



Northeastern

College of Engineering

Course: INFO7370 Designing Data Architecture and Business Intelligence

Instructor: Prof. Richard Sherman

Date: 05/04/2022

Team 14 Members:

Anshita Verma

Vini Wategaonkar

Divya Patil

Arvinda Sagiraju

Overview:

This file consist of snapshots of Talend jobs, Alteryx jobs created for creating the data pipelines for IMDb Rating project.

The snapshots shows the successful run of jobs and the final rows affected and stored into the staging tables.

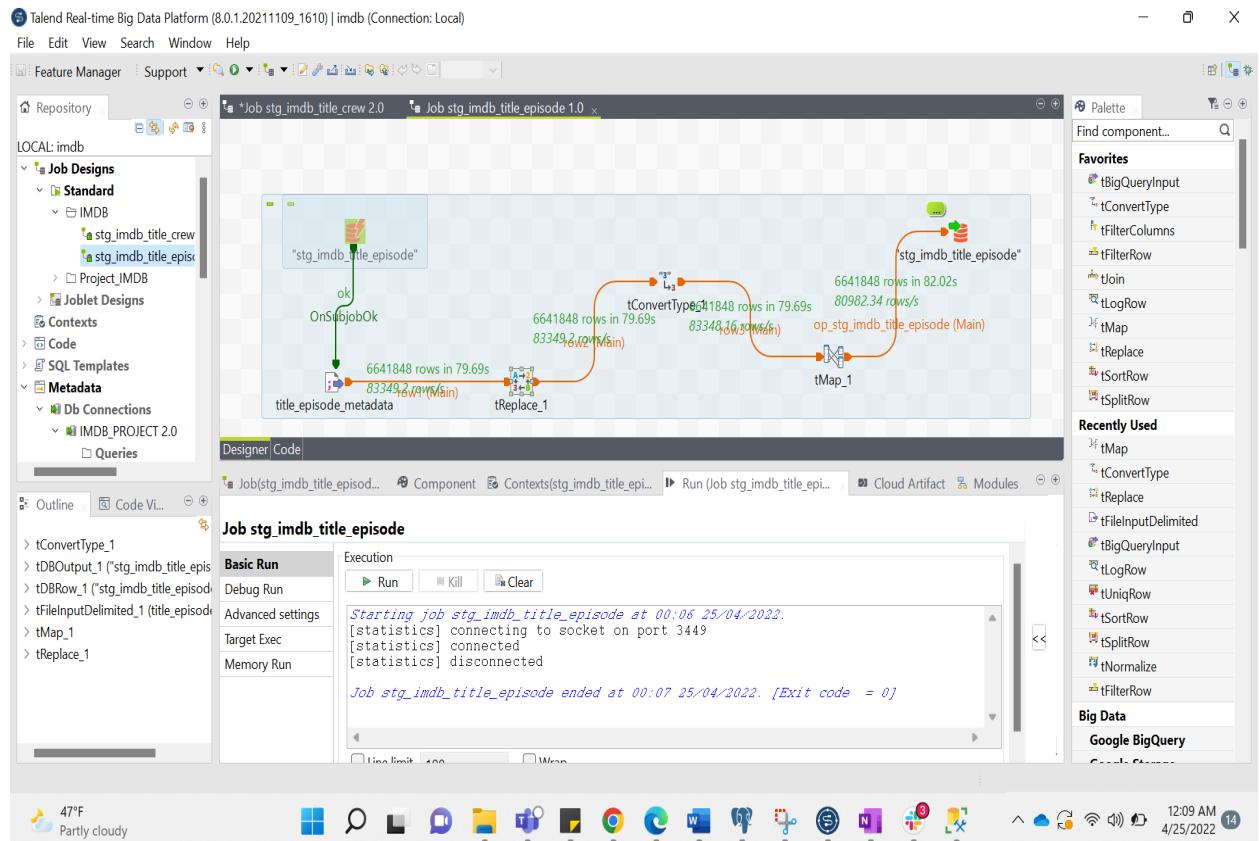
Also, consist of views of MS SQL server (which is used as a storage server for the staging and BI data) final row count stored into the database.

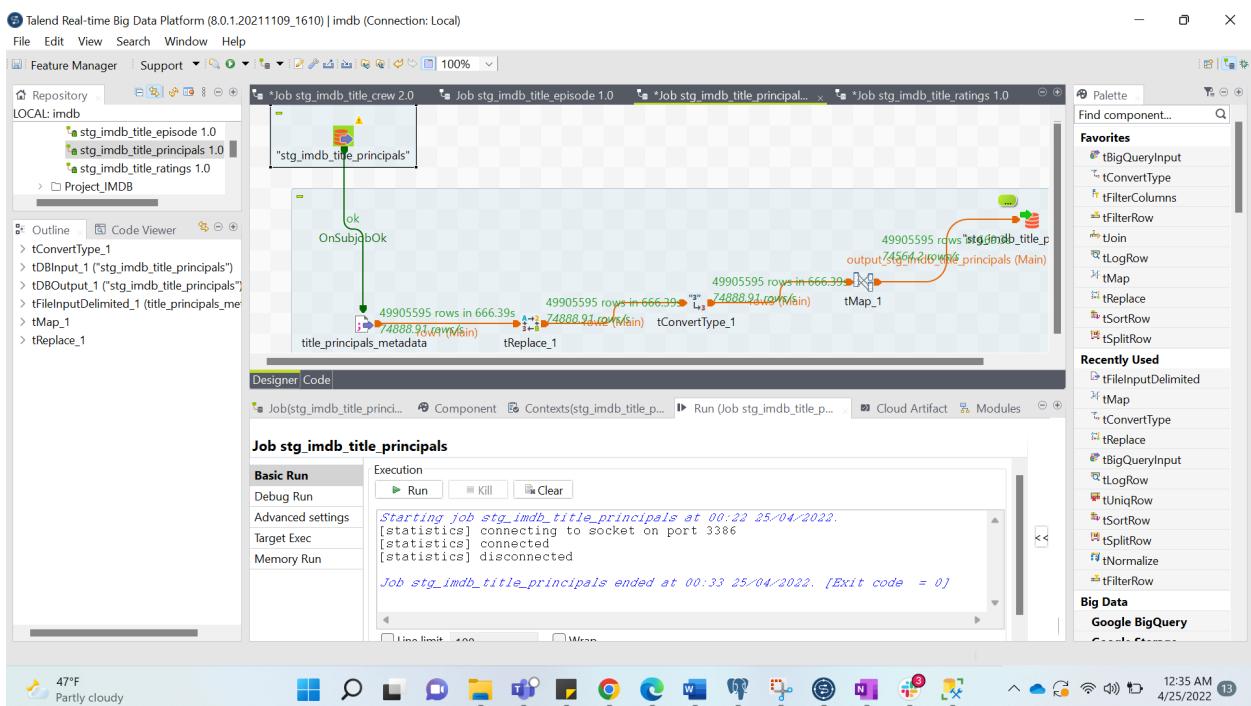
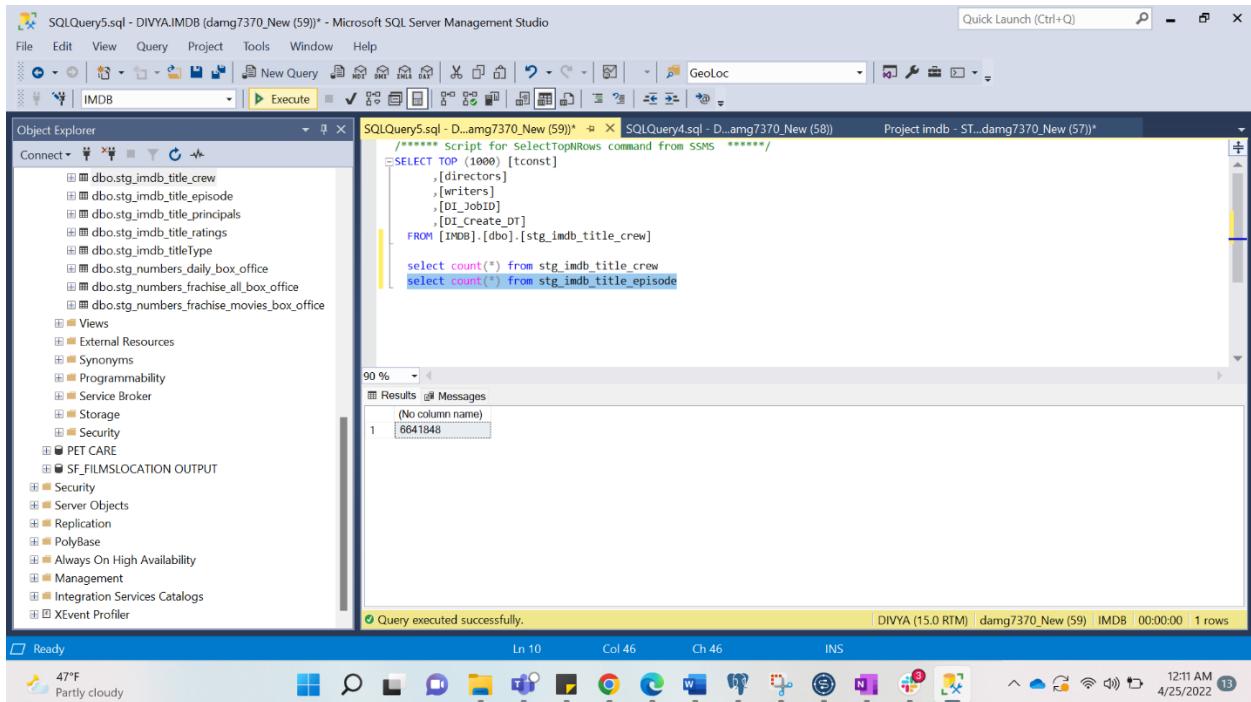
Final Row Counts of tables in Staging Schema:

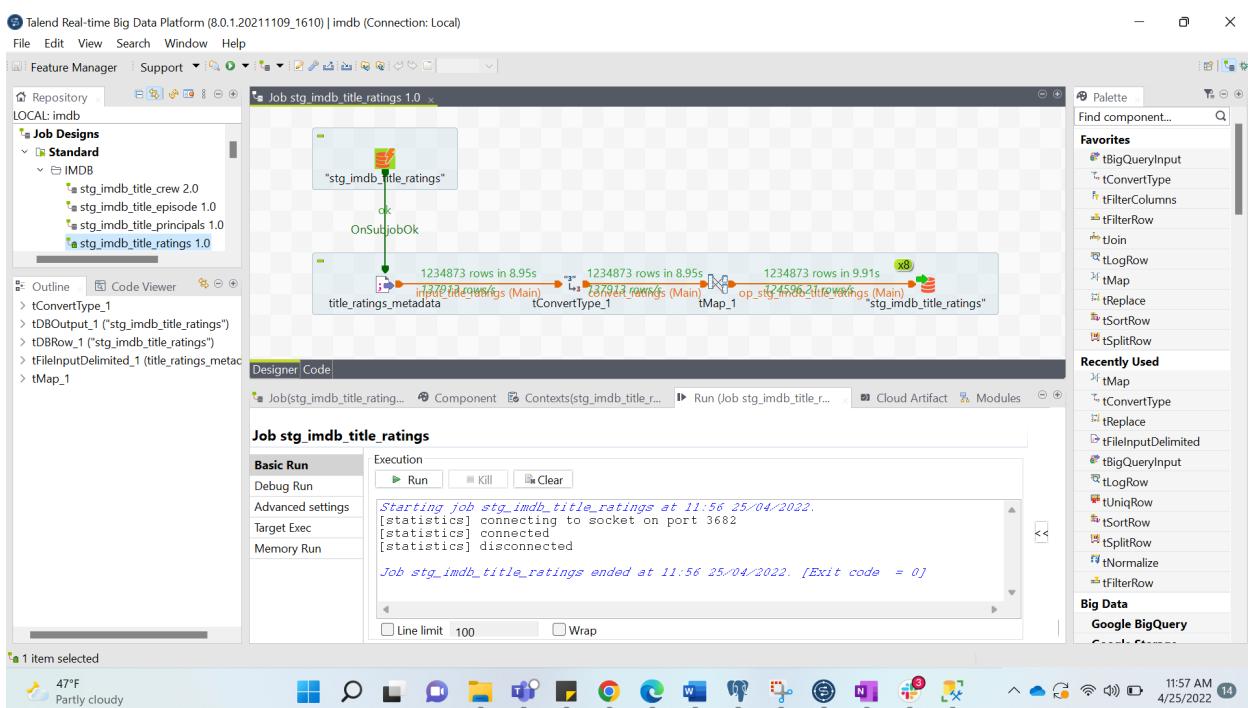
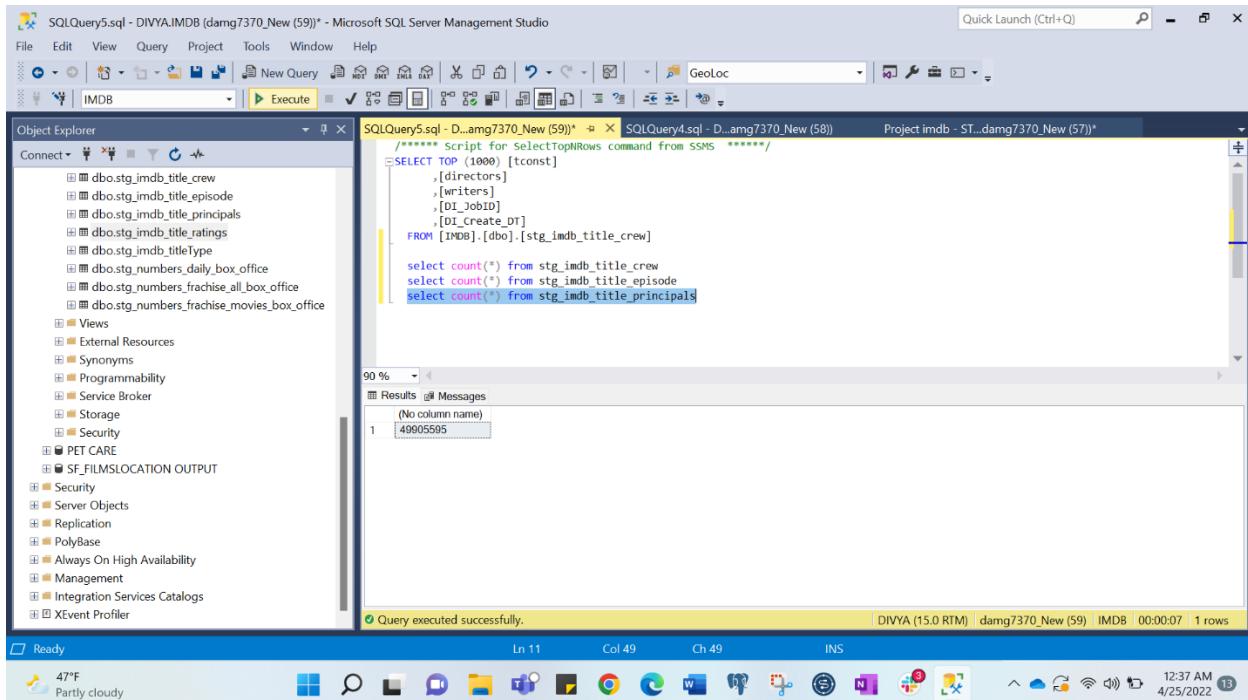
Table Name	Row_Count
stg_imdb_name_basics	11555784
stg_imdb_name_basics_knownfortitles	
stg_imdb_name_basics_primaryprofession	
stg_imdb_title_akas	31738388
stg_imdb_title_basics	8854184
stg_imdb_title_basics_genres	
stg_imdb_title_crew	8854165
stg_imdb_title_crew_directors	
stg_imdb_title_crew_writers	
stg_imdb_title_episode	664184
stg_imdb_title_principals	49905595
stg_imdb_title_ratings	1234873
stg_imdb_franchise_list	
stg_numbers_franchise_all_box_office	200
stg_numbers_franchise_movies_box_office	
stg_imdb_brand_list	
stg_imdb_titleType	39
stg_numbers_franchise_daily_box_office	
stg_imdb_brands_gross	
stg_imdb_franchises_gross	292
stg_box_office_worldwide_reject	
stg_box_office_worldwide	

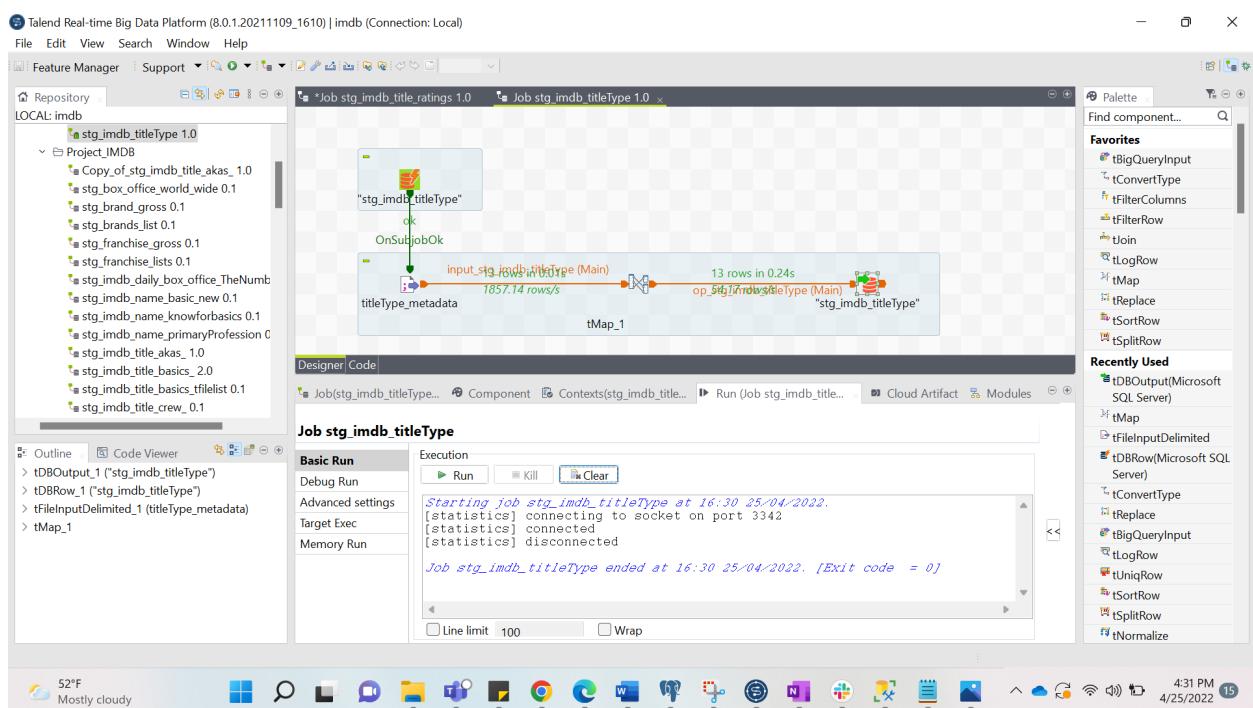
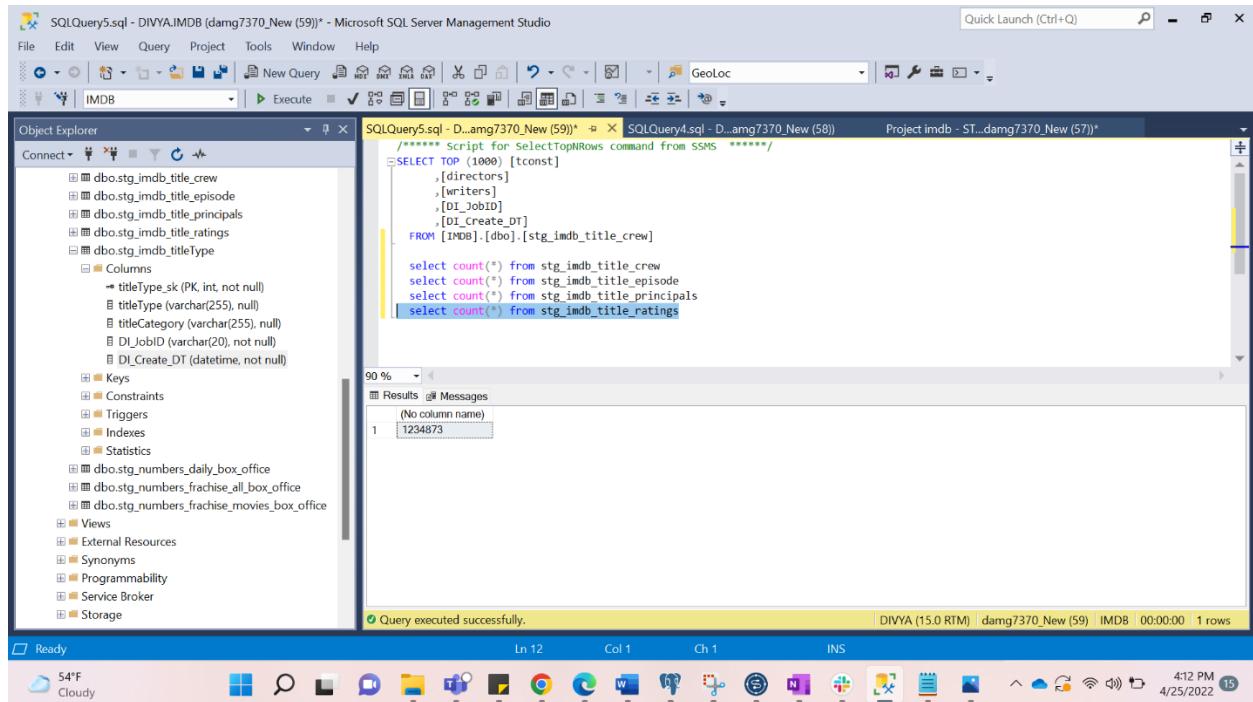
Final Row Counts of tables in Integration Schema:

Table Name	Row_Count
dim_imdb_primaryProfession	41
dim_imdb_name_basics_primaryProfession	
dim_imdb_title_principals	49880378
dim_imdb_job_category	12
dim_imdb_name_basics	11555784
dim_imdb_name_basics_knownfortitles	
dim_imdb_title_principals_rejects	25217
dim_imdb_brand	45
dim_imdb_title_crew_writers_rejects	58
dim_imdb_title_crew_writers	10412786
dim_imdb_name_basics_knownForTitles_rejects	
fct_imdb_brands_list	
fct_imdb_brands_gross	
dim_imdb_title_crew_directors	6637837
dim_imdb_title_crew_directors_rejects	48
dim_imdb_title_basics	8854184
fct_numbers_daily_box_office	
dim_imdb_titleType	11
fct_imdb_franchises_list	
fct_imdb_franchises_gross	
fct_movies_box_office_worldwide	
dim_imdb_franchise	
dim_imdb_title_basics_genres	13671953
dim_imdb_title_episode	6641716
dim_imdb_title_episodes_rejects	132
fct_numbers_franchise_all_box_office	
fct_numbers_franchise_movies_box_office	
dim_imdb_title_akas_rejects	5223
dim_imdb_genres	28
fct_imdb_title_ratings	
dim_imdb_title_akas	31733165









SQLQuery5.sql - DIVYA.IMDB (damg7370_New (59))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

```

SELECT TOP (1000) [tconst]
    ,[directors]
    ,[writers]
    ,[DI_JobID]
    ,[DI_Create_DT]
FROM [IMDB].[dbo].[stg_imdb_title_crew]

select count(*) from stg_imdb_title_crew
select count(*) from stg_imdb_title_episode
select count(*) from stg_imdb_title_principals
select count(*) from stg_imdb_title_ratings
select * from stg_imdb_title_ratings

```

Results Messages

tconst	averageRating	numVotes	DI_JobID	DI_Create_DT
tt0000001	5.7	1872	PNzgnT	2022-04-25 11:56:06.403
tt0000002	5.9	247	PNzgnT	2022-04-25 11:56:06.403
tt0000003	6.5	1646	PNzgnT	2022-04-25 11:56:06.403
tt0000004	5.8	160	PNzgnT	2022-04-25 11:56:06.403
tt0000005	6.2	2473	PNzgnT	2022-04-25 11:56:06.403
tt0000006	5.2	164	PNzgnT	2022-04-25 11:56:06.403
tt0000007	5.5	770	PNzgnT	2022-04-25 11:56:06.403
tt0000008	5.4	2014	PNzgnT	2022-04-25 11:56:06.403
tt0000009	5.4	193	PNzgnT	2022-04-25 11:56:06.403
tt0000010	6.9	6762	PNzgnT	2022-04-25 11:56:06.403
tt0000011	5.3	343	PNzgnT	2022-04-25 11:56:06.403

Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (59) | IMDB | 00:00:09 | 1,234,873 rows

Ready

52°F Mostly cloudy

Ln 13 Col 1 Ch 1 INS

4:31 PM 4/25/2022

SQLQuery5.sql - DIVYA.IMDB (damg7370_New (59))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

Execute (F5)

```

SELECT TOP (1000) [tconst]
    ,[directors]
    ,[writers]
    ,[DI_JobID]
    ,[DI_Create_DT]
FROM [IMDB].[dbo].[stg_imdb_title_crew]

select count(*) from stg_imdb_title_crew
select count(*) from stg_imdb_title_episode
select count(*) from stg_imdb_title_principals
select count(*) from stg_imdb_title_ratings
select * from stg_imdb_title_ratings
drop table stg_imdb_titleType

```

Results Messages

titleType	titleCategory	DI_JobID	DI_Create_DT
short	short	s1CXCL	2022-04-25 16:30:52.883
movie	movie	s1CXCL	2022-04-25 16:30:52.887
tvShort	tv	s1CXCL	2022-04-25 16:30:52.887
tvMovie	tv	s1CXCL	2022-04-25 16:30:52.887
tvSeries	tv	s1CXCL	2022-04-25 16:30:52.887
tvEpisode	tv	s1CXCL	2022-04-25 16:30:52.887
tvMiniseries	tv	s1CXCL	2022-04-25 16:30:52.887
tvSpecial	tv	s1CXCL	2022-04-25 16:30:52.887
video	video	s1CXCL	2022-04-25 16:30:52.887
videoGame	video	s1CXCL	2022-04-25 16:30:52.887
radioSeries	radio	s1CXCL	2022-04-25 16:30:52.887

Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (59) | IMDB | 00:00:00 | 13 rows

Ready

52°F Mostly cloudy

Ln 13 Col 1 Ch 1 INS

4:32 PM 4/25/2022

MySQL Workbench

Local instance MySQL80 × damg7370 ×

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

imdb_src_movie

Tables

Views

Stored Procedures

Functions

Administration Schemas

Information

Schema: imdb_src_movie

Query 1

imdb_src_movies*

1 • CREATE DATABASE `imdb_src_movie` /*!40100 DEFAULT CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci */ /*!

2 • select count(*) from imdb_titletype

Result Grid

Filter Rows: Export: Wrap Cell Content:

Result 13

Result Grid Form Editor Field Types

Result 8

Action Output

#	Time	Action
446	18:10:05	8871 row(s) affected Records: 8871 Duplicates: 0 Warnings: 0 Duration / Fetch: 0.188 sec
447	18:10:06	8839 row(s) affected Records: 8839 Duplicates: 0 Warnings: 0 Duration / Fetch: 0.234 sec

Object Info Session

50°F Mostly sunny

Read Only Context Help Snippets

6:21 PM 4/25/2022

This screenshot shows the MySQL Workbench interface. In the top-left, there are two tabs: 'Local instance MySQL80' and 'damg7370'. The main area is titled 'Query 1' and contains two lines of SQL code: 'CREATE DATABASE' and 'select count(*) from ...'. Below the code is a 'Result Grid' showing the result of the second query: '13'. To the right of the grid are three icons: 'Result Grid', 'Form Editor', and 'Field Types'. Below the grid is another section titled 'Result 8' with a table of execution logs. The bottom status bar shows the time as '6:21 PM' and the date as '4/25/2022'.

pgAdmin 4

File Object Tools Help

Browser

Servers (1)

PostgreSQL 14

Databases (2)

IMDB

Casts Catalogs Event Triggers Extensions Foreign Data Wrappers Languages Publications

Schemas (1)

public

Aggregates Collations Domains FTS Configurations FTS Dictionaries FTS Parsers FTS Templates Foreign Tables Functions Materialized Views Operators

Properties SQL Statistics Dependencies Dependents

General

Database: IMDB

OID: 16394

Owner: postgres

System database?: Off

Comment:

Definition

Encoding: UTF8

Tablespace: pg_default

Collation: English_United States.1252

Character type: English_United States 1252

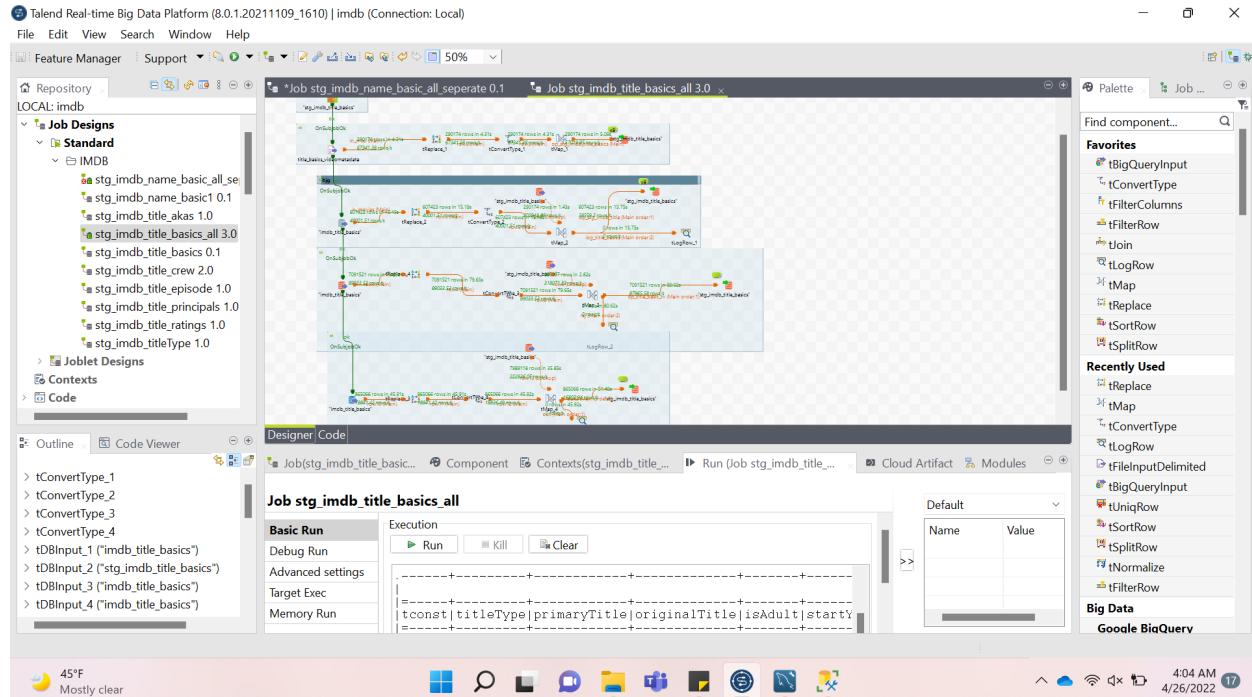
Restoring backup on the server

Restoring backup on the server 'PostgreSQL 14 (localhost:5432)'
Mon Apr 25 2022 20:47:51 GMT-0400 (Eastern Daylight Time)
72.05 seconds
More details... Stop Process
Successfully completed.

47°F Clear

8:49 PM 4/25/2022

This screenshot shows the pgAdmin 4 interface. On the left, a tree view shows 'Servers (1)', 'PostgreSQL 14', and 'Databases (2)' with 'IMDB' selected. The 'Properties' tab is active, showing the 'General' settings for the IMDB database, including the name, OID, owner, and system database status. On the right, a 'Definition' tab is open, showing encoding as 'UTF8', tablespace as 'pg_default', and collation as 'English_United States.1252'. A progress bar indicates a 'Restoring backup on the server' task has been completed successfully. The bottom status bar shows the time as '8:49 PM' and the date as '4/25/2022'.



SQLQuery1.sql - DIVYA.IMDB (damg7370_New (63)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

New Query Execute

Object Explorer Connect

IMDB

SQLQuery1.sql - D...amg7370_New (63)"

```

    FROM [IMDB].[dbo].[stg_imdb_title_basics]
    select count(*) from stg_imdb_title_basics
    select * from stg_imdb_title_basics
  
```

Results Messages

(No column name)
1 8854184

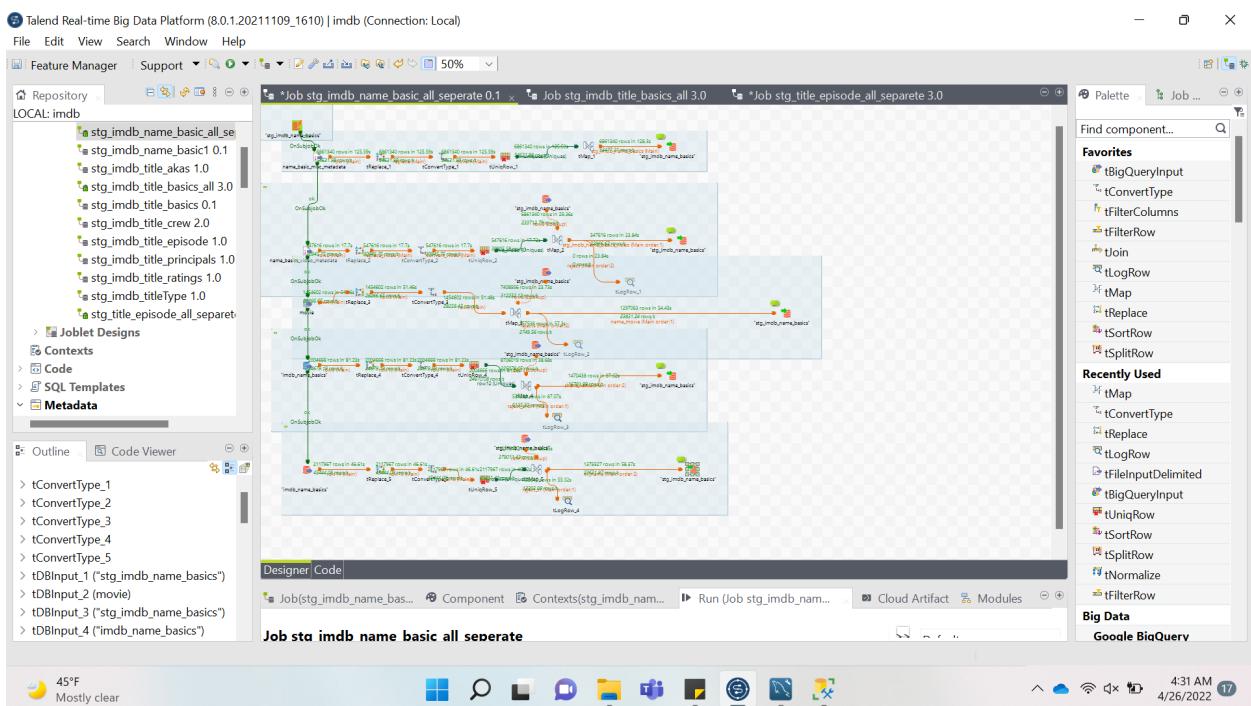
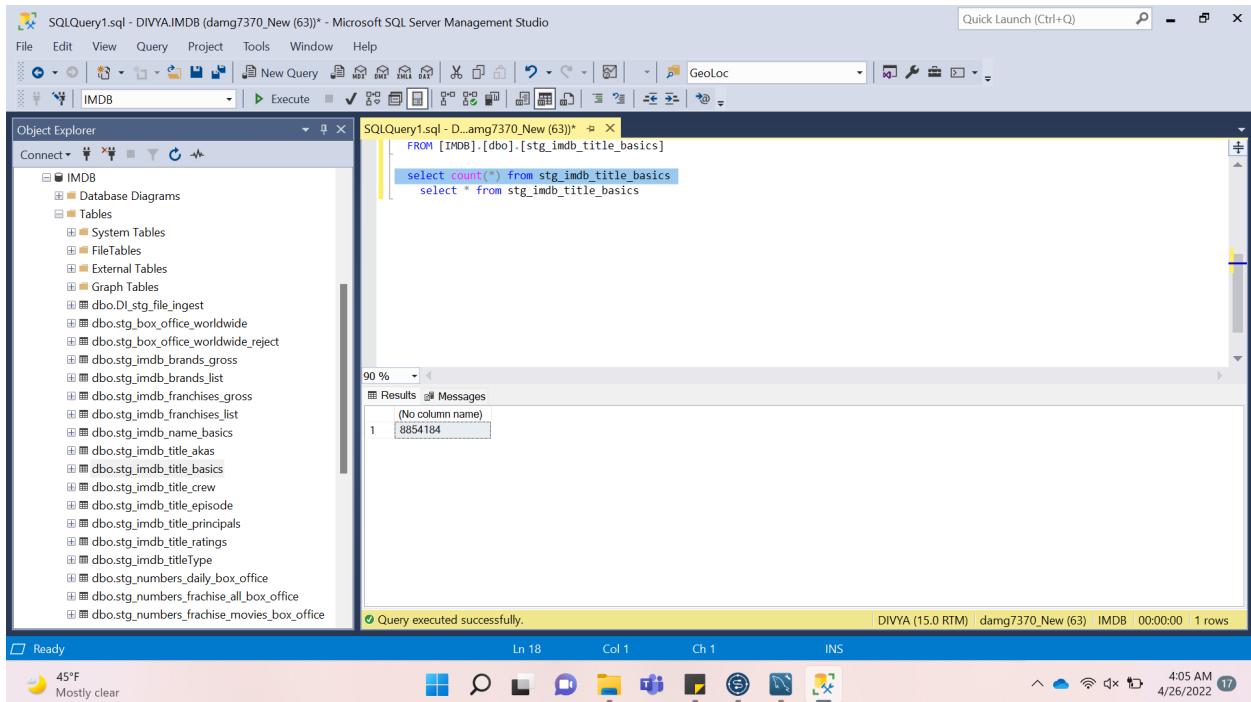
90 %

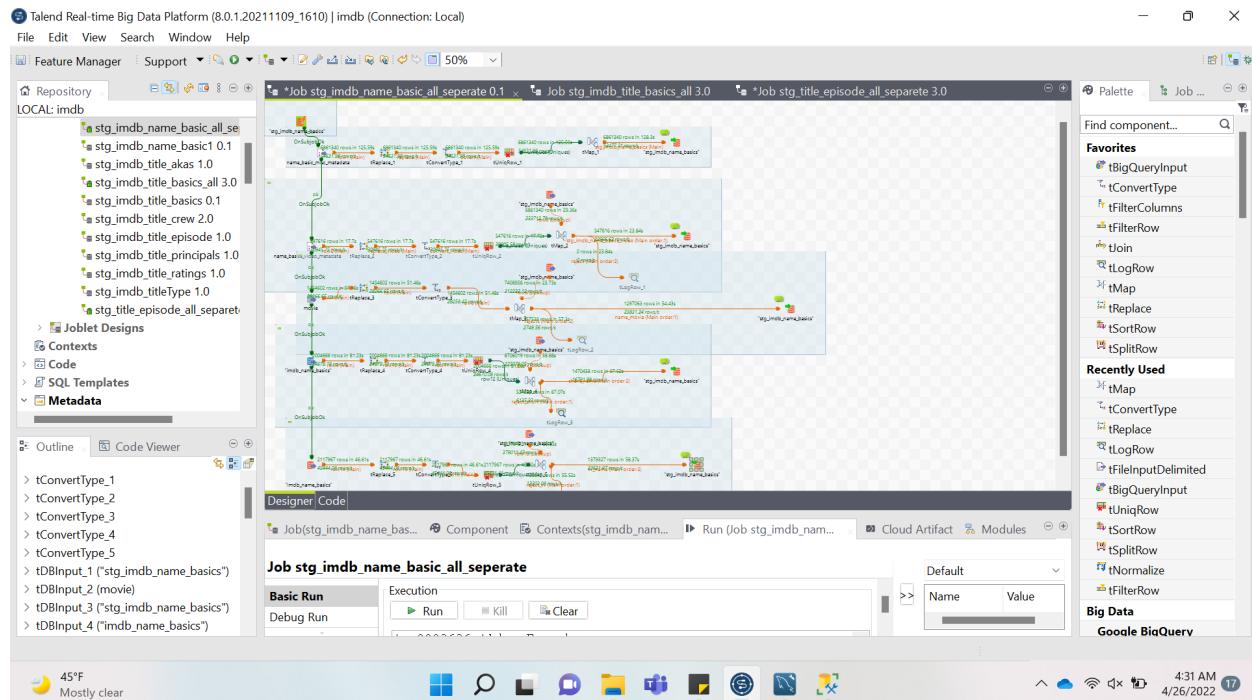
Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (63) | IMDB | 00:00:00 | 1 rows

Ready

45°F Mostly clear 4:04 AM 4/26/2022





SQLQuery1.sql - DIVYA.IMDB (damg7370_New (63)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

```
SQLQuery1.sql - D...amg7370_New (63)*
FROM [IMDB].[dbo].[stg_imdb_title_basics]
select count(*) from stg_imdb_name_basics
select * from stg_imdb_title_basics
```

Results Messages

(No column name)

1 11555784

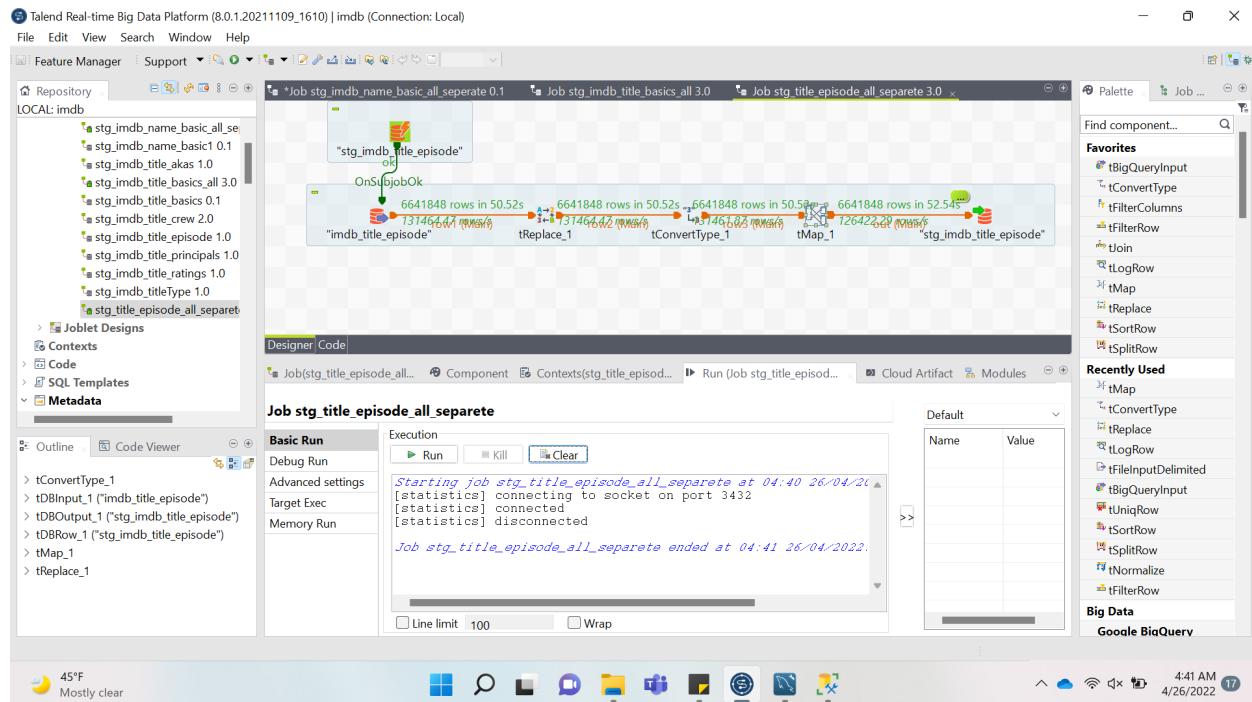
90 %

Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (63) | IMDB | 00:00:07 | 1 rows

Ready Ln 18 Col 1 Ch 1 INS

45°F Mostly clear 4:31 AM 4/26/2022



SQLQuery1.sql - DIVYA.IMDB (damg7370_New (63)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

```
SQLQuery1.sql - D...amg7370_New (63)*
FROM [IMDB].[dbo].[stg_imdb_title_basics]
select count(*) from stg_imdb_title_episode
select * from stg_imdb_title_basics
```

Results Messages

	(No column name)
1	6641848

90 %

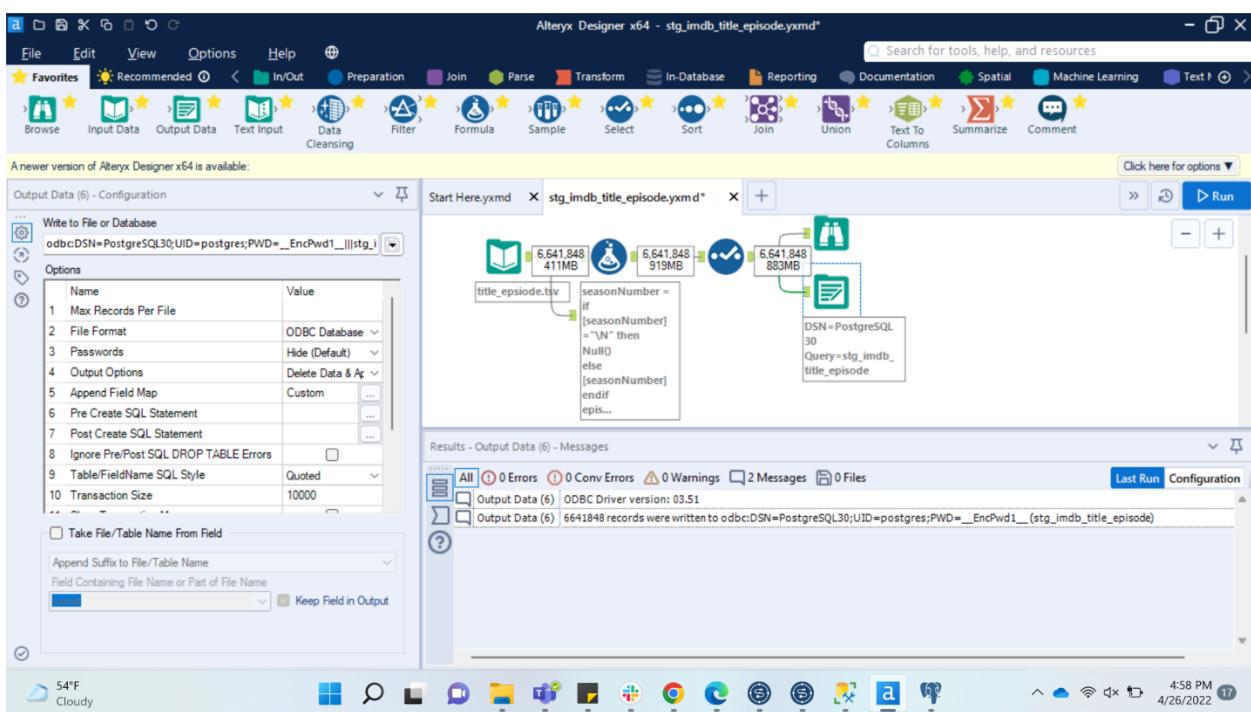
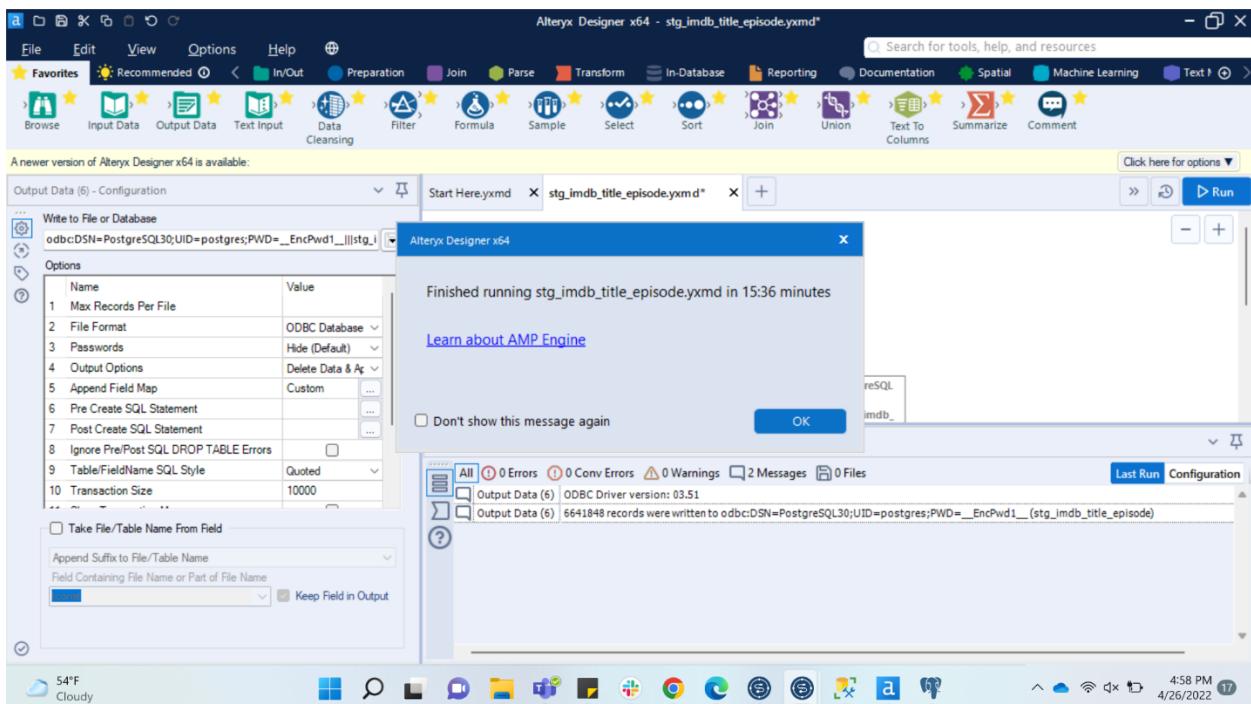
Query executed successfully.

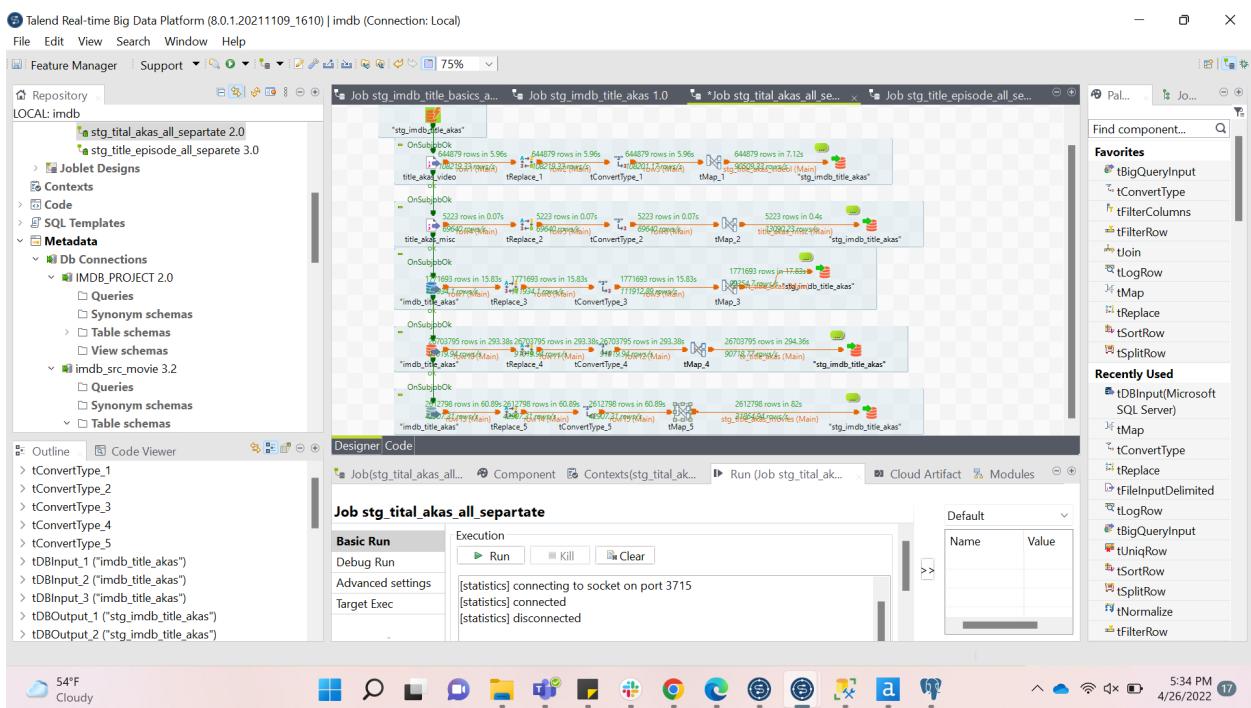
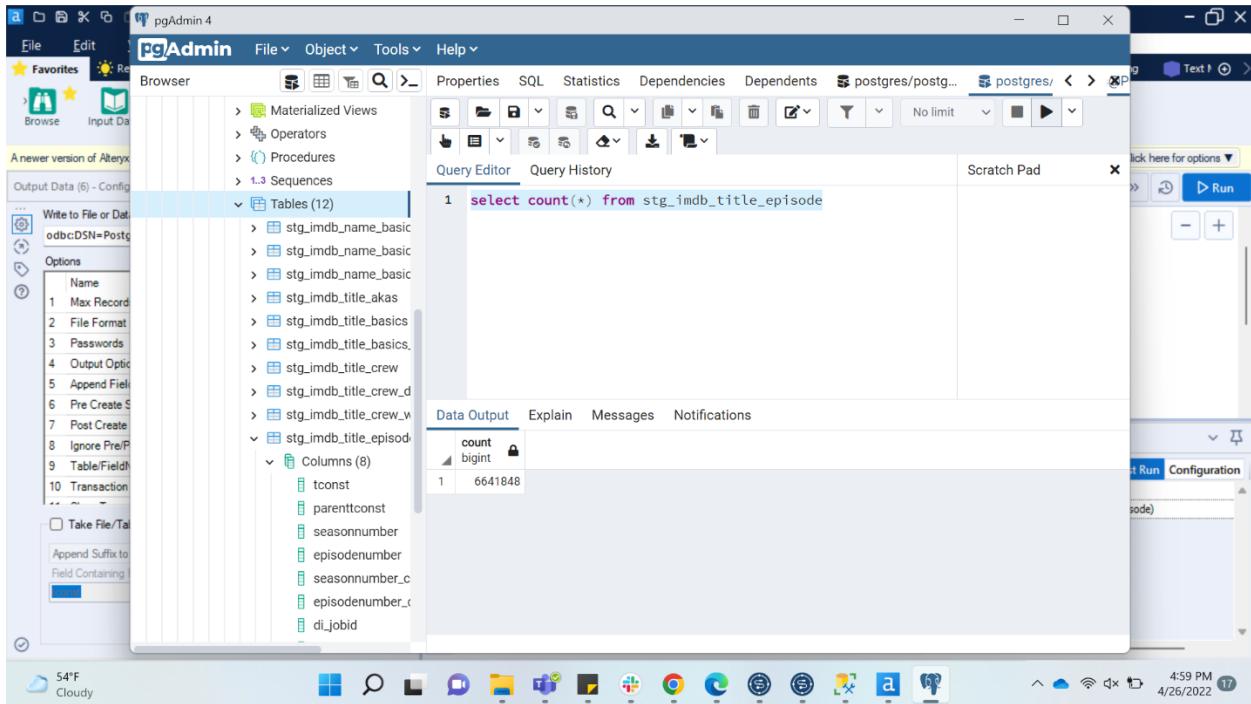
DIVYA (15.0 RTM) | damg7370_New (63) | IMDB | 00:00:00 | 1 rows

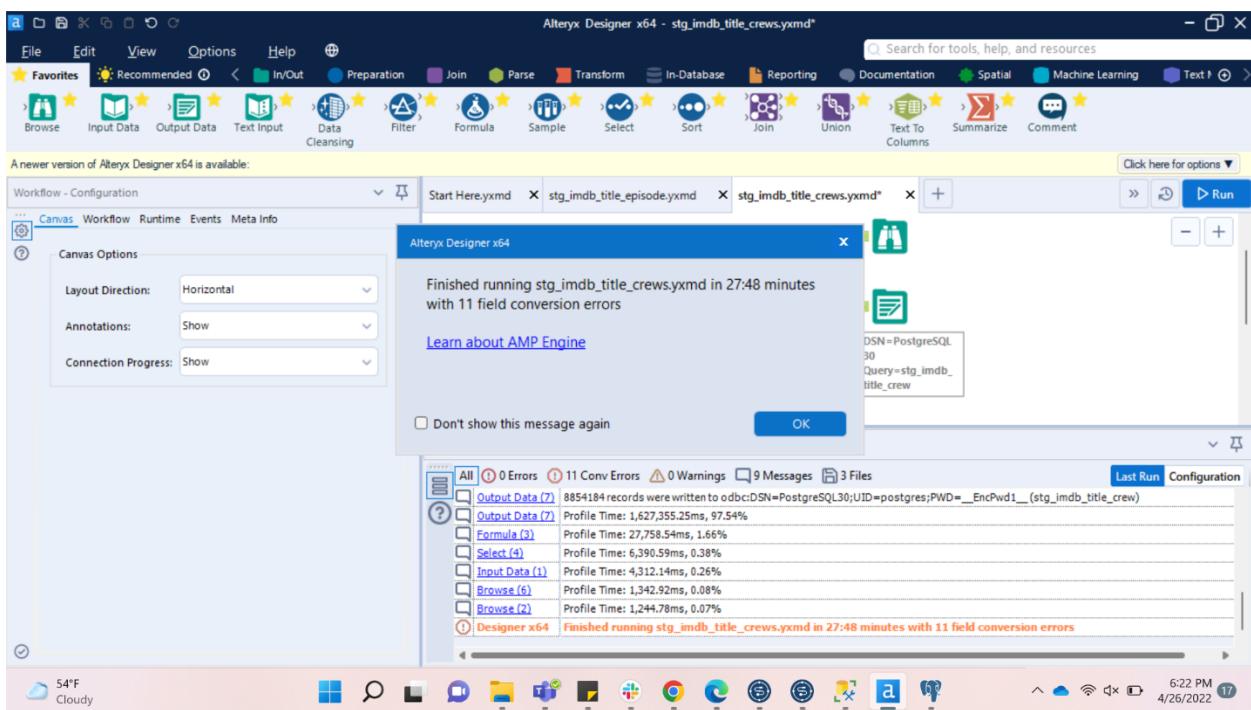
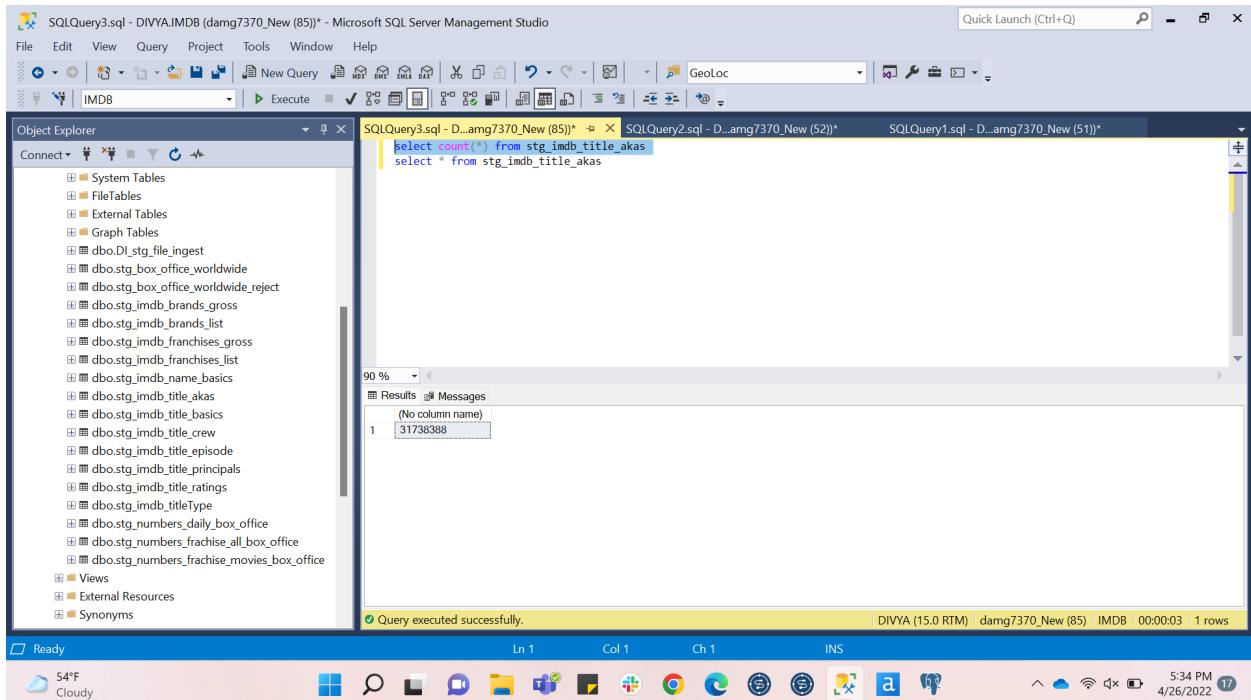
Ready

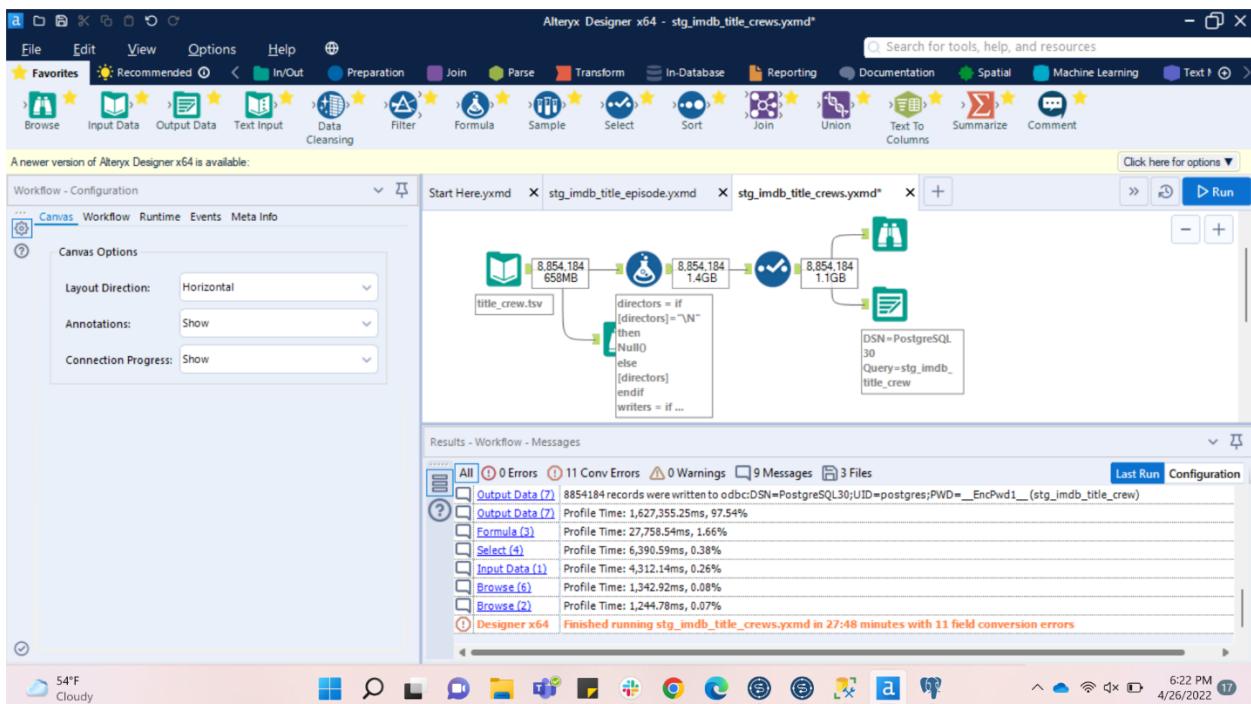
45°F Mostly clear

4:41 AM 4/26/2022









pgAdmin 4

File Edit Object Tools Help

Favorites Recommended

Workflow - Configuration

Canvas Workflow Runtime Events Meta Info

Canvas Options

Layout Direction: Horizontal
Annotations: Show
Connection Progress: Show

Browser

Properties SQL Statistics Dependencies Dependents postgres/postgres

Query Editor Query History

```
1 select count(*) from stg_imdb_title_crew
```

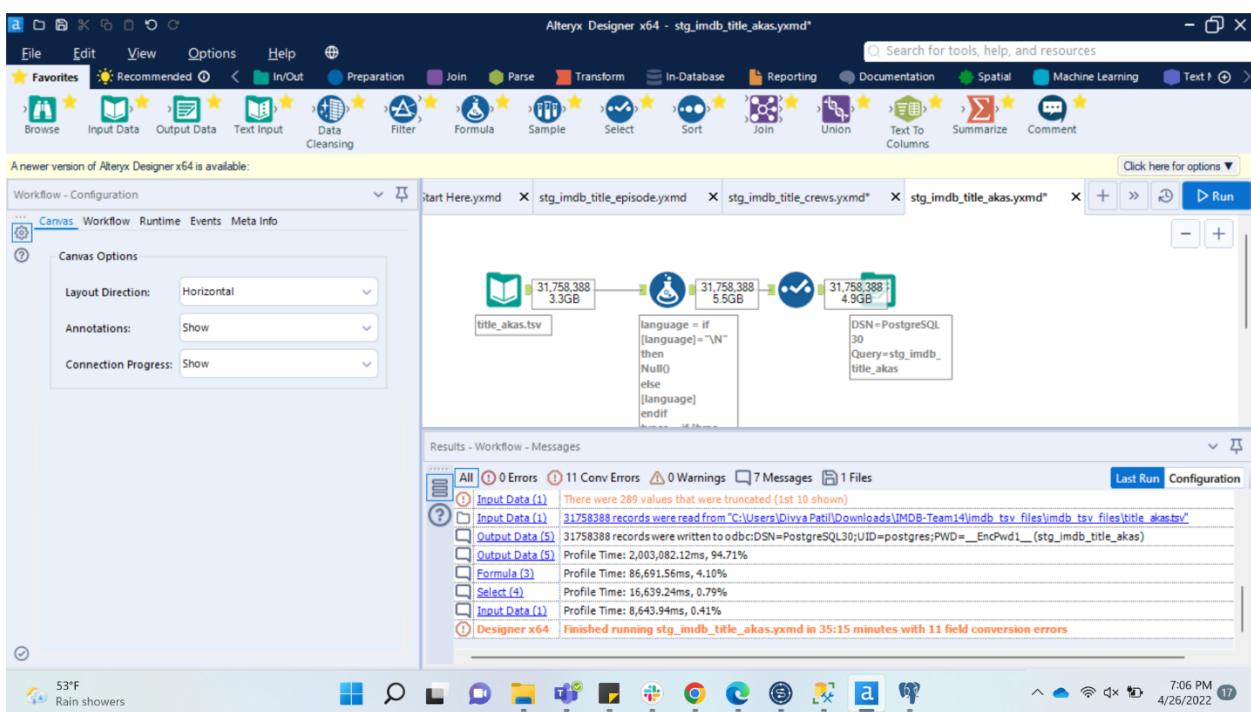
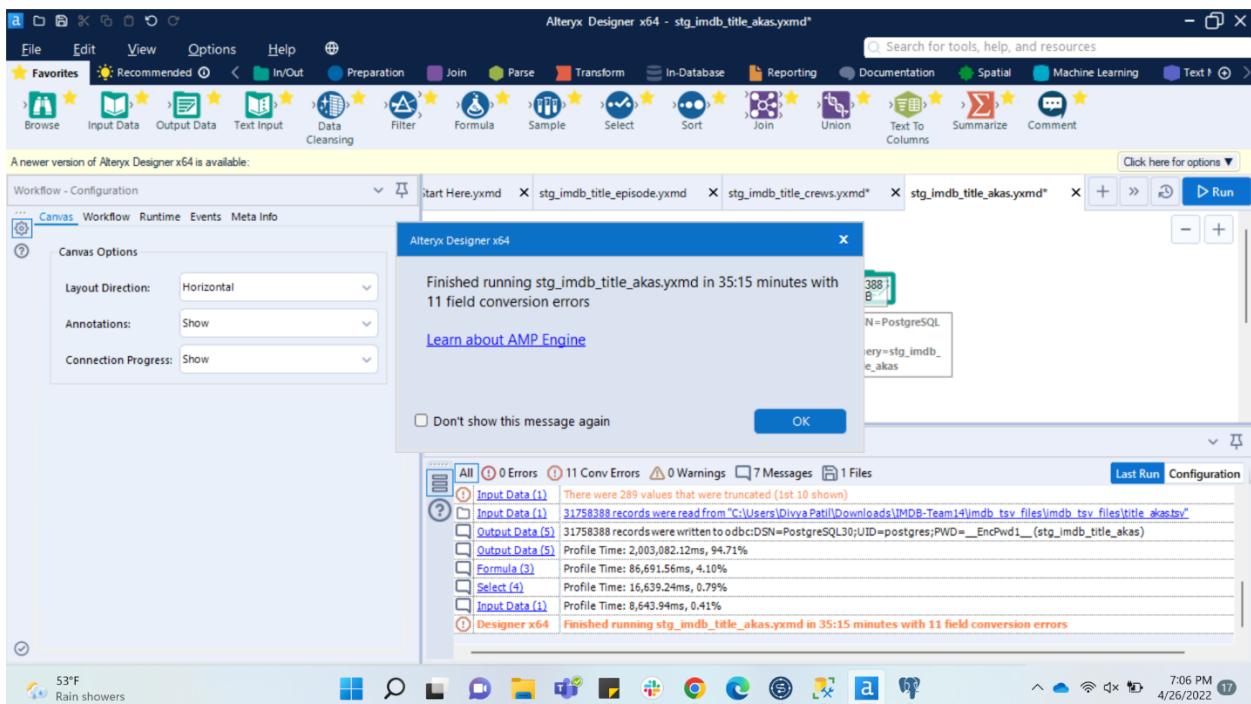
Scratch Pad

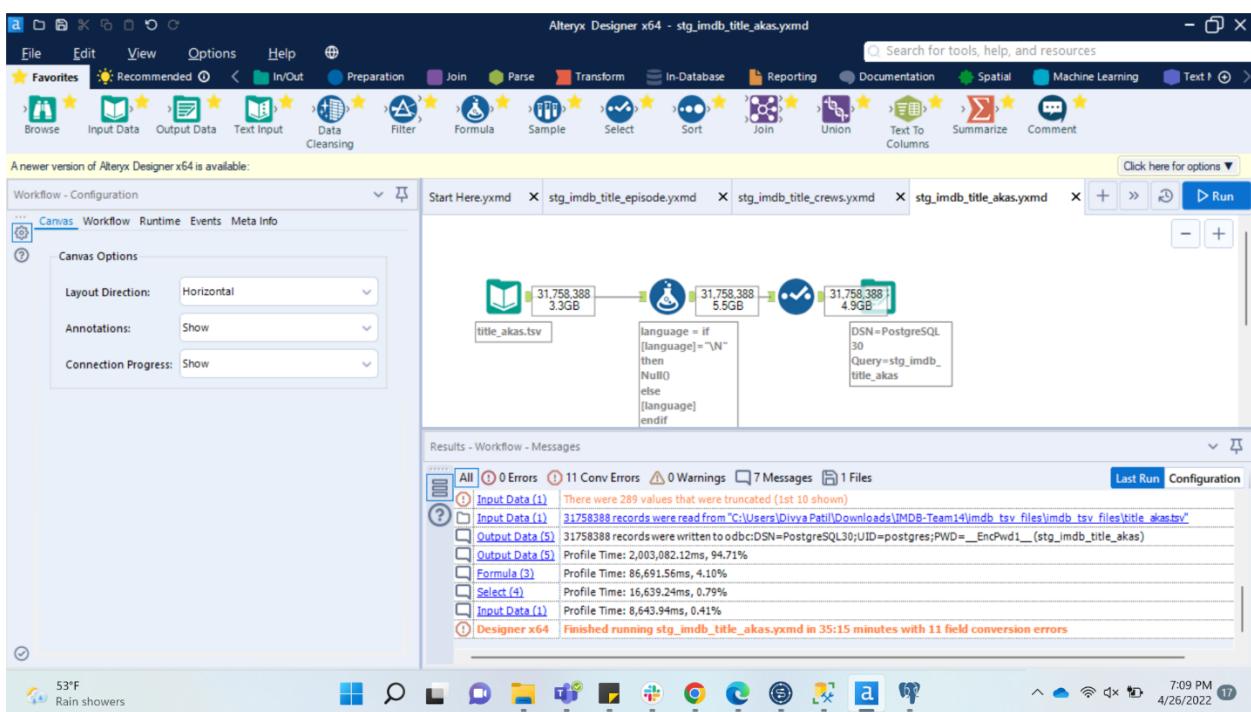
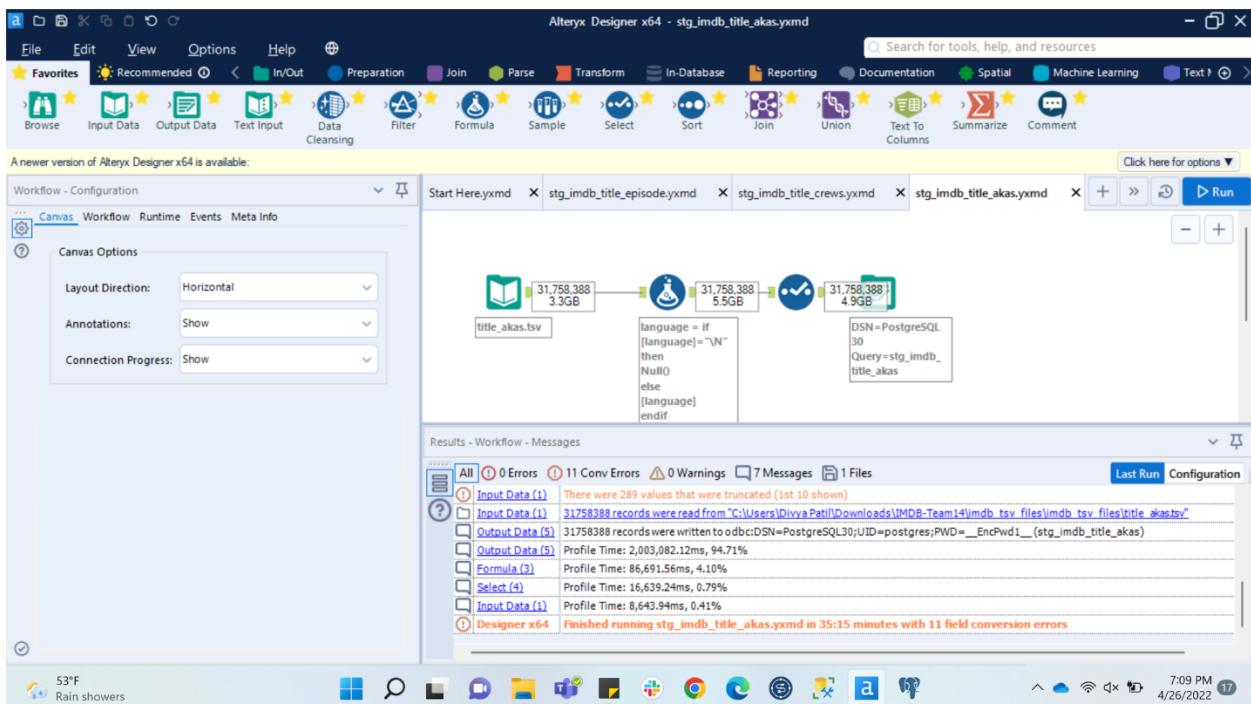
Data Output Explain Messages Notifications

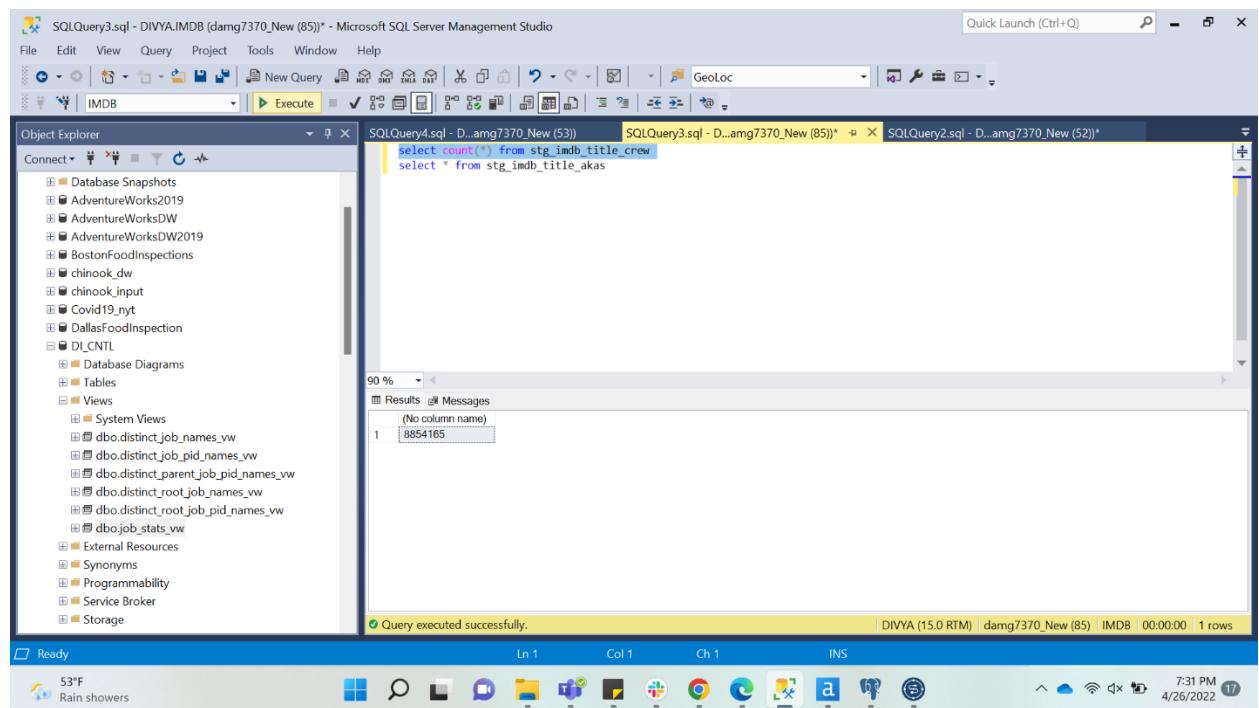
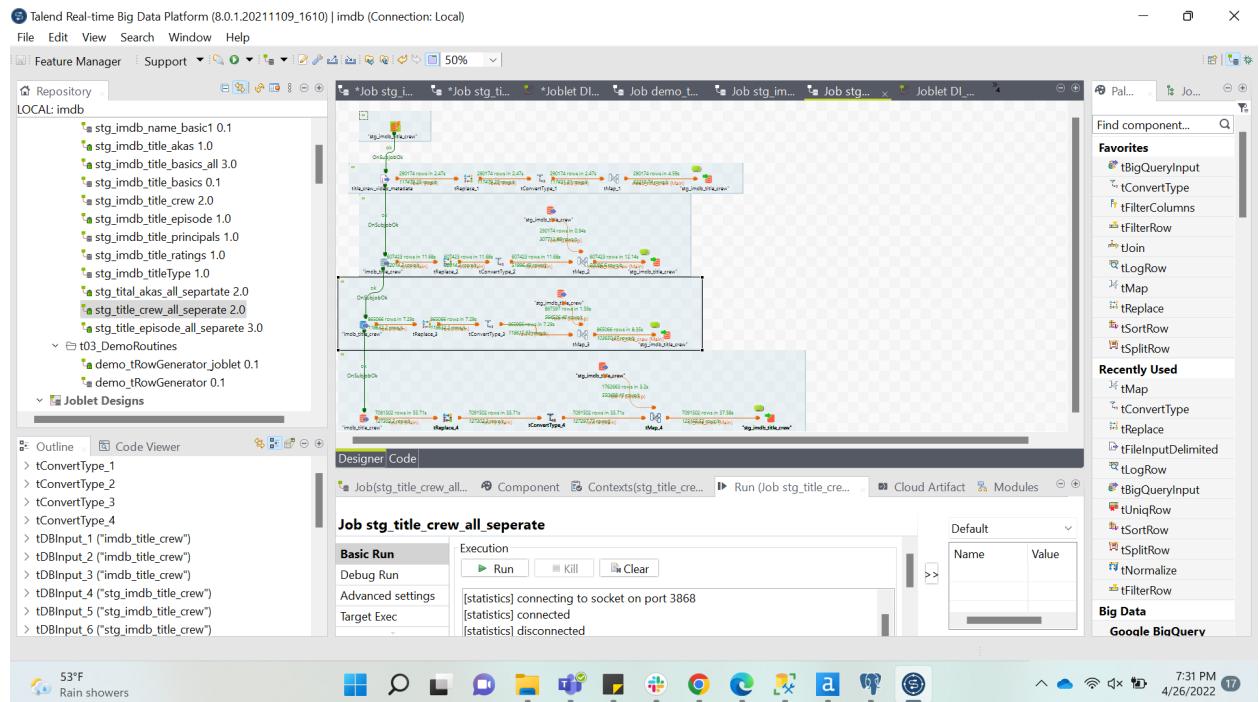
	count	bigint
1	8854184	

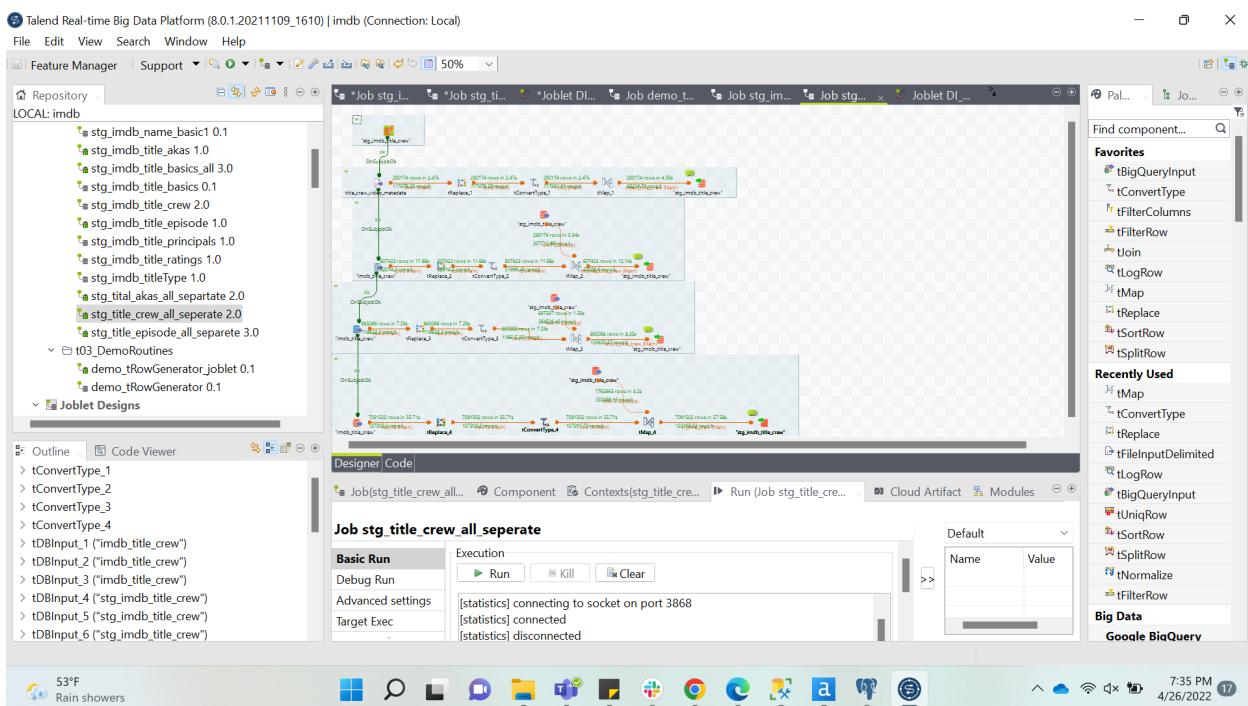
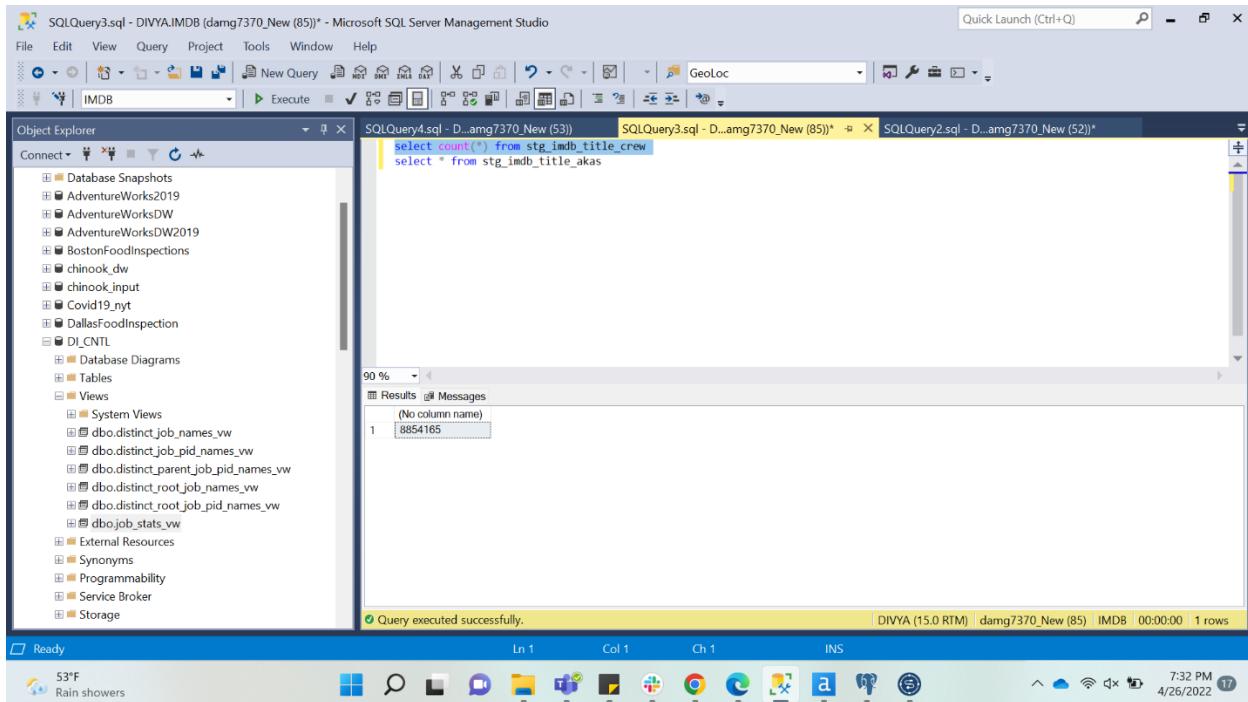
Successfully run. Total query runtime: 966 msec. 1 rows affected.

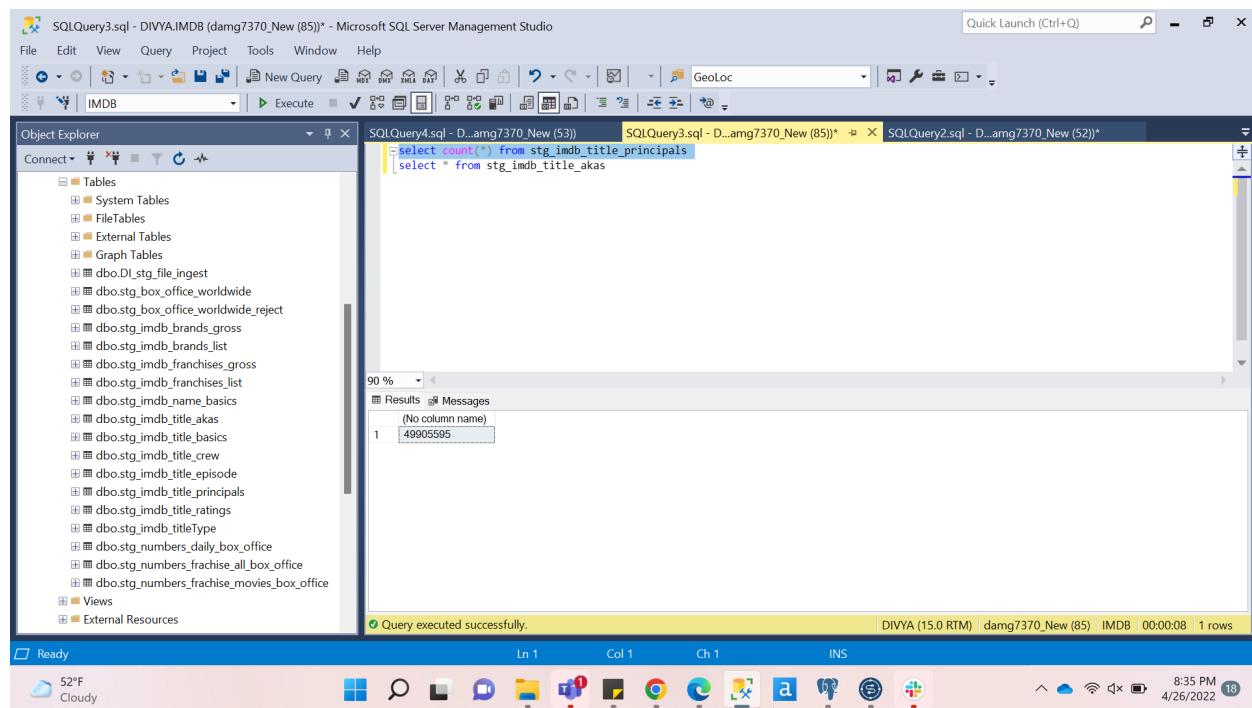
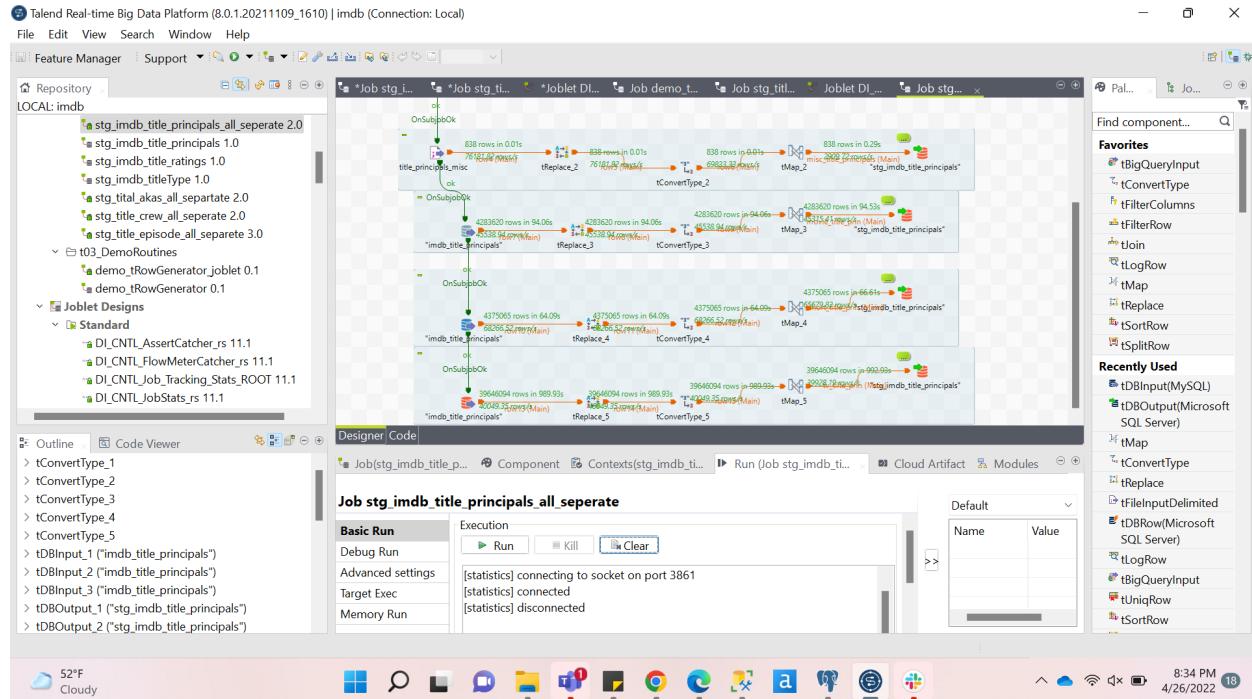
6:22 PM 4/26/2022











Talend Real-time Big Data Platform (8.0.1.20211109_1610) | imdb (Connection: Local)

File Edit View Search Window Help

Feature Manager Support

Repository LOCAL: imdb

Job stg_imdb_title_principals_all_separate 2.0

Job stg_imdb_title_principals 1.0

Job stg_imdb_title_ratings 1.0

Job stg_imdb_title_type 1.0

Job stg_tital_akas_all_separate 2.0

Job stg_title_crew_all_separate 2.0

Job stg_title_episode_all_separate 3.0

t03_DemoRoutes

demo_tRowGenerator_joblet 0.1

demo_tRowGenerator 0.1

Joblet Designs Standard

- DI_CNTL_AssertCatcher_rs 11.1
- DI_CNTL_FlowMeterCatcher_rs 11.1
- DI_CNTL_Job_Tracking_Stats_ROOT 11.1
- DI_CNTL_JobStats_rs 11.1

Outline Code Viewer

Job stg_imdb_title_principals.all_separate

Basic Run

Execution

[statistics] connecting to socket on port 3861
[statistics] connected
[statistics] disconnected

Default Name Value

52°F Cloudy

8:36 PM 4/26/2022

This screenshot shows the Talend Designer interface with a complex ETL job named "Job stg_imdb_title_principals.all_separate". The job consists of five parallel sub-jobs, each processing different parts of the "imdb_title_principals" dataset. Each sub-job involves reading from a file, performing various transformations (using tReplace, tConvertType, and tMap components), and writing to a new file. The overall process is designed to handle large amounts of data efficiently.

Alteryx Designer x64 - stg_imdb_title_principals.ymd

File Edit View Options Help

Favorites Recommended

Input Data Output Data Text Input

A newer version of Alteryx Designer x64 is available:

Workflow - Configuration

Canvas Workflow Runtime Events Meta Info

Alteryx Designer x64

Finished running stg_imdb_title_principals.ymd in 2:10 hours with 10 field conversion errors

Learn about AMP Engine

OK

Don't show this message again

title_principals.csv

job = if [job] = "N" then Null() else [job]

DSN=PostgreSQL.35W Query=stg_imdb_title_principals

Results - Workflow - Messages

All 0 Errors 10 Conv Errors 0 Warnings 7 Messages 1 Files

Input Data (1) The field "job" was truncated in record #43596737

Input Data (1) The field "characters" was truncated in record #44906707

Input Data (1) 49905595 records were read from "C:\Users\Divya Pathil\Downloads\IMDB-Team1\imdb_csv_files\imdb_csv_files\title_principals.csv"

Output Data (6) 49905595 records were written to odbc:DSN=PostgreSQL.35W;UID=postgres;PWD=_EncPwd1_(stg_imdb_title_principals)

Profile Time: 7,672,524.80ms, 97.84%

Profile Time: 128,198.73ms, 1.63%

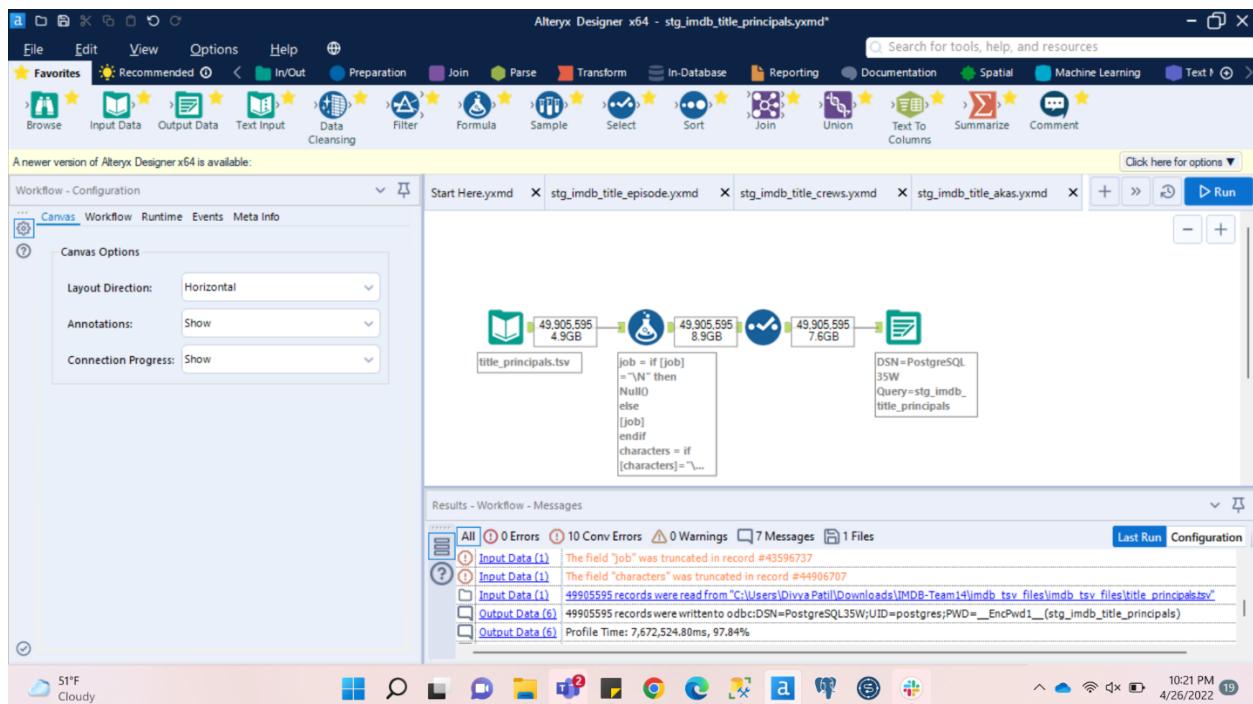
Profile Time: 26,496.13ms, 0.34%

Profile Time: 14,429.93ms, 0.18%

10:21 PM 4/26/2022

51°F Cloudy

This screenshot shows the Alteryx Designer interface with a workflow named "stg_imdb_title_principals.ymd". The workflow starts with an "Input Data" component reading a CSV file named "title_principals.csv". It then uses a script component to handle the "job" field. Finally, it writes the data to a PostgreSQL database using an "Output Data" component. The results show that the process completed in 2:10 hours with 10 field conversion errors.



pgAdmin 4

File Edit Object Tools Help

Favorites Recommended

Servers PostgreSQL 14 Databases (2) IMDB Casts Catalogs Event Triggers Extensions Foreign Data Wrappers Languages Publications Schemas (1) public Aggregates Collations Domains FTS Configurations FTS Dictionaries FTS Parsers FTS Templates Foreign Tables Functions Materialized Views

Properties SQL Statistics Dependencies Dependents postgres/postg

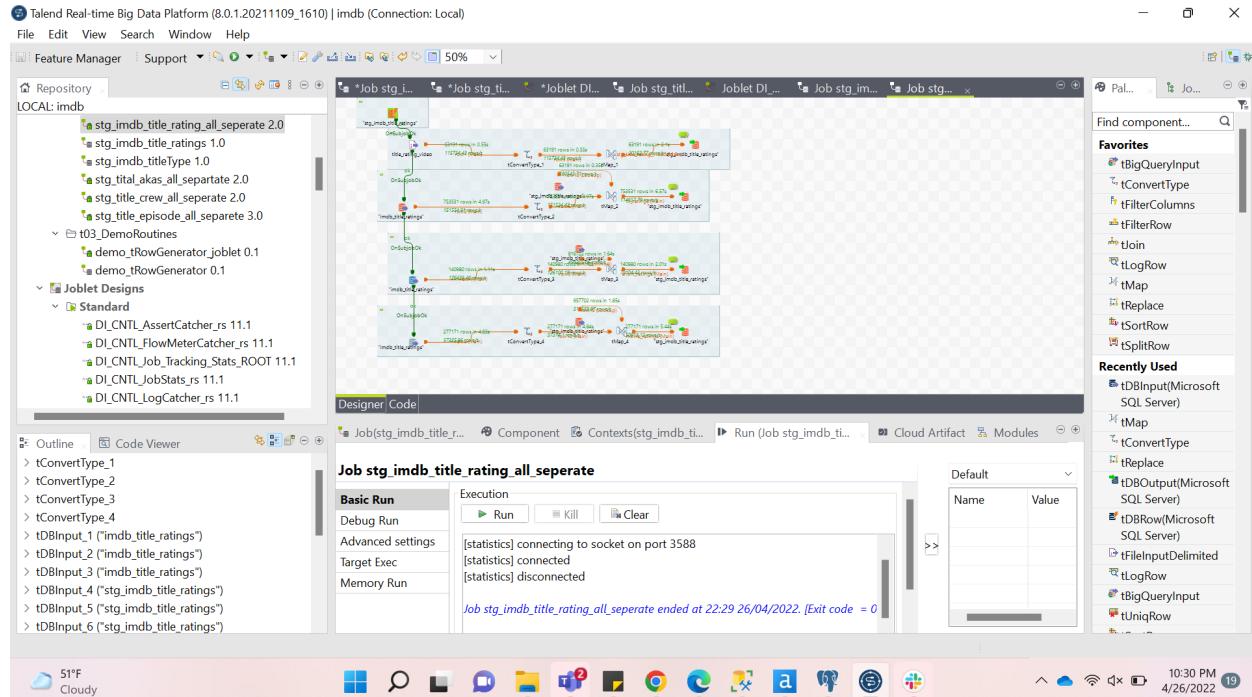
Query Editor Query History

```
1 select count(*) from stg_imdb_title_principals
```

Data Output Explain Messages Notifications

	count	bigint
1	49905595	

10:22 PM 4/26/2022



SQLQuery3.sql - DIVYA.IMDB (damg7370_New (85)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

Tables

- System Tables
- FileTables
- External Tables
- Graph Tables
- dbo.DL_stg_file_ingest
- dbo.stg_box_office_worldwide
- dbo.stg_box_office_worldwide_reject
- dbo.stg_imdb_brands_gross
- dbo.stg_imdb_brands_list
- dbo.stg_imdb_franchises_gross
- dbo.stg_imdb_franchises_list
- dbo.stg_imdb_name_bases
- dbo.stg_imdb_title_akas
- dbo.stg_imdb_title_basics
- dbo.stg_imdb_title_crew
- dbo.stg_imdb_title_episode
- dbo.stg_imdb_title_principals
- dbo.stg_imdb_title_ratings
- dbo.stg_imdb_titleType
- dbo.stg_numbers_daily_box_office
- dbo.stg_numbers_franchise_all_box_office
- dbo.stg_numbers_franchise_movies_box_office

Views

External Resources

SQLQuery4.sql - D...amg7370_New (53)) SQLQuery3.sql - D...amg7370_New (85)) SQLQuery2.sql - D...amg7370_New (52))

```
=select count(*) from stg_imdb_title_ratings
| select * from stg_imdb_title_akas
```

Results Messages

(No column name)

1	1234873
---	---------

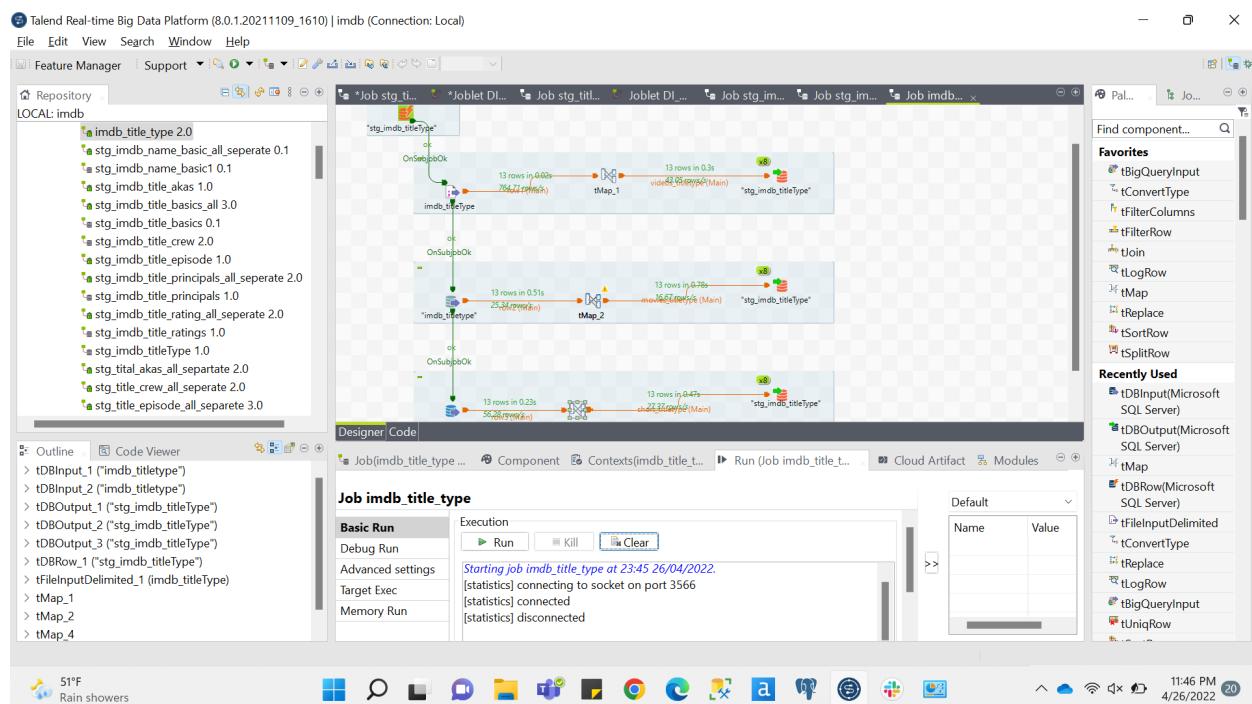
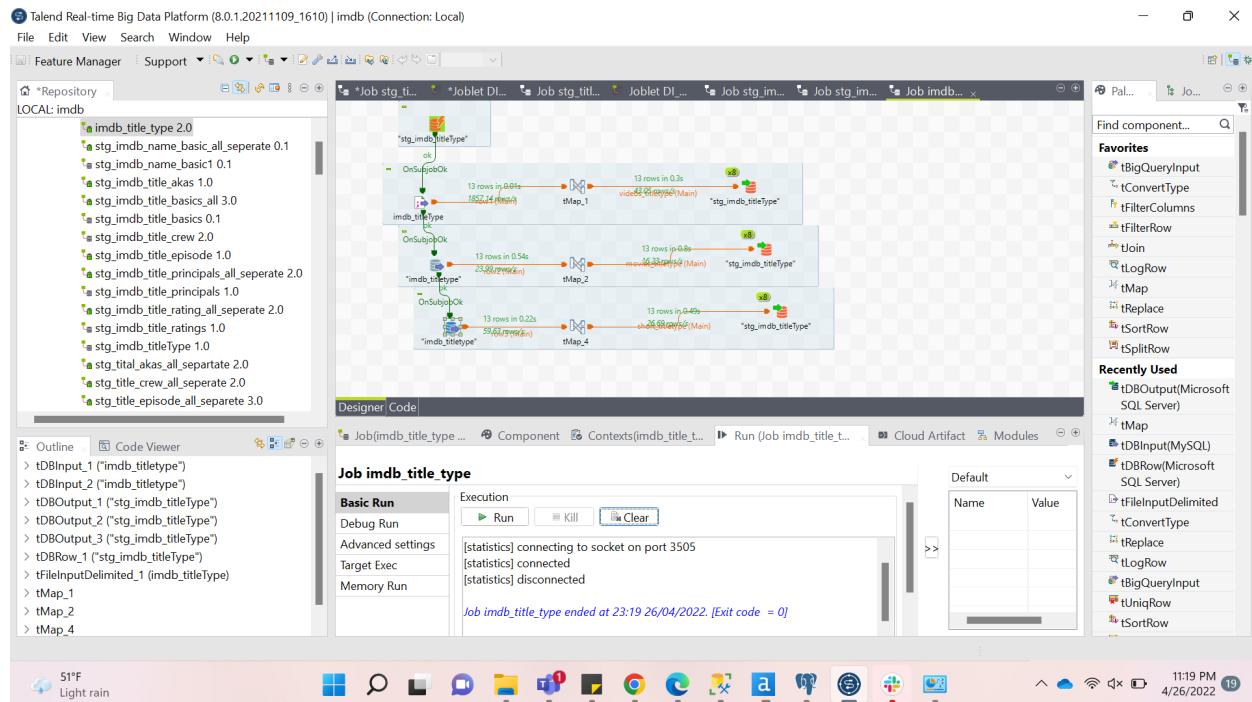
90 %

Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (85) | IMDB | 00:00:00 | 1 rows

Ready

10:33 PM 4/26/2022



SQLQuery3.sql - DIVYA.IMDB (damg7370_New (85)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB Execute GeoLoc

Object Explorer

SQLQuery4.sql - D...amg7370_New (53) SQLQuery3.sql - D...amg7370_New (85)* SQLQuery2.sql - D...amg7370_New (52)*

```
select count(*) from stg_imdb_title_ratings
SELECT * FROM stg_imdb_titleType
```

CREATE TABLE stg_imdb_titleType (
 titleType_sk int NOT NULL IDENTITY(1,1),
 titleType varchar(255) NULL,
 titleCategory varchar(255) NULL,
 DI_JobID varchar(20) NOT NULL,
 DI_Create_DT datetime DEFAULT (getdate()) NOT NULL,
 PRIMARY KEY (titleType_sk)
);

Results Messages

	titleType_sk	titleType	titleCategory	DI_JobID	DI_Create_DT
1	1	short	short	QoAB9j	2022-04-26 23:45:16.980
2	2	movie	movie	QoAB9j	2022-04-26 23:45:16.980
3	3	tvShort	tv	QoAB9j	2022-04-26 23:45:16.980
4	4	tvMovie	tv	QoAB9j	2022-04-26 23:45:16.980
5	5	tvSeries	tv	QoAB9j	2022-04-26 23:45:16.980
6	6	tvEpisode	tv	QoAB9j	2022-04-26 23:45:16.980
7	7	tvMiniSeries	tv	QoAB9j	2022-04-26 23:45:16.980
8	8	tvSpecial	tv	QoAB9j	2022-04-26 23:45:16.980
9	9	video	video	QoAB9j	2022-04-26 23:45:16.980
10	10	videoGame	video	QoAB9j	2022-04-26 23:45:16.980
11	11	radioSeries	radio	QoAB9j	2022-04-26 23:45:16.980

Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (85) | IMDB | 00:00:00 | 39 rows

SQLQuery3.sql - DIVYA.IMDB (damg7370_New (85)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB Execute GeoLoc

Object Explorer

SQLQuery4.sql - D...amg7370_New (53) SQLQuery3.sql - D...amg7370_New (85)* SQLQuery2.sql - D...amg7370_New (52)*

```
select count(*) from stg_imdb_titleType
SELECT * FROM stg_imdb_titleType
```

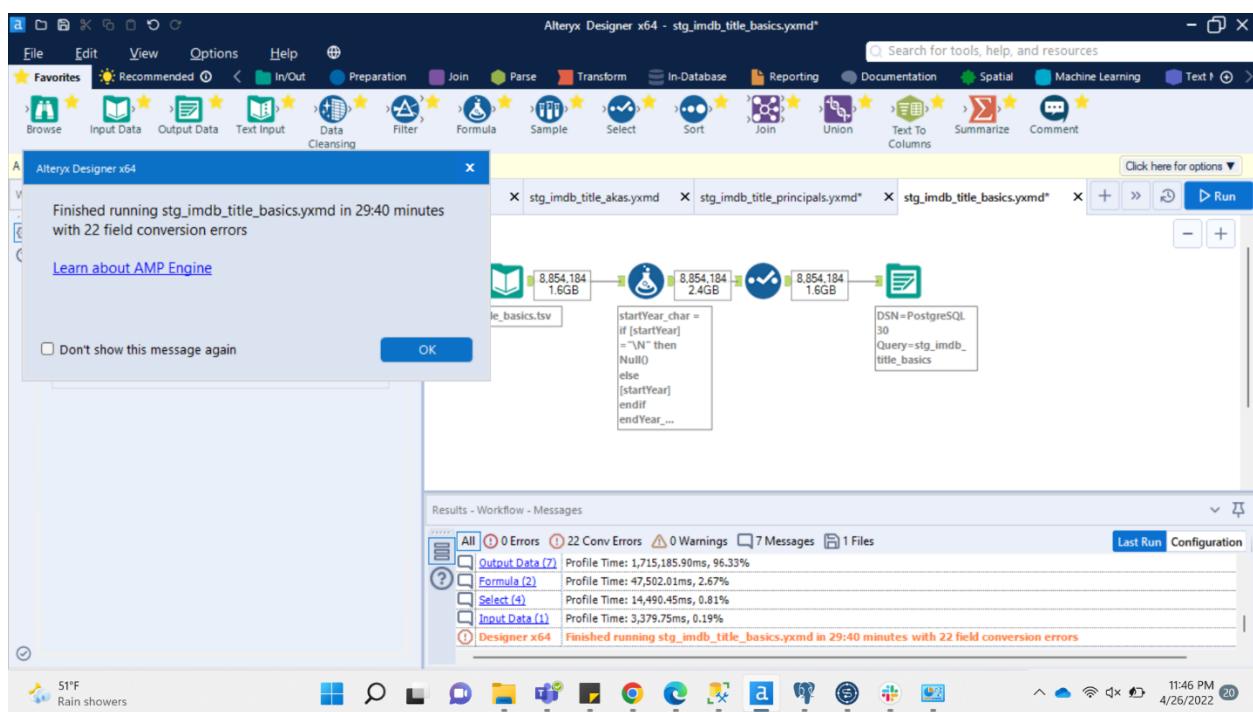
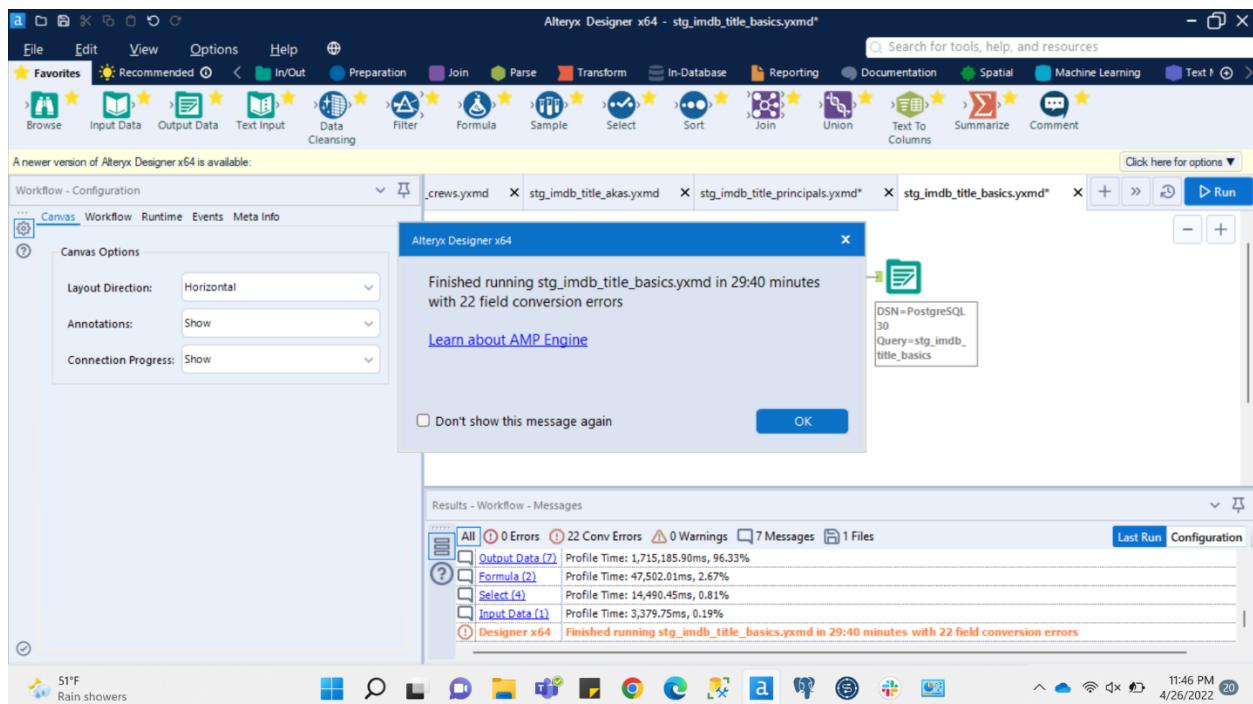
CREATE TABLE stg_imdb_titleType (
 titleType_sk int NOT NULL IDENTITY(1,1),
 titleType varchar(255) NULL,
 titleCategory varchar(255) NULL,
 DI_JobID varchar(20) NOT NULL,
 DI_Create_DT datetime DEFAULT (getdate()) NOT NULL,
 PRIMARY KEY (titleType_sk)
);

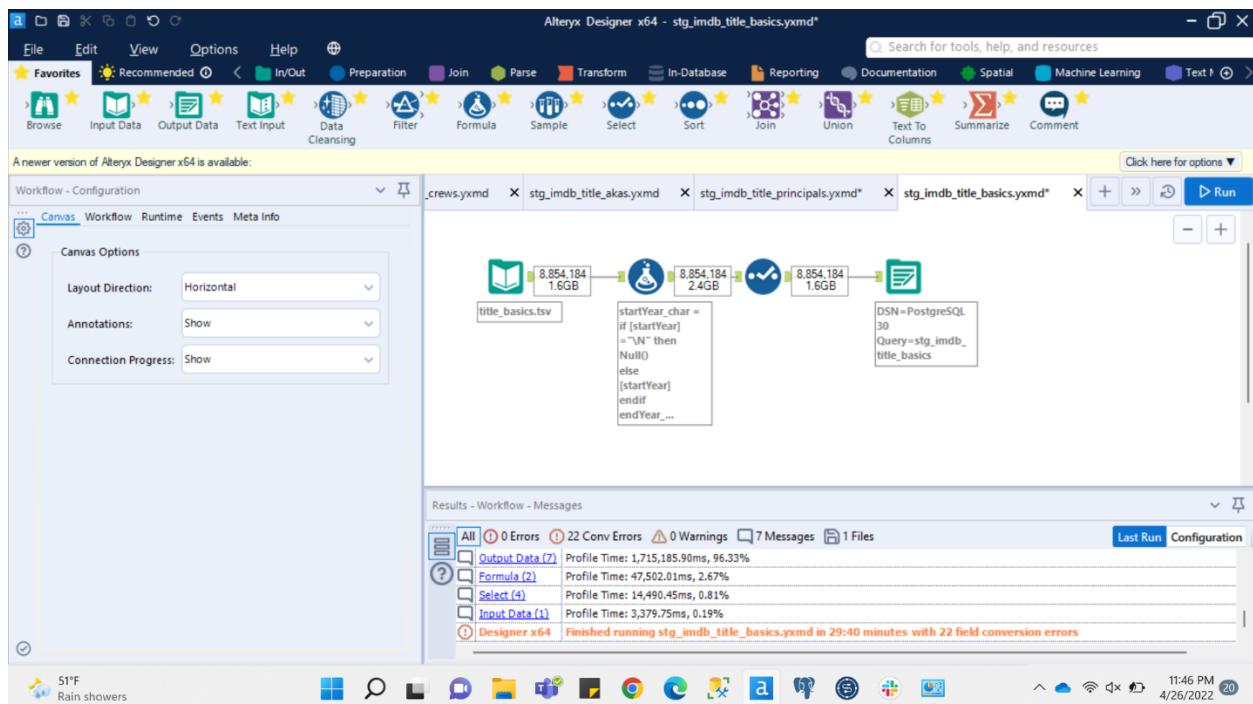
Results Messages

	(No column name)
1	39

Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (85) | IMDB | 00:00:00 | 1 rows





pgAdmin 4

File Object Tools Help

Servers (1)

PostgreSQL 14

Databases (2)

- IMDB
- postgres

Casts Catalogs Event Triggers Extensions Foreign Data Wrappers Languages Publications Schemas (1)

public

- Aggregates Collations Domains FTS Configurations FTS Dictionaries FTS Parsers FTS Templates Foreign Tables Functions Materialized Views Operators

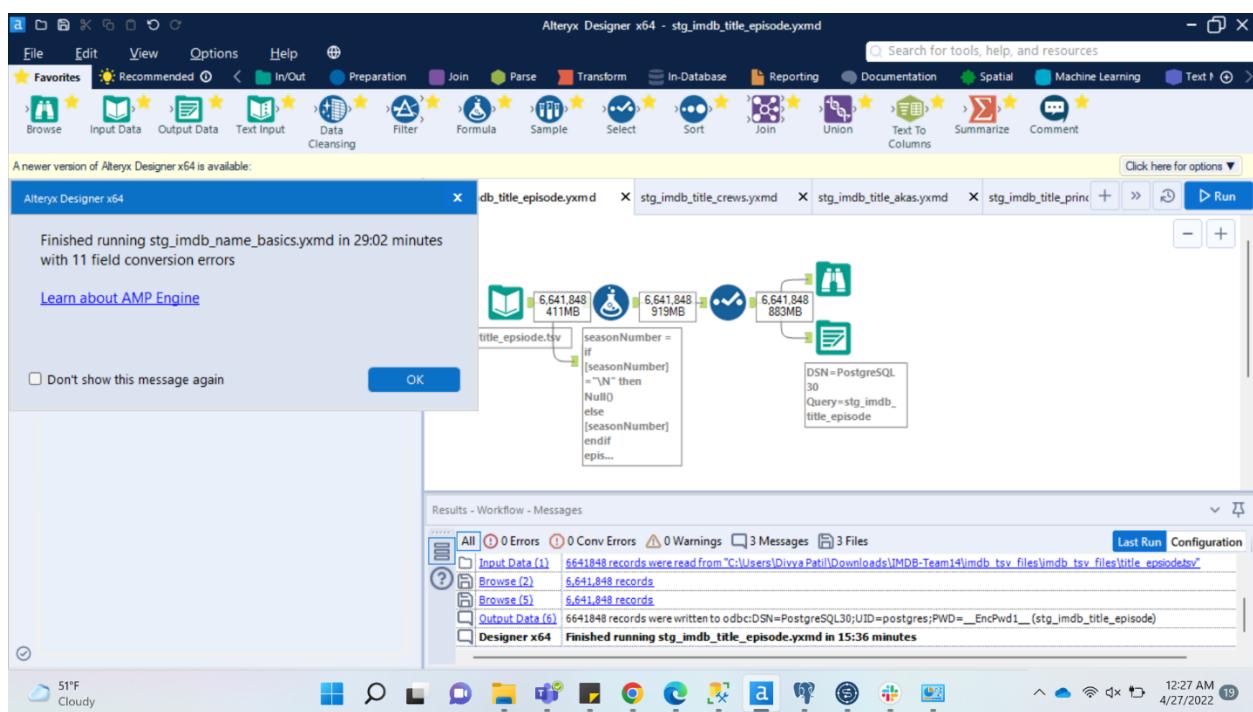
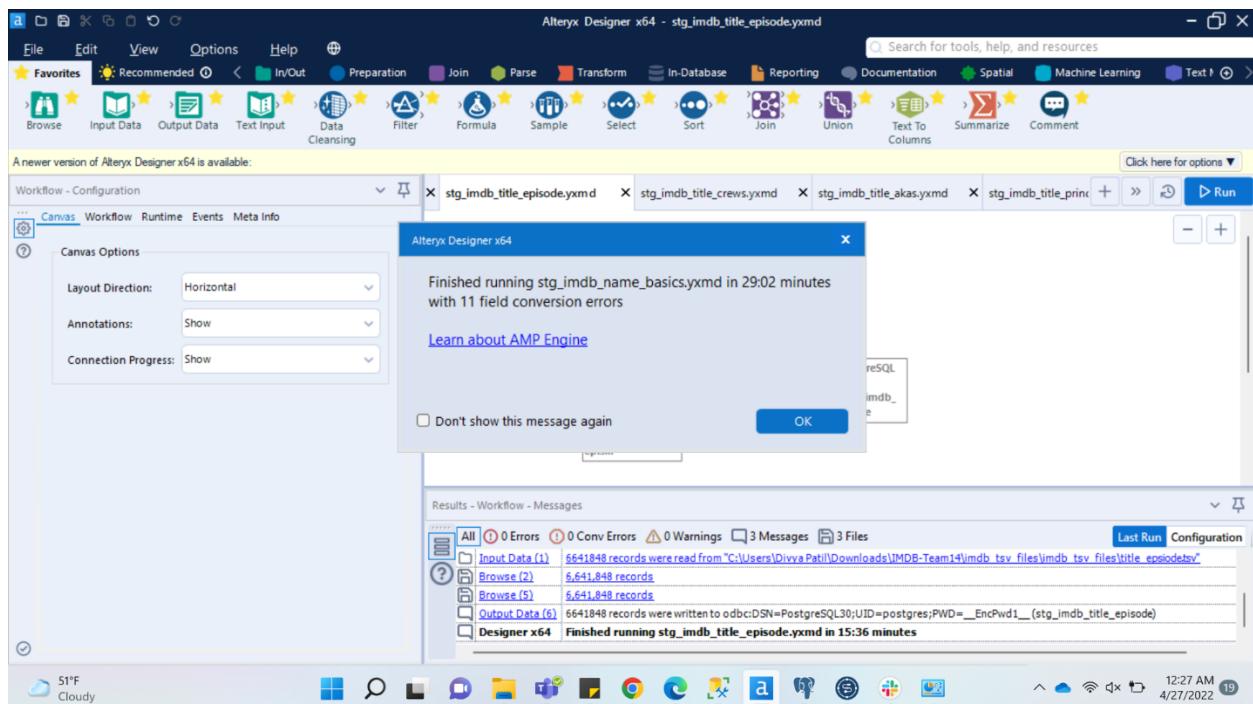
Properties SQL Statistics Dependencies Dependents postgres/postg... postgres/postg...

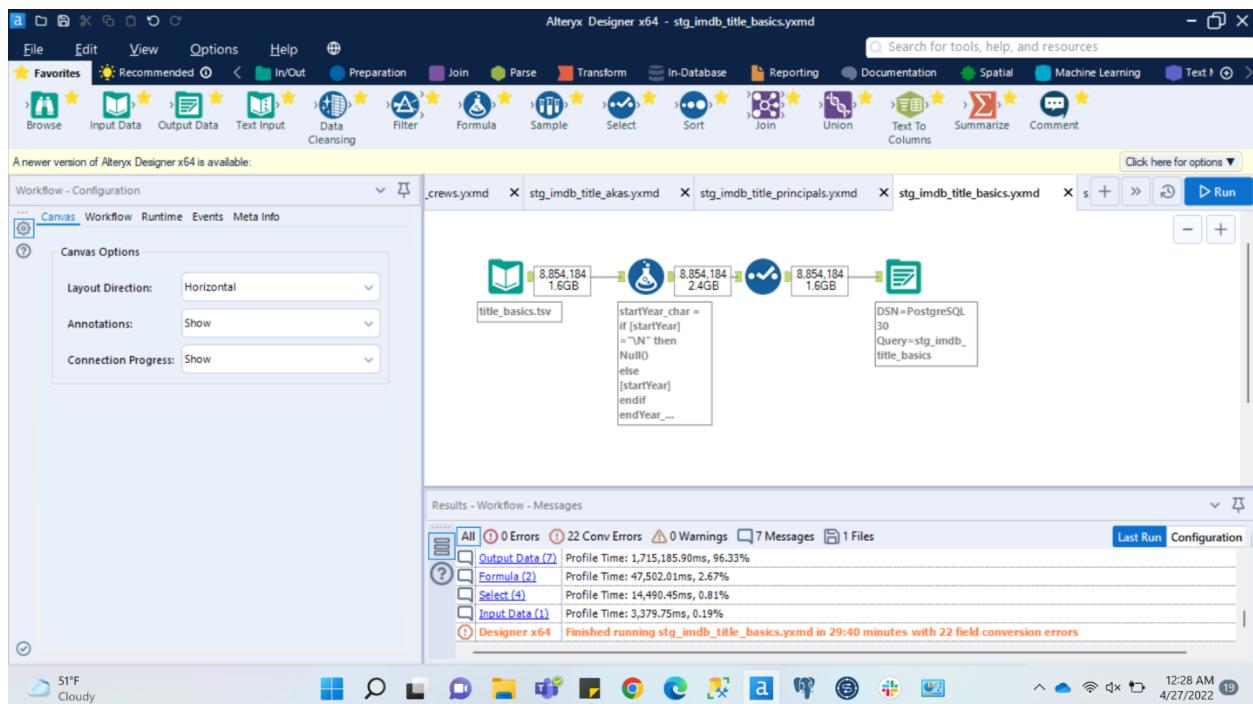
Query Editor Query History Scratch Pad

```
1 select count(*) from stg_imdb_title_basics
```

Data Output		Explain	Messages	Notifications
count	bigint			
1	8854184			

51°F Rain showers 11:47 PM 4/26/2022





pgAdmin 4

File Object Tools Help

Servers (1)

PostgreSQL 14

Databases (2)

- IMDB
- postgres

Casts Catalogs Event Triggers Extensions Foreign Data Wrappers Languages Publications Schemas (1)

public

- Aggregates Collations Domains FTS Configurations FTS Dictionaries FTS Parsers FTS Templates Foreign Tables Functions Materialized Views Operators

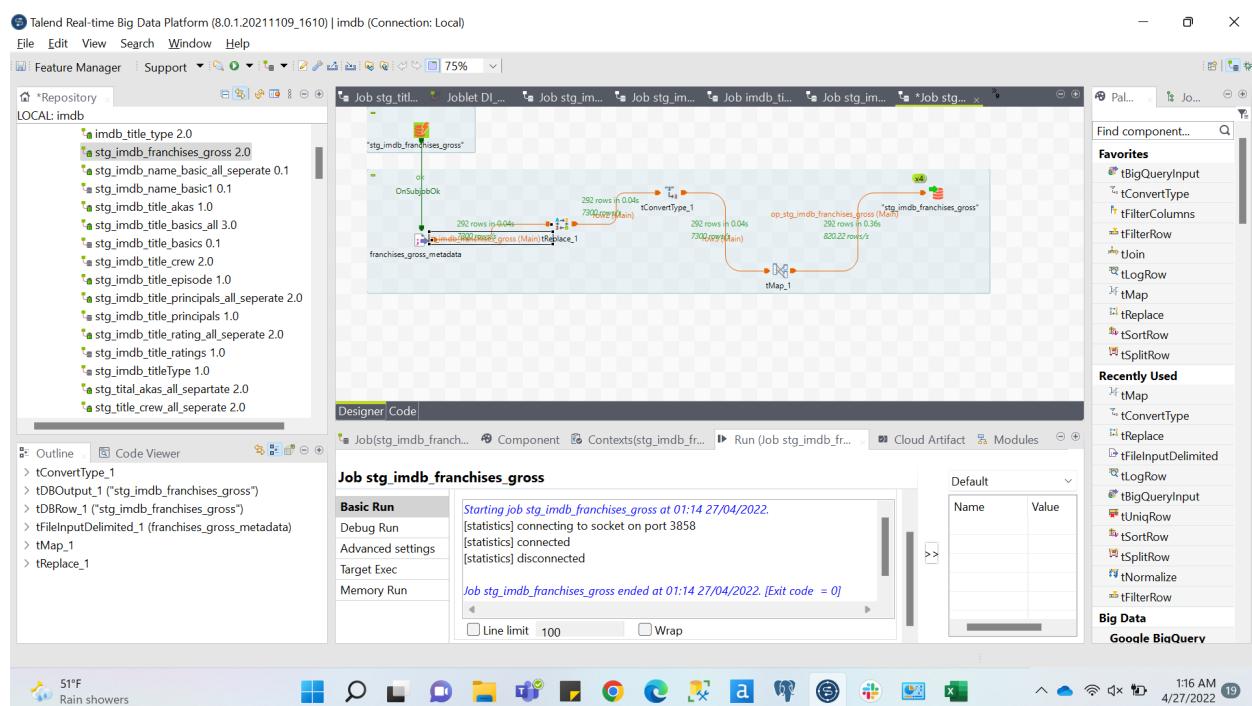
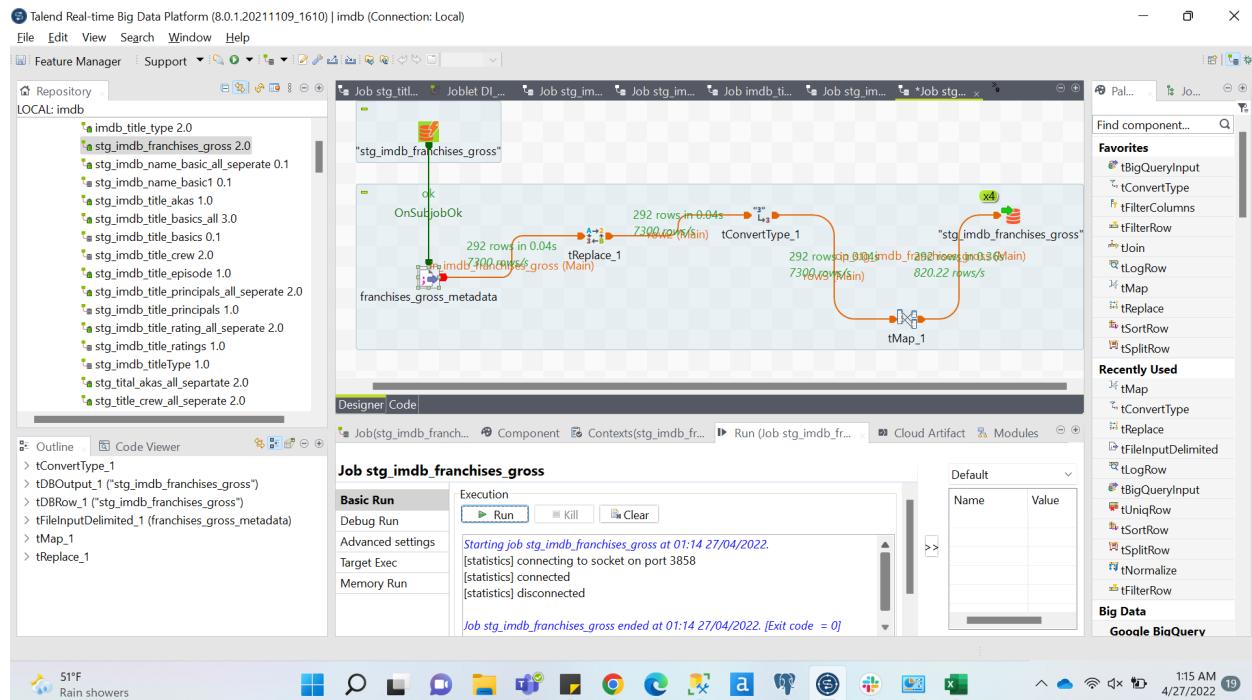
Properties SQL Statistics Dependencies Dependents postgres/postg... postgres/postg...

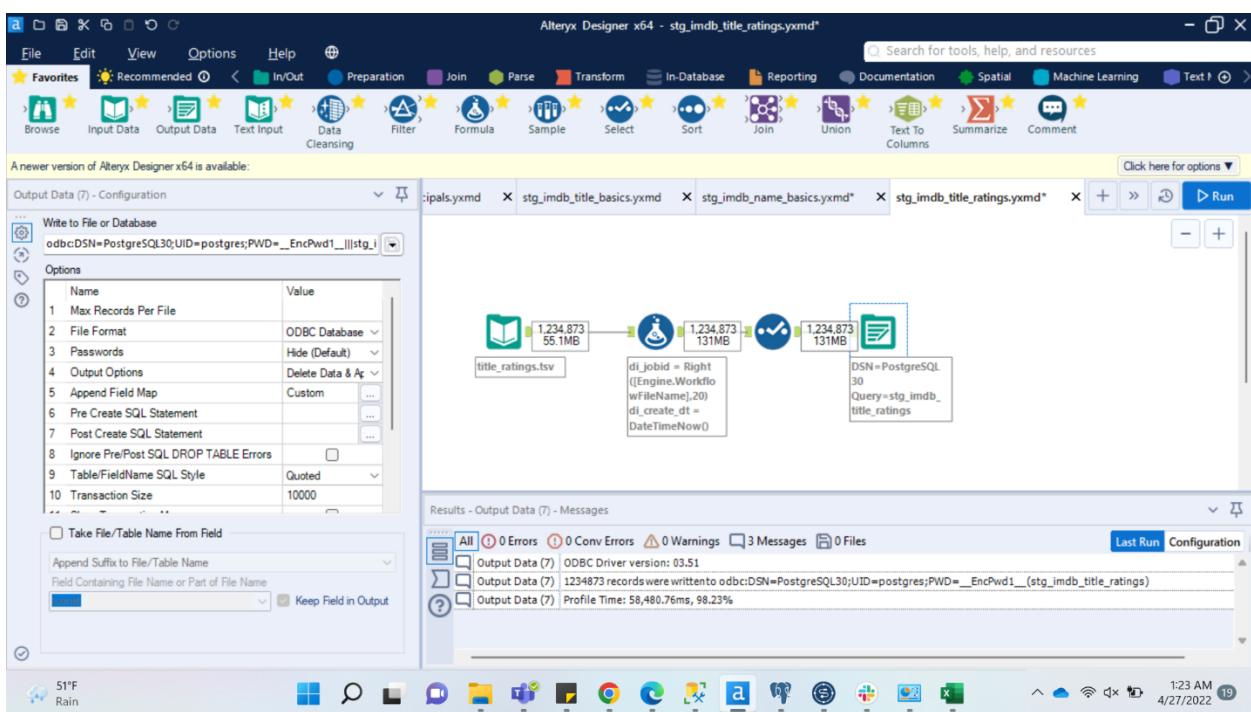
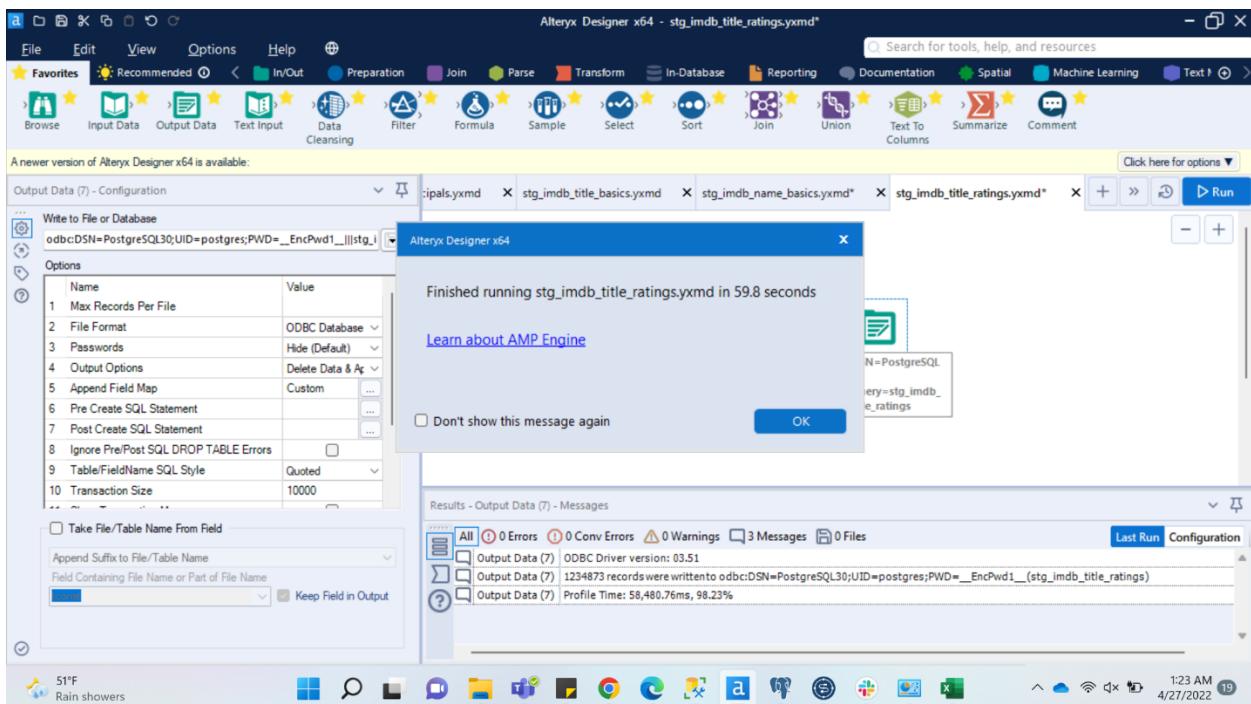
Query Editor Query History Scratch Pad

```
1 select count(*) from stg_imdb_title_basics
```

	Data Output	Explain	Messages	Notifications
count	bigint			
1	8854184			

12:29 AM 4/27/2022





pgAdmin - postgres/postgres@PostgreSQL 14 *

Browser Properties SQL Statistics Dependencies Dependents postgres/postg... postgres/postgres@PostgreSQL 14 *

Query Editor Query History Scratch Pad

```
1 select count(*) from stg_imdb_title_ratings
```

Data Output Explain Messages Notifications

count	bigint
1	1234873

Alteryx Designer x64 - stg_imdb_title_crew_writers.ymd*

File Edit View Options Help Search for tools, help, and resources

Favorites Recommended In/Out Preparation Join Parse Transform In-Database Reporting Documentation Spatial Machine Learning Text +

Browse Input Data Output Data Text Input Data Cleansing Formula Sample Select Sort Join Union Text To Columns Summarize Comment

A newer version of Alteryx Designer x64 is available.

Alteryx Designer x64 stg_imdb_name_basics.ymd* stg_imdb_title_ratings.ymd* stg_imdb_title_crew_writers.ymd* Run

Finished running stg_imdb_title_crew_writers.ymd in 19:16 minutes

[Learn about AMP Engine](#)

Don't show this message again OK

Pre Create SQL Statement Post Create SQL Statement Ignore Pre/Post SQL DROP TABLE Errors Table/FieldName SQL Style Transaction Size

Take File/Table Name From Field Append Suffix to File/Table Name Field Containing File Name or Part of File Name Keep Field in Output

Results - Output Data (8) - Input

4 of 4 Fields * 10,797 of 10,346,509 records displayed(partial results)

Record	tconst	writers	di_jobid	di_create_dt
1	tt0000009	nm0085156	le_crew_writers.ymd	2022-04-27 01:33:30
2	tt0000036	nm0410331	le_crew_writers.ymd	2022-04-27 01:33:30
3	tt0000076	nm0410331	le_crew_writers.ymd	2022-04-27 01:33:30

1:57 AM 4/27/2022

Talend Real-time Big Data Platform

pgAdmin 4

File Edit View Schema Object Tools Help

Feature Manager

Repository LOCAL: imdb

Servers (1)

- PostgreSQL 14
- Databases (2)
 - IMDB
 - postgres
- Casts
- Catalogs
- Event Triggers
- Extensions
- Foreign Data Wrappers
- Languages
- Publications
- Schemas (1)
 - public
- Aggregates
- Collations
- Domains
- FTS Configuration
- FTS Dictionaries
- FTS Parsers
- FTS Templates
- Foreign Tables
- Functions
- Materialized Views

Query Editor Query History Scratch Pad

```
1 select count(*) from stg_imdb_title_crew_writers
```

Data Output	Explain	Messages	Notifications
count bigint 1 10346509			

Component... Joins IterColumnns IterRowJoin pgRowMap ReplaceOrtRowSplitRow

Used BInput(Microsoft SQL Server) BOutput(Microsoft SQL Server) Map Delimited BRow(Microsoft SQL Server) ConvertTypeReplace pgRowigQueryInput UniqueRow

1:58 AM 4/27/2022

51°F Light rain

Alteryx Designer x64 - stg_imdb_title_crew_directors.ymd

File Edit View Options Help

Favorites Recommended In/Out Preparation Join Parse Transform In-Database Reporting Documentation Spatial Machine Learning Text

Alteryx Designer x64

Finished running stg_imdb_title_crew_directors in 14:59 minutes

Learn about AMP Engine

Don't show this message again

OK

Connection Progress: Show

DSN=PostgreSQL 30 Query="public"."stg_imdb_title_crew_writers" 8,054,184 rows 765MB

10,412,697 rows 931MB

6,614,738 rows 661MB

6,614,738 rows 204MB

6,614,738 rows 678MB

6,614,738 rows 614MB

directors = if (directors) = '\N' then Null else [directors] endif di_jobid = Ri...

DSN=PostgreSQL 30 Query="stg_imdb_title_crew_directors" 6,614,738 rows 614MB

Results - Workflow - Messages

All 0 Errors 0 Conv Errors 0 Warnings 6 Messages 0 Files Last Run Configuration

Output Data (8) ODBC Driver version: 03.51

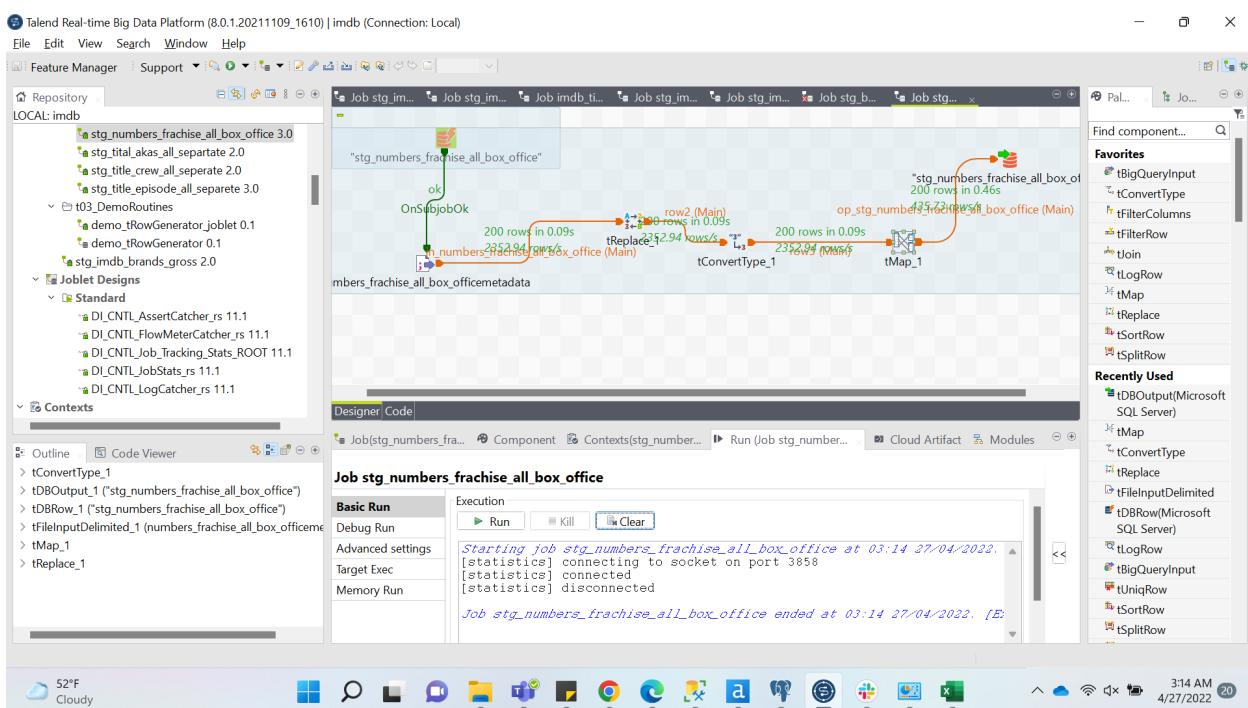
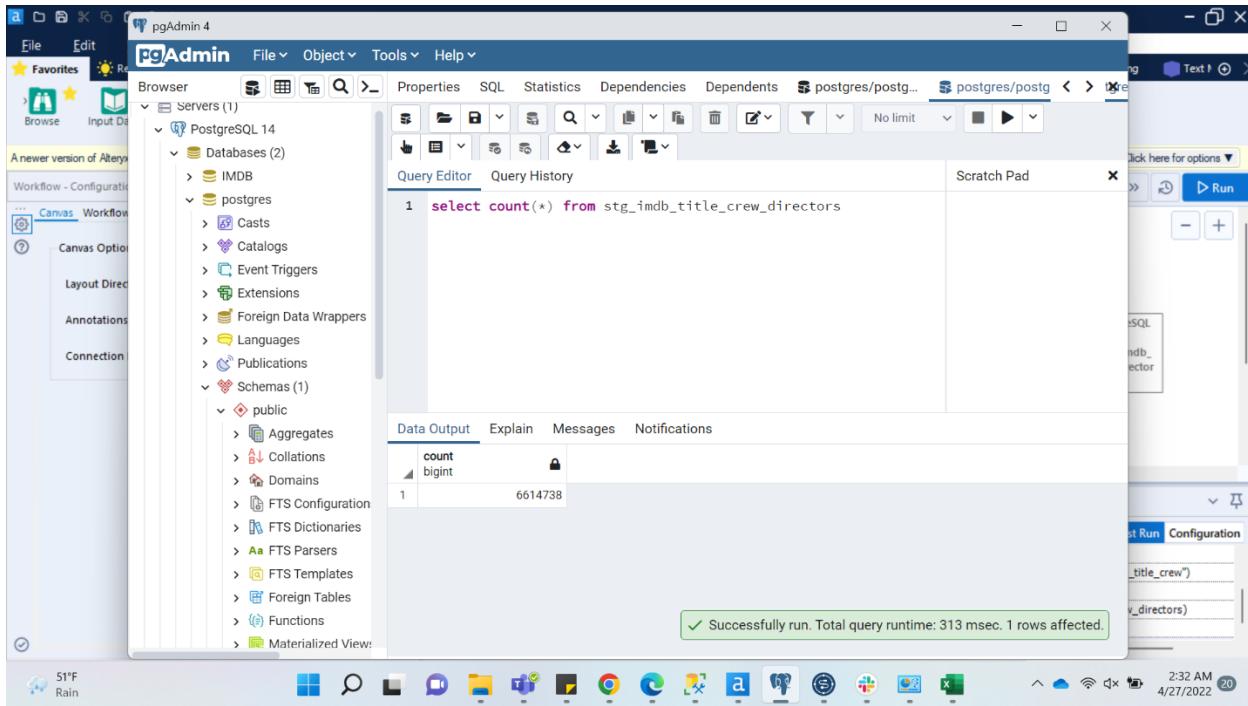
Input Data (1) 8854184 records were read from odbc:DSN=PostgreSQL30;UID=postgres;PWD=_EncPwd1_ ("public"."stg_imdb_title_crew_writers")

Filter (4) 6614738 records were True and 3797959 were False

Output Data (8) 6614738 records were written to odbc:DSN=PostgreSQL30;UID=postgres;PWD=_EncPwd1_ ("stg_imdb_title_crew_directors")

Designer x64 Finished running stg_imdb_title_crew_directors in 14:59 minutes

51°F Rain



SQLQuery5.sql - DIVYA.IMDB (damg7370_New (61)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

GeoLoc

IMDB

Execute

SQLQuery6.sql - D...amg7370_New (62)

SQLQuery5.sql - D...amg7370_New (61)*

SQLQuery4.sql - D...amg7370_New (53)

Object Explorer

Connect

dbo.stg_imdb_title_basics

dbo.stg_imdb_title_crew

dbo.stg_imdb_title_episode

dbo.stg_imdb_title_principals

dbo.stg_imdb_title_ratings

dbo.stg_imdb_titleType

dbo.stg_numbers_box_office

dbo.stg_numbers_franchise_all_box_office

dbo.stg_numbers_movies_box_office

Keys

Constraints

Triggers

Indexes

Statistics

Views

External Resources

Synonyms

Programmability

Service Broker

Storage

Security

imdb_src_tv

PET CARE

SF_FILMSLOCATION OUTPUT

```
***** script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [Franchises_Gross_SK]
,[Franchise]
,[Total_Revenue]
,[Number_of_Releases]
,[Top_Release]
,[Lifetime_Gross]
,[DI_JobID]
,[DI_create_dt]
,[SOR_fileID]
,[SOR_fileName]
FROM [IMDB].[dbo].[stg_imdb_franchises_gross]
select count(*) from stg_numbers_franchise_all_box_office
```

90 %

Results Messages

(No column name)

1 200

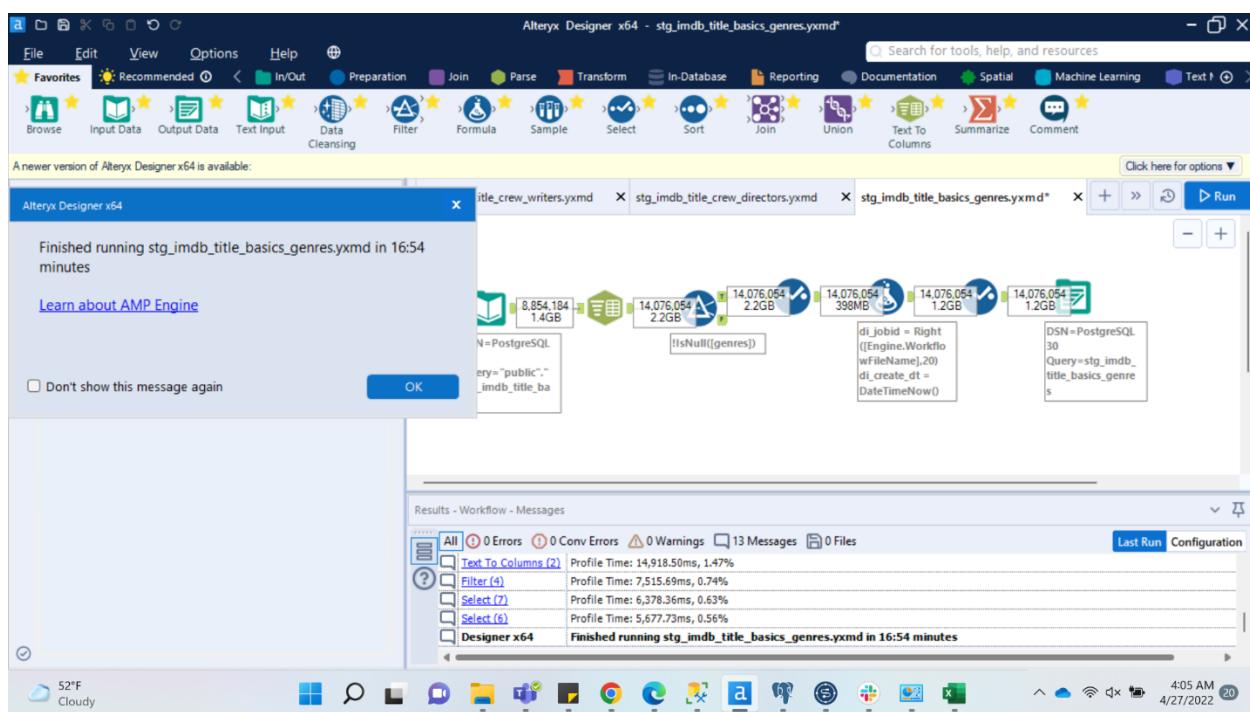
Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (61) | IMDB | 00:00:00 | 1 rows

Ready

52°F Cloudy

3:15 AM 4/27/2022



Talend Real-time Big Data Platform

pgAdmin 4

File Edit View Schema Object Tools Help

Feature Manager

Repository LOCAL: imdb

Servers (1)

- PostgreSQL 14
- Databases (2)
 - IMDB
 - postgres
- Tables (2)
- Functions (1)
- Procedures (1)
- Triggers (1)
- Events (1)
- Extensions (1)
- Foreign Data Wrappers (1)
- Languages (1)
- Publications (1)
- Schemas (1)
 - public
- Aggregates (1)
- Collations (1)
- Domains (1)
- FTS Configuration (1)
- FTS Dictionaries (1)
- FTS Parsers (1)
- FTS Templates (1)
- Foreign Tables (1)
- Functions (1)
- Materialized Views (1)

Query Editor Query History Scratch Pad

```
1 select count(*) from stg_imdb_title_genres
```

Data Output Explain Messages Notifications

count	bigint
1	14076054

4:06 AM 4/27/2022

Alteryx Designer x64 - stg_imdb_name_basics_primaryprofession.ymd*

File Edit View Options Help

Favorites Recommended In/Out Preparation Join Parse In-Database Reporting Documentation Spatial Machine Learning Text

Browse Input Data Output Data Text Input Data Filter Formula Sample Select Sort Join Union Text To Columns

Click here for options

Finished running stg_imdb_name_basics_primaryprofession.ymd in 14:44

Learn about AMP Engine

Don't show this message again

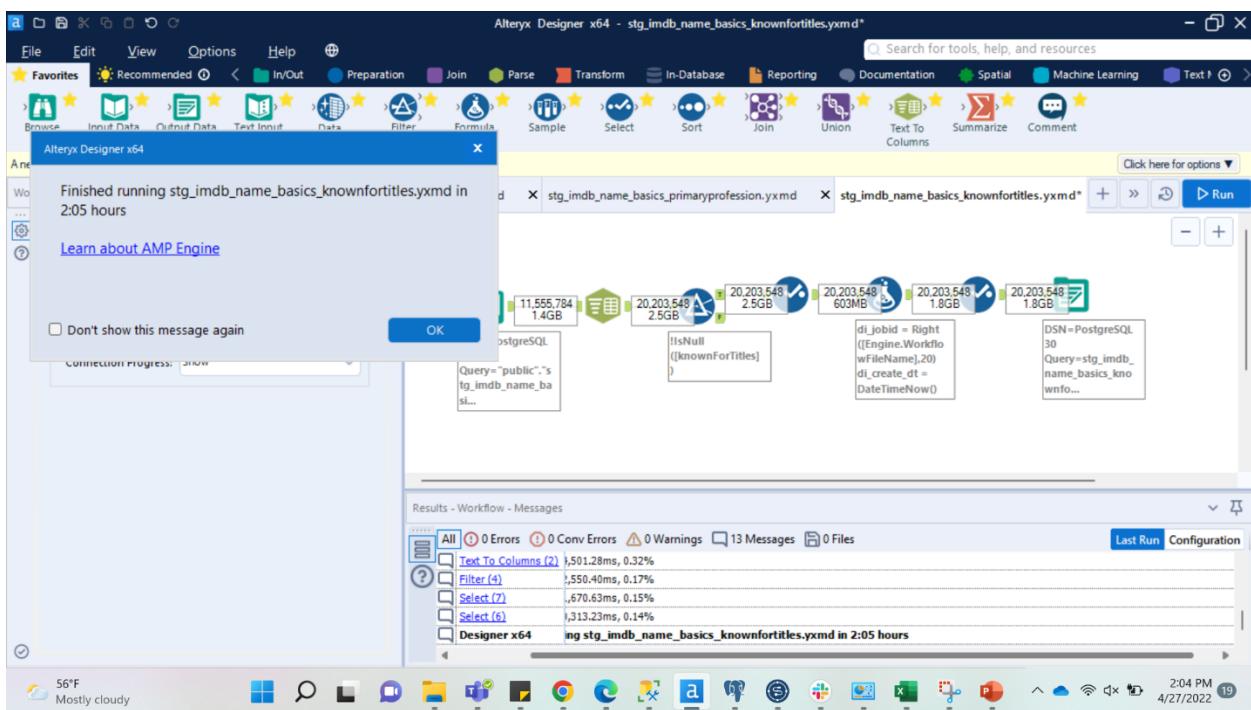
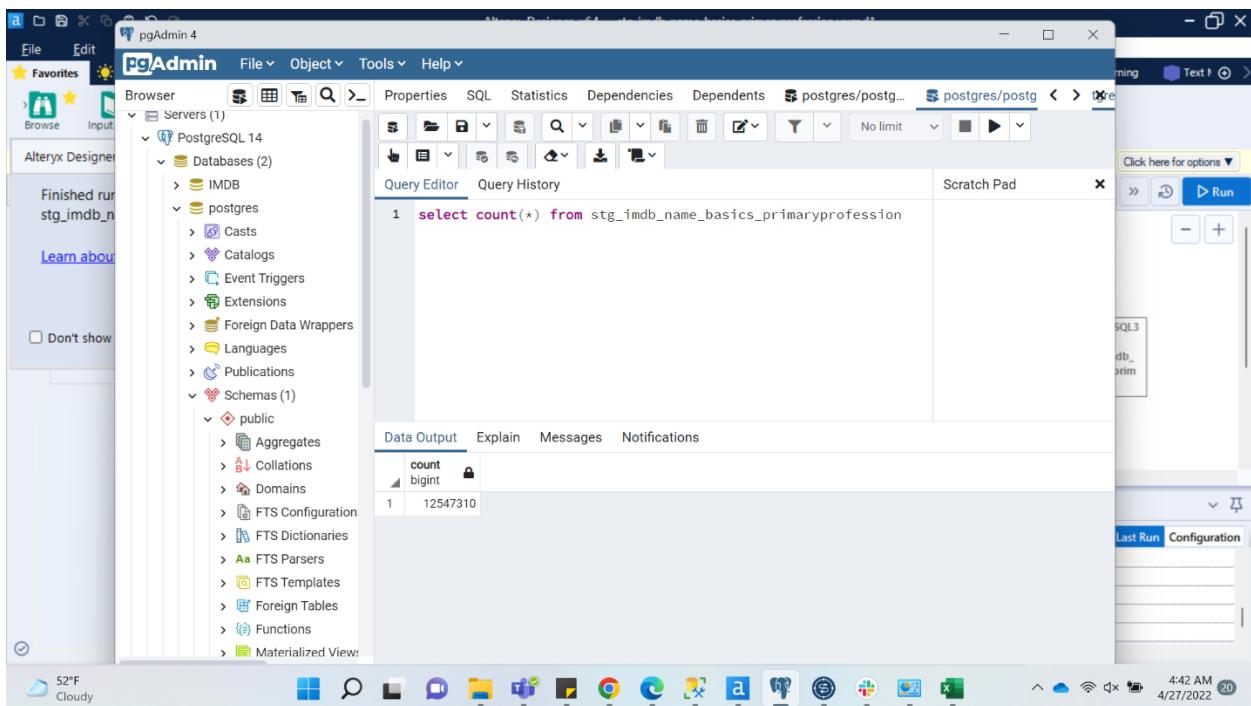
OK

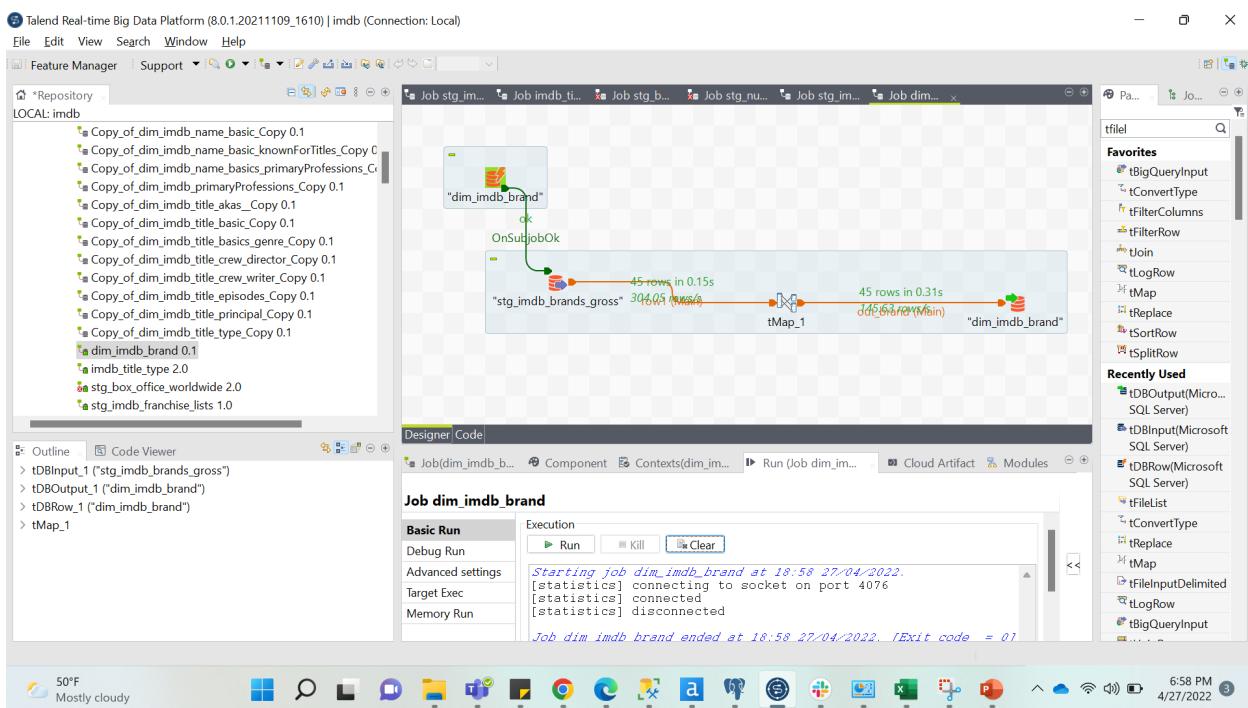
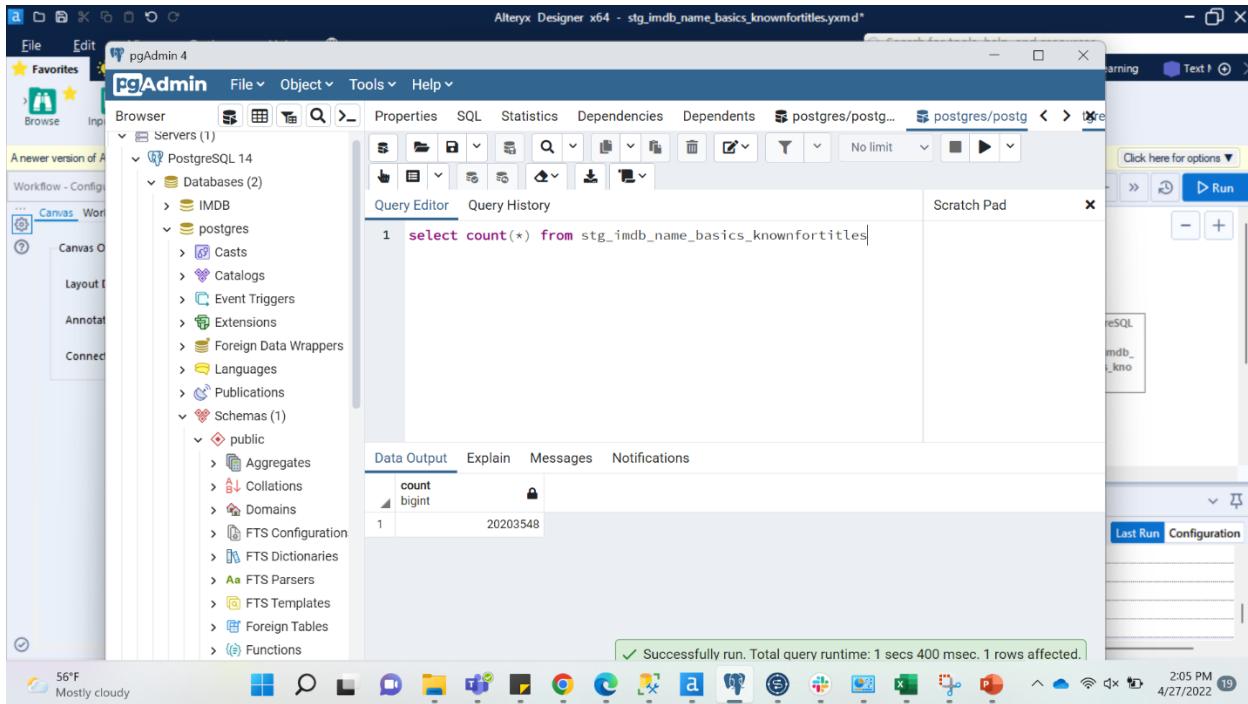
Results - Workflow - Messages

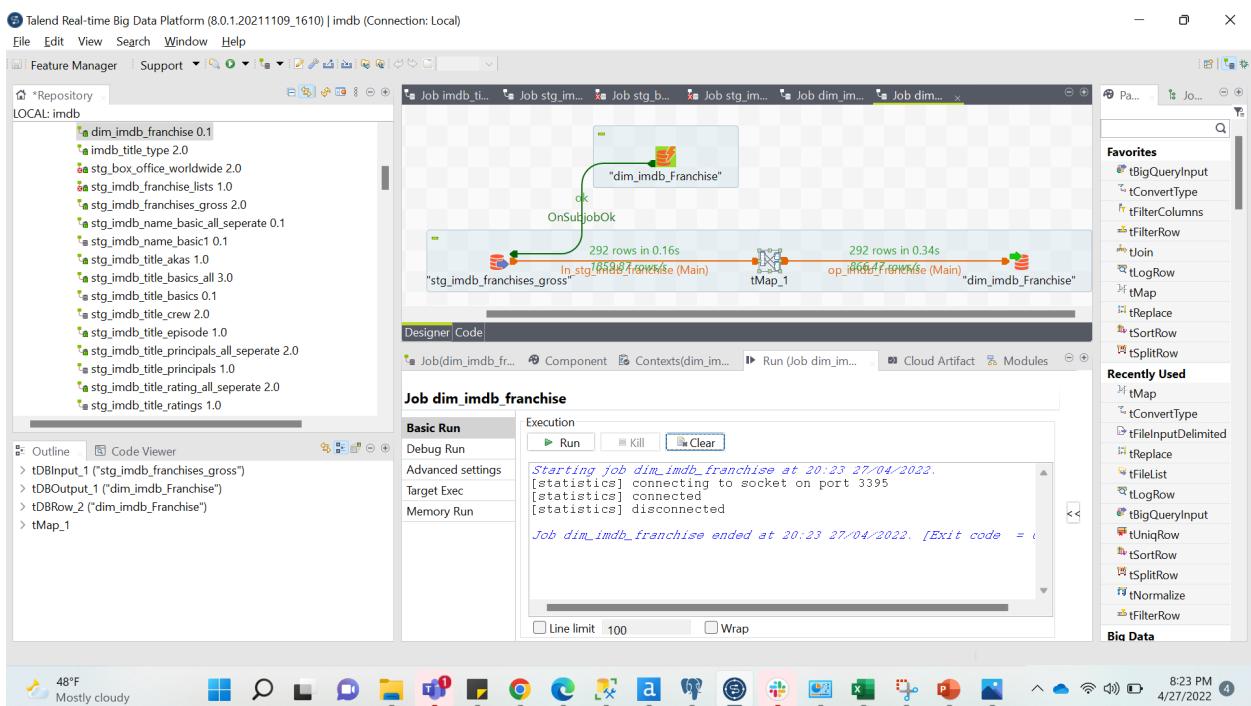
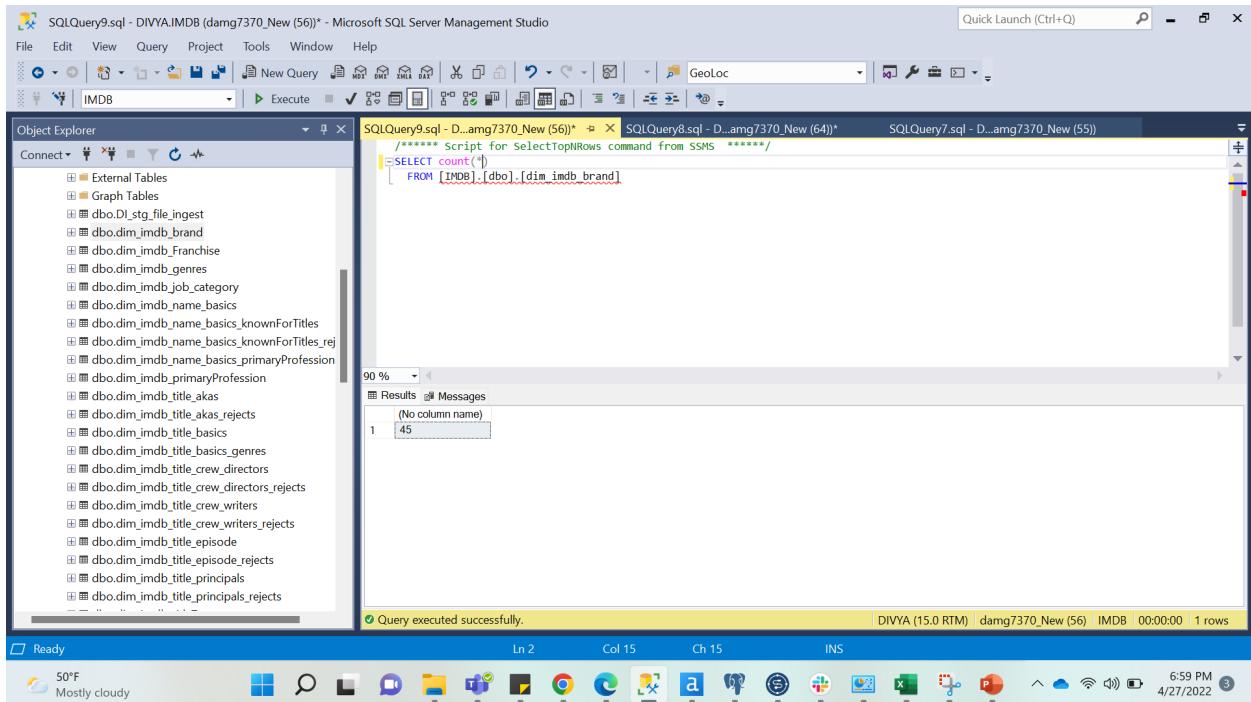
All 0 Errors 0 Conv Errors 0 Warnings 13 Messages 0 Files Last Run Configuration

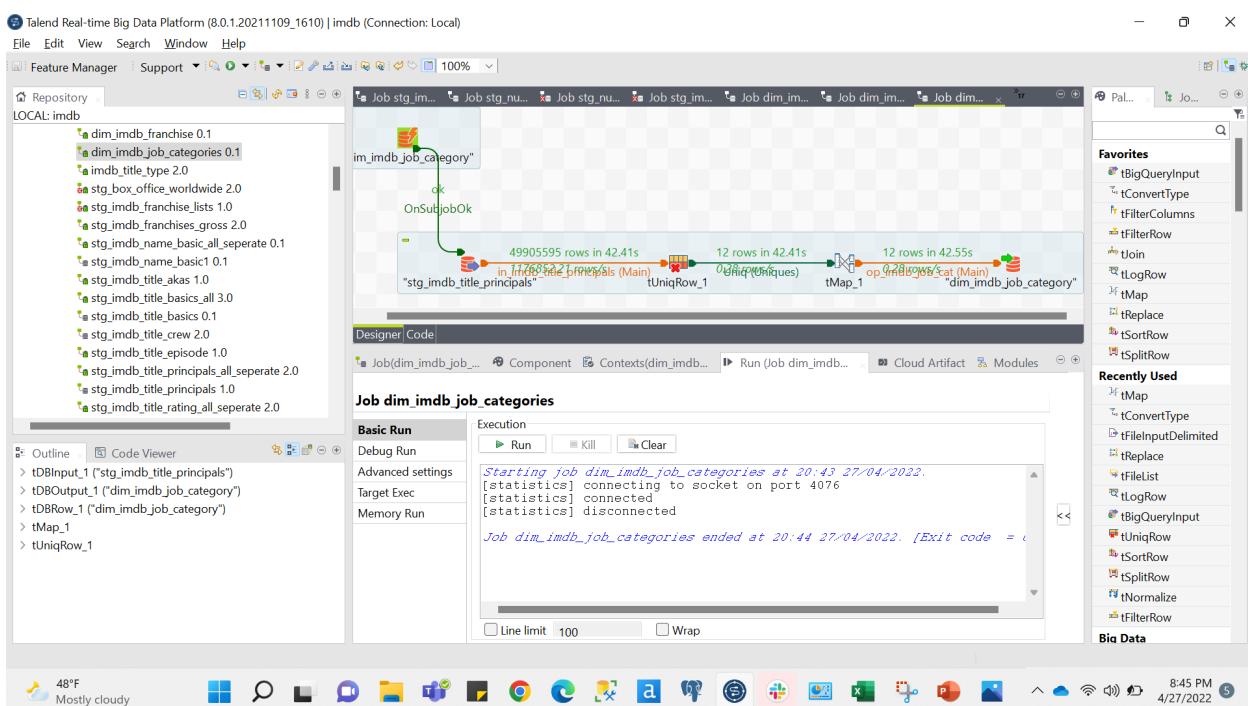
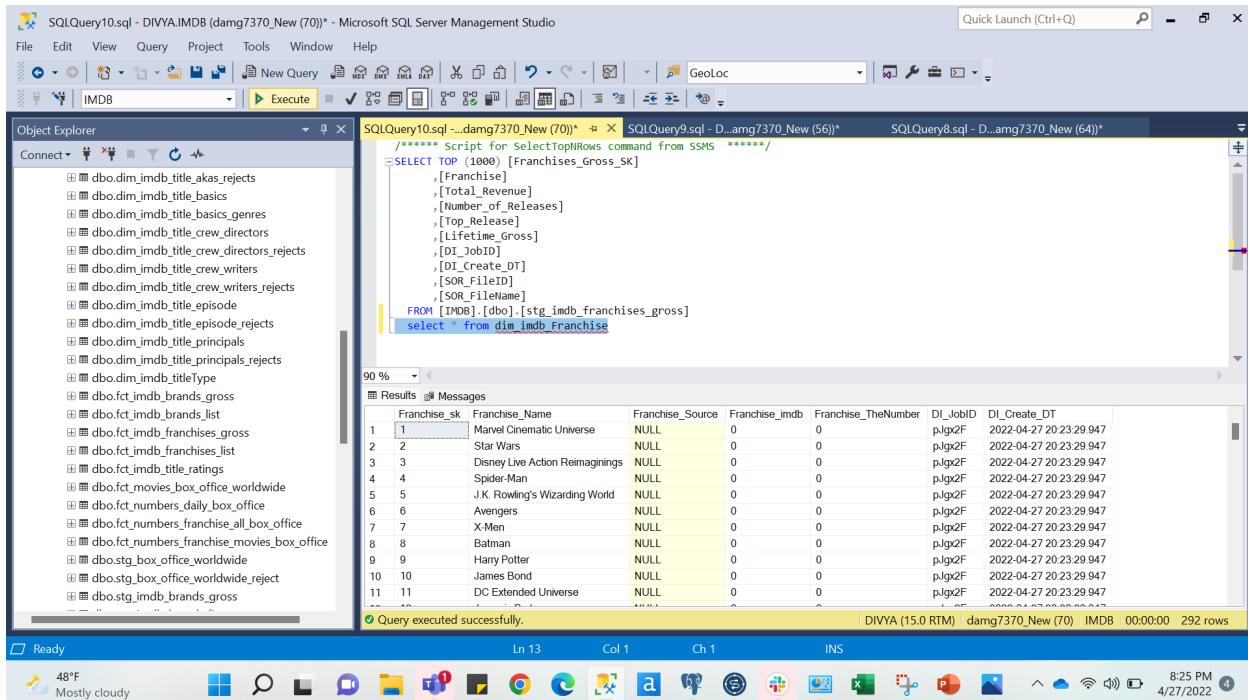
	Profile Time
Formula (5)	19,213.53ms, 2.17%
Filter (4)	7,317.75ms, 0.83%
Select (7)	5,779.05ms, 0.65%
Select (6)	5,243.53ms, 0.59%
Designer x64	Finished running stg_imdb_name_basics_primaryprofession.ymd in 14:44 minutes

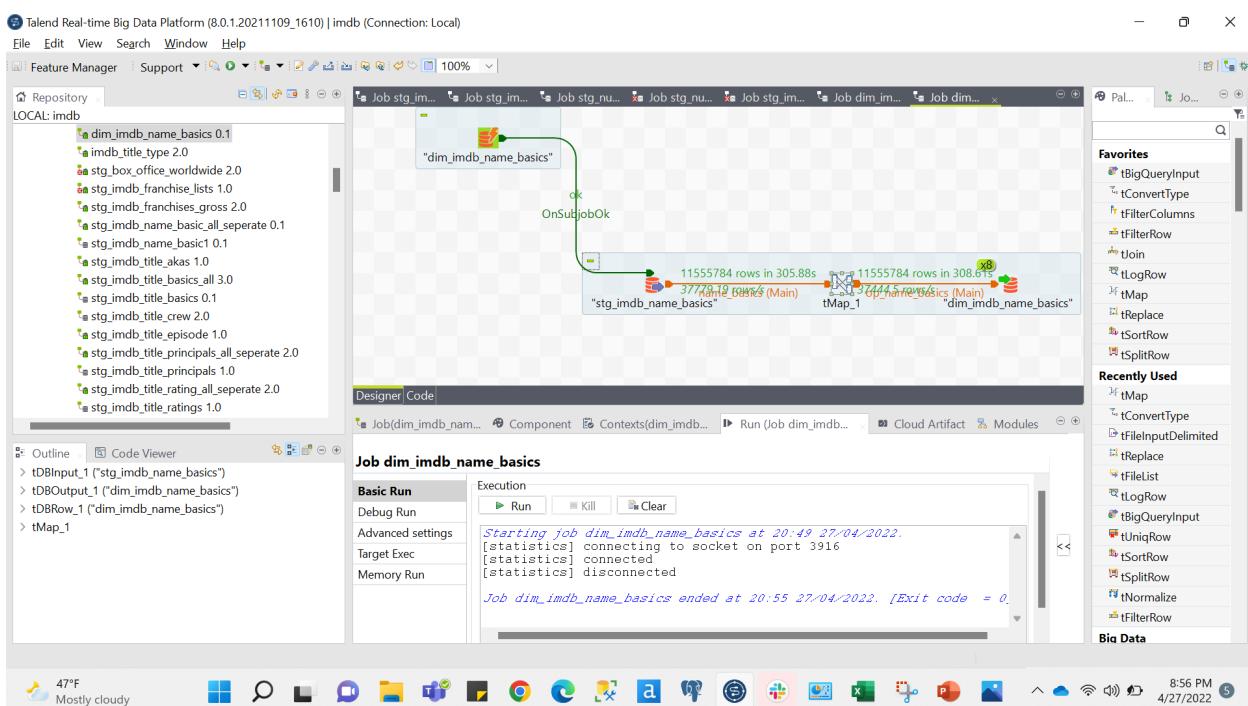
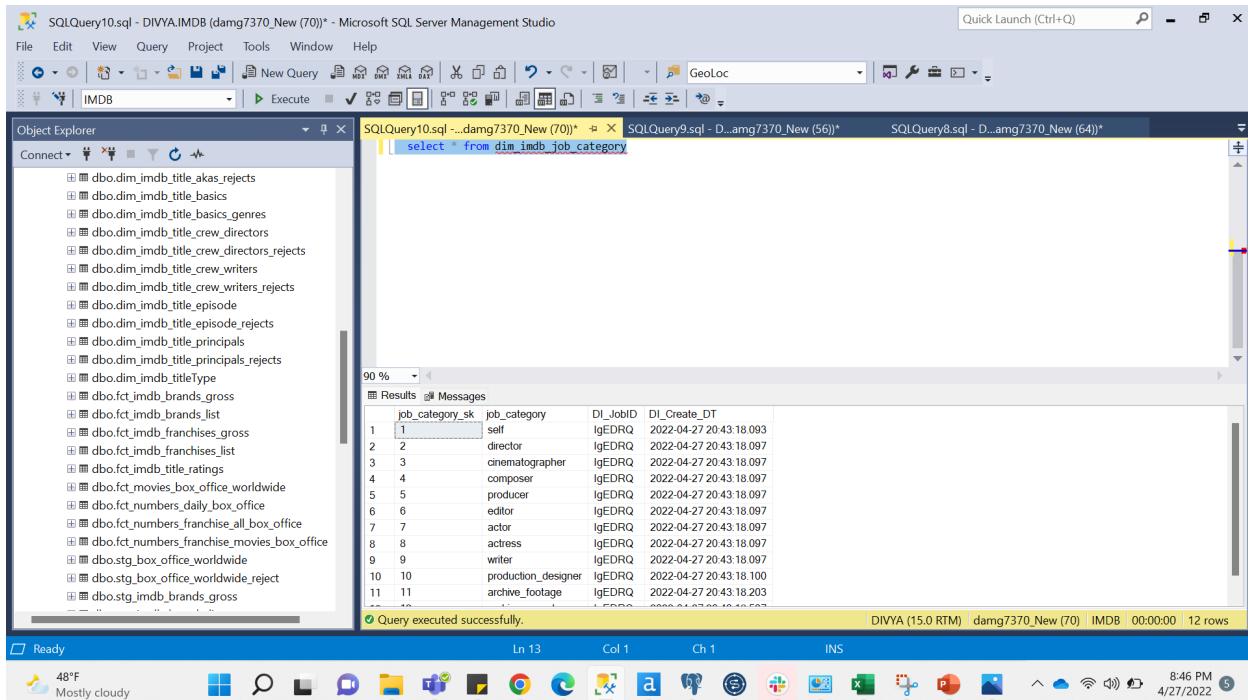
4:40 AM 4/27/2022











SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

GeoLoc

IMDB

Object Explorer

SQLQuery10.sql - damg7370_New (70)* SQLQuery9.sql - D...amg7370_New (56)* SQLQuery8.sql - D...amg7370_New (64)*

```
select * from dim_imdb_name_basics
select count(*) from dim_imdb_name_basics
```

Results Messages

(No column name)

1 11555784

Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready

47°F Mostly cloudy

Ln 14 Col 1 Ch 1 INS

8:57 PM 4/27/2022

SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

GeoLoc

IMDB

Object Explorer

SQLQuery10.sql - damg7370_New (70)* SQLQuery9.sql - D...amg7370_New (56)* SQLQuery8.sql - D...amg7370_New (64)*

```
select * from dim_imdb_name_basics
select count(*) from dim_imdb_name_basics
```

Results Messages

(No column name)

1 11555784

Query executed successfully.

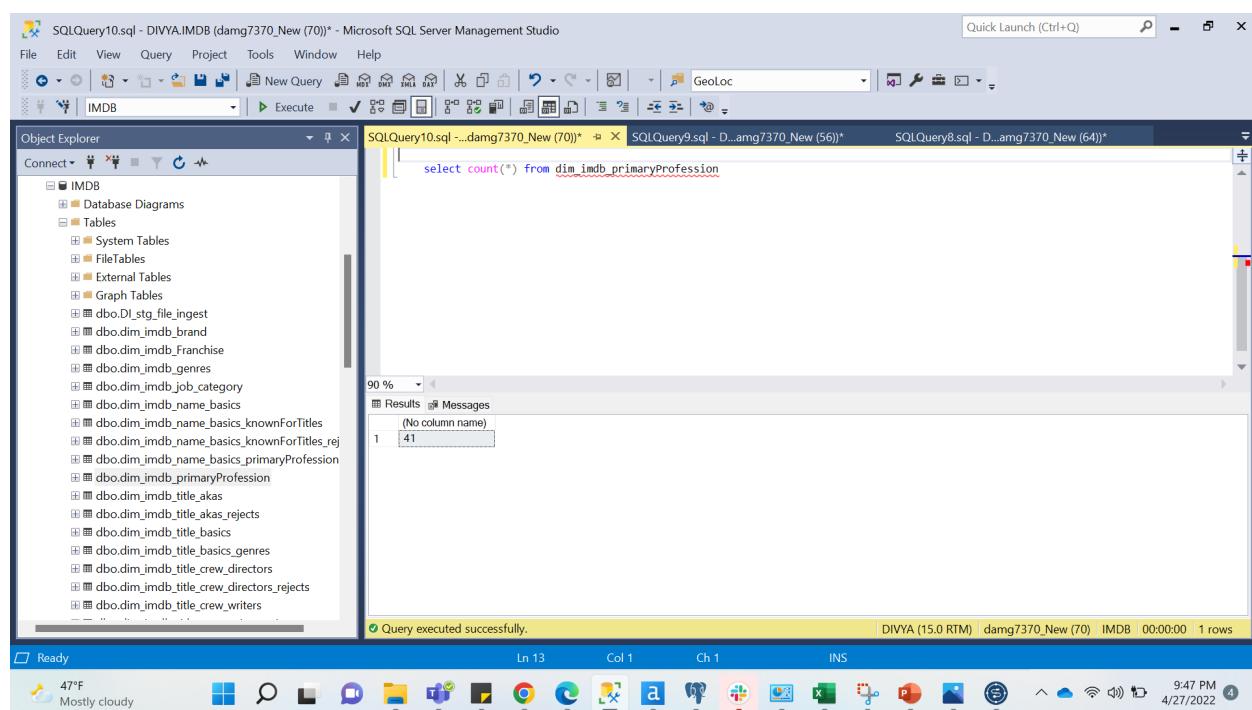
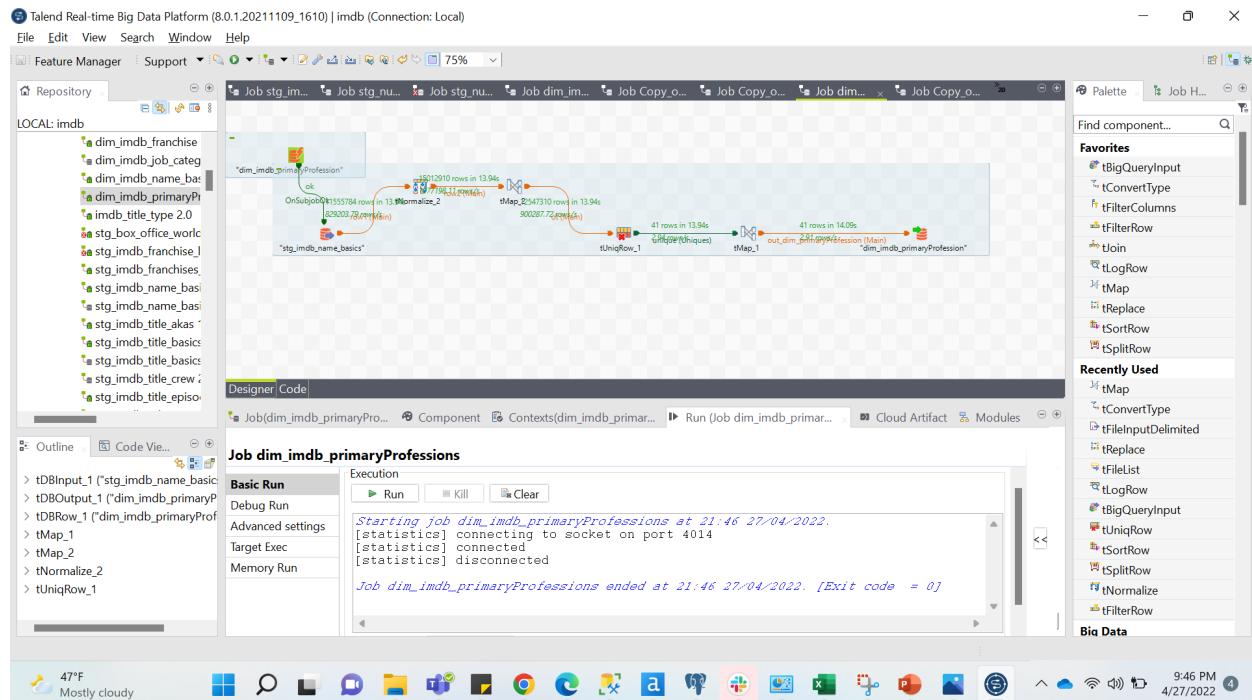
DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

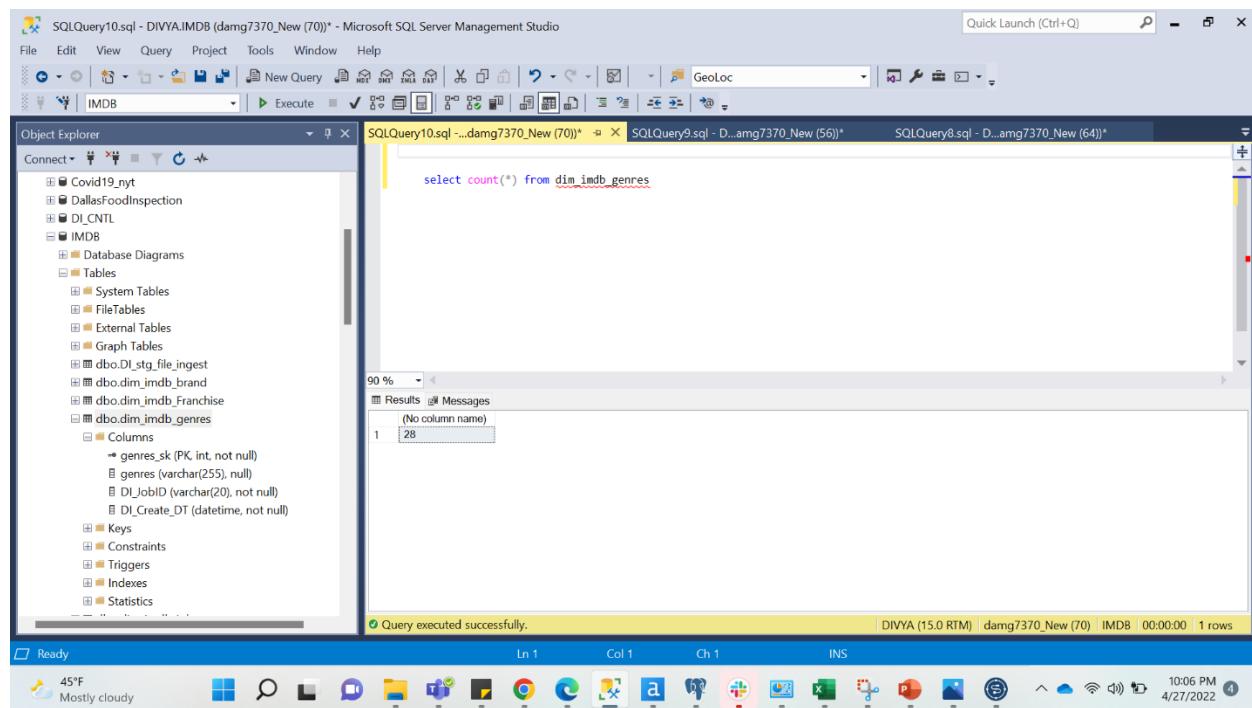
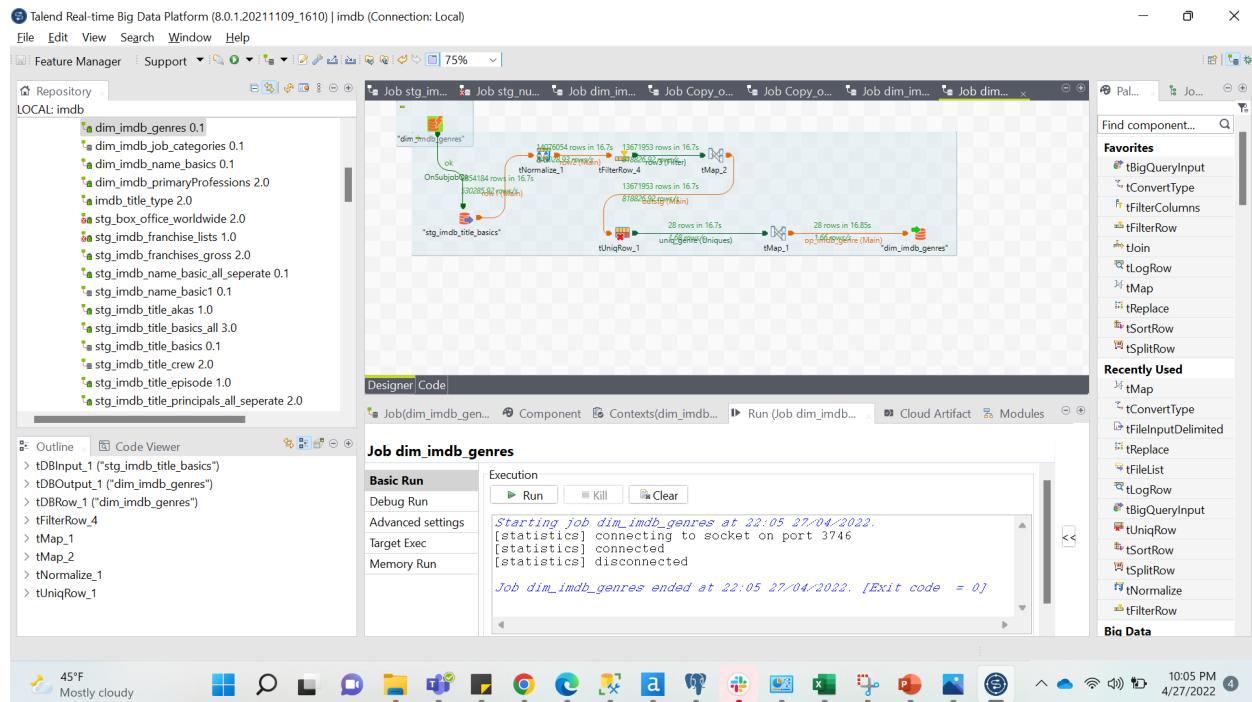
Ready

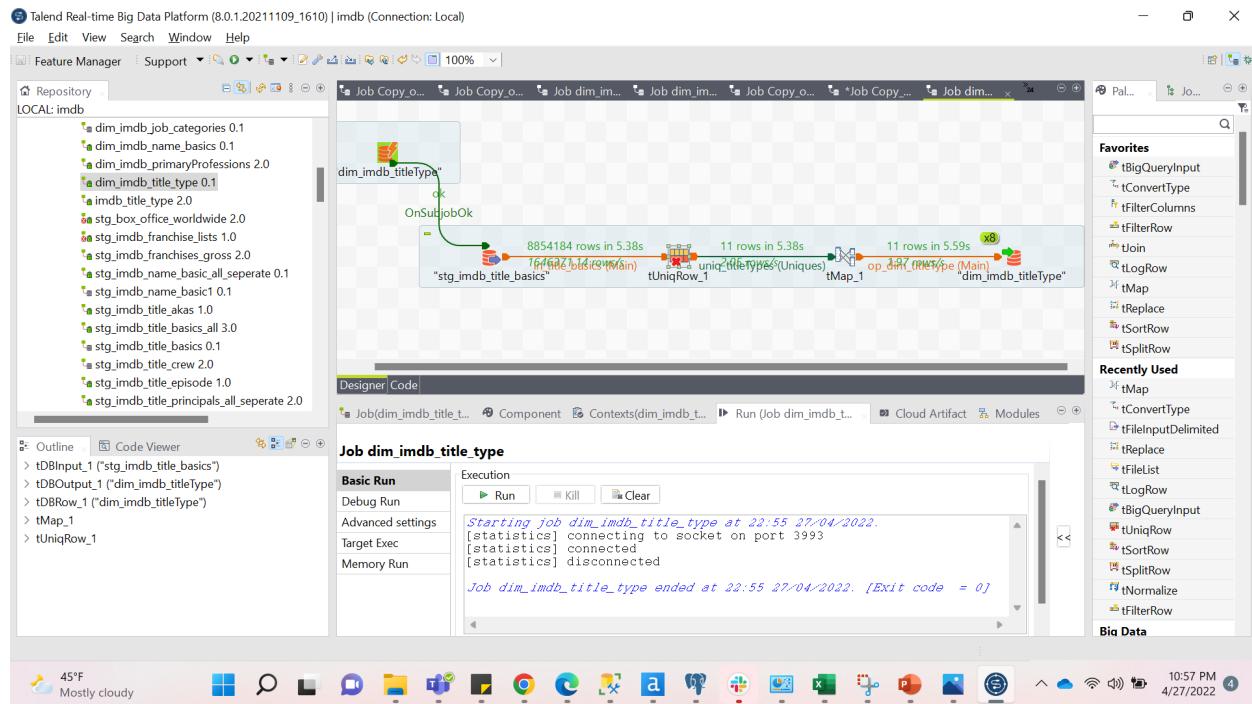
47°F Mostly cloudy

Ln 14 Col 1 Ch 1 INS

8:57 PM 4/27/2022







45°F Mostly cloudy 10:57 PM 4/27/2022

SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

- Covid19_nyt
- DallasFoodInspection
- DL_CNTL
- IMDB
 - Database Diagrams
 - Tables
 - System Tables
 - FileTables
 - External Tables
 - Graph Tables
 - dbo.DL_stg_file_ingest
 - dbo.dim_imdb_brand
 - dbo.dim_imdb_franchise
 - dbo.dim_imdb_genres
 - Columns
 - genres_sk (PK, int, not null)
 - genres (varchar(255), null)
 - DI_JobID (varchar(20), not null)
 - DI_Create_DT (datetime, not null)
 - Keys
 - Constraints
 - Triggers
 - Indexes
 - Statistics

SQLQuery10.sql - damg7370_New (70)* SQLQuery9.sql - D...amg7370_New (56)* SQLQuery8.sql - D...amg7370_New (64)*

```
select count(*) from dim_imdb_titleType
```

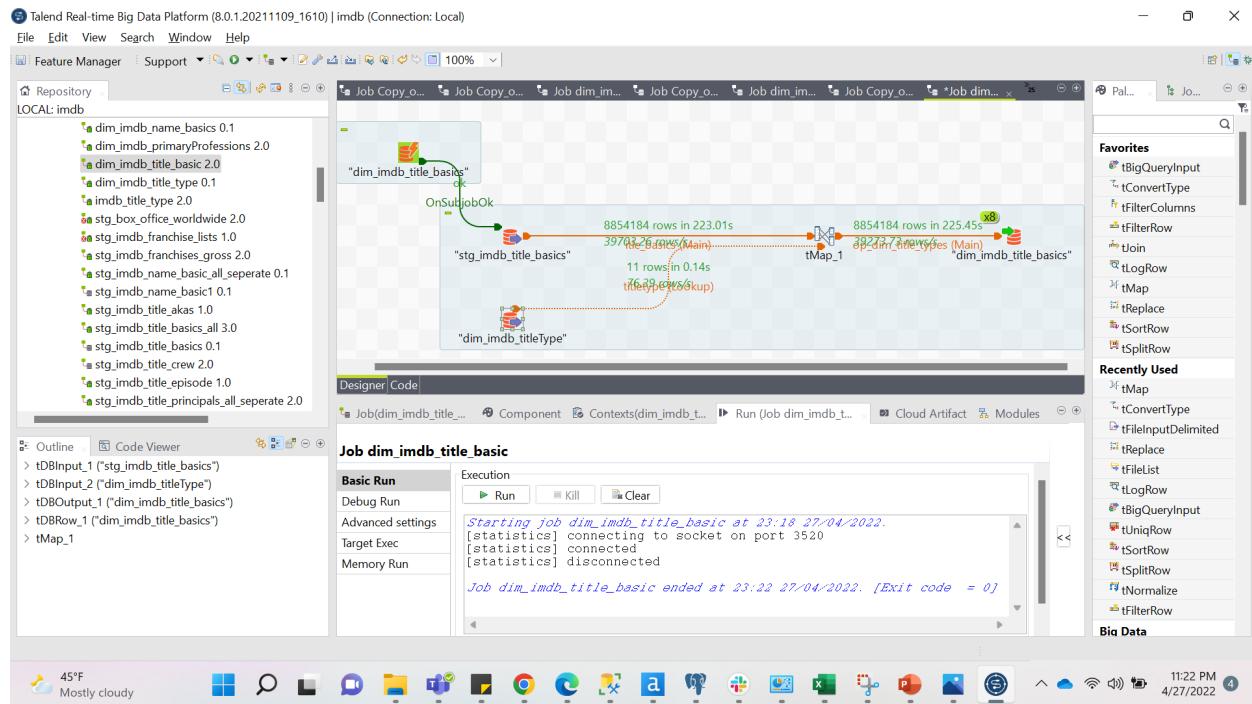
Results Messages

(No column name)
11

90 %

Query executed successfully. DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready 45°F Mostly cloudy 11:00 PM 4/27/2022



45°F Mostly cloudy

SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

```
select count(*) from dim_imdb_title_basics
```

Results Messages

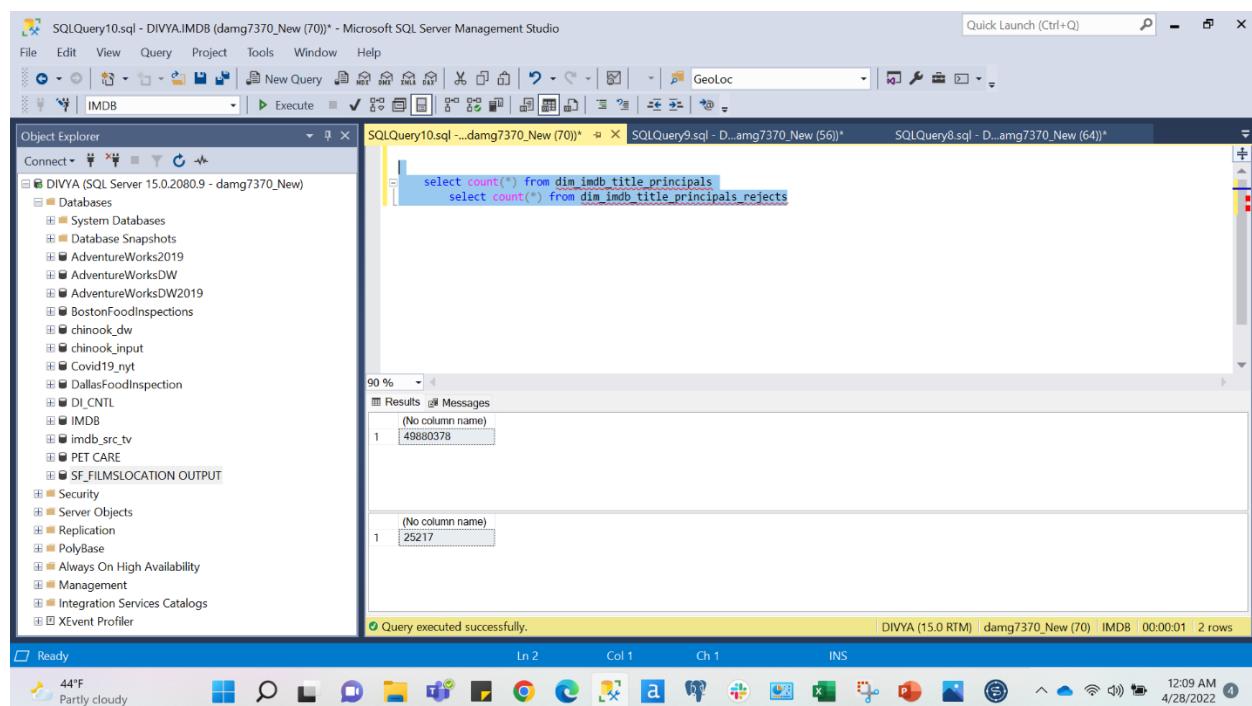
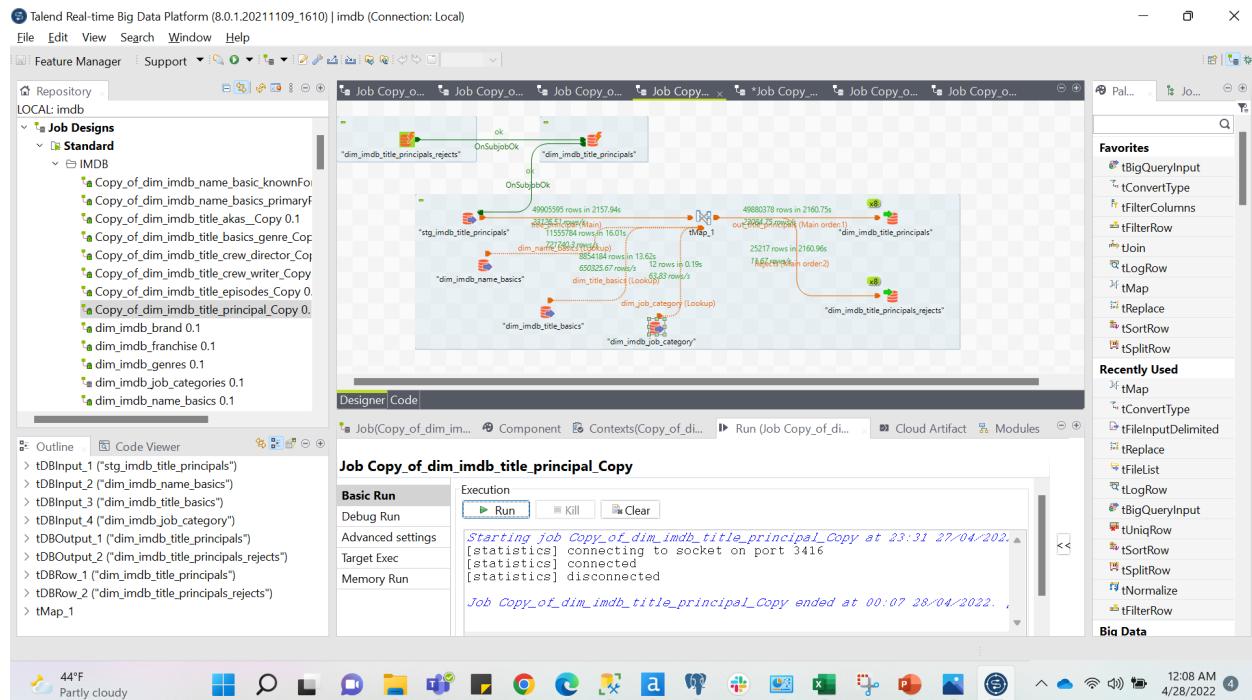
(No column name)

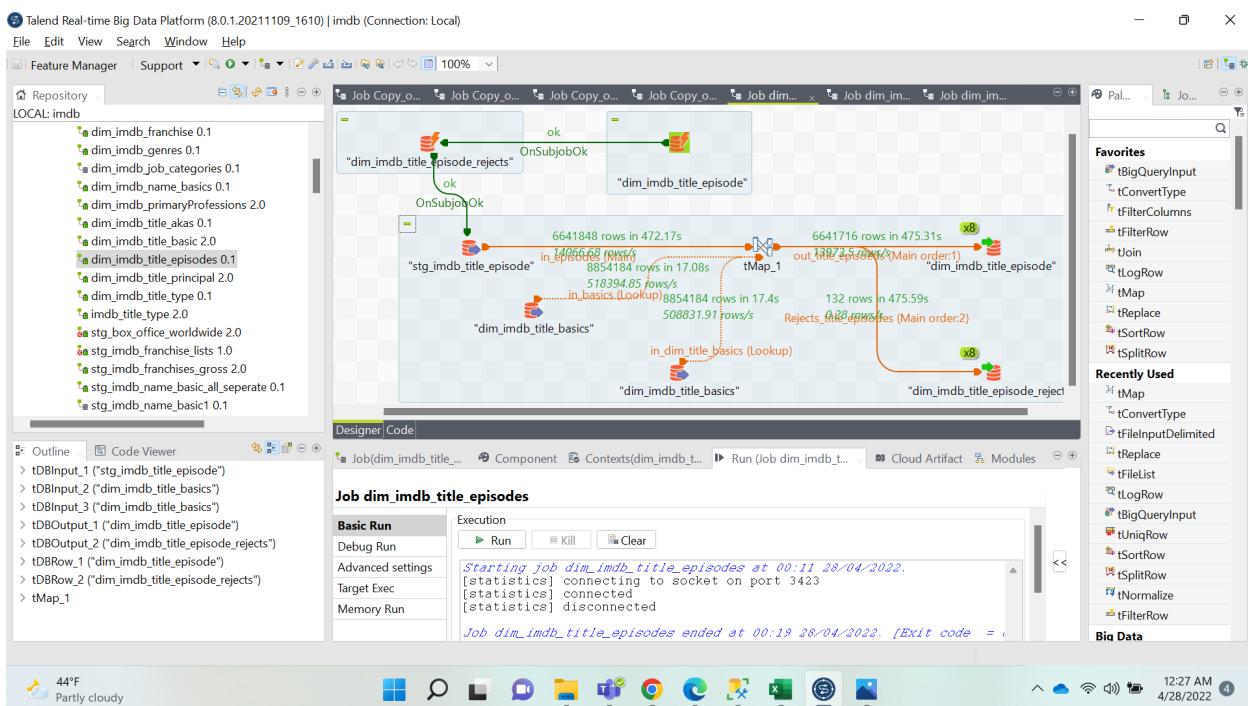
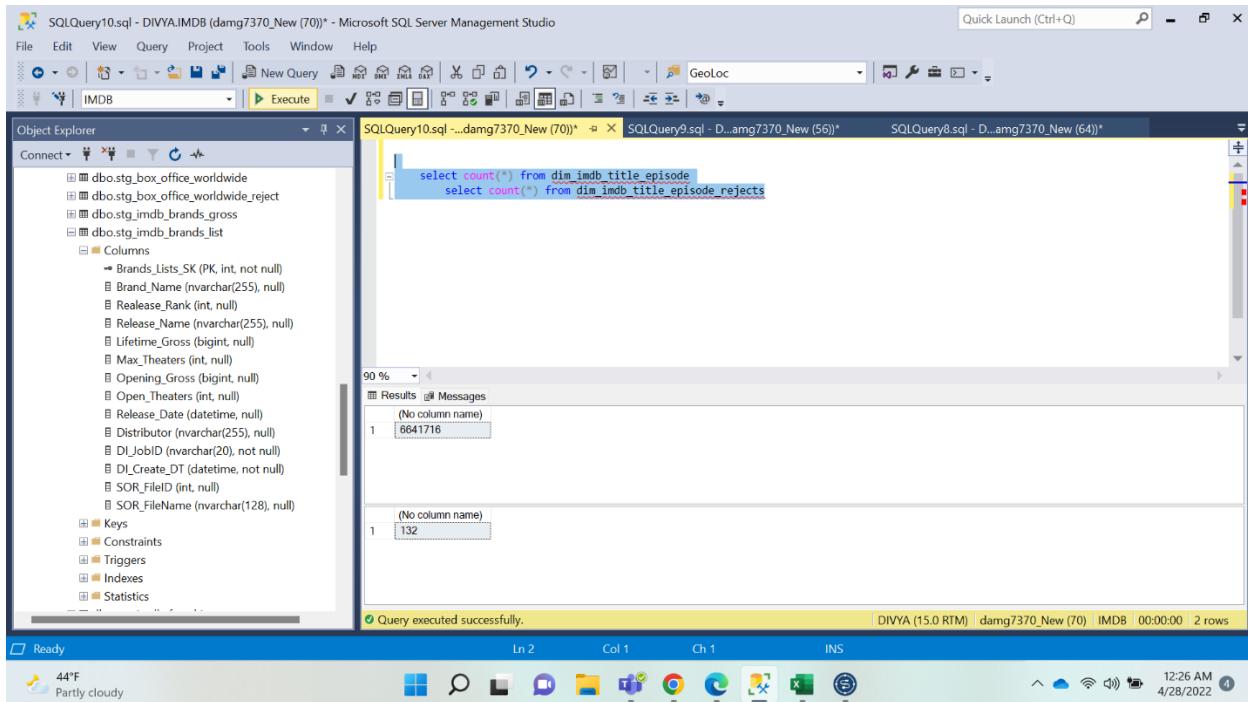
1 8854184

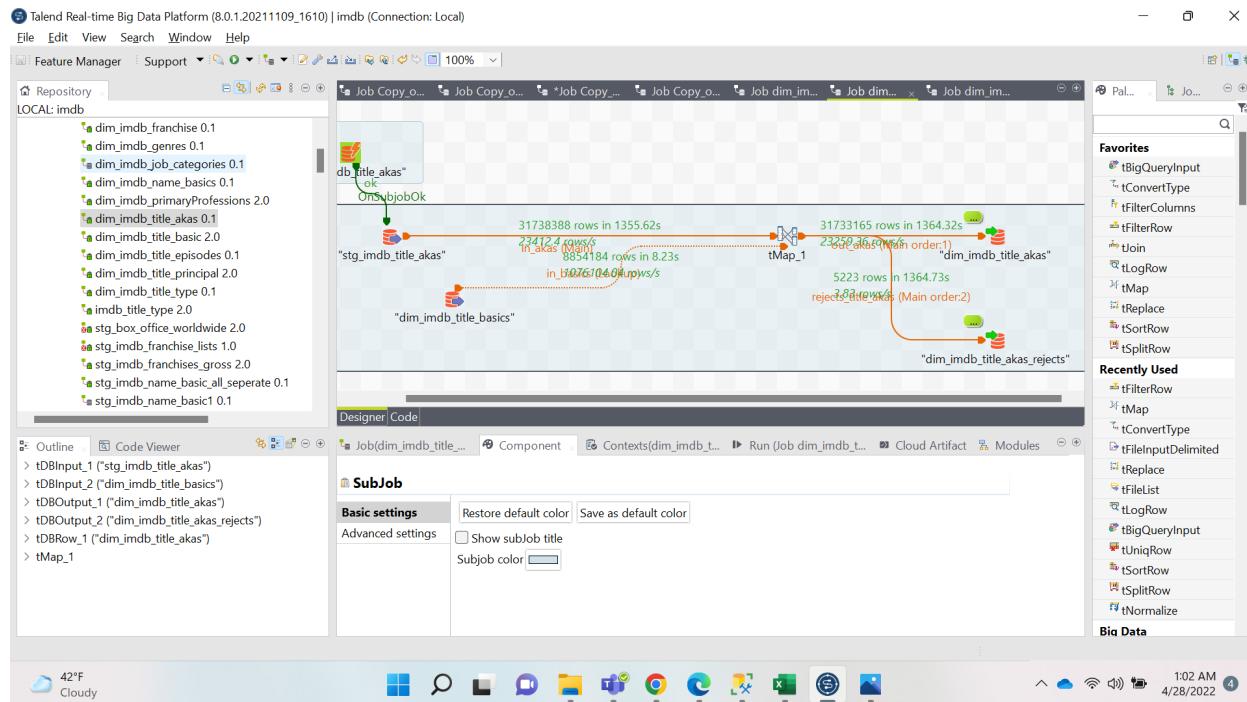
Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

11:22 PM 4/27/2022







SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Execute

Object Explorer

SQLQuery10.sql - damg7370_New (70)* SQLQuery9.sql - D...amg7370_New (56)* SQLQuery8.sql - D...amg7370_New (64)*

```
select count(*) from dim_imdb_title_akas
select count(*) from dim_imdb_title_akas_rejects
```

Results Messages

(No column name)

1 31733165

(No column name)

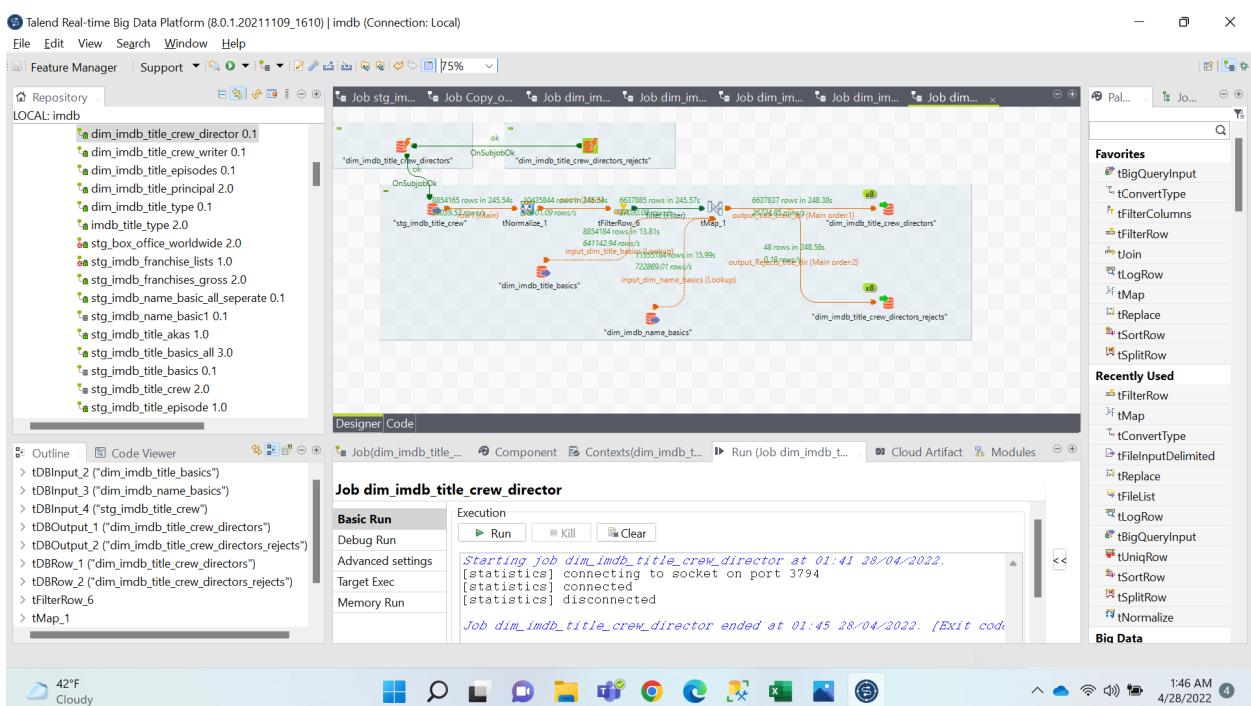
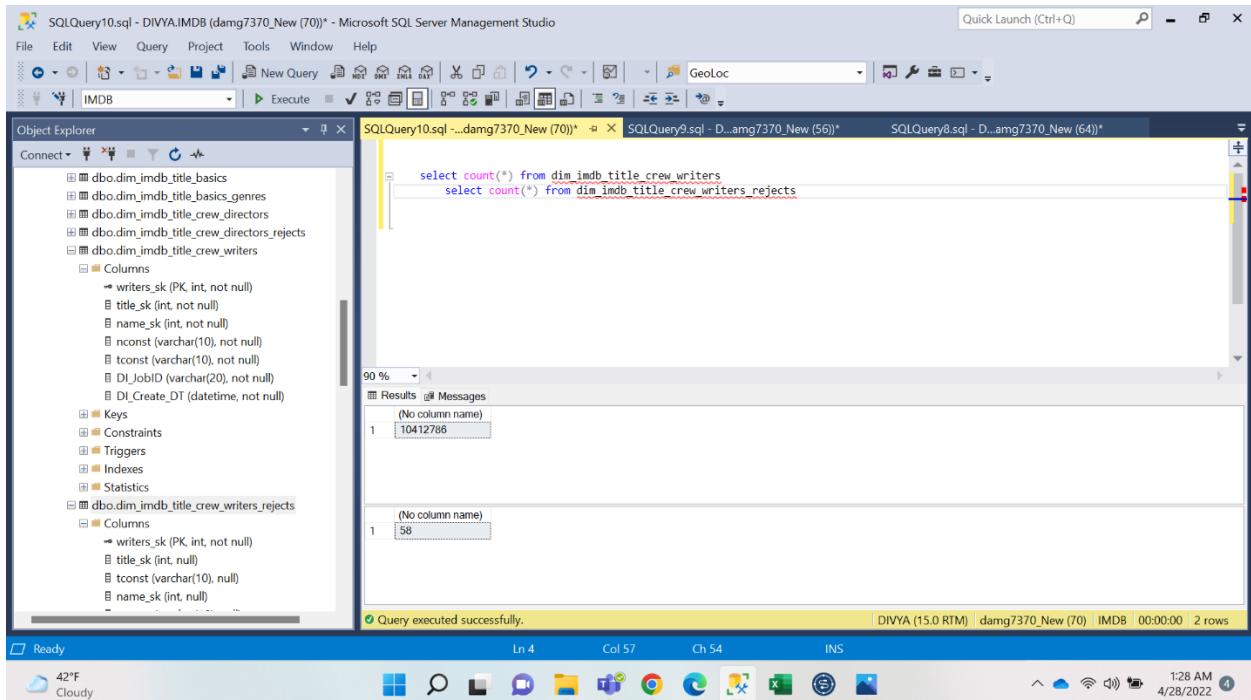
1 5223

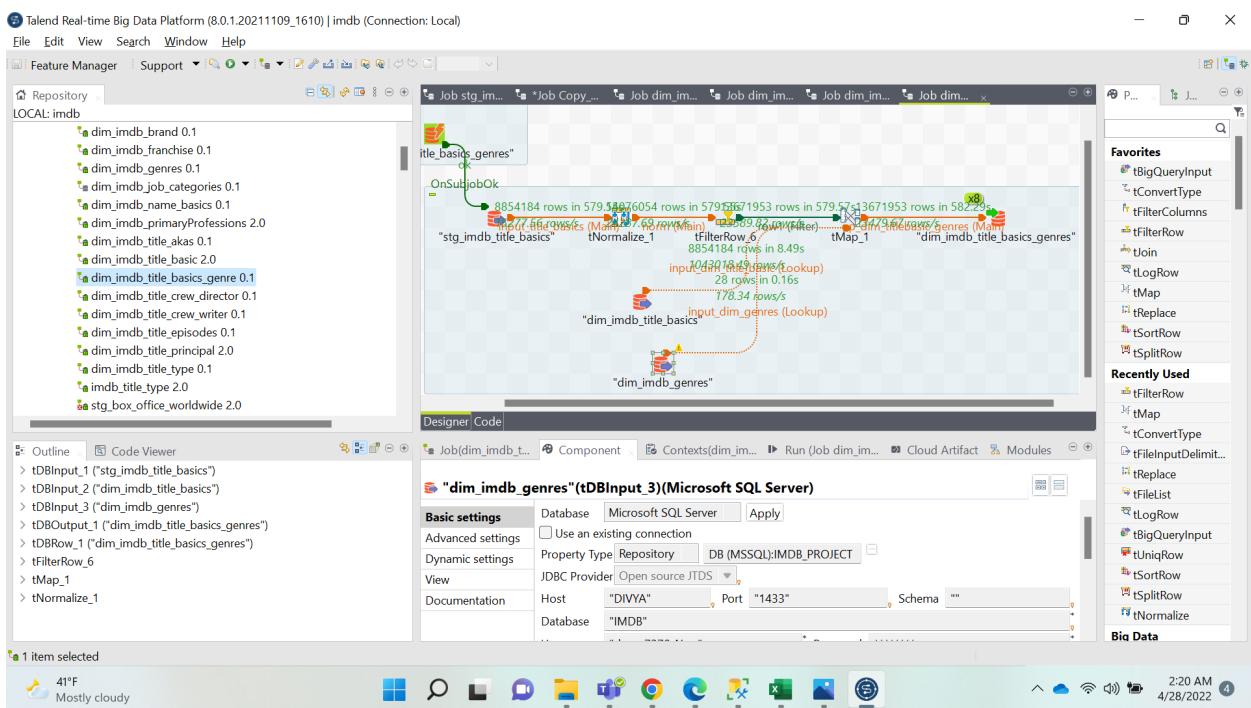
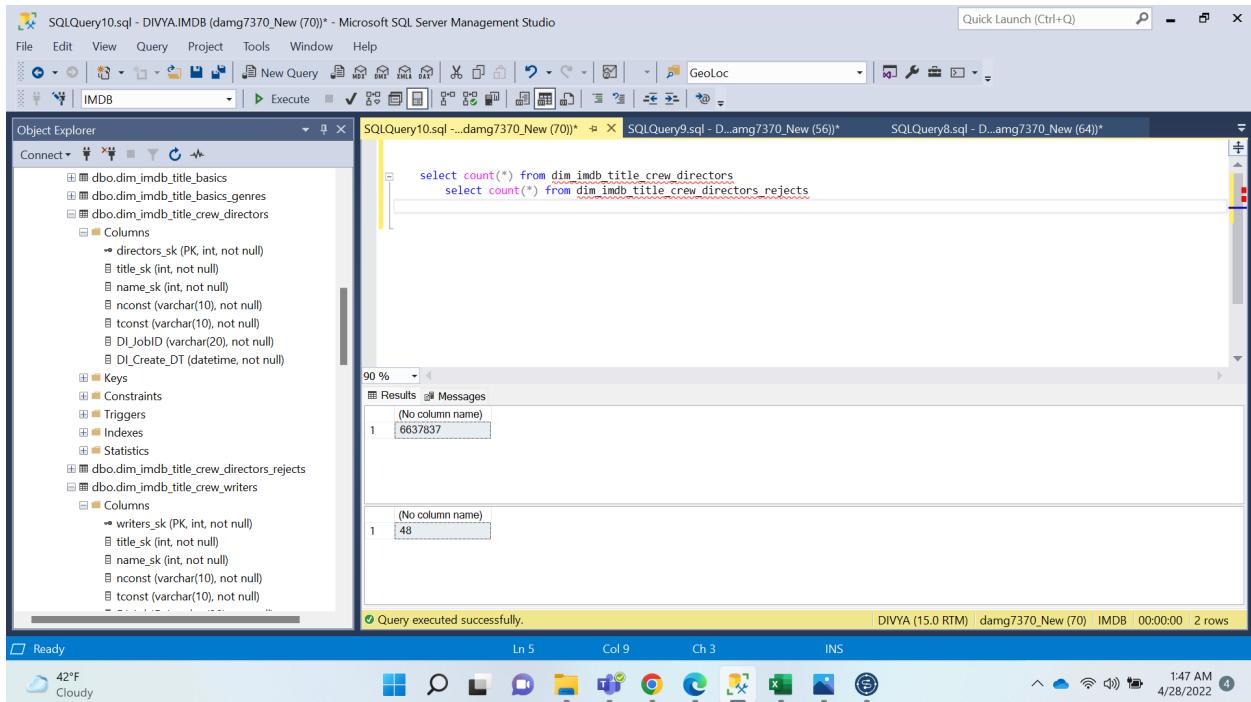
Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 2 rows

Ready Ln 5 Col 9 Ch 3 INS

42°F Cloudy 10:03 AM 4/28/2022





SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Quick Launch (Ctrl+Q)

IMDB

Object Explorer

SQLQuery10.sql - damg7370_New (70)* SQLQuery9.sql - D...amg7370_New (56)* SQLQuery8.sql - D...amg7370_New (64)*

select count(*) from dim_imdb_title_genres

Results Messages

(No column name)

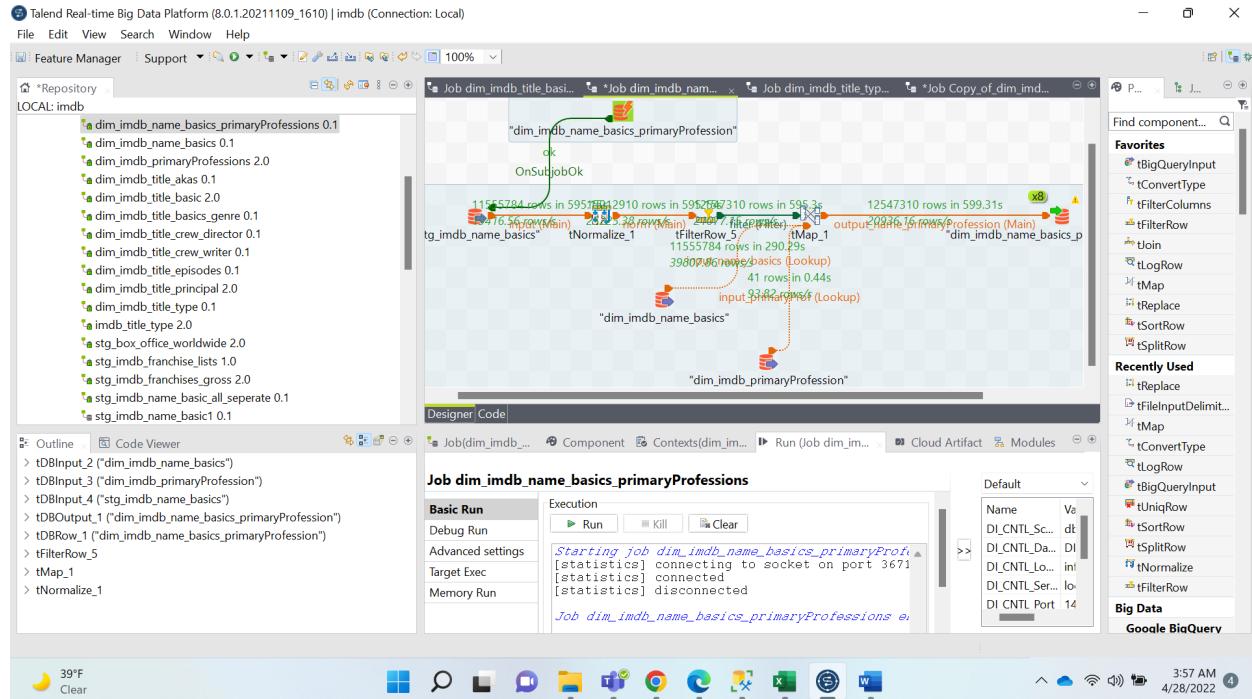
1 13671953

90 %

Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready 41°F Mostly cloudy 2:22 AM 4/28/2022



SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

Connect

dbo.dim_imdb_genres
dbo.dim_imdb_job_category
dbo.dim_imdb_name_basics
dbo.dim_imdb_name_basics_knownForTitles
dbo.dim_imdb_name_basics_rejects
dbo.dim_imdb_name_basics_primaryProfession
dbo.dim_imdb_primaryProfession
dbo.dim_imdb_title_akas
dbo.dim_imdb_title_akas_rejects
dbo.dim_imdb_title_basics
dbo.dim_imdb_title_basics_genres
dbo.dim_imdb_title_crew_directors
dbo.dim_imdb_title_crew_rejects
dbo.dim_imdb_title_crew_writer
dbo.dim_imdb_title_crew_writers_rejects
dbo.dim_imdb_title_episode
dbo.dim_imdb_title_episode_rejects
dbo.dim_imdb_title_principals
dbo.dim_imdb_title_principals_rejects
dbo.dim_imdb_title_type
dbo.fct_imdb_brands_gross
dbo.fct_imdb_brands_list
dbo.fct_imdb_franchises_gross
dbo.fct_imdb_franchises_list
dbo.fct_imdb_title_ratings

SQLQuery11.sql - ...damg7370_New (72)

```
select count(*) from dim_imdb_name_basics_primaryProfession
select top 100 from dim_imdb_title_akas
```

Results Messages

(No column name)

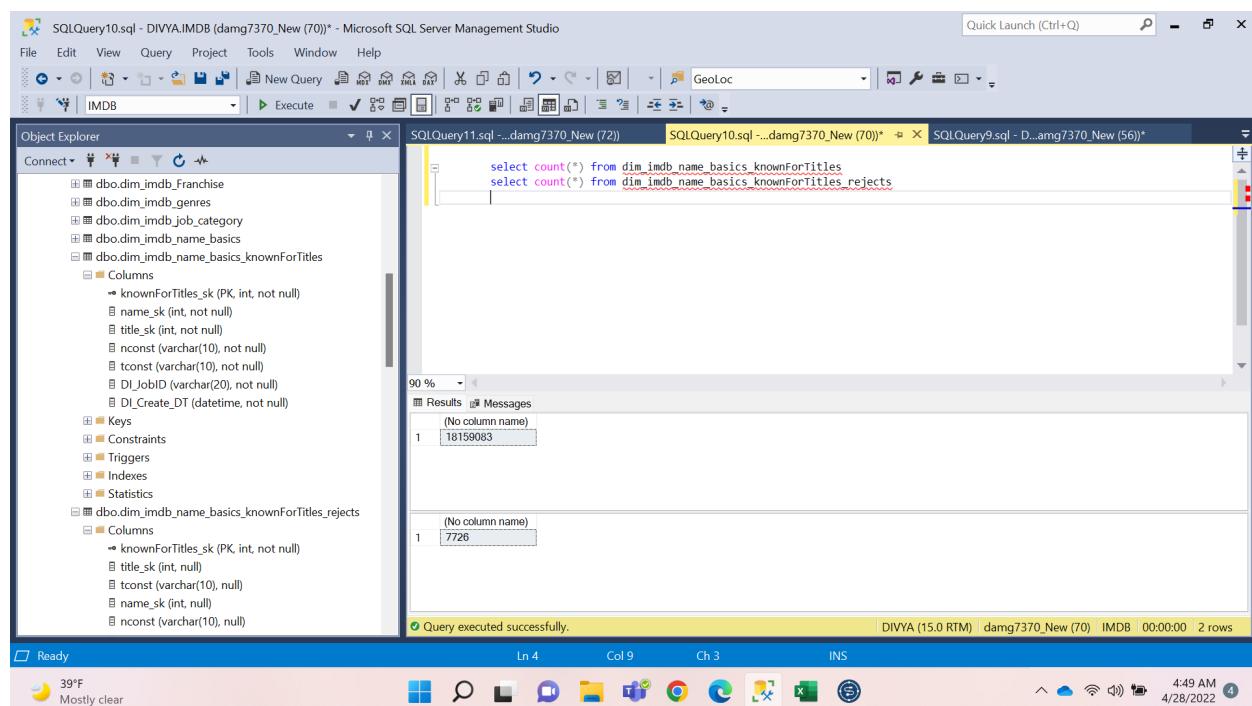
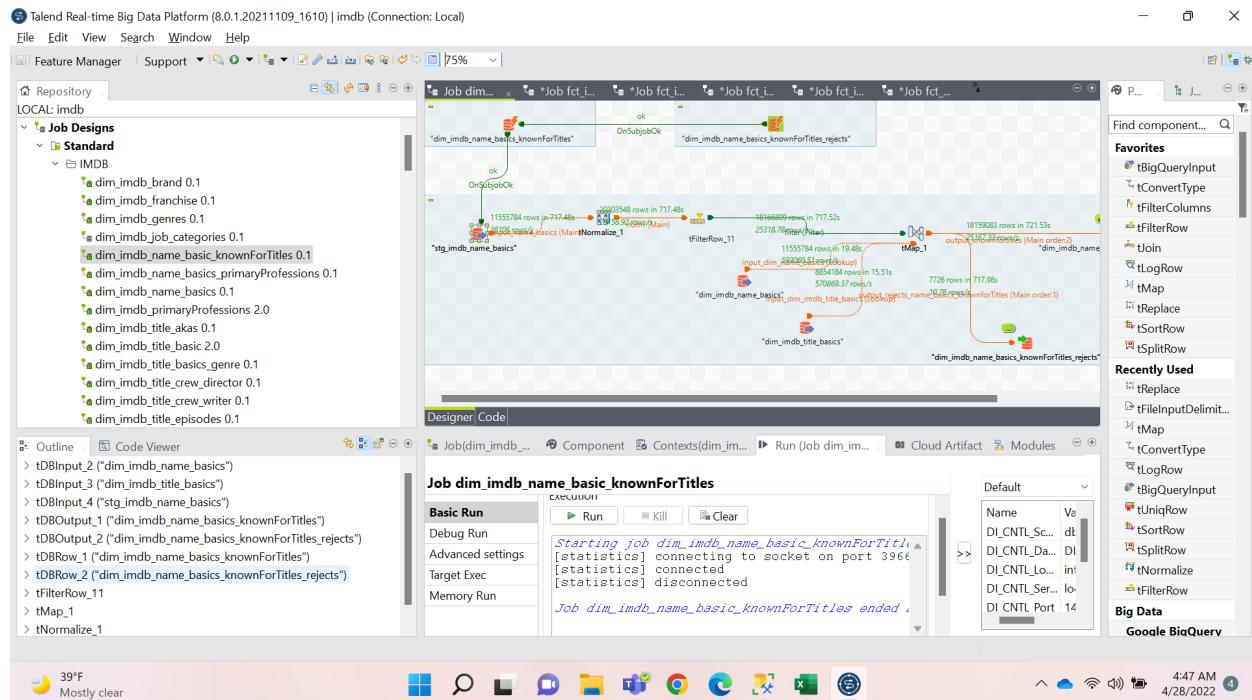
1 12547310

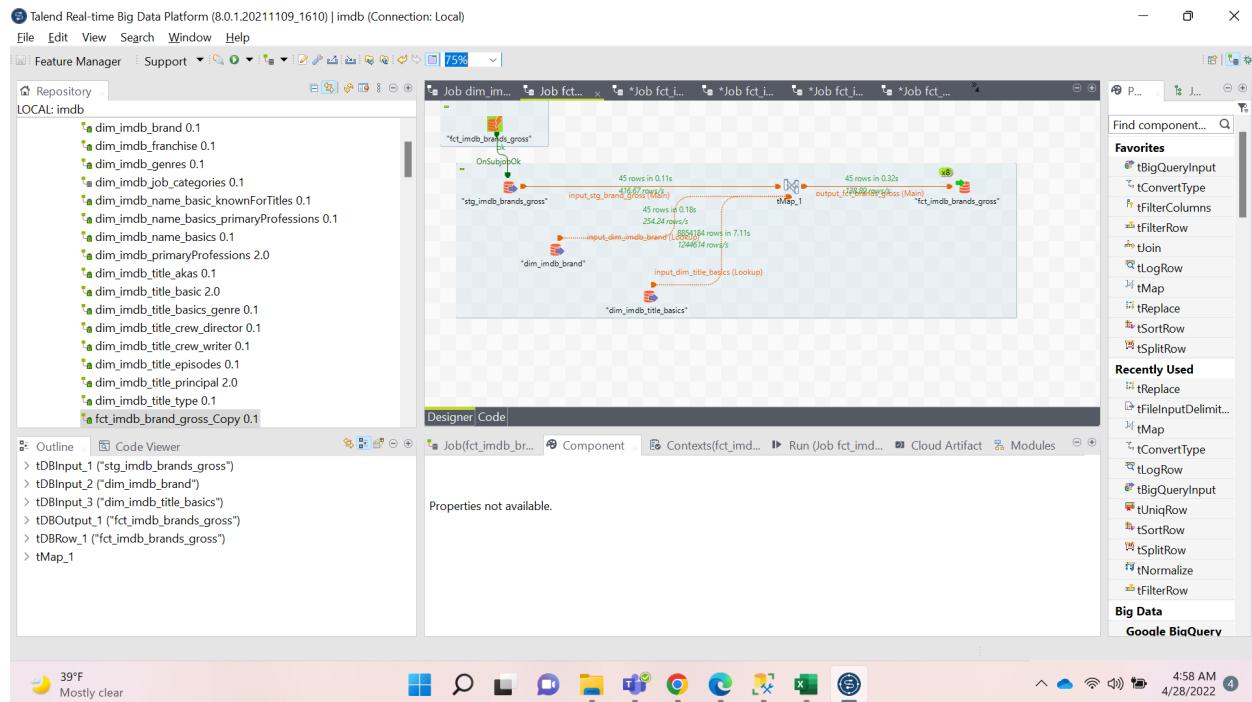
90 %

Query executed successfully. DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready Ln 2 Col 9 Ch 6 INS

39°F Clear 3:57 AM 4/28/2022





SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70)* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

```
select count(*) From fct_imdb_brands_gross
```

Results Messages

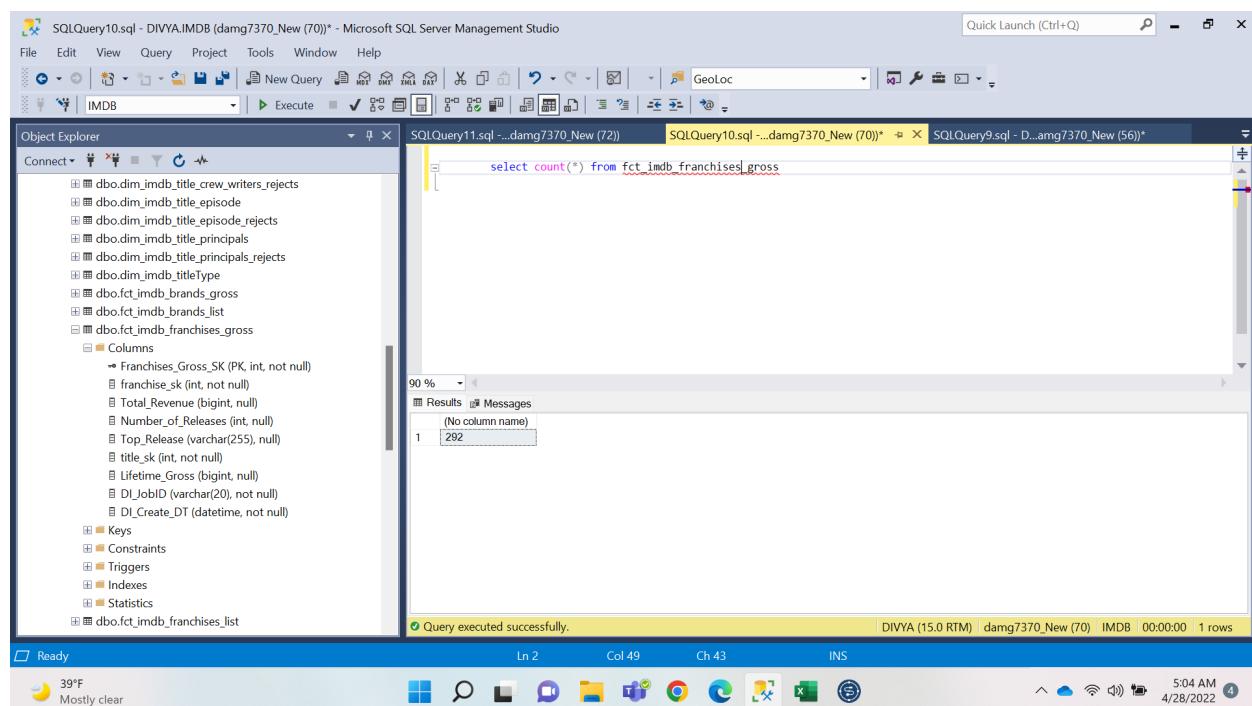
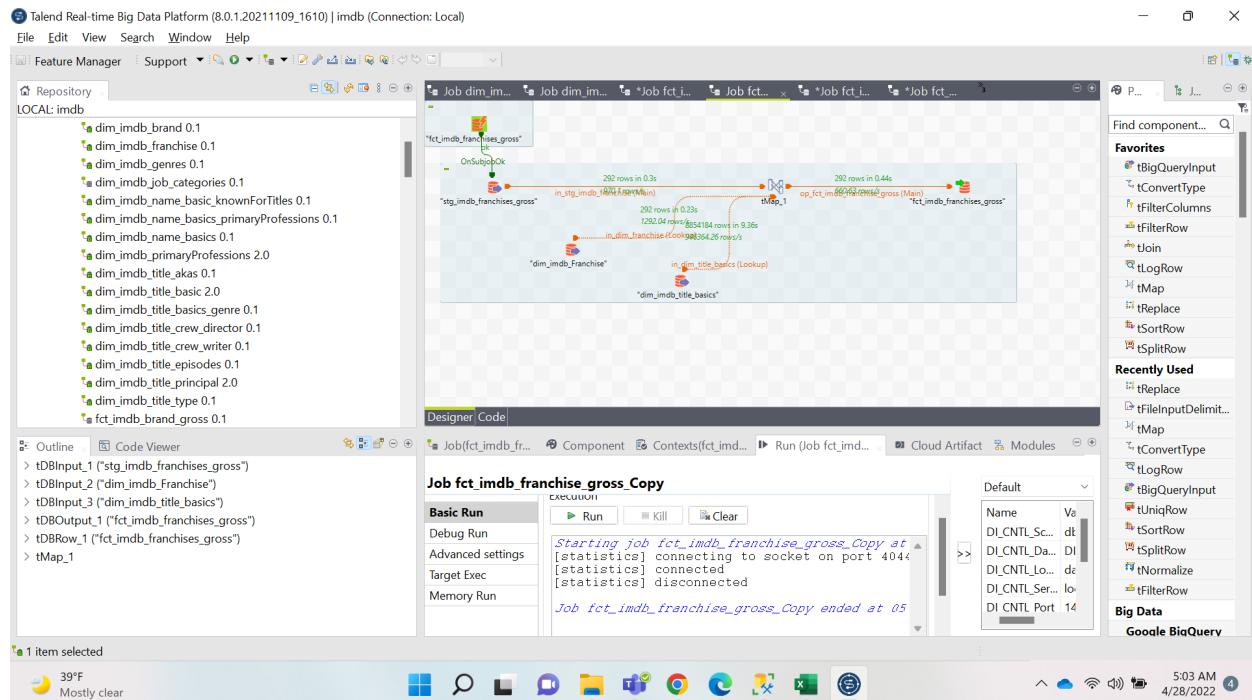
(No column name)	1	45
------------------	---	----

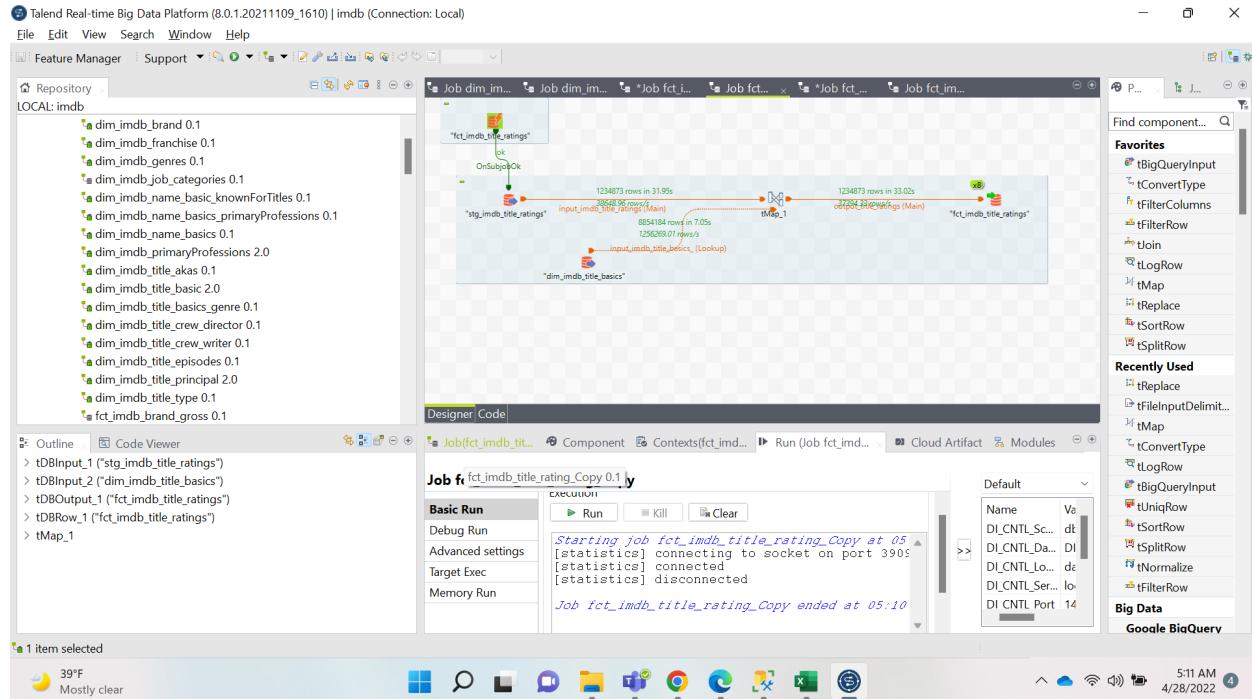
90 %

Query executed successfully. DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready Ln 2 Col 33 Ch 27 INS

39°F Mostly clear 4:58 AM 4/28/2022





SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

SQLQuery11.sql - ...damg7370_New (72)

SQLQuery10.sql - ...damg7370_New (70)*

SQLQuery9.sql - D...amg7370_New (56)*

select count(*) from fct_imdb_title_ratings

Results

(No column name)

1 1234873

Query executed successfully.

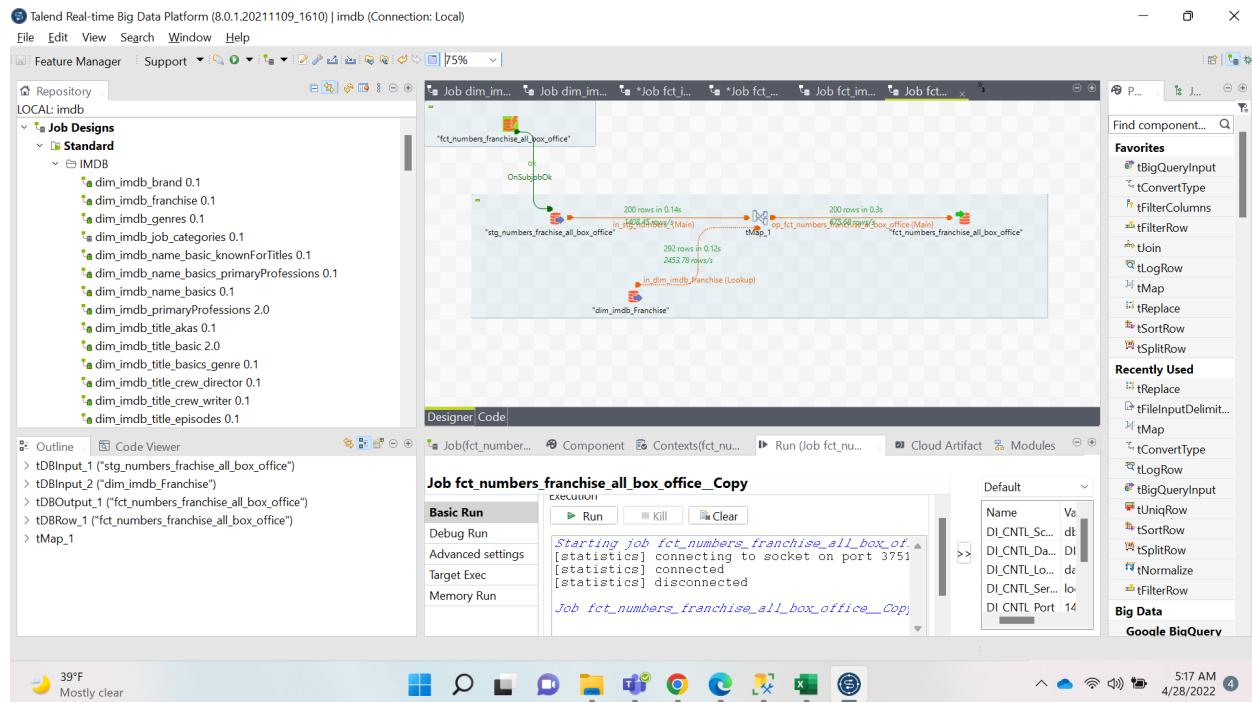
DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready

Ln 3 Col 9 Ch 3 INS

39°F Mostly clear

5:11 AM 4/28/2022



SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

New Query Execute

Object Explorer

IMDB

SQLQuery12.sql - damg7370_New (58) SQLQuery11.sql - damg7370_New (72) SQLQuery10.sql - damg7370_New (70)*

```
select count(*) from fct_numbers_franchise_all_box_office
```

Results Messages

(No column name)

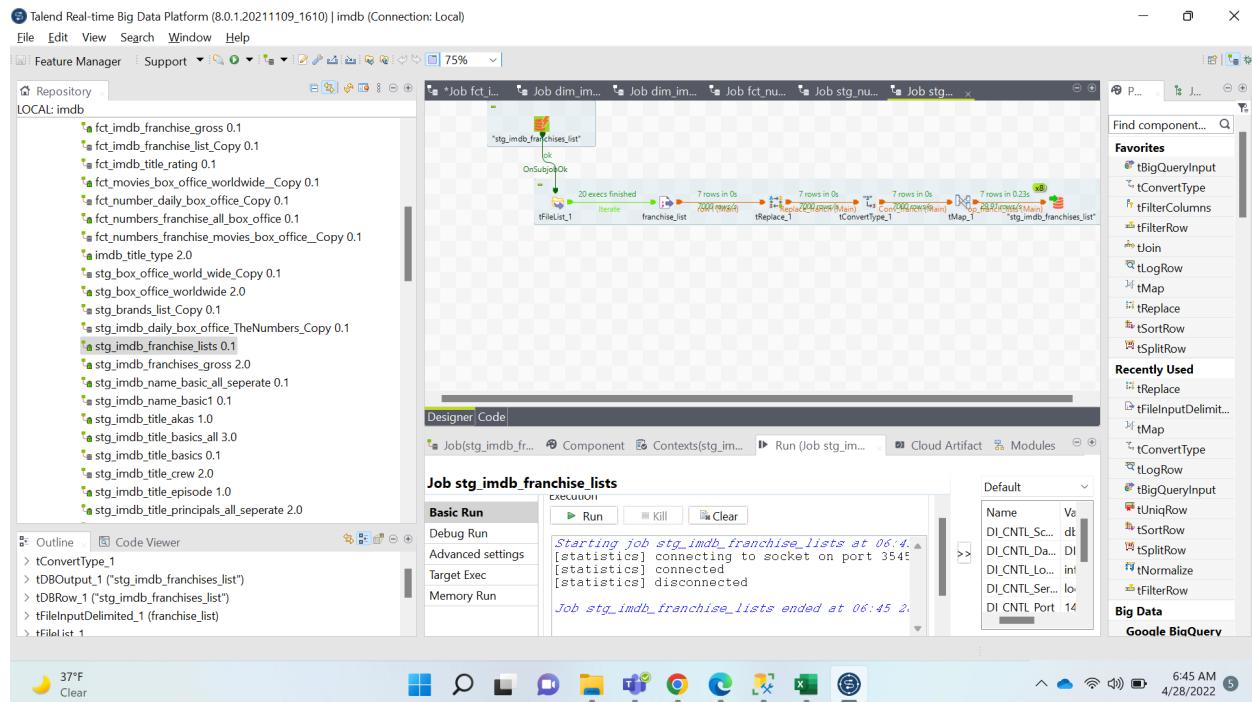
1 200

Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready Ln 17 Col 11 Ch 5 INS

39°F Mostly clear 5:29 AM 4/28/2022



SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Object Explorer

Connect

IMDB

SQLQuery12.sql - damg7370_New (58) SQLQuery11.sql - damg7370_New (72) SQLQuery10.sql - damg7370_New (70)*

select count(*) from stg_imdb_franchise_list

Results Messages

(No column name)

1 276

Query executed successfully.

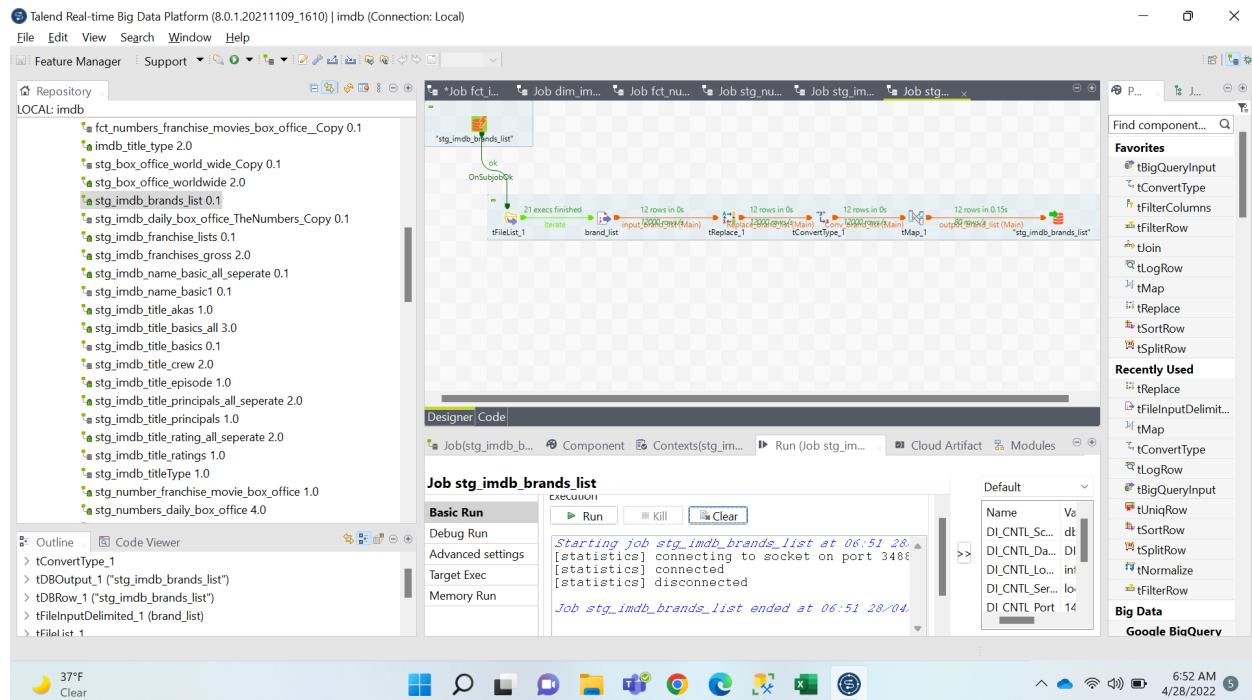
DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready

Ln 2 Col 9 Ch 3 INS

37°F Clear

6:46 AM 4/28/2022



SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Object Explorer

IMDB

SQLQuery12.sql - damg7370_New (58) SQLQuery11.sql - damg7370_New (72) SQLQuery10.sql - damg7370_New (70)*

select count(*) from stg_imdb_brands_list

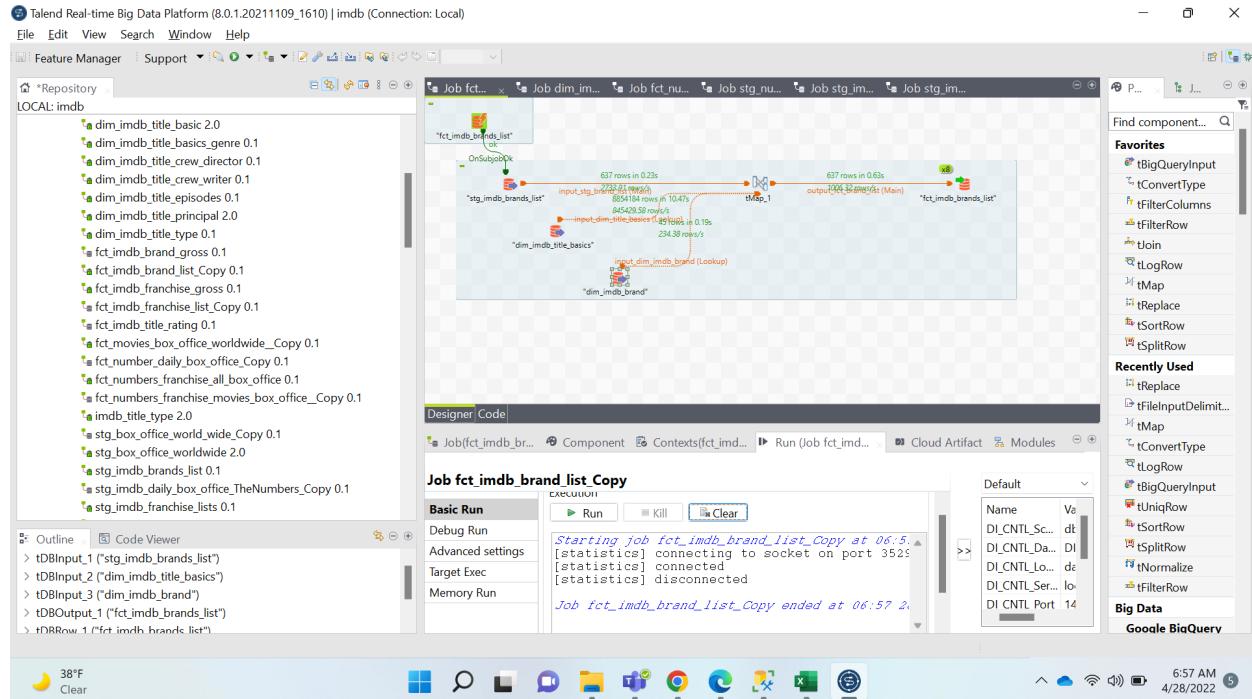
Results Messages

(No column name)

1 637

Query executed successfully. DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready Ln 2 Col 50 Ch 44 INS 6:53 AM 4/28/2022



SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

- dbo.dim_imdb_titleType
- dbo.dim_imdb_gross
- dbo.fct_imdb_brands_gross
- dbo.fct_imdb_franchises_gross
- dbo.fct_imdb_franchises_list
- dbo.fct_imdb_title_ratings
- dbo.fct_movies_box_office_worldwide
- dbo.fct_numbers_daily_box_office
- dbo.fct_numbers_franchise_all_box_office
- dbo.fct_numbers_franchise_movies_box_office
- dbo.stg_box_office_worldwide
- dbo.stg_box_office_worldwide_reject
- dbo.stg_imdb_brands_gross
- dbo.stg_imdb_brands_list
- dbo.stg_imdb_franchises_gross
- dbo.stg_imdb_franchises_list
- dbo.stg_imdb_name_basics
- dbo.stg_imdb_title_akas
- dbo.stg_imdb_title_basics
- dbo.stg_imdb_title_crew
- dbo.stg_imdb_title_episode
- dbo.stg_imdb_title_principals
- dbo.stg_imdb_title_ratings
- dbo.stg_imdb_title_type
- dbo.stg_numbers_daily_box_office

SQLQuery12.sql - damg7370_New (58) SQLQuery11.sql - damg7370_New (72) SQLQuery10.sql - damg7370_New (70)*

select count(*) From fct_imdb.brands.list

Results Messages

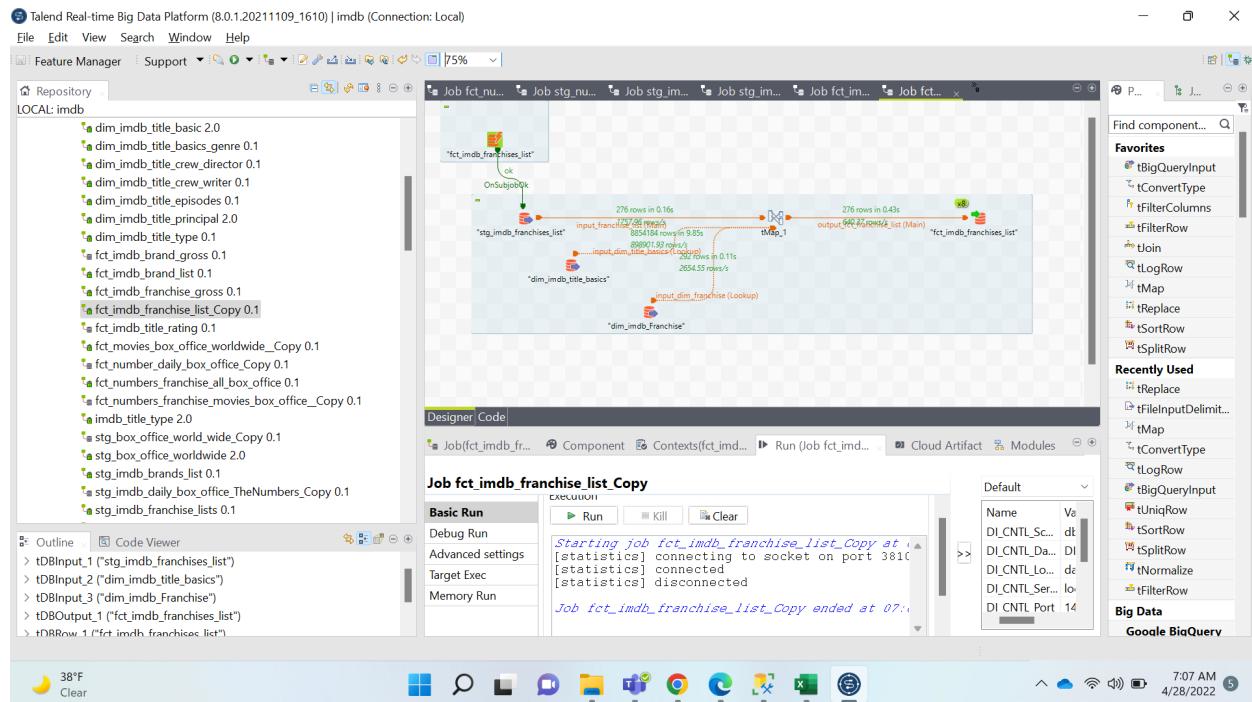
(No column name)

1 637

Query executed successfully. DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready Ln 2 Col 5 Ch 2 INS 6:58 AM 4/28/2022

38°F Clear



SQLQuery10.sql - DIVYA.IMDB (damg7370_New (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

IMDB

Object Explorer

SQLQuery12.sql - damg7370_New (58) SQLQuery11.sql - damg7370_New (72) SQLQuery10.sql - damg7370_New (70)*

select count(*) from ft_imdb_franchise_list

Results Messages

(No column name)

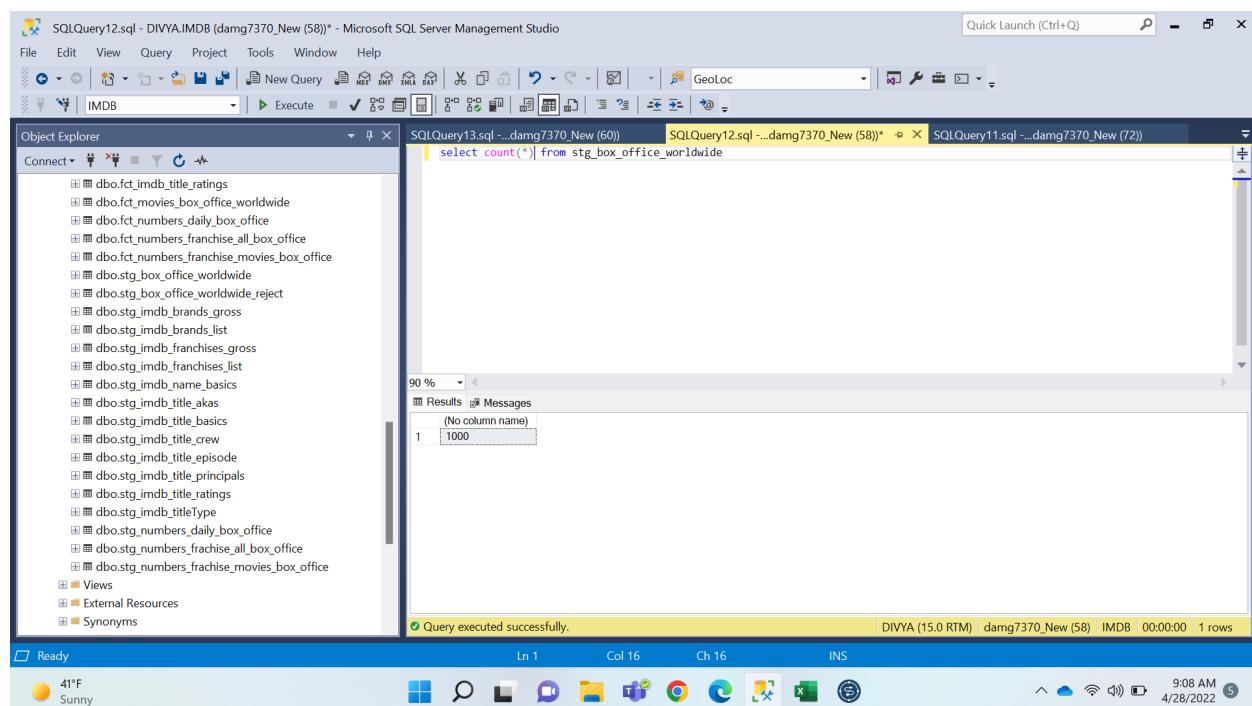
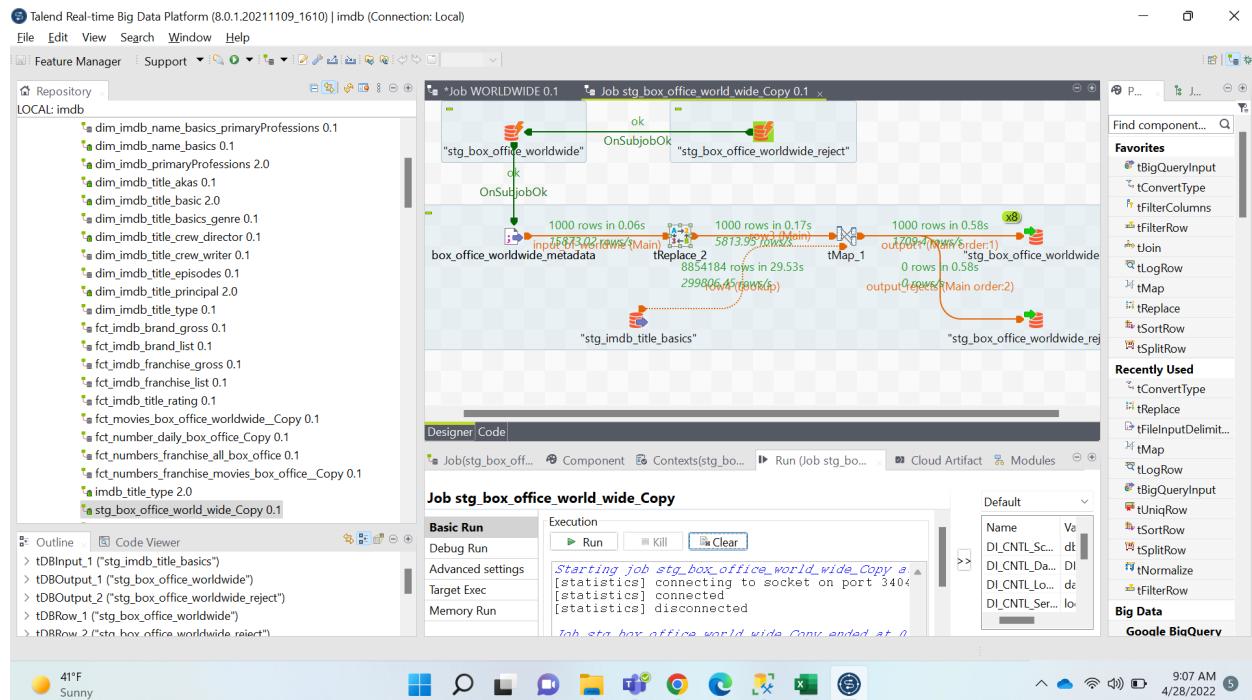
1 276

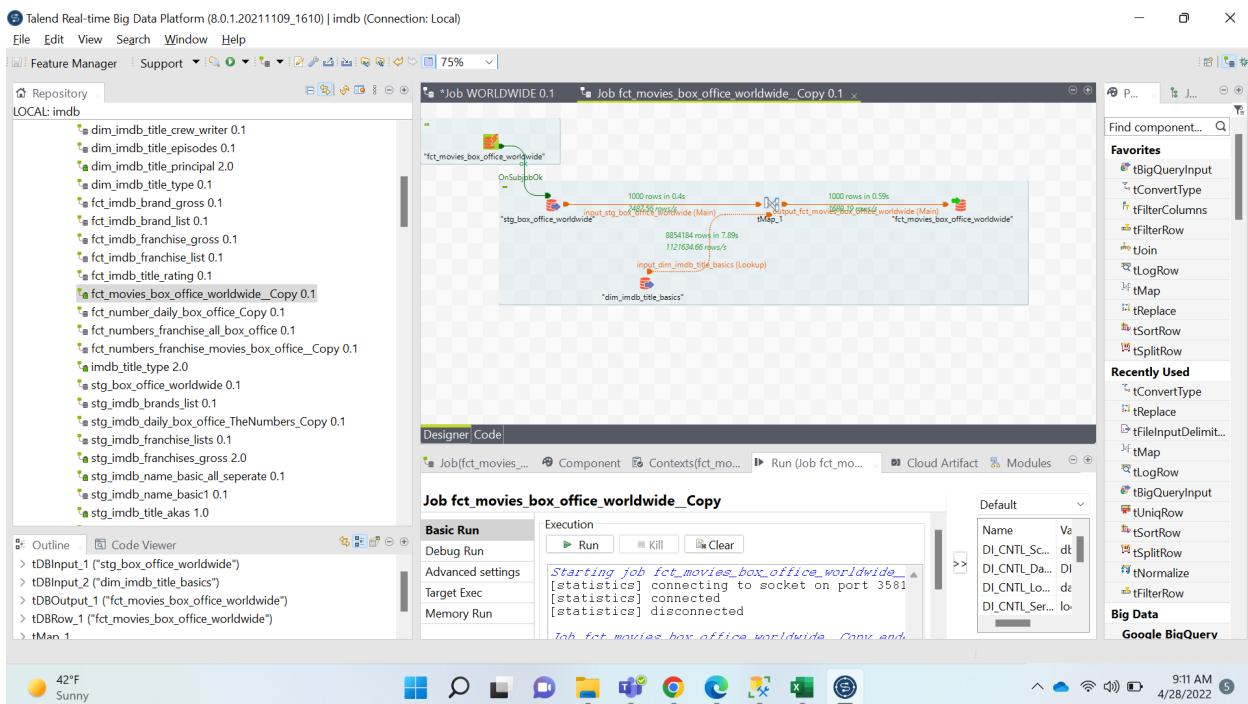
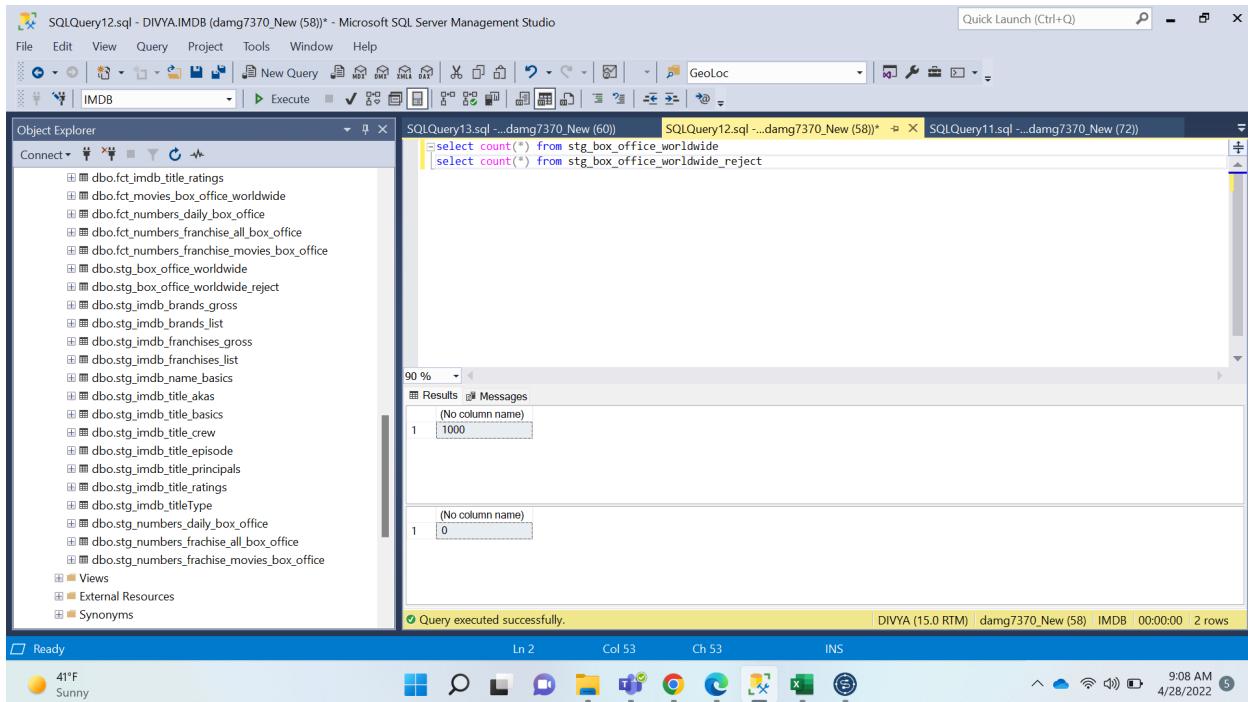
Query executed successfully.

DIVYA (15.0 RTM) | damg7370_New (70) | IMDB | 00:00:00 | 1 rows

Ready Ln 2 Col 9 Ch 3 INS 7:08 AM 4/28/2022

38°F Clear





SQLQuery12.sql - DIVYA.IMDB (damg7370_New (58)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Quick Launch (Ctrl+Q)

IMDB

Object Explorer

SQLQuery14.sql - damg7370_New (57) SQLQuery13.sql - damg7370_New (60) SQLQuery12.sql - damg7370_New (58)*

```
select count(*) from fact_movies_box_office_worldwide
```

Results (No column name)

1	1000
---	------

90 %

Query executed successfully. DIVYA (15.0 RTM) | damg7370_New (58) | IMDB | 00:00:00 | 1 rows

Ready 9:13 AM 4/28/2022

42°F Sunny