM.Fact_revenue r
join imdb_DM.Dim_Date d on
d.Date=r.date
join imdb_DM.Dim_Movie m
on m.title_id=r.Title_ID
group by 1,2)
select * from cte where rw<4
```

100% 29:11 1 error found

Result Grid Filter Rows: Search Export:

	season	primary_title	max(total_gross)	rw	
	Fall	Avengers: Endgame (2019)	858373000	1	
	Fall	Avatar (2009)	760507625	2	
	Fall	Avengers: Infinity War (2018)	678815482	3	
	Spring	Star Wars: Episode VII - The Force Awakens (2...	935642689	1	
	Spring	Spider-Man: No Way Home (2021)	804337488	2	
	Spring	Avatar (2009)	746830110	3	
	Summer	Avengers: Endgame (2019)	858301342	1	
	Summer	Avatar (2009)	754612474	2	
	Summer	Black Panther (2018)	700059566	3	

Result 5

Release Region:

```

1 SELECT primary_title,count(region) FROM imdb_DM.Movie_region_bridge b
2 join imdb_DM.dim_movie m
3 on m.title_id=b.title_id
4 group by 1
5 order by 2 desc
6

```

100% 11:4

Result Grid Filter Rows: Search Export: Fetch rows:

	primary_title	count(regio...	
	The Three Musketeers	380	
	Mother	377	
	Les Misérables	351	
	Cinderella	285	
	Beauty and the Beast	283	
	Hamlet	259	
	Home	244	
	Love	243	
	Desire	242	
	A Christmas Carol	239	
	The Stranger	236	
	Revenge	233	
	Alone	230	
	Paradise	228	

Matrix

```
with cte as (  
select f.primary_title,Category_Name,count(person_id) p_count from imdb_DM.Fact_movie_info f  
join imdb_DM.Dim_category c on  
f.category_id=c.category_id  
where title_id in (select distinct(Title_ID) from imdb_DM.fact_revenue) and Category_Name  
in('actress','actor','director','writer')  
group by 1,2)  
  
select sum(p_count) from cte  
where Category_Name='director'
```

Director

```
1 with cte as (  
2   select f.primary_title,Category_Name,count(person_id) p_count from imdb_DM.Fact_movie_info f  
3   join imdb_DM.Dim_category c on  
4     f.category_id=c.category_id  
5   where title_id in (select distinct(Title_ID) from imdb_DM.fact_revenue) and Category_Name in('actress','actor','director','writer')  
6   group by 1,2)  
7  
8   select sum(p_count) from cte  
9   where Category_Name='director'
```

100% 31:9

Result Grid Filter Rows: Search Export

sum(p_count)
11

Result Grid Form Editor

Actor:

```
1 with cte as (  
2   select f.primary_title,Category_Name,count(person_id) p_count from imdb_DM.Fact_movie_info f  
3   join imdb_DM.Dim_category c on  
4     f.category_id=c.category_id  
5   where title_id in (select distinct(Title_ID) from imdb_DM.fact_revenue) and Category_Name in('actress','actor','director','writer')  
6   group by 1,2)  
7  
8   select sum(p_count) from cte  
9   where Category_Name='writer'
```

100% 1:10

Result Grid Filter Rows: Search Export

sum(p_count)
45

Result Grid Form Editor

Similarly count for
Actress: 8, writers:24

Total_World Wide Gross Earning

```
1 SELECT title,max(total_gross),sum(gross) FROM imdb_DM.Fact_revenue fr
2 group by 1;
```

100% 12:2

Result Grid Filter Rows: Search: Export:

title	max(total_gross)	sum(gross)
Avatar	760507625	760508625
Avengers: Age of Ultron	455998418	483598418
Avengers: Endgame	858373000	918373000
Avengers: Infinity War	678815482	717815482
Black Panther	700059566	725259566
Spider-Man: No Way Home	804337488	854337488
Star Wars: Episode VII - The Force Awakens	935642689	992642689
The Avengers	621623219	637748207
Titanic	659363944	644078638

5.Average_Rating

```
1 SELECT primary_title,round(avg(rating),2) FROM imdb_DM.Fact_movie_Info
2 where title_id in (select Distinct(Title_ID) from imdb_DM.Fact_revenue)
3 group by 1
4
```

100% 1:4

Result Grid Filter Rows: Search: Export:

primary_title	round(avg(rating)...
Avatar (2009)	7.8
Avengers: Age of Ultron (2015)	7.3
Avengers: Endgame (2019)	8.4
Avengers: Infinity War (2018)	8.5
Black Panther (2018)	7.3
Spider-Man: No Way Home (2021)	8.5
Star Wars: Episode VII - The Force Awakens (2...	7.9
The Avengers (2012)	8
Titanic (1997) remake	7.9

No of Genre Per Movie

```
1 SELECT primary_title,count(genre_name) FROM imdb_DM.Movie_genre_bridge b
2 join imdb_DM.dim_movie m
3 on m.title_id=b.title_id
4 where m.title_id in (select Distinct(Title_ID) from imdb_DM.Fact_revenue)
5 group by 1
```

100% 11:5

Result Grid Filter Rows: Search Export:

primary_title	count(genre_name)
Avatar (2009)	3
Avengers: Age of Ultron (2015)	3
Avengers: Endgame (2019)	3
Avengers: Infinity War (2018)	3
Black Panther (2018)	3
Spider-Man: No Way Home...	3
Star Wars: Episode VII - The...	3
The Avengers (2012)	3
Titanic (1997) remake	2

6 No of region,movie was released

```
1 SELECT primary_title,count(region) FROM imdb_DM.Movie_region_bridge b
2 join imdb_DM.dim_movie m
3 on m.title_id=b.title_id
4 where m.title_id in (select Distinct(Title_ID) from imdb_DM.Fact_revenue)
5 group by 1
6
```

100% 1:6 1 error found

Result Grid Filter Rows: Search Export:

primary_title	count(regio...
Avatar (2009)	58
Avengers: Age of Ultron (2015)	59
Avengers: Endgame (2019)	56
Avengers: Infinity War (2018)	55
Black Panther (2018)	48
Spider-Man: No Way Home (2021)	60
Star Wars: Episode VII - The Force Awakens (2...	76
The Avengers (2012)	67
Titanic (1997) remake	69

ALTERYX Profiling

imdb_name_basics

column	max	min	Null%
nconst	10	9	0
Primary name	95	1	0
Birthyear	4	2	0
deathyear	4	2	0
Primary profession	65	5	0
Known for titles	10	5	0

There are 265 birthyears which are below 1,800 and there are 1,208,402 records where the birthyear is '\n'

There are 1,35,378 records where there are '/n' death years

Imdb_title_akas

column	max	min	Null%
Title id	10	9	0
ordering	149	1	0
title	273	1	0
region	4	2	0
language	3	2	0

LANGUAGE CONTAINS 2177430 NULL VALUES

Imdb_title_basics

column	max	min	Null%
Tconst	10	9	0
Titletype	5	5	0
Primary title	242	1	0
Original title	242	1	0
startyear	4	2	0
endyear	2	1	0
runtimes	5	1	0

THERE ARE 62 RECORDS WHICH HAVE A MOVIE RUNTIME OF MORE THAN 300 MINUTES WHICH IS AN ABBERRATION

Imdb_title_crew

Column	Max	Min	Null%
--------	-----	-----	-------

directors	29	2	0
writers	1189	2	0

The names for the writers and directors are not the conventional names its just abberations
 Imdb_title_principle

Column	Max	Min	Null%
ordering	10	1	0
category	19	4	0
job	157	1	0
titleID	10	9	0