Data Warehousing & Business Intelligence (IT)
3rd Year,1st Semester

Assignment 1

Submitted to
Sri Lanka Institute of Information Technology

IT20100698
BRITTO T.A
Weekday Batch

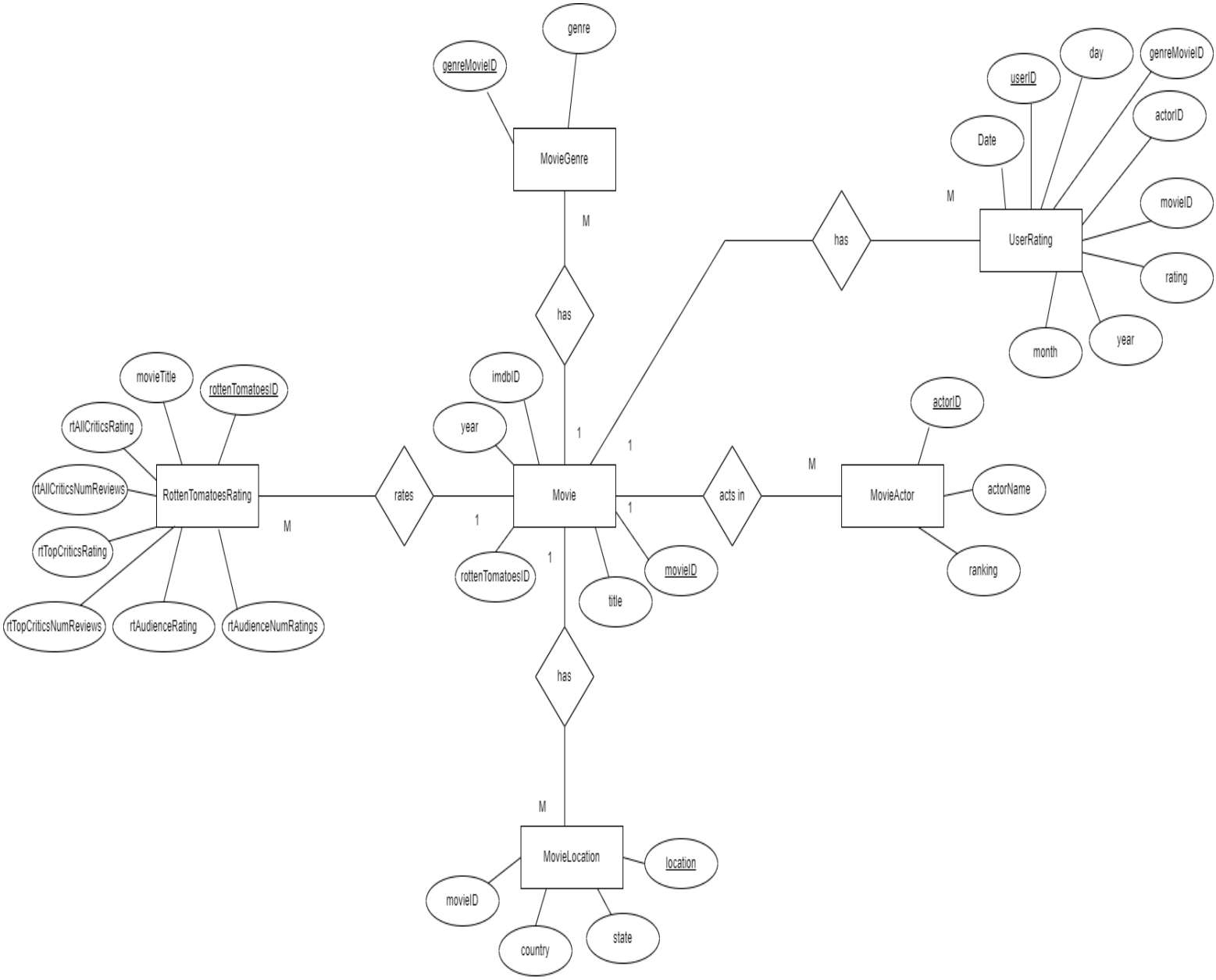
STEP 01: Data set selection

This dataset is an extension of MovieLens10M dataset, published by GroupLeans research group. It connects the movies in the MovieLens dataset to their web pages on the Internet Movie Database (IMDb) and the Rotten Tomatoes movie review systems. Only those users with both rating and tagging information were maintained from the original dataset. The dataset is released in the framework of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011).
According to the assignment principles, the data set was started with enough data.
In my data set which I have selected there are transactional data, and it has data of more than a year.
I can receive two sources of data from the data set that I selected (CSV and a text file).
And the data set is sufficient to build a data warehouse. I was able to perform ETL functions with this data set.
And I can correctly identify hierarchies, dimensions, and aggregates in this data collection. I also realized that with this data collection, I will be capable of creating correct reports.
So, I selected this MovieLens + IMDb/Rotten Tomatoes details data set for my assignment.

Data Set Link:

https://grouplens.org/datasets/hetrec-2011/

The ER Diagram of the Data set



STEP 02: Preparation of Data Sources

There were ten csv files available in the data set. They are movie_actors.csv, movie_countries.csv, movie_directors.csv, movie_genres.csv, movie_locations.csv, movie_tags.csv, movies.csv, tags.csv, user_ratedmovies.csv and user_taggedmovies.csv. Among these csv files movie_countries.csv, movie_directors.csv, movie_tags.csv, tags.csv and user_taggedmovies.csv were not taken to create tables.

I created separate two tables for my source dataset by dividing movies.csv. They are movies table and rottenTomatoesRating table. And I decided to convert movies data csv into text file format as movies.txt. I renamed rottenTomatoesRating table as rotten_tomatoes.csv.

Data Source Name	Data Source Type	Description
Movies	TEXT	This file contains basic information about the movies of the database
Rotten Tomatoes	CSV	This file contains information about the Rotten Tomatoes ratings of the movies
Movie Locations	CSV	This file contains filming locations of the movies
Movie Genres	CSV	This file contains the genres of the movies.
User Rated Movies	CSV	This file contains the information about how users have rated the movies, and the date that a specific rating has done
Movie Actors	CSV	This file contains the main actors and actresses of the movies. A ranking is given to the actors of each movie according to the order in which they appear on the movie IMDb cast web page.

Then I imported those csv files into my newly created HetRecMovies_SourceDB. After I imported my csv files into sourceDB, I created Data warehouse named HetRecMovies_DW and created my dimension tables and fact tables inside the data warehouse.
The tables which I have imported to my database source is shown below.

actorID	actorName	ranking
1	Annie Potts	10
2	Bill Farmer	20
3	Don Rickles	3
4	Erik von Detten	13
5	Greg Berg	17
6	Jack Angel	6
7	Jan Rabson	19
8	Jim Varney	4
9	Joan Cusack	24
10	Joe Ranft	16
11	John Morris	23
12	John Ratzenberger	12
13	Kendall Cunningham	21
14	Laurie Metcalf	8
15	Patrick Pinney	9
16	Penn Jillette	15
17	Philip Proctor	11
18	R. Lee Ermey	14
19	Sarah Freeman	22
20	Scott McAfee	18
21	Sherry Lynn	7
22	Tim Allen	2
23	Tom Hanks	1
24	Wallace Shawn	5
25	Peter Bryant	16
26	Adam Hann-Byrd	13
27	Bebe Neuwirth	5

movie_actors

movie_actors.csv

genreMovieID	genre
1	Adventure
2	Animation
3	Children
4	Comedy
5	Fantasy
6	Adventure
7	Children
8	Fantasy
9	Comedy
10	Romance
11	Comedy
12	Drama
13	Romance
14	Comedy
15	Action
16	Crime
17	Thriller
18	Comedy
19	Romance
20	Adventure
21	Children
22	Action
23	Action
24	Adventure
25	Thriller
26	Comedy
27	Drama

movie_genres

movie_genres.csv

movieID	country	state	location
1	Canada	British Columbia	
2	Canada	British Columbia	Delta
3	Canada	British Columbia	Maple Ridge
4	Canada	British Columbia	Vancouver
5	USA	Maine	
6	USA	Maine	Kennebunk
7	USA	Maine	North Berwick
8	USA	New Hampshire	
9	USA	New Hampshire	Keene
10	USA	New Hampshire	Keene
11	USA	New Hampshire	Swanzey
12	USA	California	Burbank
13	USA	Minnesota	Center City
14	USA	Minnesota	Chanhassen
15	USA	Minnesota	Chanhassen
16	USA	Minnesota	Faribault
17	USA	Minnesota	Red Wing
18	USA	Minnesota	Rockford
19	USA	Minnesota	South St. Paul
20	USA	Minnesota	St. Paul
21	USA	Minnesota	Stillwater
22	USA	Arizona	Chandler
23	USA	Arizona	Fountain Hills
24	USA	Arizona	Paradise Valley
25	USA	Arizona	Phoenix
26	USA	Utah	Monument Valley
27	USA	California	Altadena

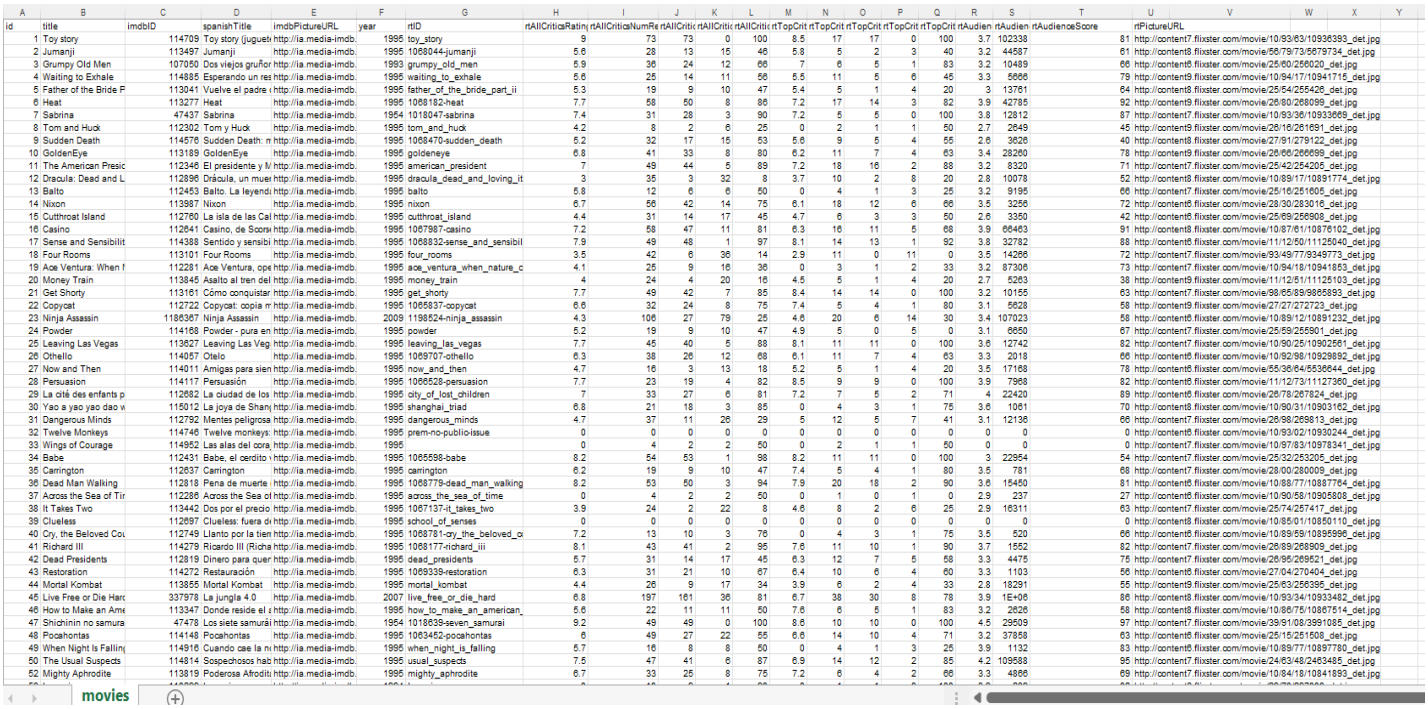
movie_locations

movie_locations.csv

userID	movieID	rating	year	month	day	Date	actorID	genreMovieID
1	3	1	2006	10	29	10/29/2006	1	1
2	32	4.5	2006	10	29	10/29/2006	2	2
3	110	4	2006	10	29	10/29/2006	3	3
4	160	2	2006	10	29	10/29/2006	4	4
5	163	4	2006	10	29	10/29/2006	5	5
6	165	4.5	2006	10	29	10/29/2006	6	6
7	173	3.5	2006	10	29	10/29/2006	7	7
8	296	5	2006	10	29	10/29/2006	8	8
9	353	3.5	2006	10	29	10/29/2006	9	9
10	420	2	2006	10	29	10/29/2006	10	10
11	589	4	2006	10	29	10/29/2006	11	11
12	653	3	2006	10	29	10/29/2006	12	12
13	832	4.5	2006	10	29	10/29/2006	13	13
14	920	0.5	2006	10	29	10/29/2006	14	14
15	996	4.5	2006	10	29	10/29/2006	15	15
16	1036	4	2006	10	29	10/29/2006	16	16
17	1127	3.5	2006	10	29	10/29/2006	17	17
18	1215	4.5	2006	10	29	10/29/2006	18	18
19	1233	4	2006	10	29	10/29/2006	19	19
20	1304	2.5	2006	10	29	10/29/2006	20	20
21	1370	4	2006	10	29	10/29/2006	21	21
22	1374	4	2006	10	29	10/29/2006	22	22
23	1485	4	2006	10	29	10/29/2006	23	23
24	1527	4.5	2006	10	29	10/29/2006	24	24
25	1917	2.5	2006	10	29	10/29/2006	25	25
26	2011	2	2006	10	29	10/29/2006	26	26
27	2054	1.5	2006	10	29	10/29/2006	27	27

user_rated_movies

user_rated_movies.csv



id	title	imdbID	spanishTitle	imdbPictureURL	year	rtID	rtAllCriticsRating	rtAllCriticsNumRe	rtAllCritic	rtAllCritic	rtTopCrit	rtTopCrit	rtTopCrit	rtTopCrit	rtTopCrit	rtAudien	rtAudienc	rtAudienceScore	rtPictureURL		
1	Toy story	114709	Toy story (juguet	http://ia.media-imdb.	1995	toy_story	9	73	73	0	100	8.5	17	17	0	100	3.7	102338	81	http://content7.flixster.com/movie/10/93/63/10936393_det.jpg	
2	Jumanji	113497	Jumanji	http://ia.media-imdb.	1995	1068044-jumanji	5.6	28	13	15	46	5.8	5	2	3	40	3.2	44587	61	http://content8.flixster.com/movie/56/79/73/5679734_det.jpg	
3	Grumpy Old Men	107050	Dos viejos gruñor	http://ia.media-imdb.	1993	grumpy_old_men	5.9	36	24	12	66	7	6	5	1	83	3.2	10489	66	http://content6.flixster.com/movie/25/60/256020_det.jpg	
4	Waiting to Exhale	114885	Esperando un res	http://ia.media-imdb.	1995	waiting_to_exhale	5.6	25	14	11	56	5.5	11	5	6	45	3.3	5666	79	http://content9.flixster.com/movie/10/94/17/10941715_det.jpg	
5	Father of the Bride P	113041	Vuelve el padre (	http://ia.media-imdb.	1995	father_of_the_bride_part_ii	5.3	19	9	10	47	5.4	5	1	4	20	3	13761	64	http://content8.flixster.com/movie/25/54/255428_det.jpg	
6	Heat	113277	Heat	http://ia.media-imdb.	1995	1068182-heat	7.7	58	50	8	86	7.2	17	14	3	82	3.9	42785	92	http://content9.flixster.com/movie/26/80/268099_det.jpg	
7	Sabrina	47437	Sabrina	http://ia.media-imdb.	1954	1018047-sabrina	7.4	31	28	3	90	7.2	5	5	0	100	3.8	12812	87	http://content7.flixster.com/movie/10/93/36/10933669_det.jpg	
8	Tom and Huck	112302	Tom y Huck	http://ia.media-imdb.	1995	tom_and_huck	4.2	8	2	6	25	0	2	1	1	50	2.7	2649	45	http://content9.flixster.com/movie/26/16/261691_det.jpg	
9	Sudden Death	114576	Sudden Death: n	http://ia.media-imdb.	1995	1068470-sudden_death	5.2	32	17	15	53	5.6	9	5	4	55	2.6	3626	40	http://content8.flixster.com/movie/27/91/279122_det.jpg	
10	GoldenEye	113189	GoldenEye	http://ia.media-imdb.	1995	goldeneye	6.8	41	33	8	80	6.2	11	7	4	63	3.4	28260	78	http://content9.flixster.com/movie/26/66/266699_det.jpg	
11	The American Presic	112346	El presidente y M	http://ia.media-imdb.	1995	american_president	7	49	44	5	89	7.2	18	16	2	88	3.2	8320	71	http://content7.flixster.com/movie/25/42/254205_det.jpg	
12	Dracula: Dead and L	112896	Drácula, un muer	http://ia.media-imdb.	1995	dracula_dead_and_loving_it	3	35	3	32	8	3.7	10	2	8	20	2.8	10078	52	http://content8.flixster.com/movie/10/89/17/10891774_det.jpg	
13	Balto	112453	Balto. La leyenda	http://ia.media-imdb.	1995	balto	5.8	12	6	6	50	0	4	1	3	25	3.2	9195	66	http://content7.flixster.com/movie/25/16/251605_det.jpg	
14	Nixon	113987	Nixon	http://ia.media-imdb.	1995	nixon	6.7	56	42	14	75	6.1	18	12	6	66	3.5	3256	72	http://content6.flixster.com/movie/28/30/283016_det.jpg	
15	Cutthroat Island	112760	La isla de las Cal	http://ia.media-imdb.	1995	cutthroat_island	4.4	31	14	17	45	4.7	6	3	3	50	2.6	3350	42	http://content6.flixster.com/movie/25/69/256908_det.jpg	
16	Casino	112641	Casino, de Scors	http://ia.media-imdb.	1995	1067987-casino	7.2	58	47	11	81	6.3	16	11	5	68	3.9	66463	91	http://content8.flixster.com/movie/10/87/61/10876102_det.jpg	
17	Sense and Sensibilit	114388	Sentido y sensibi	http://ia.media-imdb.	1995	1068832-sense_and_sensibil	7.9	49	48	1	97	8.1	14	13	1	92	3.8	32782	88	http://content6.flixster.com/movie/11/12/50/11125040_det.jpg	
18	Four Rooms	113101	Four Rooms	http://ia.media-imdb.	1995	four_rooms	3.5	42	6	36	14	2.9	11	0	11	0	3.5	14266	72	http://content7.flixster.com/movie/93/49/77/9349773_det.jpg	
19	Ace Ventura: When I	112281	Ace Ventura, ope	http://ia.media-imdb.	1995	ace_ventura_when_nature_c	4.1	25	9	16	36	0	3	1	2	33	3.2	87306	73	http://content7.flixster.com/movie/10/94/18/10941853_det.jpg	
20	Money Train	113845	Asalto al tren del	http://ia.media-imdb.	1995	money_train	4	24	4	20	16	4.5	5	1	4	20	2.7	5263	38	http://content9.flixster.com/movie/11/12/51/11125103_det.jpg	
21	Get Shorty	113161	Cómo conquistar	http://ia.media-imdb.	1995	get_shorty	7.7	49	42	7	85	8.4	14	14	0	100	3.2	10155	63	http://content7.flixster.com/movie/98/65/89/9865893_det.jpg	
22	Copycat	112722	Copycat: copia m	http://ia.media-imdb.	1995	1065837-copycat	6.6	32	24	8	75	7.4	5	4	1	80	3.1	5628	58	http://content9.flixster.com/movie/27/27/272723_det.jpg	
23	Ninja Assassin	1186367	Ninja Assassin	http://ia.media-imdb.	2009	1198524-ninja_assassin	4.3	106	27	79	25	4.6	20	6	14	30	3.4	107023	58	http://content6.flixster.com/movie/10/89/12/10891232_det.jpg	
24	Powder	114168	Powder - pura en	http://ia.media-imdb.	1995	powder	5.2	19	9	10	47	4.9	5	0	5	0	3.1	6650	67	http://content7.flixster.com/movie/25/59/255901_det.jpg	
25	Leaving Las Vegas	113627	Leaving Las Veg	http://ia.media-imdb.	1995	leaving_las_vegas	7.7	45	40	5	88	8.1	11	11	0	100	3.6	12742	82	http://content6.flixster.com/movie/10/90/25/10902561_det.jpg	
26	Othello	114057	Otelo	http://ia.media-imdb.	1995	1069707-othello	6.3	38	26	12	68	6.1	11	7	4	63	3.3	2018	66	http://content6.flixster.com/movie/10/92/98/10929892_det.jpg	
27	Now and Then	114011	Amigas para sien	http://ia.media-imdb.	1995	now_and_then	4.7	16	3	13	18	5.2	5	1	4	20	3.5	17168	78	http://content6.flixster.com/movie/55/36/64/5536644_det.jpg	
28	Persuasion	114117	Persuasión	http://ia.media-imdb.	1995	1066528-persuasion	7.7	23	19	4	82	8.5	9	9	0	100	3.9	7968	82	http://content6.flixster.com/movie/11/12/73/11127360_det.jpg	
29	La cité des enfants p	112682	La ciudad de los	http://ia.media-imdb.	1995	city_of_lost_children	7	33	27	6	81	7.2	7	5	2	71	4	22420	89	http://content6.flixster.com/movie/26/78/267824_det.jpg	
30	Yao a yao yao dao w	115012	La joya de Shan	http://ia.media-imdb.	1995	shanghai_triad	6.8	21	18	3	85	0	4	3	1	75	3.6	1061	70	http://content8.flixster.com/movie/10/90/31/10903162_det.jpg	
31	Dangerous Minds	112792	Mentes peligrosa	http://ia.media-imdb.	1995	dangerous_minds	4.7	37	11	26	29	5	12	5	7	41	3.1	12136	66	http://content7.flixster.com/movie/26/98/269813_det.jpg	
32	Twelve Monkeys	114746	Twelve monkeys:	http://ia.media-imdb.	1995	prem-no-publio-issue	0	0	0	0	0	0	0	0	0	0	0	0	0	http://content6.flixster.com/movie/10/93/02/10930244_det.jpg	
33	Wings of Courage	114952	Las alas del cora	http://ia.media-imdb.	1995		0	4	2	2	50	0	2	1	1	50	0	0	0	http://content7.flixster.com/movie/10/97/83/10978341_det.jpg	
34	Babe	112431	Babe, el cerdito	http://ia.media-imdb.	1995	1065598-babe	8.2	54	53	1	98	8.2	11	11	0	100	3	22954	54	http://content7.flixster.com/movie/25/32/253205_det.jpg	
35	Carrington	112637	Carrington	http://ia.media-imdb.	1995	carrington	6.2	19	9	10	47	7.4	5	4	1	80	3.5	781	68	http://content7.flixster.com/movie/28/00/280009_det.jpg	
36	Dead Man Walking	112818	Pena de muerte	http://ia.media-imdb.	1995	1068779-dead_man_walking	8.2	53	50	3	94	7.9	20	18	2	90	3.6	15460	81	http://content6.flixster.com/movie/10/88/77/10887764_det.jpg	
37	Across the Sea of Tir	112286	Across the Sea of	http://ia.media-imdb.	1995	across_the_sea_of_time	0	4	2	2	50	0	1	0	1	0	2.9	237	27	http://content6.flixster.com/movie/10/90/58/10905808_det.jpg	
38	It Takes Two	113442	Dos por el precio	http://ia.media-imdb.	1995	1067137-it_takes_two	3.9	24	2	22	8	4.6	8	2	6	25	2.9	16311	63	http://content7.flixster.com/movie/25/74/257417_det.jpg	
39	Clueless	112697	Clueless: fuera de	http://ia.media-imdb.	1995	school_of_senses	0	0	0	0	0	0	0	0	0	0	0	0	0	http://content8.flixster.com/movie/10/85/01/10850110_det.jpg	
40	Cry, the Beloved Cou	112749	Llanto por la tier	http://ia.media-imdb.	1995	1068781-cry_the_beloved_c	7.2	13	10	3	76	0	4	3	1	75	3.5	520	66	http://content6.flixster.com/movie/10/89/59/10895996_det.jpg	
41	Richard III	114279	Ricardo III (Richa	http://ia.media-imdb.	1995	1068177-richard_iii	8.1	43	41	2	95	7.6	11	10	1	90	3.7	1552	82	http://content7.flixster.com/movie/26/89/268909_det.jpg	
42	Dead Presidents	112819	Dinero para quer	http://ia.media-imdb.	1995	dead_presidents	5.7	31	14	17	45	6.3	12	7	5	58	3.3	4475	75	http://content7.flixster.com/movie/26/95/269521_det.jpg	
43	Restoration	114272	Restauración	http://ia.media-imdb.	1995	1069339-restoration	6.3	31	21	10	67	6.4	10	6	4	60	3.3	1103	56	http://content6.flixster.com/movie/27/04/270404_det.jpg	
44	Mortal Kombat	113855	Mortal Kombat	http://ia.media-imdb.	1995	mortal_kombat	4.4	26	9	17	34	3.9	6	2	4	33	2.8	18291	55	http://content9.flixster.com/movie/25/63/256395_det.jpg	
45	Live Free or Die Hard	337978	La jungla 4.0	http://ia.media-imdb.	2007	live_free_or_die_hard	6.8	197	161	36	81	6.7	38	30	8	78	3.9	1E+06	86	http://content8.flixster.com/movie/10/93/34/10933482_det.jpg	
46	How to Make an Ame	113347	Donde reside el c	http://ia.media-imdb.	1995	how_to_make_an_american_	5.6	22	11	11	50	7.6	6	5	1	83	3.2	2626	58	http://content8.flixster.com/movie/10/88/75/10887514_det.jpg	
47	Shichinin no samura	47478	Los siete samurái	http://ia.media-imdb.	1954	1018639-seven_samurai	9.2	49	49	0	100	8.6	10	10	0	100	4.5	29509	97	http://content7.flixster.com/movie/39/91/08/3991085_det.jpg	
48	Pocahontas	114148	Pocahontas	http://ia.media-imdb.	1995	1063452-pocahontas	6	49	27	22	55	6.6	14	10	4	71	3.2	37858	63	http://content6.flixster.com/movie/25/15/251508_det.jpg	
49	When Night Is Fallin	114916	Cuando cae la n	http://ia.media-imdb.	1995	when_night_is_falling	5.7	16	8	8	50	0	4	1	3	25	3.9	1132	83	http://content6.flixster.com/movie/10/89/77/10897780_det.jpg	
50	The Usual Suspects	114814	Sospechosos hab	http://ia.media-imdb.	1995	usual_suspects	7.5	47	41	6	87	6.9	14	12	2	85	4.2	109588	95	http://content7.flixster.com/movie/24/63/48/2463485_det.jpg	
52	Mighty Aphrodite	113819	Poderosa Afrodit	http://ia.media-imdb.	1995	mighty_aphrodite	6.7	33	25	8	75	7.2	6	4	2	66	3.3	4866	69	http://content7.flixster.com/movie/10/84/18/10841893_det.jpg	

movies

movies.csv

created separate two tables for my source dataset by dividing main movies.csv

rotten_tomatoes.csv

rottenTomatoesID	movieTitle	rtAllCriticsRating	rtAllCriticsNumR	rtTopCriticsRating	rtTopCriticsNumReviews	rtAudienceRating	rtAudienceNumRatings
1	toy_story	9	73	8.5	17	3.7	102338
2	1068044-jumanji	5.6	28	5.8	5	3.2	44587
3	grumpy_old_men	5.9	36	7	6	3.2	10489
4	waiting_to_exhale	5.6	25	5.5	11	3.3	5666
5	father_of_the_bride_part_ii	5.3	19	5.4	5	3	13761
6	1068182-heat	7.7	58	7.2	17	3.9	42785
7	1018047-sabrina	7.4	31	7.2	5	3.8	12812
8	tom_and_huck	4.2	8	0	2	2.7	2649
9	1068470-sudden_death	5.2	32	5.6	9	2.6	3626
10	goldeneye	6.8	41	6.2	11	3.4	28260
11	american_president	7	49	7.2	18	3.2	8320
12	dracula_dead_and_loving_it	3	35	3.7	10	2.8	10078
13	balto	5.8	12	0	4	3.2	9195
14	nixon	6.7	56	6.1	18	3.5	3256
15	cutthroat_island	4.4	31	4.7	6	2.6	3350
16	1067987-casino	7.2	58	6.3	16	3.9	66463
17	1068832-sense_and_sensibility	7.9	49	8.1	14	3.8	32782
18	four_rooms	3.5	42	2.9	11	3.5	14266
19	ace_ventura_when_nature_call	4.1	25	0	3	3.2	87306
20	money_train	4	24	4.5	5	2.7	5263
21	get_shorty	7.7	49	8.4	14	3.2	10155
22	1065837-copycat	6.6	32	7.4	5	3.1	5628
23	1198524-ninja_assassin	4.3	106	4.6	20	3.4	107023
24	powder	5.2	19	4.9	5	3.1	6650
25	leaving_las_vegas	7.7	45	8.1	11	3.6	12742
26	1069707-othello	6.3	38	6.1	11	3.3	2018
27	now_and_then	4.7	16	5.2	5	3.5	17168

rotten_tomatoes

movies.csv

movieID	title	imdbID	year	rottenTomatoesID
1	Toy story	114709	1995	1
2	Jumanji	113497	1995	2
3	Grumpy Old Men	107050	1993	3
4	Waiting to Exhale	114885	1995	4
5	Father of the Bride Part II	113041	1995	5
6	Heat	113277	1995	6
7	Sabrina	47437	1954	7
8	Tom and Huck	112302	1995	8
9	Sudden Death	114576	1995	9
10	GoldenEye	113189	1995	10
11	The American President	112346	1995	11
12	Dracula: Dead and Loving	112896	1995	12
13	Balto	112453	1995	13
14	Nixon	113987	1995	14
15	Cutthroat Island	112760	1995	15
16	Casino	112641	1995	16
17	Sense and Sensibility	114388	1995	17
18	Four Rooms	113101	1995	18
19	Ace Ventura: When Natur	112281	1995	19
20	Money Train	113845	1995	20
21	Get Shorty	113161	1995	21
22	Copycat	112722	1995	22
23	Ninja Assassin	1186367	2009	23
24	Powder	114168	1995	24
25	Leaving Las Vegas	113627	1995	25
26	Othello	114057	1995	26
27	Now and Then	114011	1995	27

movies

The movie details were taken as a text file, and I used it as another source type.

movieID	title	imdbID	year	rottenTomatoesID
1	Toy story	114709	1995	1
2	Jumanji	113497	1995	2
3	Grumpy Old Men	107050	1993	3
4	Waiting to Exhale	114885	1995	4
5	Father of the Bride Part II	113041	1995	5
6	Heat	113277	1995	6
7	Sabrina	47437	1954	7
8	Tom and Huck	112302	1995	8
9	Sudden Death	114576	1995	9
10	GoldenEye	113189	1995	10
11	The American President	112346	1995	11
12	Dracula: Dead and Loving	112896	1995	12
13	Balto	112453	1995	13
14	Nixon	113987	1995	14
15	Cutthroat Island	112760	1995	15
16	Casino	112641	1995	16
17	Sense and Sensibility	114388	1995	17
18	Four Rooms	113101	1995	18
19	Ace Ventura: When Natur	112281	1995	19
20	Money Train	113845	1995	20
21	Get Shorty	113161	1995	21
22	Copycat	112722	1995	22
23	Ninja Assassin	1186367	2009	23
24	Powder	114168	1995	24
25	Leaving Las Vegas	113627	1995	25
26	Othello	114057	1995	26
27	Now and Then	114011	1995	27

movies - Notepad

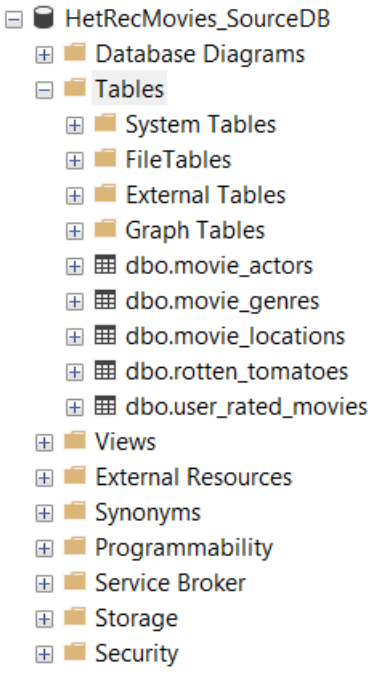
File Edit View

```
id      title imdbID       year  rottenTomatoesID
1       Toy story    114709       1995  1
2       Jumanji      113497       1995  2
3       Grumpy Old Men     107050       1993  3
4       Waiting to Exhale 114885       1995  4
5       Father of the Bride Part II    113041       1995  5
6       Heat   113277       1995  6
7       Sabrina      47437 1954  7
8       Tom and Huck       112302       1995  8
9       Sudden Death       114576       1995  9
10      GoldenEye    113189       1995  10
11      The American President  112346       1995  11
12      Dracula: Dead and Loving It    112896       1995  12
13      Balto 112453       1995  13
14      Nixon 113987       1995  14
15      Cutthroat Island   112760       1995  15
16      Casino       112641       1995  16
17      Sense and Sensibility    114388       1995  17
18      Four Rooms   113101       1995  18
19      Ace Ventura: When Nature Calls       112281       1995  19
20      Money Train 113845       1995  20
21      Get Shorty   113161       1995  21
22      Copycat      112722       1995  22
23      Ninja Assassin     1186367      2009  23
24      Powder       114168       1995  24
25      Leaving Las Vegas 113627       1995  25
26      Othello      114057       1995  26
27      Now and Then       114011       1995  27
28      Persuasion   114117       1995  28
29      La cité des enfants perdus     112682       1995  29
30      Yao a yao yao dao waipo qiao   115012       1995  30
31      Dangerous Minds    112792       1995  31
32      Twelve Monkeys     114746       1995  32
33      Wings of Courage   114952       1995  33
34      Babe   112431       1995  34
35      Carrington   112637       1995  35
36      Dead Man Walking   112818       1995  36
37      Across the Sea of Time   112286       1995  37
38      It Takes Two       113442       1995  38
```

Ln 1, Col 1

movies.csv → **movies.txt**

I have loaded movie_actors.csv, movie_genres.csv, movie_locations.csv, user_rated_movies.csv, rotten_tomatoes.csv files to the DB called HetRecMovies_SourceDB.



Data Sources Data Types

LAPTOP-B183G8T7.H...dbo.movie_actors | SQLQuery3.sql - not connected | SQLQ

Column Name	Data Type	Allow Nulls
actorID	int	☐
actorName	nvarchar(50)	☑
ranking	int	☑
		☐

LAPTOP-B183G8T7.H...o.movie_locations | LAPTOP-B183G8T7....dbo.movie_genres | LA

Column Name	Data Type	Allow Nulls
movieID	int	☐
country	nvarchar(50)	☑
state	nvarchar(50)	☑
location	nvarchar(100)	☑
		☐

LAPTOP-B183G8T7....dbo.movie_genres | LAPTOP-B183G8T7....eDB - dbo.Table_2 | LAPT

Column Name	Data Type	Allow Nulls
genreMovieID	int	☐
genre	nvarchar(50)	☑
		☐

LAPTOP-B183G8T7....o.rotten_tomatoes

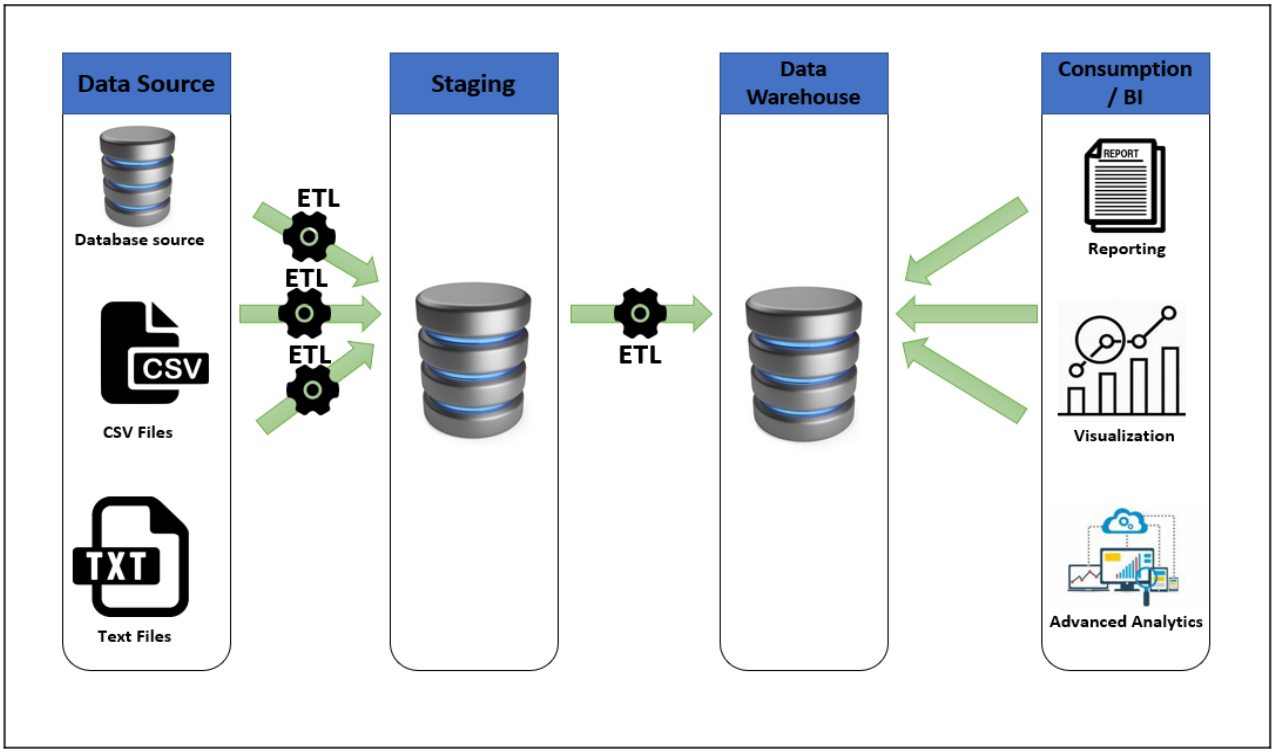
Column Name	Data Type	Allow Nulls
rottenTomatoesID	int	☐
movieTitle	nvarchar(100)	☑
rtAllCriticsRating	numeric(18, 1)	☑
rtAllCriticsNumReviews	int	☑
rtTopCriticsRating	numeric(18, 1)	☑
rtTopCriticsNumReviews	int	☑
rtAudienceRating	numeric(18, 1)	☑
rtAudienceNumRatings	int	☑
		☐

LAPTOP-B183G8T7....user_rated_movies*

Column Name	Data Type	Allow Nulls
userID	int	☐
movieID	int	☐
rating	numeric(18, 1)	☑
year	int	☑
month	int	☑
day	int	☑
Date	datetime	☐
actorID	int	☐
genreMovieID	int	☐
		☐

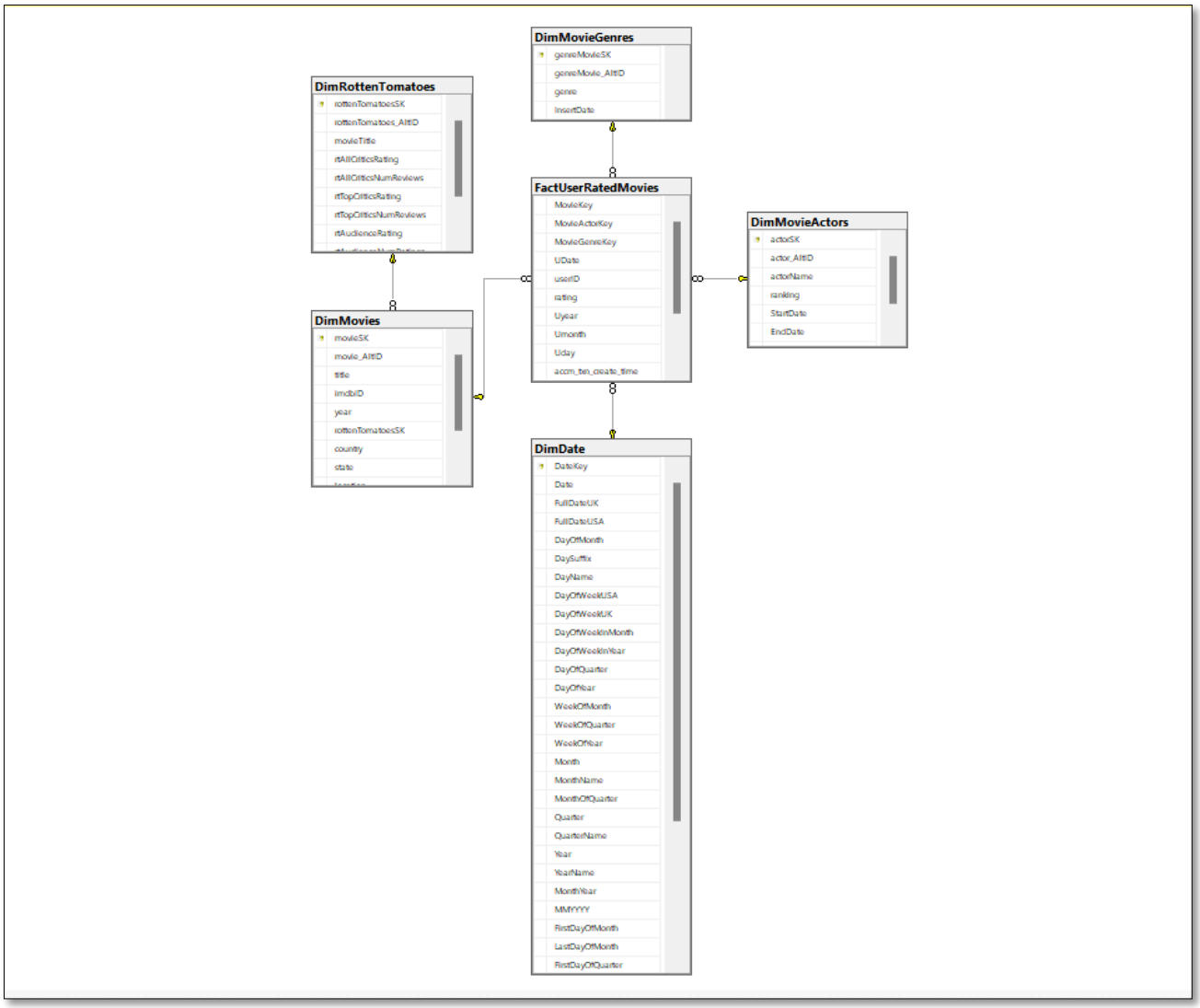
STEP 03: SOULUTION ARCHITECTURE

Architectural diagram



The data warehouse is the core of the BI system. A data warehouse is a database built for the purpose of data analysis and reporting. This purpose changes the design of this database as well. This architecture shows the high-level BI solution to the warehouse.

STEP 04: Data Warehouse design and development

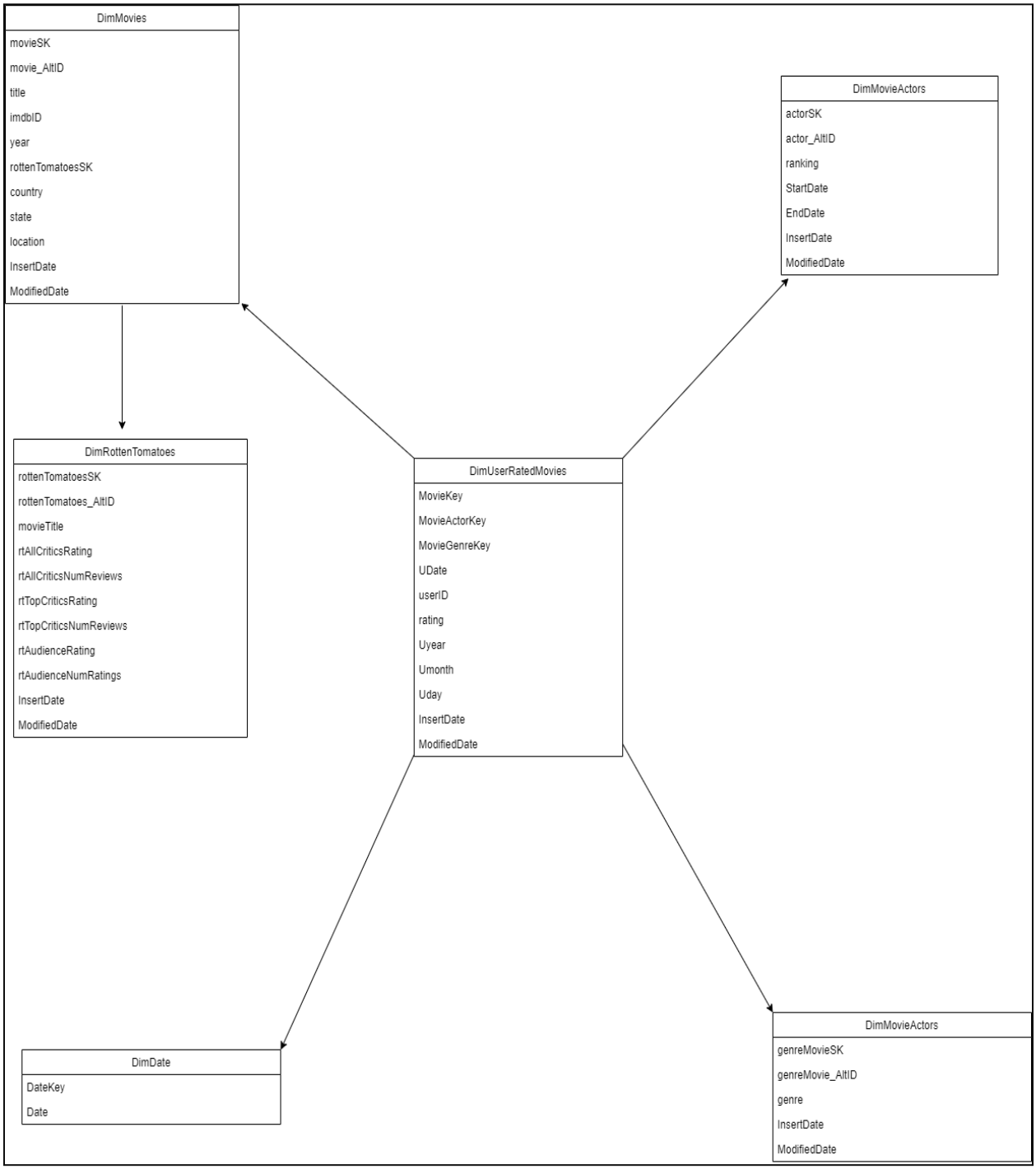


Assumptions

I have taken **DimMovieActors as slowly changing dimension**, Actor Ranking can change time to time, and we need to keep track of their historical data.

For my data set the schema is a snowflake schema. In my data set there are 4-dimension tables, date dimension and a fact table. The slowly changing dimension is Movie Actors table.

I have implemented data warehouse tables in the SQL server and the tables that I have created are shown below.


DimMovies
movieSK
movie_AltID
title
imdbID
year
rottenTomatoesSK
country
state
location
InsertDate
ModifiedDate
DimMovieActors
actorSK
actor_AltID
ranking
StartDate
EndDate
InsertDate
ModifiedDate
DimRottenTomatoes
rottenTomatoesSK
rottenTomatoes_AltID
movieTitle
rtAllCriticsRating
rtAllCriticsNumReviews
rtTopCriticsRating
rtTopCriticsNumReviews
rtAudienceRating
rtAudienceNumRatings
InsertDate
ModifiedDate
DimUserRatedMovies
MovieKey
MovieActorKey
MovieGenreKey
UDate
userID
rating
Uyear
Umonth
Uday
InsertDate
ModifiedDate
DimDate
DateKey
Date
DimMovieActors
genreMovieSK
genreMovie_AltID
genre
InsertDate
ModifiedDate

Data Warehouse Date Types

Before creating the FactUserRatedMovies fact table and other dimensions, start by creating the Date dimension. For that I used the code in the file “DateMaster.sql” file.

Column Name	Data Type	Allow Nulls
MovieKey	int	☐
MovieActorKey	int	☐
MovieGenreKey	int	☐
UDate	int	☐
userID	int	☐
rating	varchar(50)	☑
Uyear	int	☑
Umonth	int	☑
Uday	int	☑
accm_txn_create_time	datetime	☑
accm_txn_complete_time	datetime	☑
txn_process_time_hours	int	☑
ModifiedDate	datetime	☑
		☐

FactUserRatedMovies

Column Name	Data Type	Allow Nulls
genreMovieSK	int	☐
genreMovie_AltID	int	☐
genre	nvarchar(50)	☑
InsertDate	datetime	☑
ModifiedDate	datetime	☑
		☐

DimMovieGenres

Column Name	Data Type	Allow Nulls
rottenTomatoesSK	int	☐
rottenTomatoes_AltID	int	☐
movieTitle	nvarchar(100)	☑
rtAllCriticsRating	numeric(18, 1)	☑
rtAllCriticsNumReviews	int	☑
rtTopCriticsRating	numeric(18, 1)	☑
rtTopCriticsNumReviews	int	☑
rtAudienceRating	numeric(18, 1)	☑
rtAudienceNumRatings	int	☑
InsertDate	datetime	☑
ModifiedDate	datetime	☑
		☐

DimRottenTomatoes

Column Name	Data Type	Allow Nulls
actorSK	int	☐
actor_AltID	int	☐
actorName	nvarchar(50)	☑
ranking	int	☑
StartDate	datetime	☑
EndDate	datetime	☑
InsertDate	datetime	☑
ModifiedDate	datetime	☑
		☐

DimMovieActors

Column Name	Data Type	Allow Nulls
movieSK	int	☐
movie_AltID	int	☐
title	varchar(500)	☑
imdbID	int	☑
year	int	☑
rottenTomatoesSK	int	☑
country	nvarchar(50)	☑
state	nvarchar(50)	☑
location	nvarchar(100)	☑
InsertDate	datetime	☑
ModifiedDate	datetime	☑
		☐

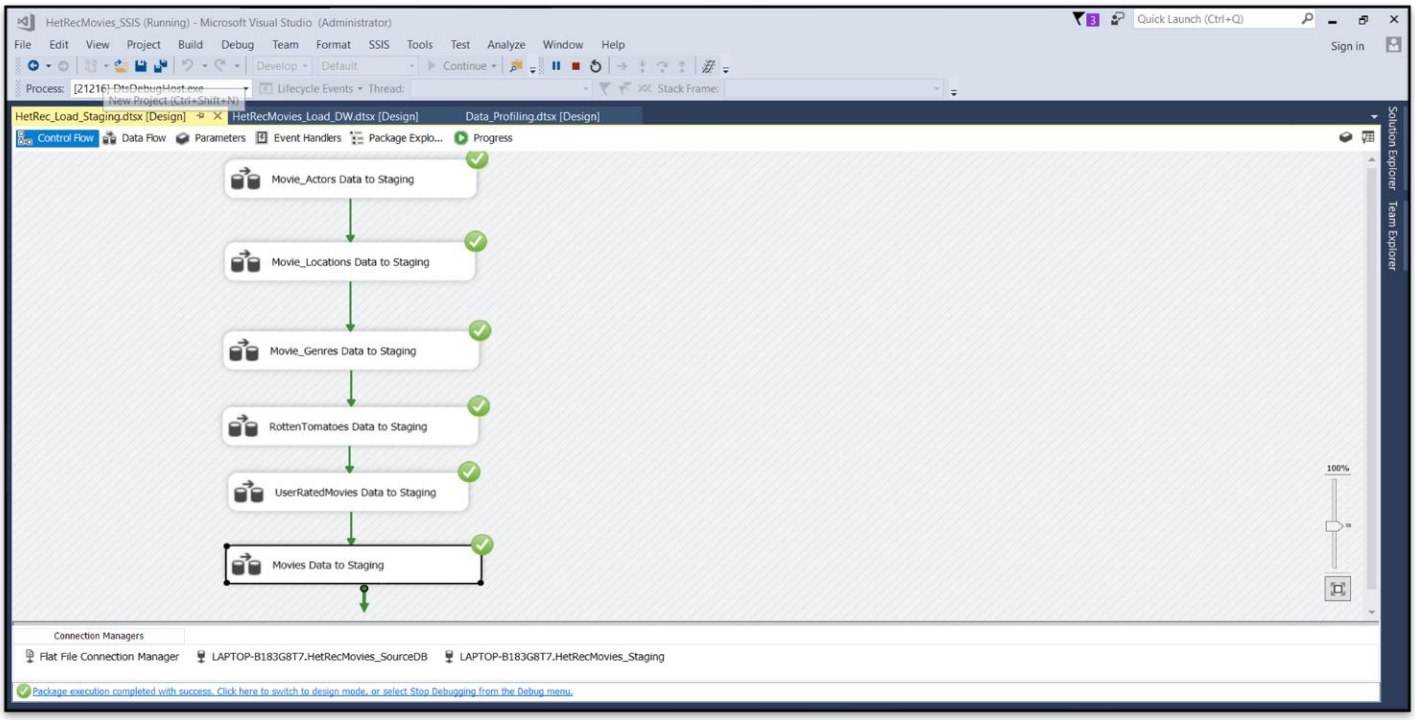
DimMovies

LAPTOP-B183G8T7....DW - dbo.DimDate

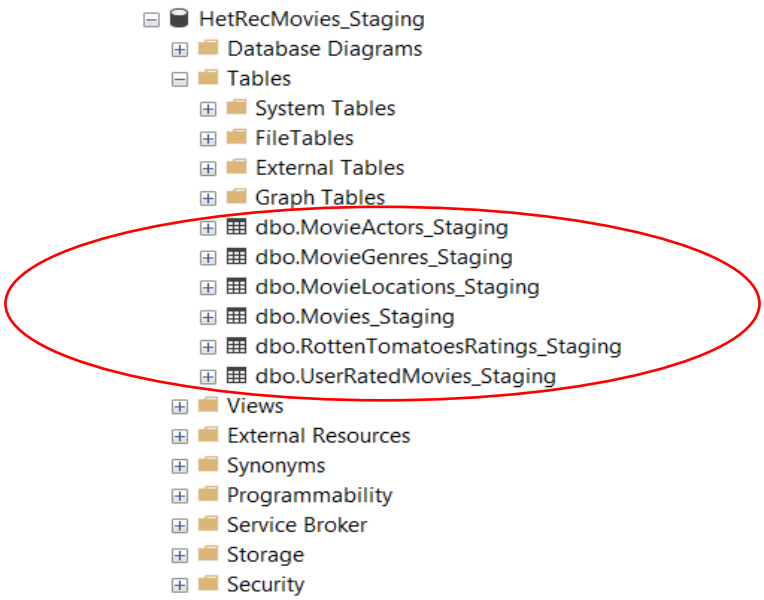
Column Name	Data Type	Allow Nulls
DateKey	int	☐
Date	datetime	☑
FullDateUK	char(10)	☑
FullDateUSA	char(10)	☑
DayOfMonth	varchar(2)	☑
DaySuffix	varchar(4)	☑
DayName	varchar(9)	☑
DayOfWeekUSA	char(1)	☑
DayOfWeekUK	char(1)	☑
DayOfWeekInMonth	varchar(2)	☑
DayOfWeekInYear	varchar(2)	☑
DayOfQuarter	varchar(3)	☑
DayOfYear	varchar(3)	☑
WeekOfMonth	varchar(1)	☑
WeekOfQuarter	varchar(2)	☑
WeekOfYear	varchar(2)	☑
Month	varchar(2)	☑
MonthName	varchar(9)	☑
MonthOfQuarter	varchar(2)	☑
Quarter	char(1)	☑
QuarterName	varchar(9)	☑
Year	char(4)	☑
YearName	char(7)	☑

DimDate

STEP 05: ETL development

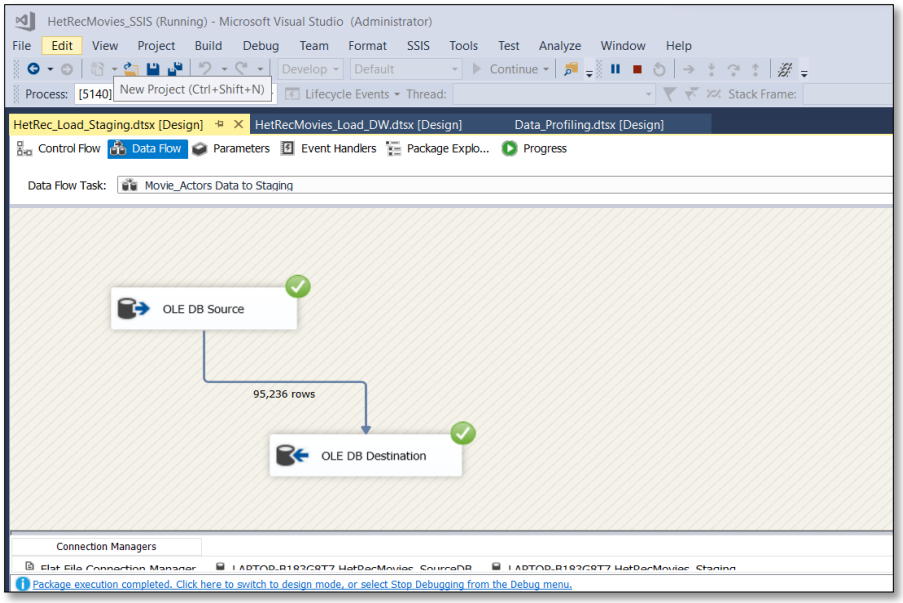


First using the SQL Server Integration Services Software, I have extracted all the data from the tables which were in the HetRecMovies_SourceDB and Extract movies.txt to separate staging DB called HetRecMovies _Staging as shown in the below.

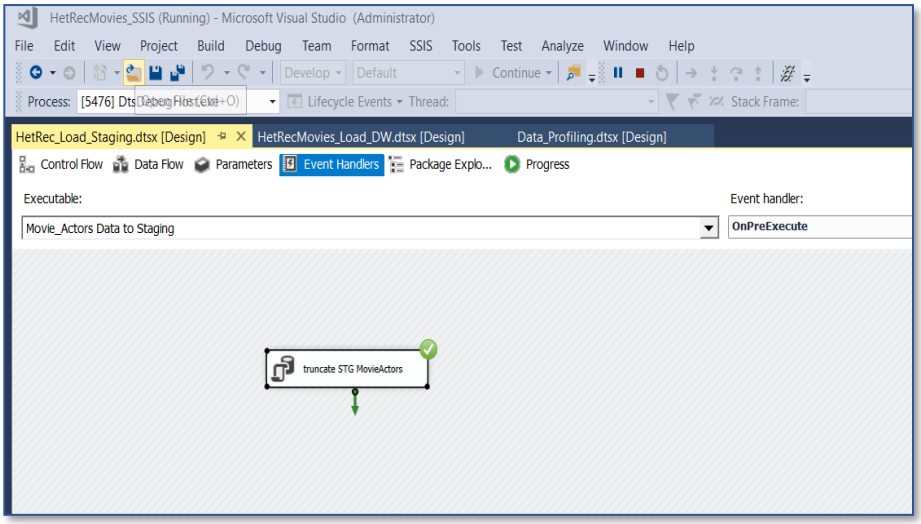


HetRecMovies_Staging

Extract Movie Actors Data to Staging

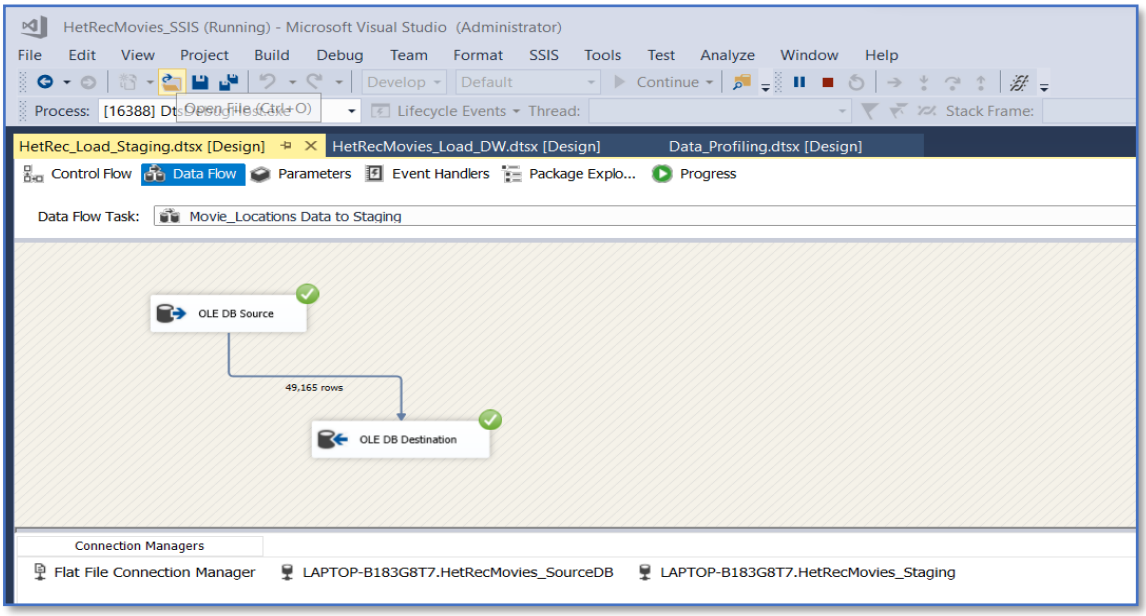


Used OLE DB Source as dbo. movie_actors data table in HetRecMovies_SourceDB.OLE DB Destination for create new table MovieActors_Staging in the HetRecMovies _Staging database.

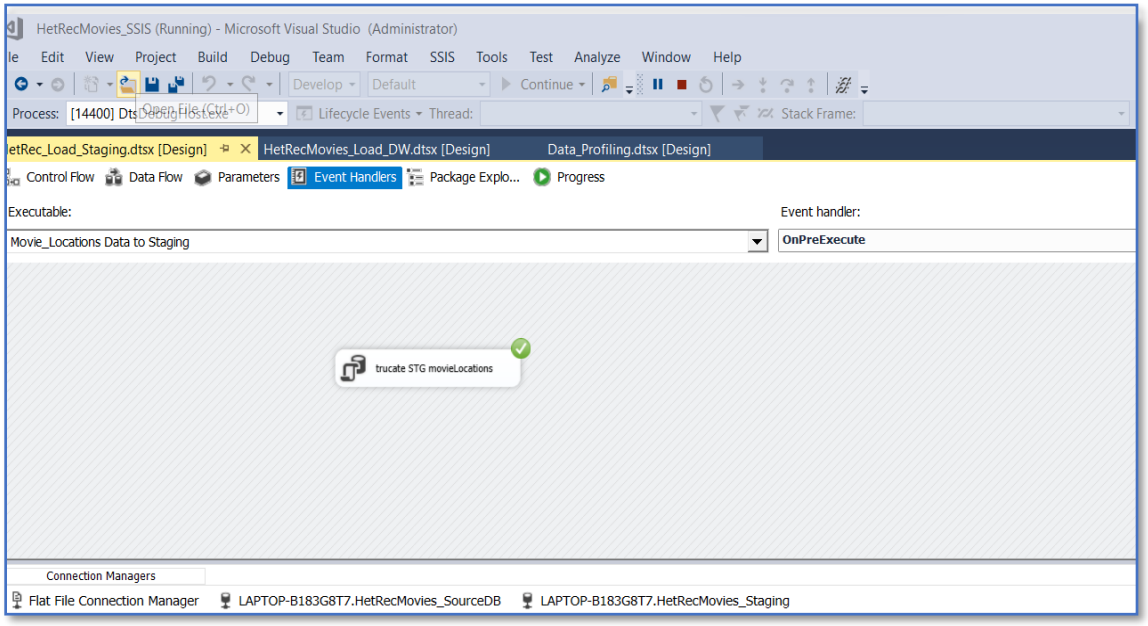


Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table [HetRecMovies_Staging]. [dbo]. [MovieActors_Staging] in HetRecMovies_Staging database.

Extract Movie Locations Data to Staging

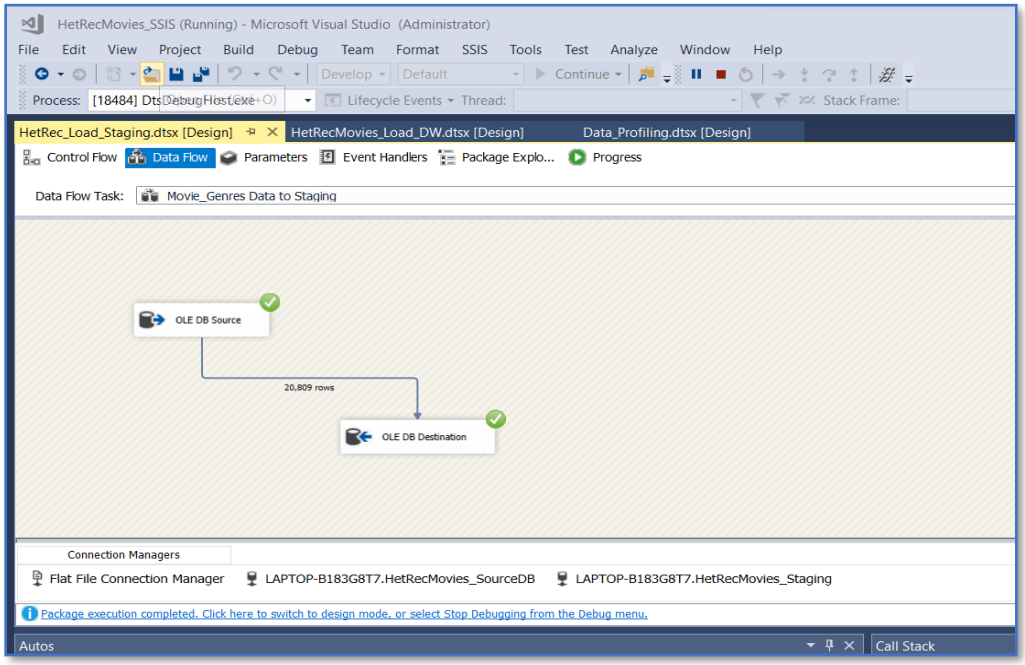


Used OLE DB Source as dbo. movie_locations data table in HetRecMovies_SourceDB.OLE DB Destination for create new table MovieLocations_Staging in the HetRecMovies _Staging database.

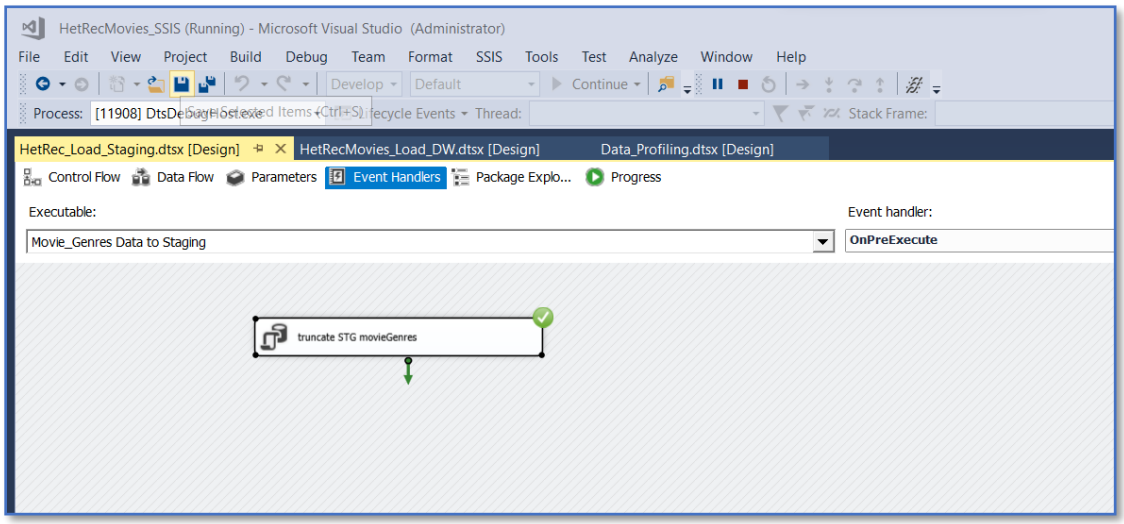


Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table [HetRecMovies_Staging]. [dbo]. [MovieLocations_Staging] in HetRecMovies_Staging database.

Extract Movie Genres Data to Staging

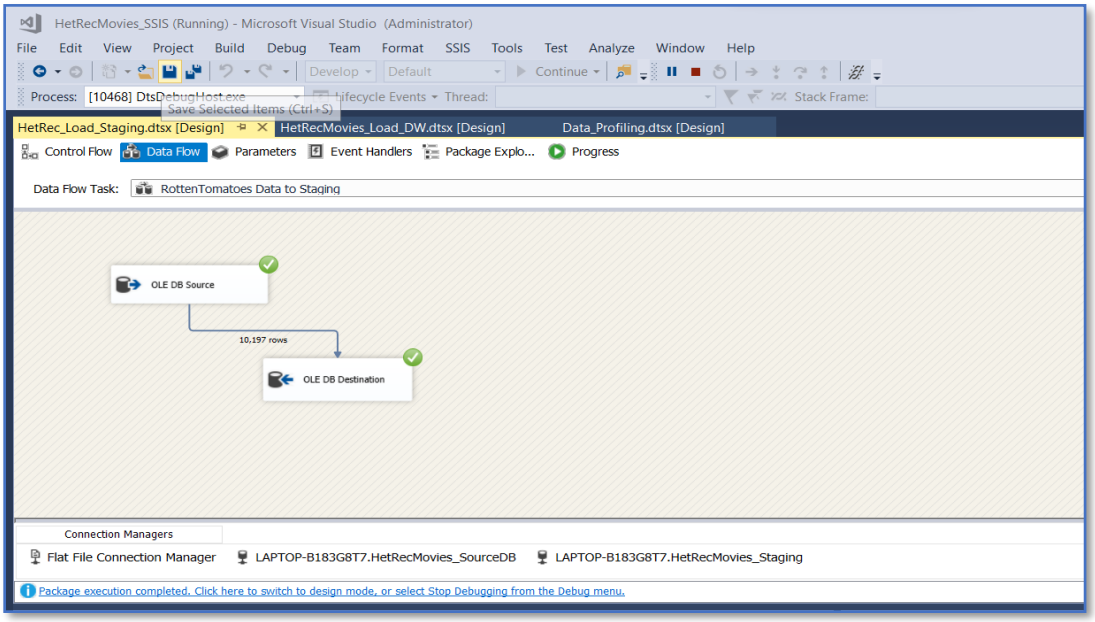


Used OLE DB Source as dbo. movie_genres data table in HetRecMovies_SourceDB.OLE DB Destination for create new table MovieGenres_Staging in the HetRecMovies _Staging database.

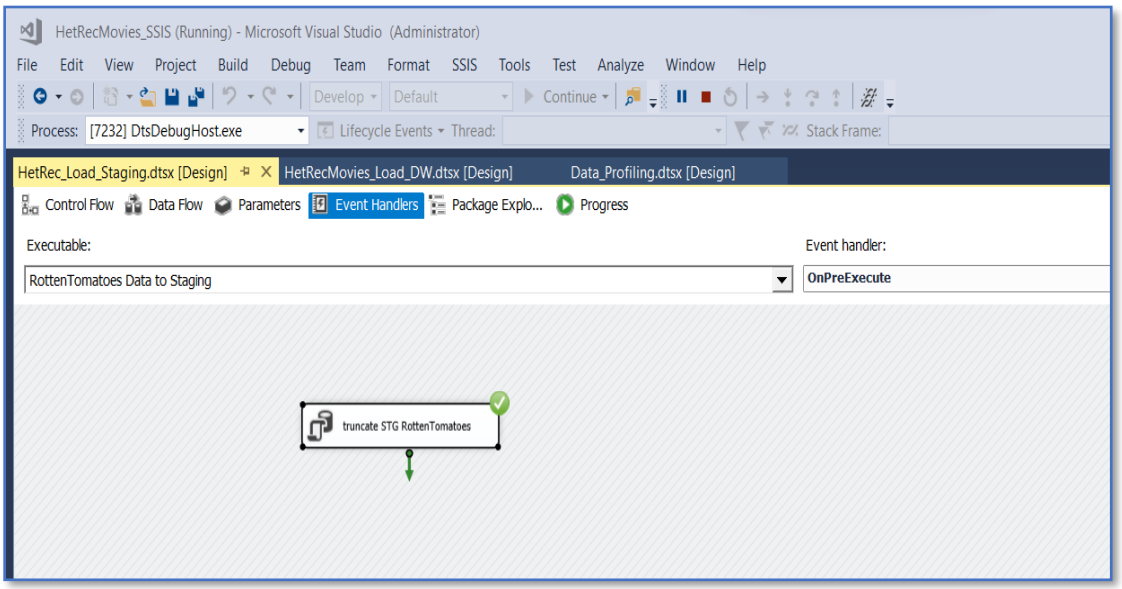


Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table [HetRecMovies_Staging]. [dbo]. [MovieGenres_Staging]in HetRecMovies_Staging database.

Extract RottenTomatoesRating Data to Staging

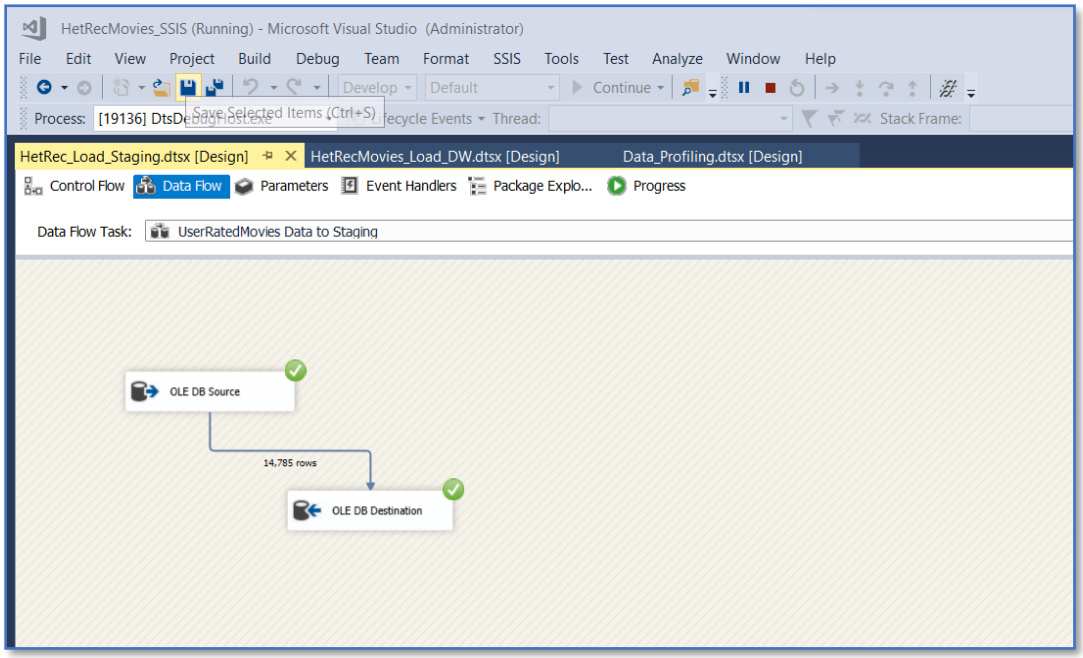


Used OLE DB Source as dbo. rotten_tomatoes data table in HetRecMovies_SourceDB.OLE DB Destination for create new table RottenTomatoesRatings_Staging in the HetRecMovies _Staging database.

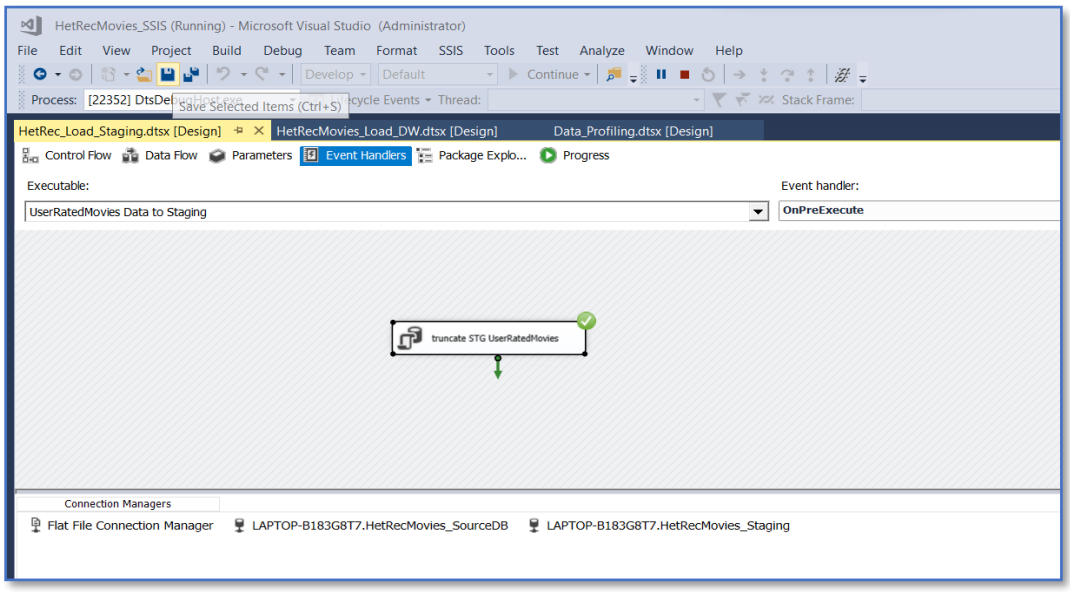


Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table [HetRecMovies_Staging]. [dbo].[RottenTomatoesRatings_Staging] in HetRecMovies_Staging database.

Extract User Rated Movies Data to Staging

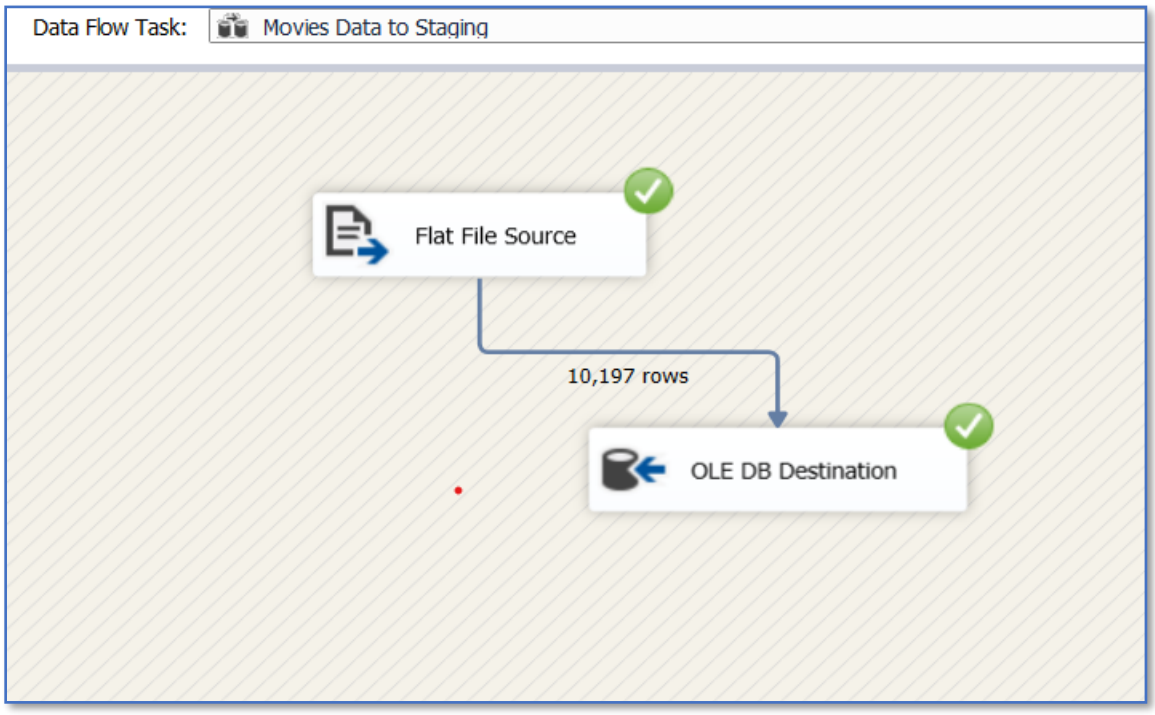


Used OLE DB Source as dbo. User_rated_movies data table in HetRecMovies_SourceDB.OLE DB Destination for create new table UserRatedMovies_Staging in the HetRecMovies _Staging database.

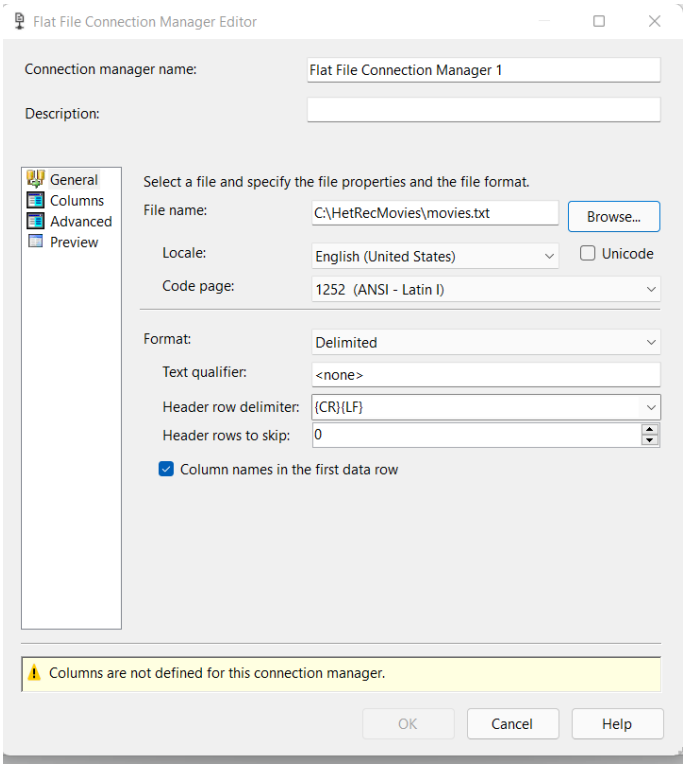


Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table [HetRecMovies_Staging]. [dbo]. [UserRatedMovies_Staging] in HetRecMovies_Staging database.

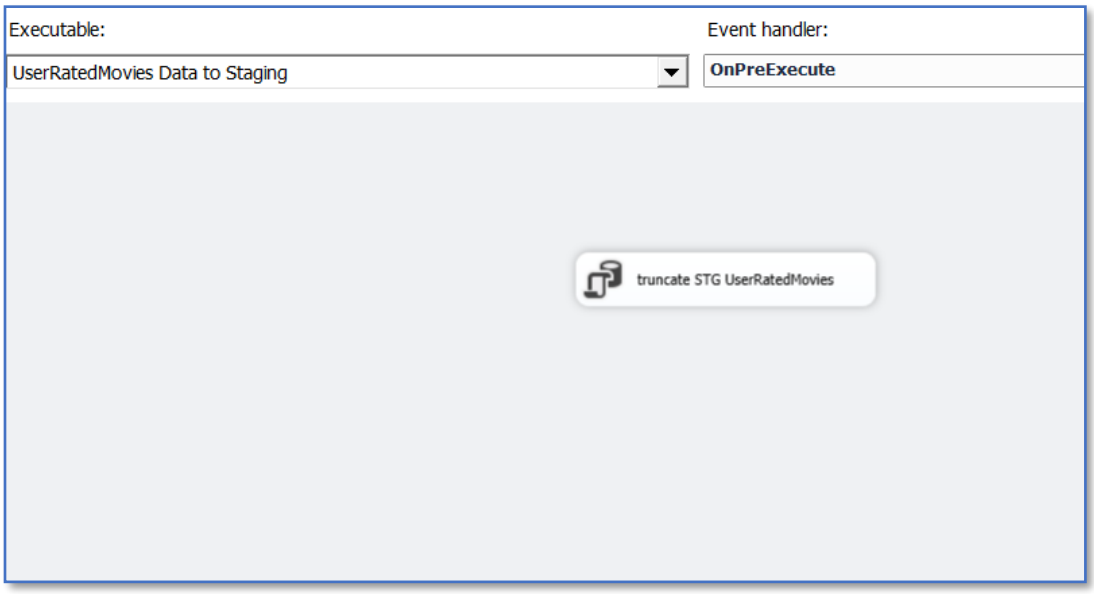
Extract Movies Data to Staging



Used Flat file Source SSIS tool, to extract movies.txt data.



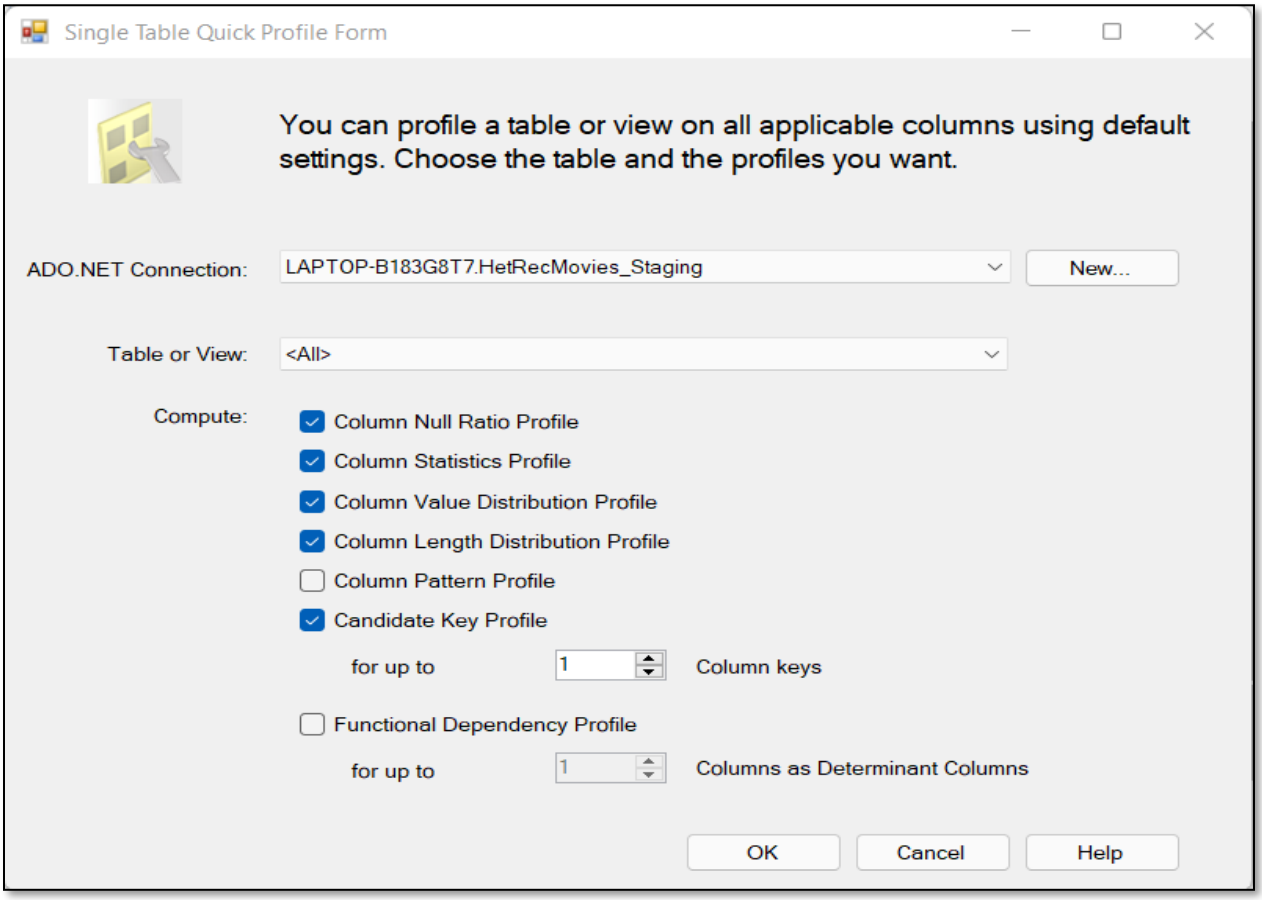
Used OLE DO Destination SSIS tool to create new table as movies.txt load next file data to HetRecMovies_Staging database.

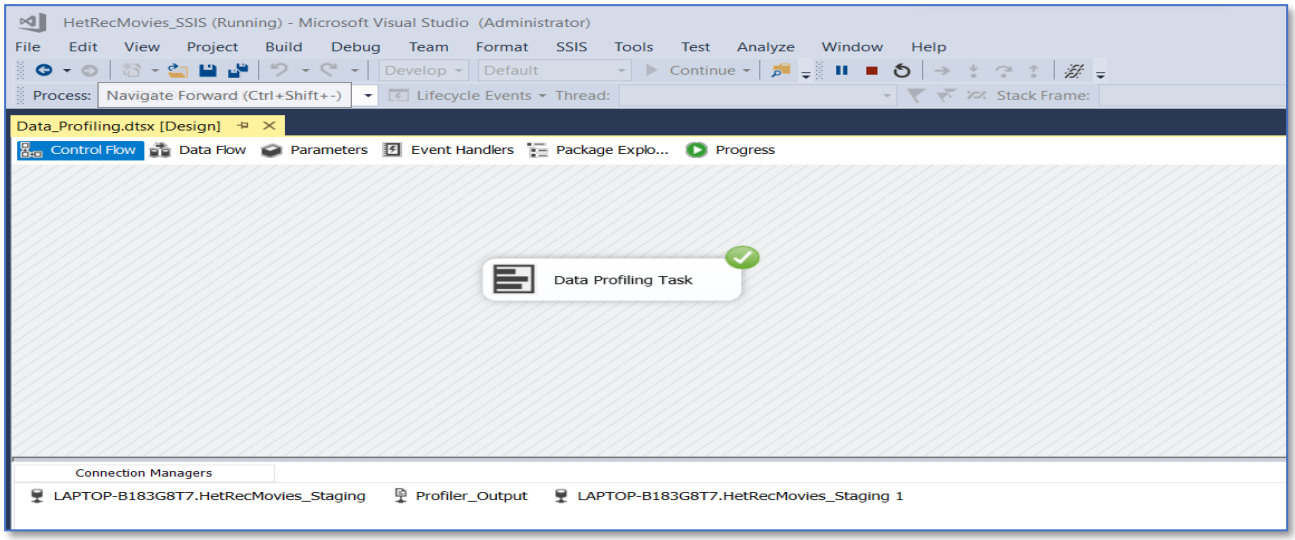


Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table [HetRecMovies_Staging]. [dbo]. [UserRatedMovies_Staging] in HetRecMovies_Staging database.

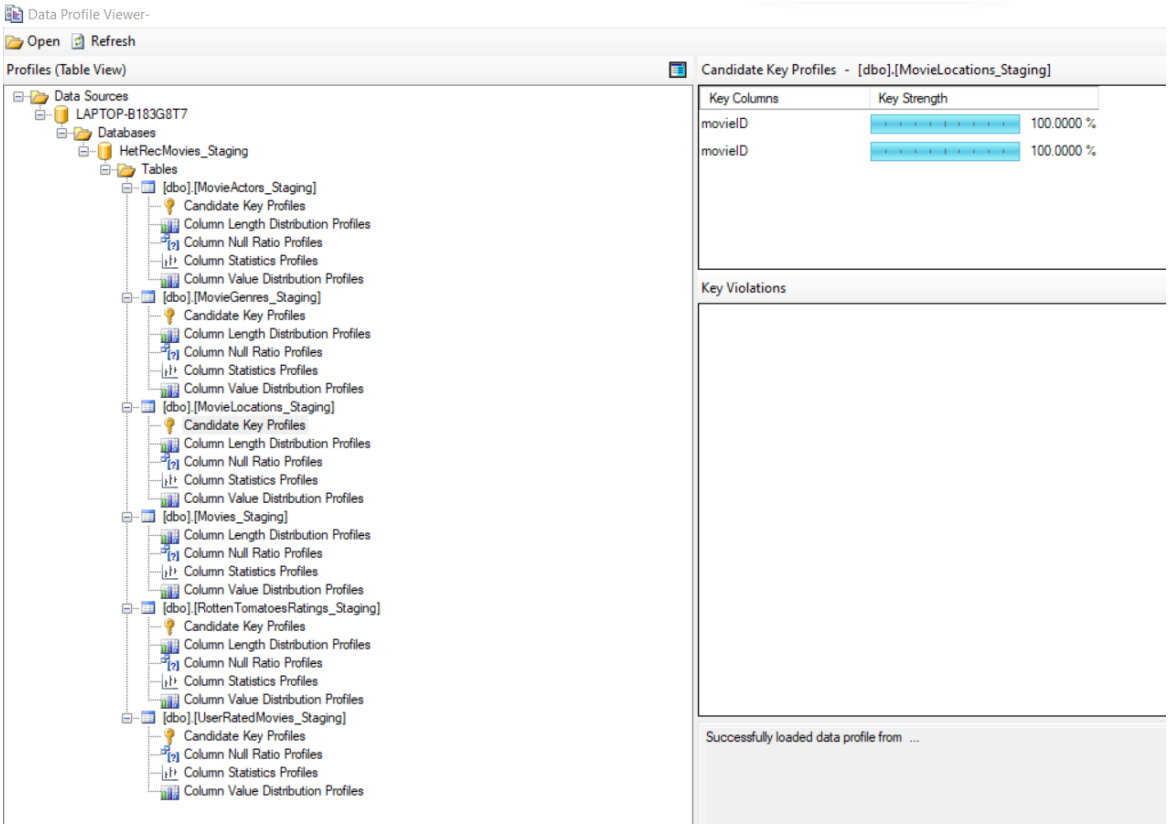
Data Profiling

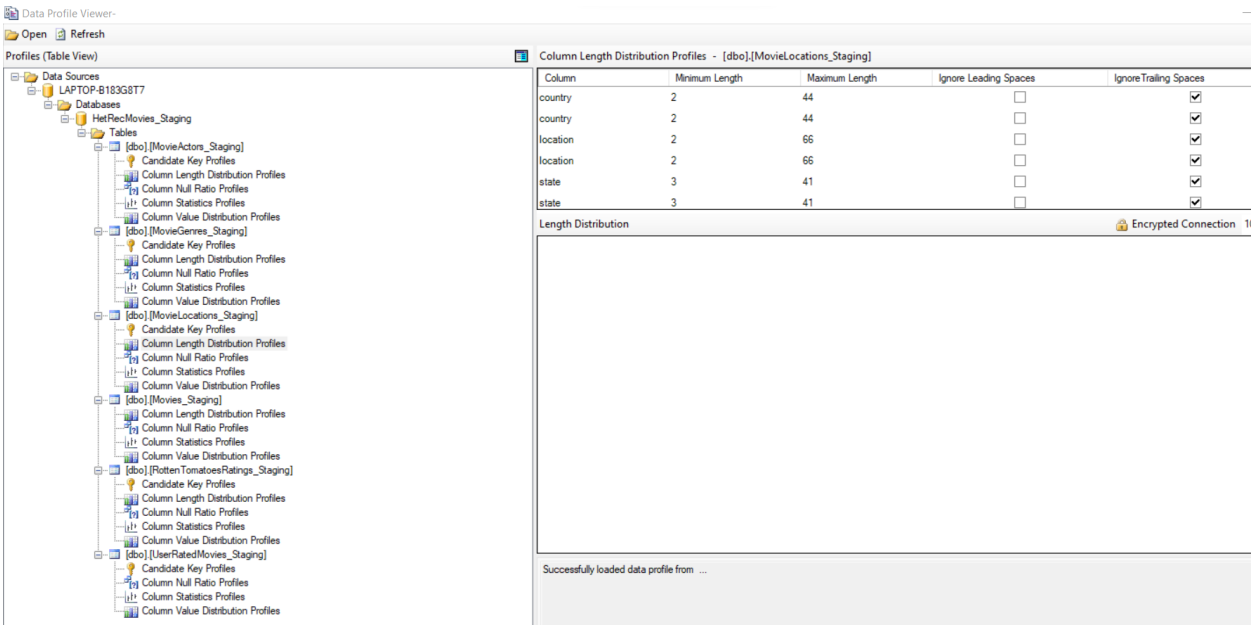
I used the staging table data to analyze how the data looks like to determine what type of transformations I need to perform on the data.

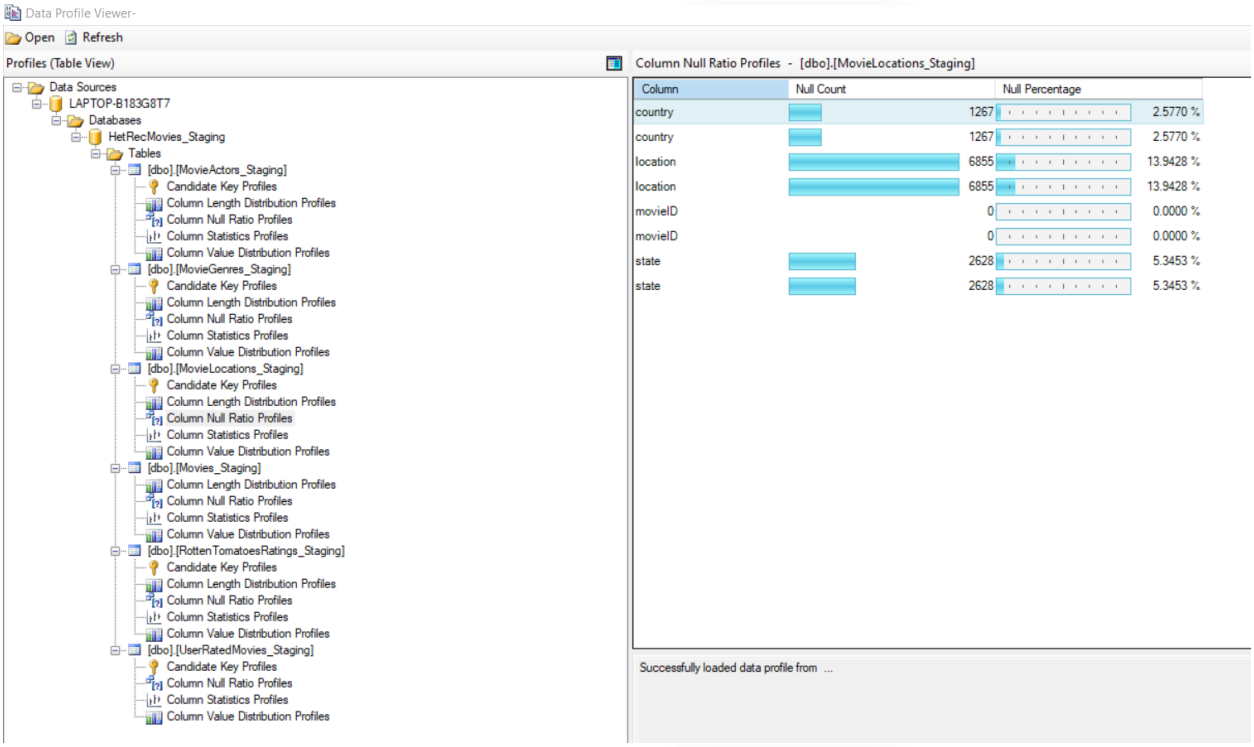


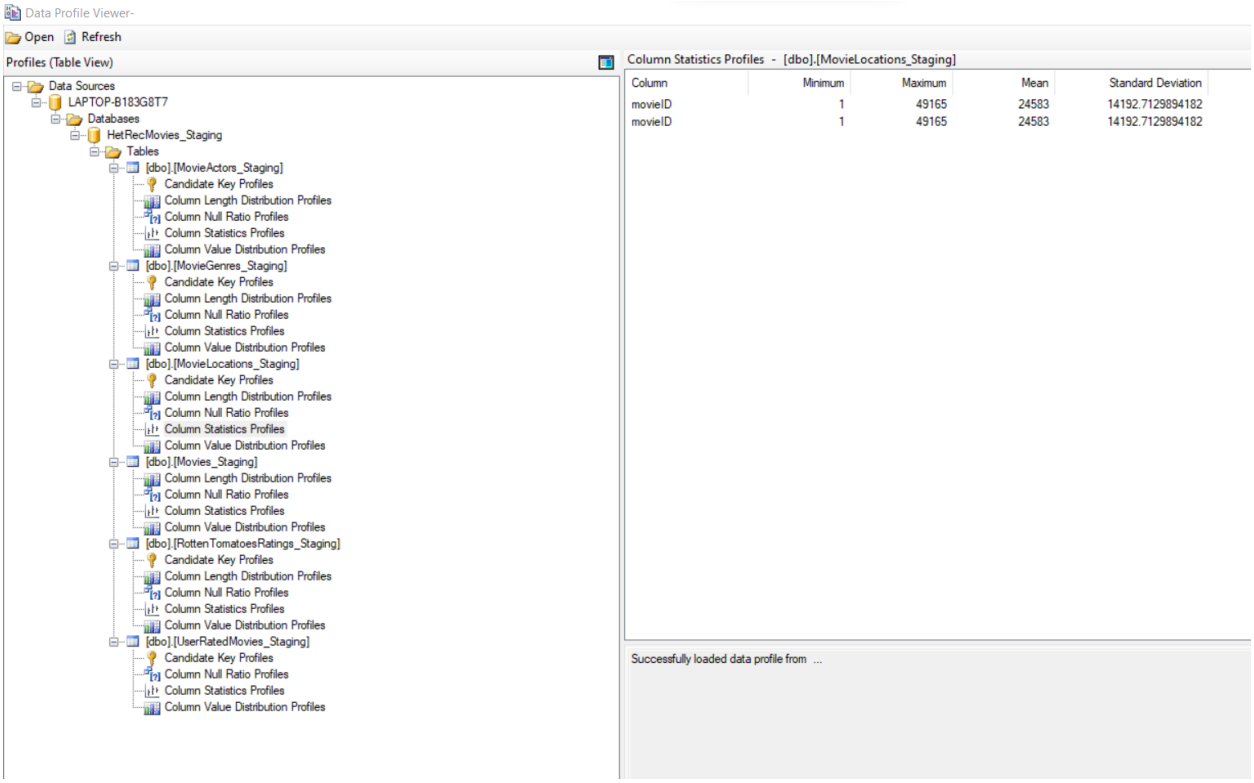


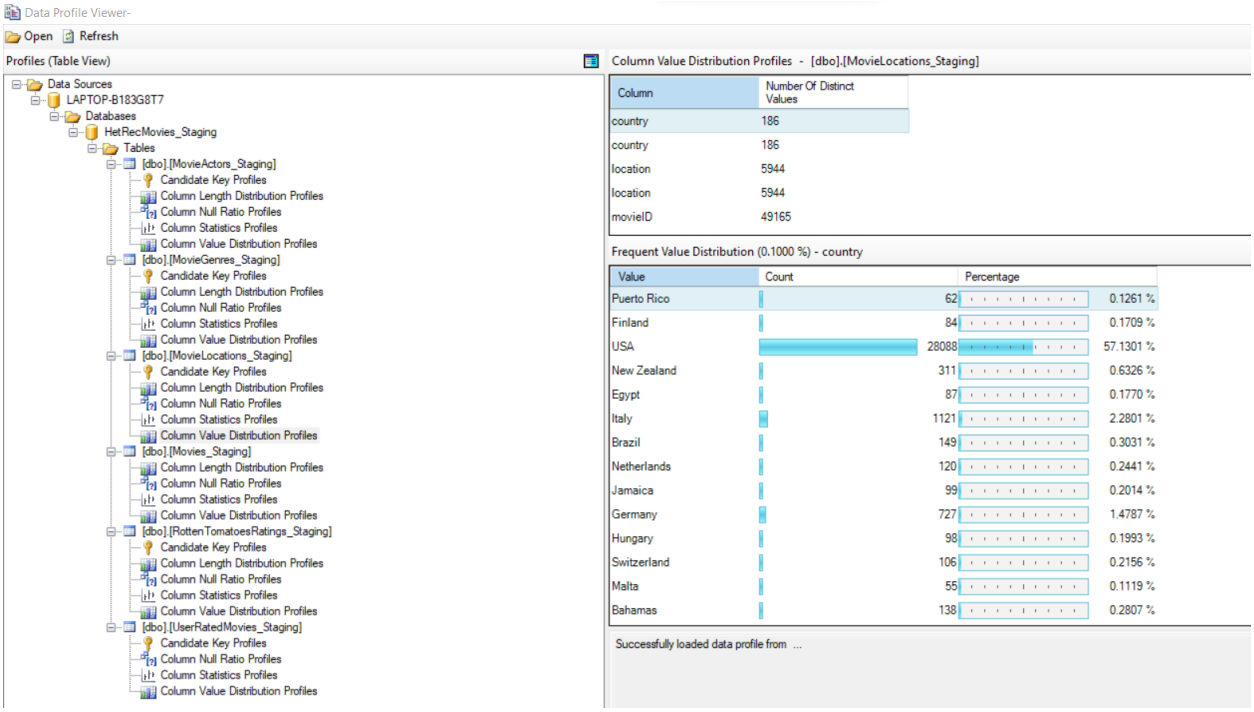
When double click the Data Profiling task and click on Open Profile Viewer to view the analyzed data.






Data Profile Viewer-
Open
Refresh
Profiles (Table View)
Data Sources
LAPTOP-B183G8T7
Databases
HetRecMovies_Staging
Tables
[dbo].[MovieActors_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[MovieGenres_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[MovieLocations_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[Movies_Staging]
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[RottenTomatoesRatings_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[UserRatedMovies_Staging]
Candidate Key Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
Column Null Ratio Profiles - [dbo].[MovieLocations_Staging]
Column
Null Count
Null Percentage
country 1267 2.5770 %
country 1267 2.5770 %
location 6855 13.9428 %
location 6855 13.9428 %
movieID 0 0.0000 %
movieID 0 0.0000 %
state 2628 5.3453 %
state 2628 5.3453 %
Successfully loaded data profile from ...


Data Profile Viewer-
Open
Refresh
Profiles (Table View)
Data Sources
LAPTOP-B183G8T7
Databases
HetRecMovies_Staging
Tables
[dbo].[MovieActors_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[MovieGenres_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[MovieLocations_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[Movies_Staging]
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[RottenTomatoesRatings_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[UserRatedMovies_Staging]
Candidate Key Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
Column Statistics Profiles - [dbo].[MovieLocations_Staging]
Column
Minimum
Maximum
Mean
Standard Deviation
movieID 1 49165 24583 14192.7129894182
movieID 1 49165 24583 14192.7129894182
Successfully loaded data profile from ...

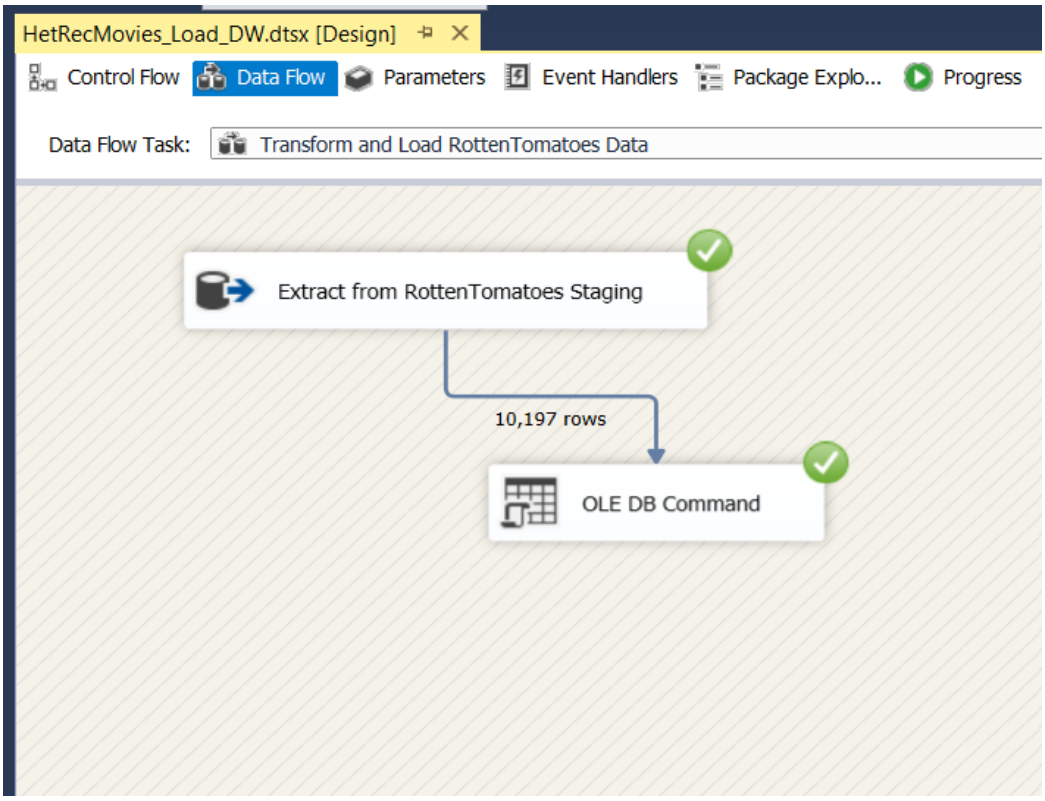

Data Profile Viewer-
Open
Refresh
Profiles (Table View)
Data Sources
LAPTOP-B183G8T7
Databases
HetRecMovies_Staging
Tables
[dbo].[MovieActors_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[MovieGenres_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[MovieLocations_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[Movies_Staging]
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[RottenTomatoesRatings_Staging]
Candidate Key Profiles
Column Length Distribution Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
[dbo].[UserRatedMovies_Staging]
Candidate Key Profiles
Column Null Ratio Profiles
Column Statistics Profiles
Column Value Distribution Profiles
Column Value Distribution Profiles - [dbo].[MovieLocations_Staging]
Column
Number Of Distinct Values
country 186
country 186
location 5944
location 5944
movieID 49165
Frequent Value Distribution (0.1000 %) - country
Value
Count
Percentage
Puerto Rico 62 0.1261 %
Finland 84 0.1709 %
USA 28088 57.1301 %
New Zealand 311 0.6326 %
Egypt 87 0.1770 %
Italy 1121 2.2801 %
Brazil 149 0.3031 %
Netherlands 120 0.2441 %
Jamaica 99 0.2014 %
Germany 727 1.4787 %
Hungary 98 0.1993 %
Switzerland 106 0.2156 %
Malta 55 0.1119 %
Bahamas 138 0.2807 %
Successfully loaded data profile from ...

Data Transformation

Transform and Load Rotten Tomatoes Data

I created Rotten Tomatoes data Transformation by below mentioned steps.

- Created new package called HetRecMovies_Load_DW.dtsx.
- Then Dragged and dropped a Data Flow Task, renamed it as Transform and Load_Rotten Tomatoes Data_and go the Data Flow tab.
- Dragged and dropped OLE DB Source, renamed as Extract from RottenTomatoes Staging and configure it to access the Rotten Tomatoes Staging table.
- After that I dragged and dropped OLE DB Command rotten tomatoes and connect the OLE DB source_In the OLE DB Command, I set the configurations as below.



First, I have created a procedure called **UpdateDimRottenTomatoes** and executed in the HetRecMovies_DW database.
OLE DB Command SSIS tool used to execute, **UpdateDimRottenTomatoes procedure**, it is used to insert data into rotten tomatoes staging to DimRottenTomatoes without data duplication.

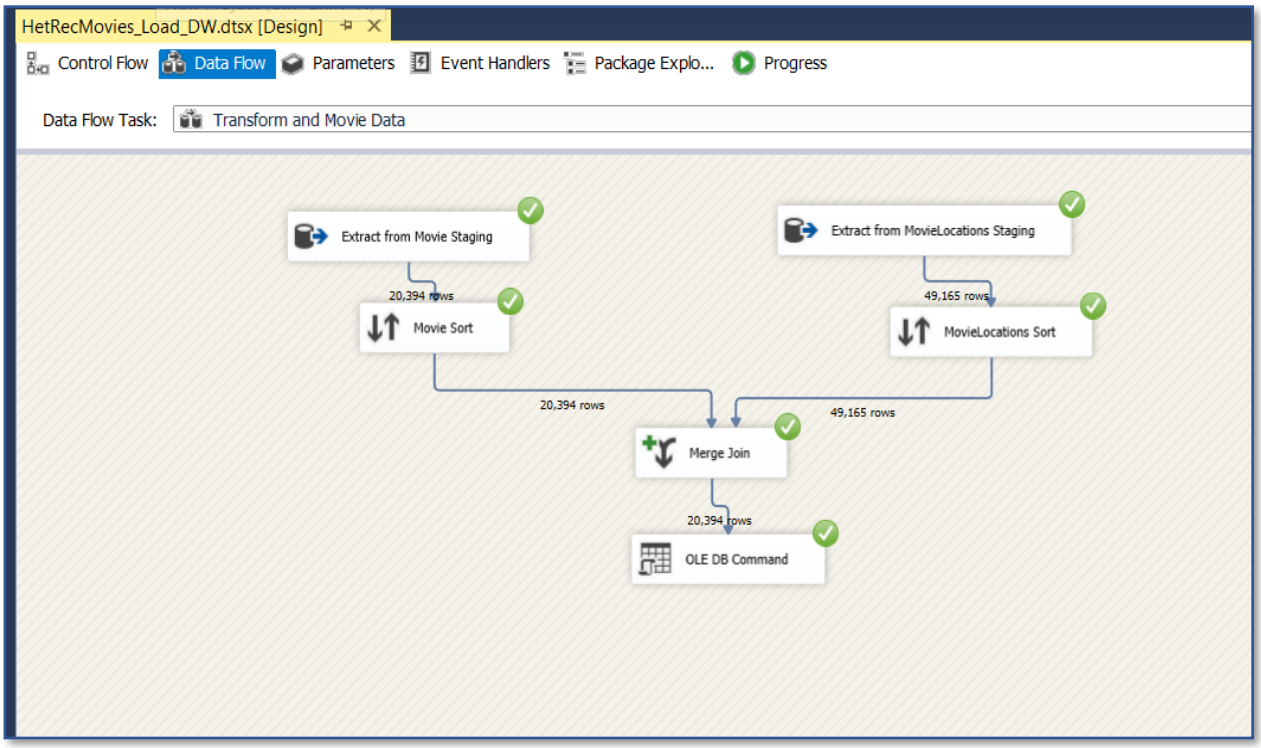
SQLQuery2.sql - LAP...3G8T7\Thilini (66))* | SQLQuery1.sql - not connected* | LAPTOP-B183G8T7....DW - dbo.DimDate

```
CREATE PROCEDURE [dbo].[UpdateDimRottenTomatoes]
@rottenTomatoesID int,
@movieTitle nvarchar(100),
@rtAllCriticsRating numeric(18, 1),
@rtAllCriticsNumReviews int,
@rtTopCriticsRating numeric(18, 1),
@rtTopCriticsNumReviews int,
@rtAudienceRating numeric(18, 1),
@rtAudienceNumRatings int

AS
BEGIN
if not exists (select rottenTomatoesSK
from [dbo].[DimRottenTomatoes]
where rottenTomatoes_AltID = @rottenTomatoesID)
BEGIN
insert into dbo.DimRottenTomatoes
(rottenTomatoes_AltID,movieTitle,rtAllCriticsRating,rtAllCriticsNumReviews,rtTopCriticsRating,rtTopCriticsNumReviews,rtAudienceRating,rtAudienceNumRatings,InsertDate,ModifiedDate)
values
(@rottenTomatoesID,@movieTitle,@rtAllCriticsRating,@rtAllCriticsNumReviews,@rtTopCriticsRating,@rtTopCriticsNumReviews,@rtAudienceRating,@rtAudienceNumRatings,GETDATE(), GETDATE())
END;
if exists (select rottenTomatoesSK
from dbo.DimRottenTomatoes
where rottenTomatoes_AltID = @rottenTomatoesID)
BEGIN
update dbo.DimRottenTomatoes
set rottenTomatoes_AltID = @rottenTomatoesID,
movieTitle = @movieTitle,
rtAllCriticsRating = @rtAllCriticsRating,
rtAllCriticsNumReviews = @rtAllCriticsNumReviews,
rtTopCriticsRating = @rtTopCriticsRating,
rtTopCriticsNumReviews = @rtTopCriticsNumReviews,
rtAudienceRating = @rtAudienceRating,
rtAudienceNumRatings = @rtAudienceNumRatings,
ModifiedDate = GETDATE()
where rottenTomatoes_AltID = @rottenTomatoesID
END;
END;
```

Then I did same process to Movies table and MovieGenres table as well.

Transform and Movie Data

Movies table and movie locations table primary key is movieID therefore, I created **DimMovies** dimension table by connecting Movies_Staging and MovieLocations_Staging.

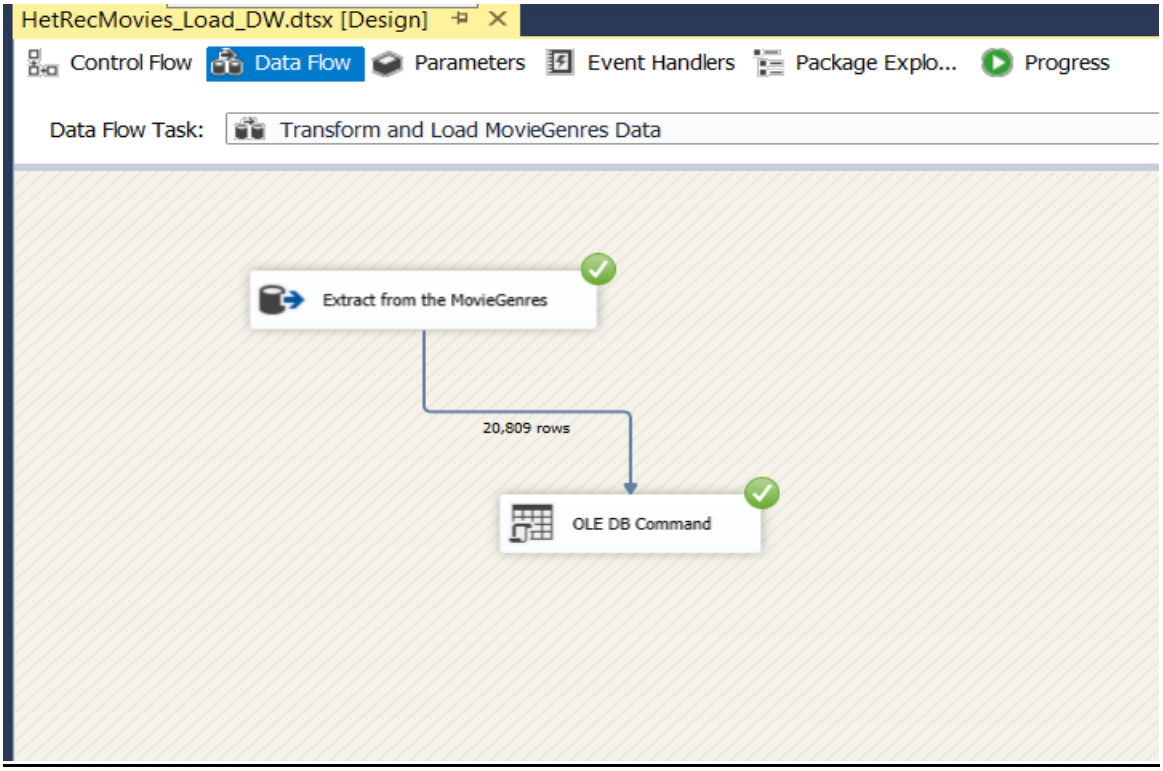


UpdateDimMovies Procedure

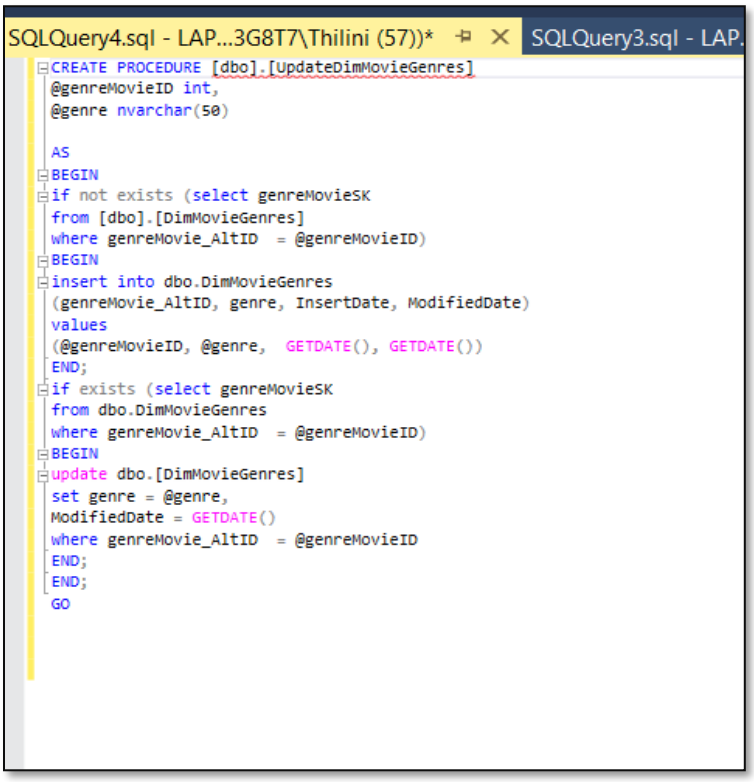
SQLQuery3.sql - LAP...3G8T7\Thilini (66))*

```
CREATE PROCEDURE [dbo].[UpdateDimMovies]
@movieID int,
@title varchar(500),
@imdbID int,
@year int,
@country nvarchar(50),
@state nvarchar(50),
@location nvarchar(50),
@rottenTomatoesID int

AS
BEGIN
if not exists (select movieSK
from [dbo].[DimMovies]
where movie_AltID = @movieID)
BEGIN
insert into dbo.DimMovies
(movie_AltID,title,imdbID,year,country,state,location,rottenTomatoesSK,InsertDate,ModifiedDate)
values
(@movieID,@title,@imdbID,@year,@country,@state,@location,@rottenTomatoesID,GETDATE(), GETDATE())
END;
if exists (select movieSK
from [dbo].[DimMovies]
where movie_AltID = @movieID)
BEGIN
update dbo.DimMovies
set title = @title,
imdbID = @imdbID,
year = @year,
country = @country,
state = @state,
location = @location,
rottenTomatoesSK = @rottenTomatoesID,
ModifiedDate = GETDATE()
where movie_AltID = @movieID
END;
END;
GO
```

Transform and Load MovieGenres Data



UpdateDimMovieGenres Procedure

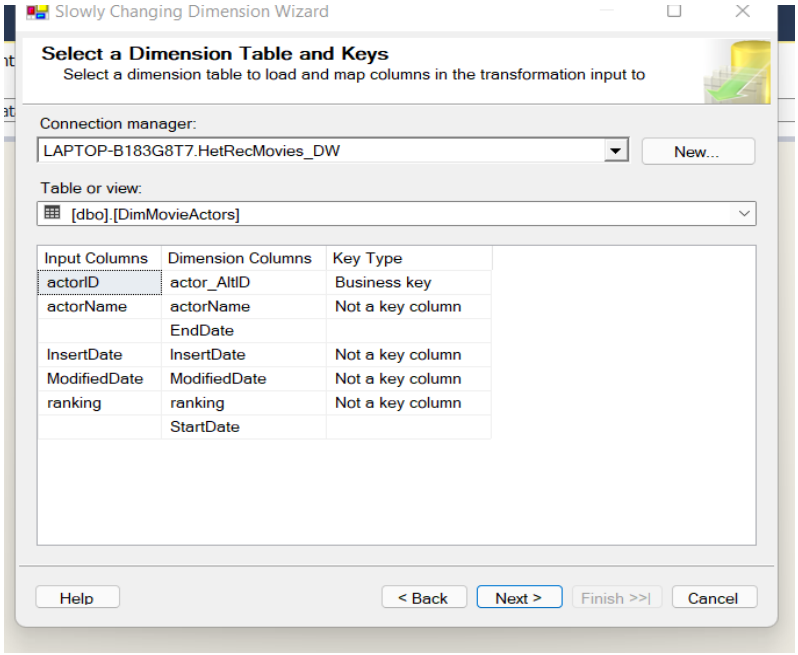


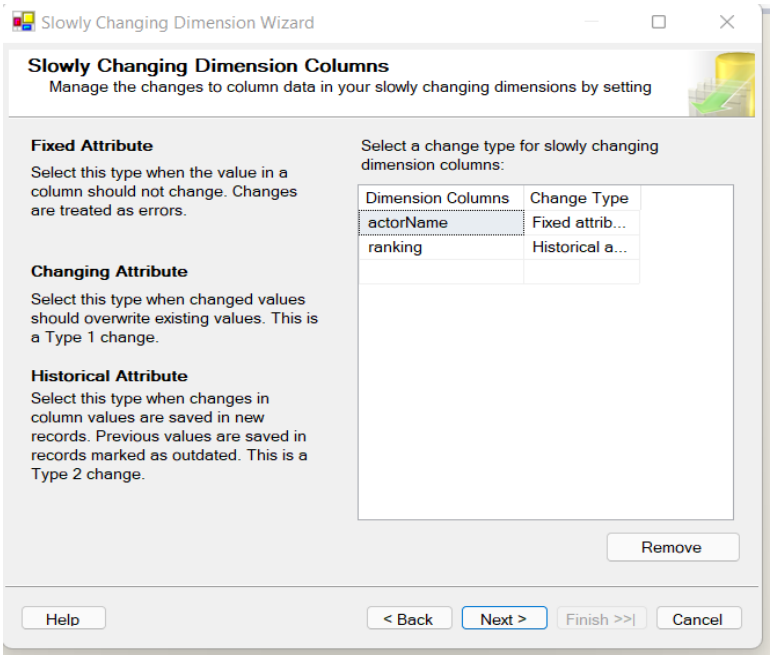

```
CREATE PROCEDURE [dbo].[UpdateDimMovieGenres]
@genreMovieID int,
@genre nvarchar(50)

AS
BEGIN
if not exists (select genreMovieSK
from [dbo].[DimMovieGenres]
where genreMovie_AltID  = @genreMovieID)
BEGIN
insert into dbo.DimMovieGenres
(genreMovie_AltID, genre, InsertDate, ModifiedDate)
values
(@genreMovieID, @genre,  GETDATE(), GETDATE())
END;
if exists (select genreMovieSK
from dbo.DimMovieGenres
where genreMovie_AltID  = @genreMovieID)
BEGIN
update dbo.[DimMovieGenres]
set genre = @genre,
ModifiedDate = GETDATE()
where genreMovie_AltID  = @genreMovieID
END;
END;
GO
```

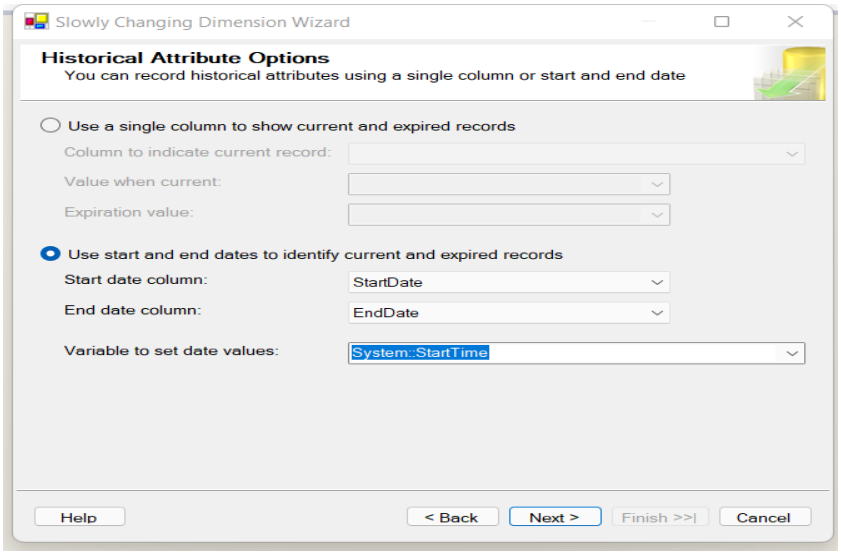
Transform and Load MovieActors Data (Slowly Changing Dimension)

I created Movie Actors Data Transformation by below mentioned steps.

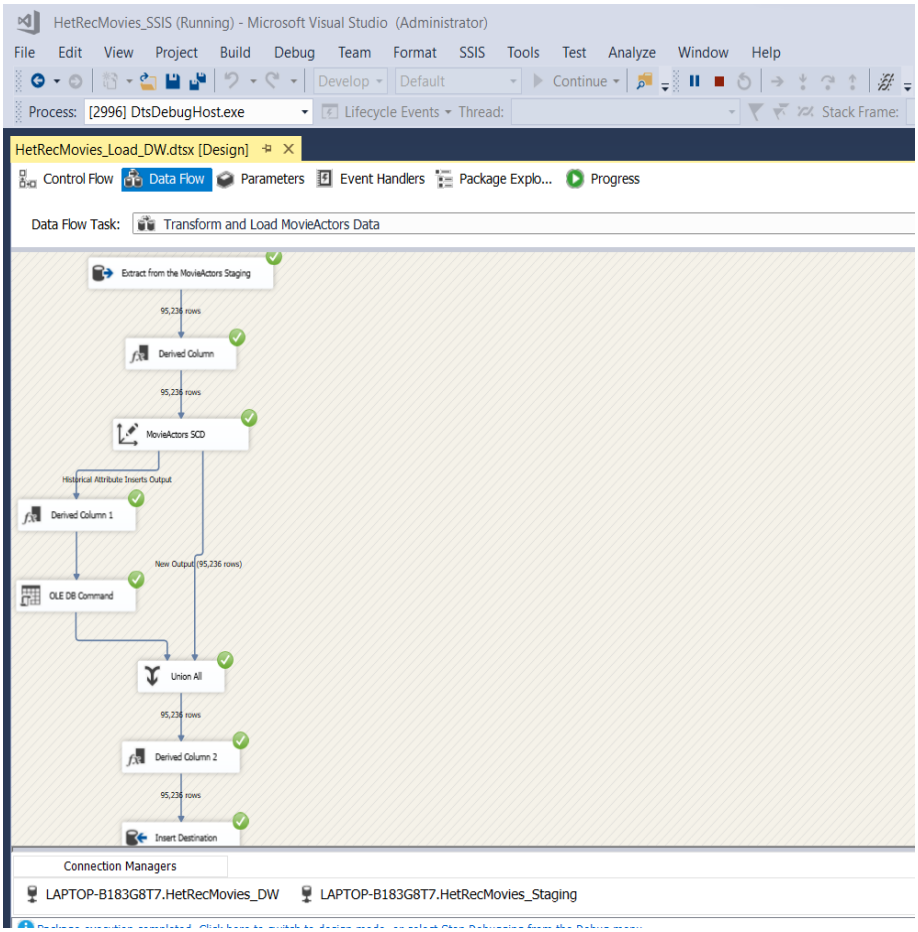
- Created new package called HetRecMovies_Load_DW.dtsx.
- Then dragged and dropped a Data Flow Task, renamed it as Transform and Load Movie Actors data details and go the Data Flow tab.
- Dragged and dropped OLE DB Source, renamed as Extract from Movie Actors Staging and configure it to access the Movie Actors Staging table.
- After that I dragged and dropped Slowly Changing Dimension movie actors and connect the OLE DB source
- In the SCD Configuration Wizard I set the configurations as below.



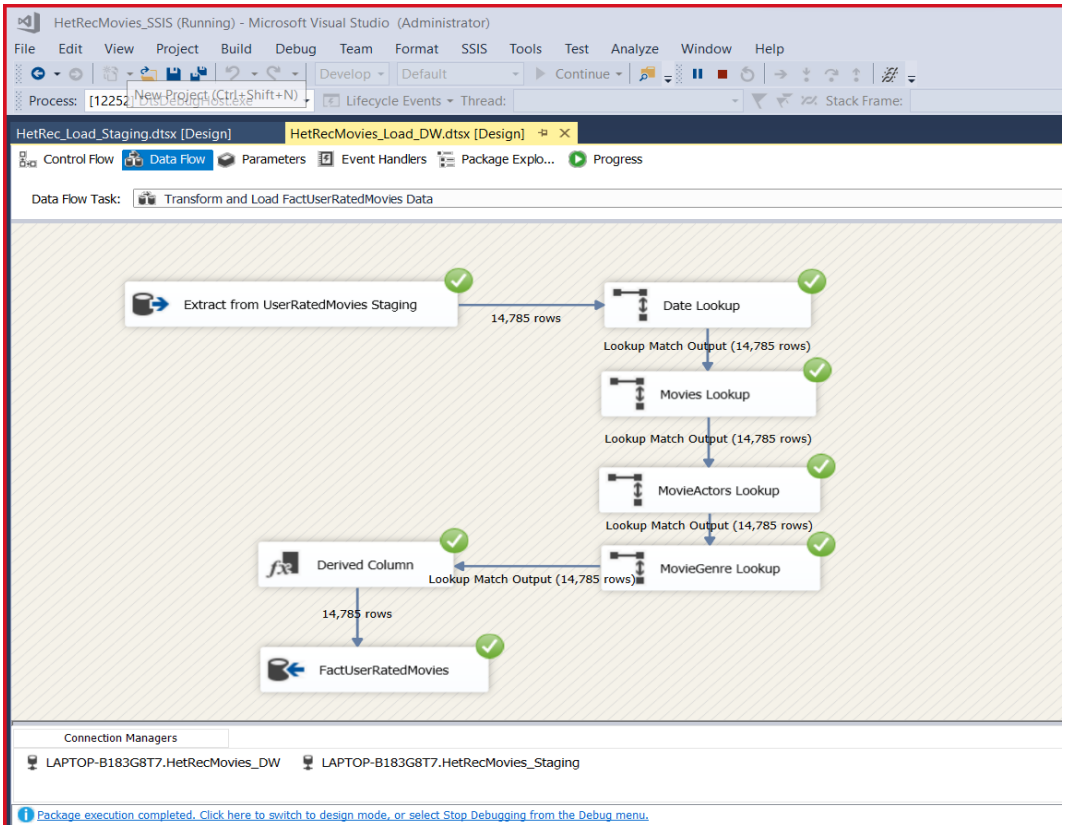




Once all configurations are done properly, it will automatically create the slowly changing dimension as shown below.



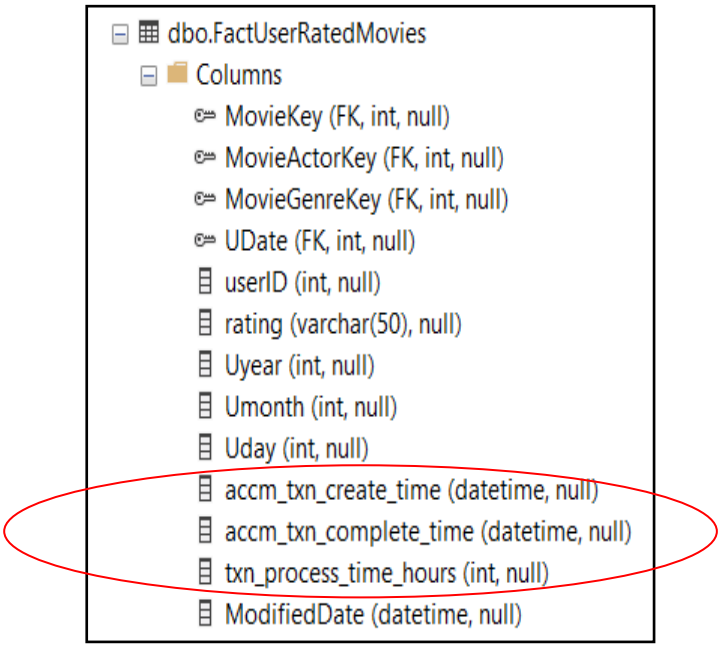
Transform and Load FactUserRatedMovies Data



STEP 06: ETL Development -Accumulating Fact Table

First, I extended my fact table (FactUserRatedMovies Table) with following 03 columns.

accm_txn_create_time
accm_txn_complete_time
txn_process_time_hours

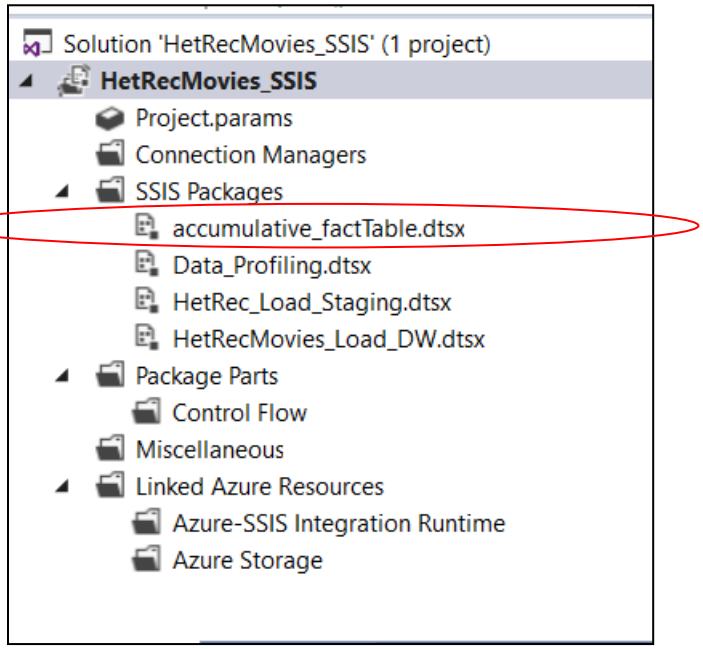


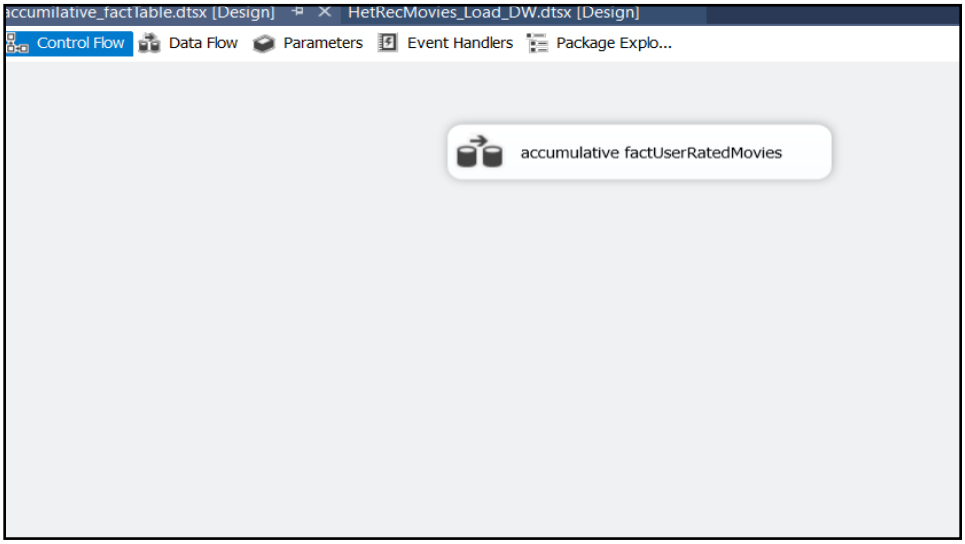
Then I prepared a dataset which contains fact table natural key(userID) and accm_txn_complete time.

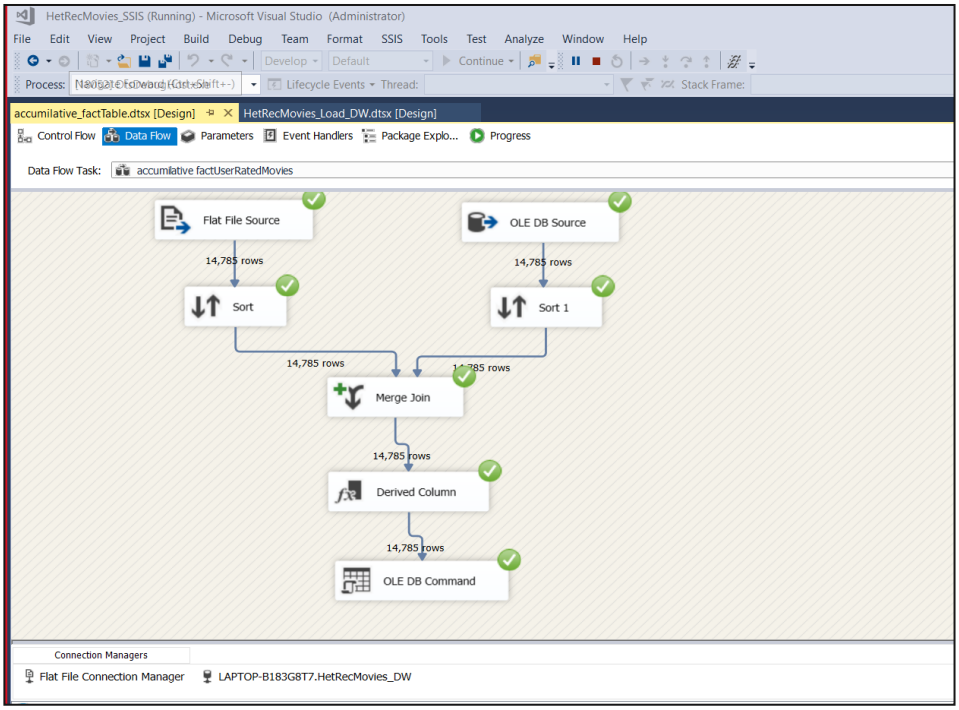
	A	B
1	userID	accm_txn_complete_time
2	1	5/13/2022 9:26
3	2	5/12/2022 19:05
4	3	5/14/2022 5:16
5	4	5/13/2022 14:06
6	5	5/12/2022 22:27
7	6	5/12/2022 7:57
8	7	5/12/2022 7:24
9	8	5/14/2022 6:22
10	9	5/13/2022 8:14
11	10	5/12/2022 3:13
12	11	5/13/2022 5:32
13	12	5/14/2022 20:41
14	13	5/13/2022 21:27
15	14	5/12/2022 16:19
16	15	5/14/2022 4:10
17	16	5/13/2022 2:31
18	17	5/12/2022 8:52
19	18	5/12/2022 21:56
20	19	5/13/2022 0:22
21	20	5/12/2022 14:54
22	21	5/13/2022 13:54
23	22	5/13/2022 20:43
24	23	5/14/2022 4:27
25	24	5/14/2022 1:45
26	25	5/14/2022 21:41
27	26	5/13/2022 9:49
28	27	5/12/2022 23:53

accumulating factTable

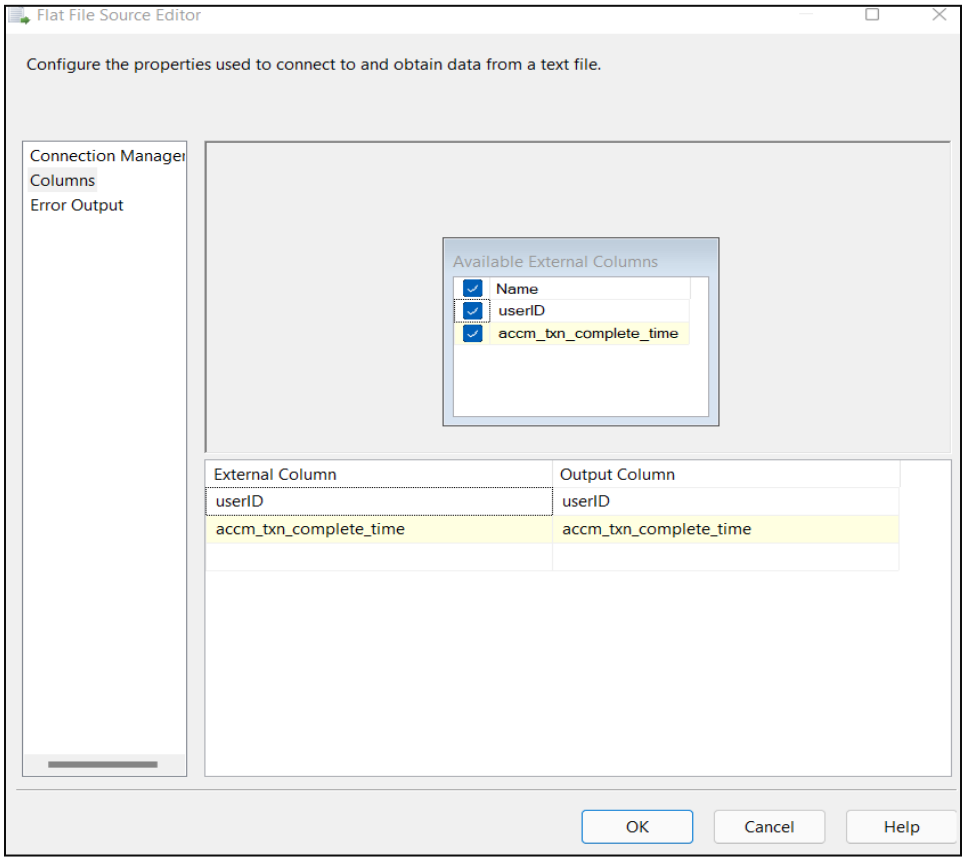
After that, I generated a new ETL SSIS package called accumulative_factTable.dtsx. Which receives data from this file and updates the accm txn complete time in my DW Fact table accordingly.



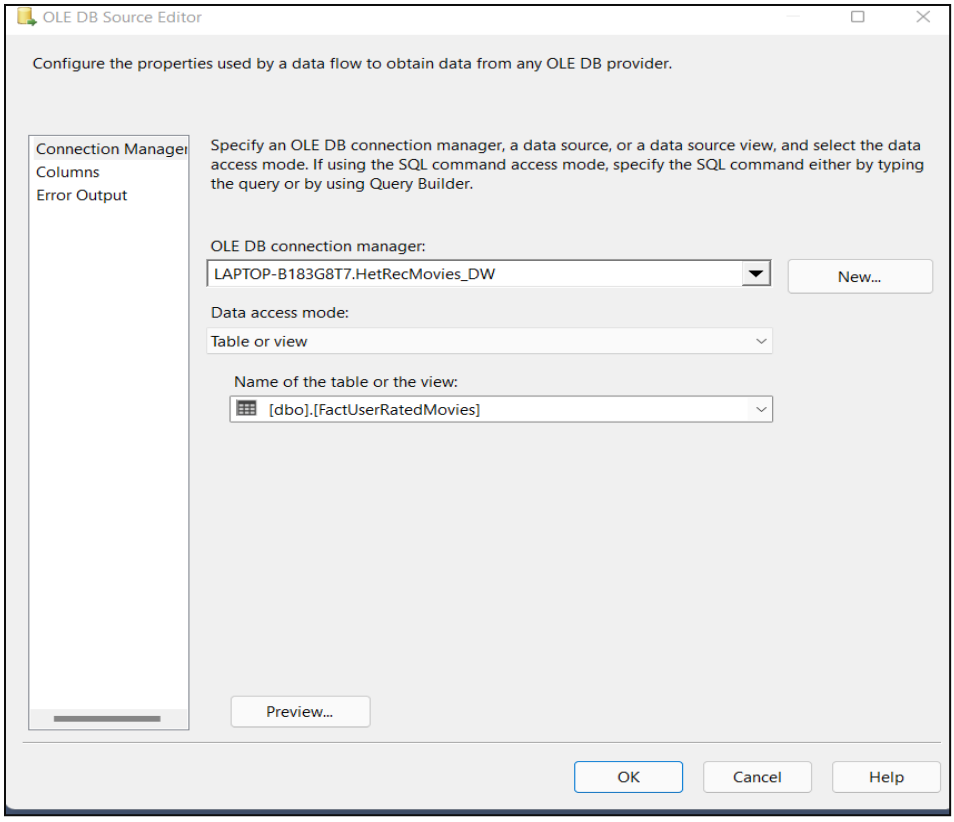


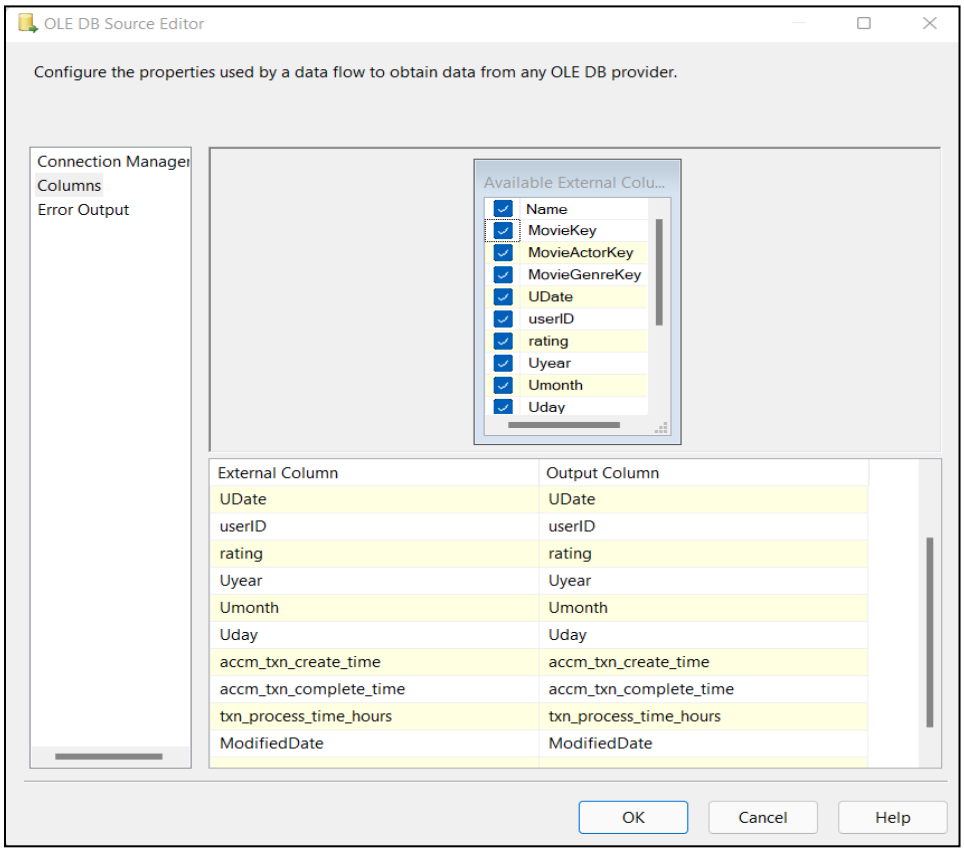


Flat File Source

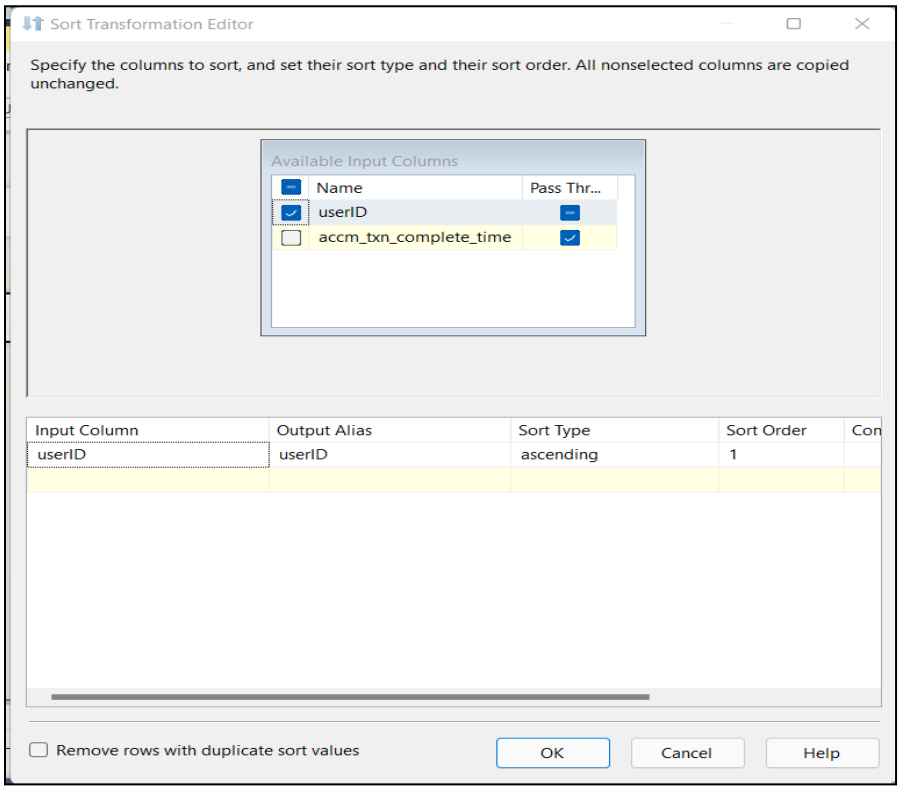


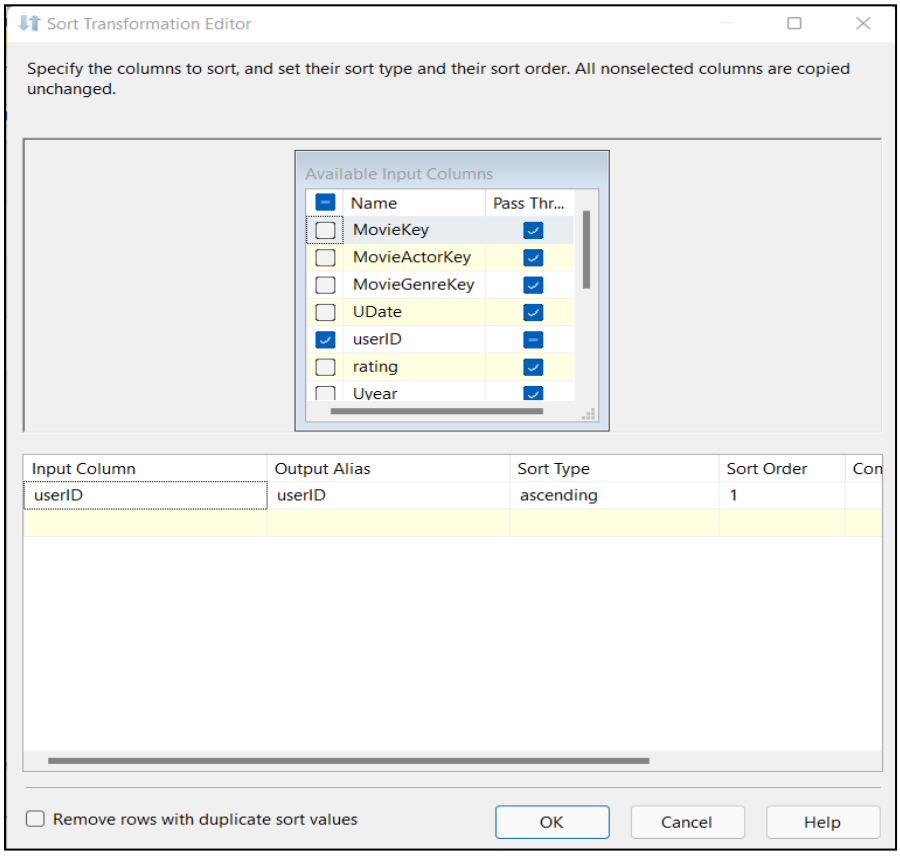
OLE DB Source



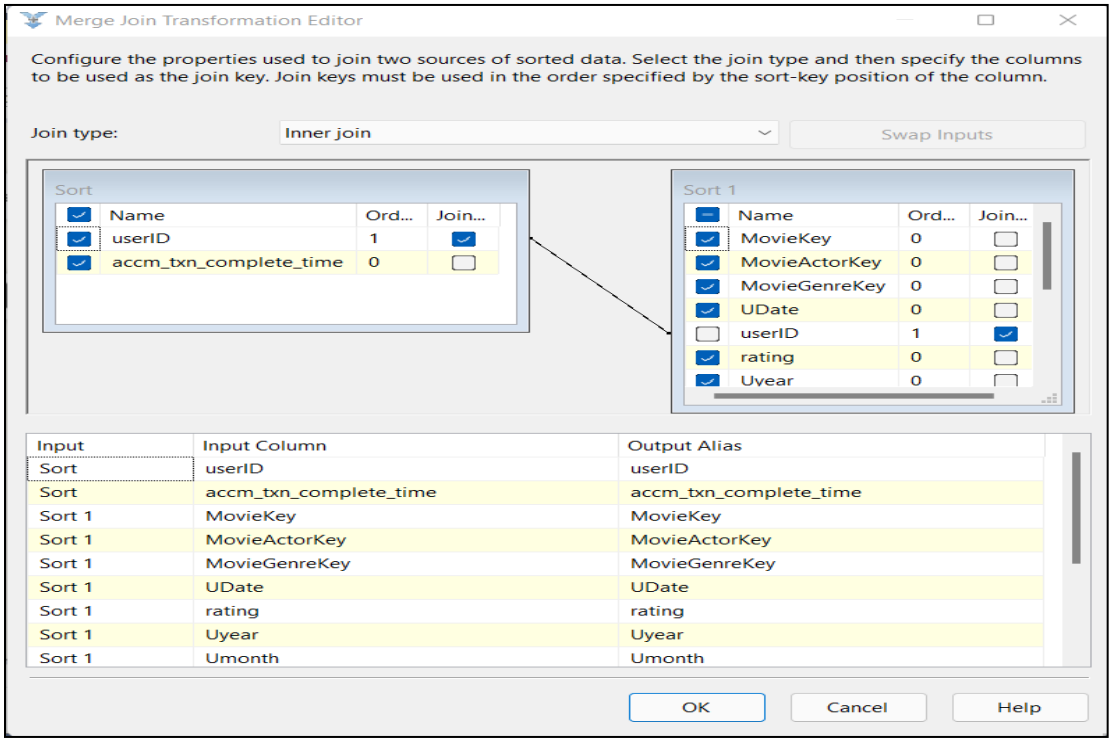


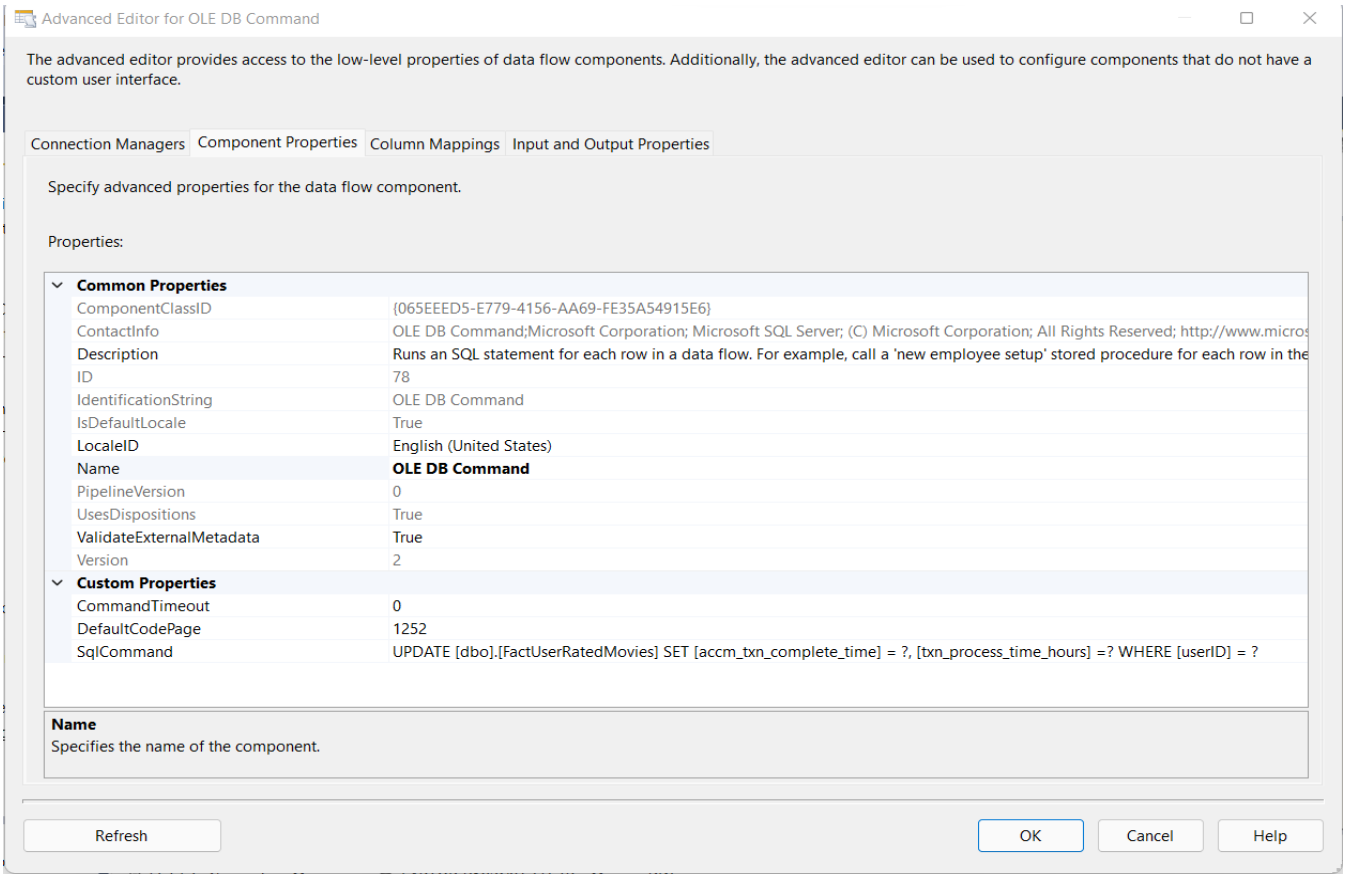
Sort with FactUserRatedMovies natural key. (userID)





Merge Join


Merge Join Transformation Editor
Configure the properties used to join two sources of sorted data. Select the join type and then specify the columns to be used as the join key. Join keys must be used in the order specified by the sort-key position of the column.
Join type:
Inner join
Swap Inputs
Sort
Name
Ord...
Join...
userID
1
accm_txn_complete_time
0
Sort 1
Name
Ord...
Join...
MovieKey
0
MovieActorKey
0
MovieGenreKey
0
UDate
0
userID
1
rating
0
Uyear
0
Input
Input Column
Output Alias
Sort
userID
userID
Sort
accm_txn_complete_time
accm_txn_complete_time
Sort 1
MovieKey
MovieKey
Sort 1
MovieActorKey
MovieActorKey
Sort 1
MovieGenreKey
MovieGenreKey
Sort 1
UDate
UDate
Sort 1
rating
rating
Sort 1
Uyear
Uyear
Sort 1
Umonth
Umonth
OK
Cancel
Help


Advanced Editor for OLE DB Command
The advanced editor provides access to the low-level properties of data flow components. Additionally, the advanced editor can be used to configure components that do not have a custom user interface.
Connection Managers
Component Properties
Column Mappings
Input and Output Properties
Specify advanced properties for the data flow component.
Properties:
Common Properties
ComponentClassID
{065EEED5-E779-4156-AA69-FE35A54915E6}
ContactInfo
OLE DB Command;Microsoft Corporation; Microsoft SQL Server; (C) Microsoft Corporation; All Rights Reserved; http://www.micros
Description
Runs an SQL statement for each row in a data flow. For example, call a 'new employee setup' stored procedure for each row in the
ID
78
IdentificationString
OLE DB Command
IsDefaultLocale
True
LocaleID
English (United States)
Name
OLE DB Command
PipelineVersion
0
UsesDispositions
True
ValidateExternalMetadata
True
Version
2
Custom Properties
CommandTimeout
0
DefaultCodePage
1252
SqlCommand
UPDATE [dbo].[FactUserRatedMovies] SET [accm_txn_complete_time] = ?, [txn_process_time_hours] =? WHERE [userID] = ?
Name
Specifies the name of the component.
Refresh
OK
Cancel
Help

Final FactUserRatedMovies Table with Updated txn_process_time.

Results Messages

	MovieKey	MovieActorKey	MovieGenreKey	UDate	userID	rating	Uyear	Umonth	Uday	accm_txn_create_time	accm_txn_complete_time	txn_process_time_hours	ModifiedDate
1	3	66326	1	20061029	1	1.0	2006	10	29	2022-05-11 20:30:00.713	2022-05-13 09:26:00.000	37	2022-05-11 20:30:00.713
2	32	66327	2	20061029	2	4.5	2006	10	29	2022-05-11 20:30:00.713	2022-05-12 19:05:00.000	23	2022-05-11 20:30:00.713
3	106	66328	3	20061029	3	4.0	2006	10	29	2022-05-11 20:30:00.713	2022-05-14 05:16:00.000	57	2022-05-11 20:30:00.713
4	152	66329	4	20061029	4	2.0	2006	10	29	2022-05-11 20:30:00.713	2022-05-13 14:06:00.000	42	2022-05-11 20:30:00.713
5	155	66330	5	20061029	5	4.0	2006	10	29	2022-05-11 20:30:00.713	2022-05-12 22:27:00.000	26	2022-05-11 20:30:00.713
6	157	66331	6	20061029	6	4.5	2006	10	29	2022-05-11 20:30:00.713	2022-05-12 07:57:00.000	11	2022-05-11 20:30:00.713
7	165	66332	7	20061029	7	3.5	2006	10	29	2022-05-11 20:30:00.713	2022-05-12 07:24:00.000	11	2022-05-11 20:30:00.713
8	286	66333	8	20061029	8	5.0	2006	10	29	2022-05-11 20:30:00.713	2022-05-14 06:22:00.000	58	2022-05-11 20:30:00.713
9	341	66334	9	20061029	9	3.5	2006	10	29	2022-05-11 20:30:00.713	2022-05-13 08:14:00.000	36	2022-05-11 20:30:00.713
10	404	66335	10	20061029	10	2.0	2006	10	29	2022-05-11 20:30:00.713	2022-05-12 03:13:00.000	7	2022-05-11 20:30:00.713
11	567	66336	11	20061029	11	4.0	2006	10	29	2022-05-11 20:30:00.713	2022-05-13 05:32:00.000	33	2022-05-11 20:30:00.713
12	624	66337	12	20061029	12	3.0	2006	10	29	2022-05-11 20:30:00.713	2022-05-14 20:41:00.000	72	2022-05-11 20:30:00.713
13	767	66338	13	20061029	13	4.5	2006	10	29	2022-05-11 20:30:00.713	2022-05-13 21:27:00.000	49	2022-05-11 20:30:00.713
14	840	66339	14	20061029	14	.5	2006	10	29	2022-05-11 20:30:00.713	2022-05-12 16:19:00.000	20	2022-05-11 20:30:00.713
15	910	66340	15	20061029	15	4.5	2006	10	29	2022-05-11 20:30:00.713	2022-05-14 04:10:00.000	56	2022-05-11 20:30:00.713
16	950	66341	16	20061029	16	4.0	2006	10	29	2022-05-11 20:30:00.713	2022-05-13 02:31:00.000	30	2022-05-11 20:30:00.713
17	1029	66342	17	20061029	17	3.5	2006	10	29	2022-05-11 20:30:00.713	2022-05-12 08:52:00.000	12	2022-05-11 20:30:00.713
18	1104	66343	18	20061029	18	4.5	2006	10	29	2022-05-11 20:30:00.713	2022-05-12 21:56:00.000	25	2022-05-11 20:30:00.713
19	1121	66344	19	20061029	19	4.0	2006	10	29	2022-05-11 20:30:00.713	2022-05-13 00:22:00.000	28	2022-05-11 20:30:00.713
20	1190	66345	20	20061029	20	2.5	2006	10	29	2022-05-11 20:30:00.713	2022-05-12 14:54:00.000	18	2022-05-11 20:30:00.713

Query executed successfully. LAPTOP-B183G8T7 (15.0 RTM) LAPTOP-B183G8T7\Thilin... HetRecMovies_DW

Final HetRecMovies_Load_DW Control Flow.