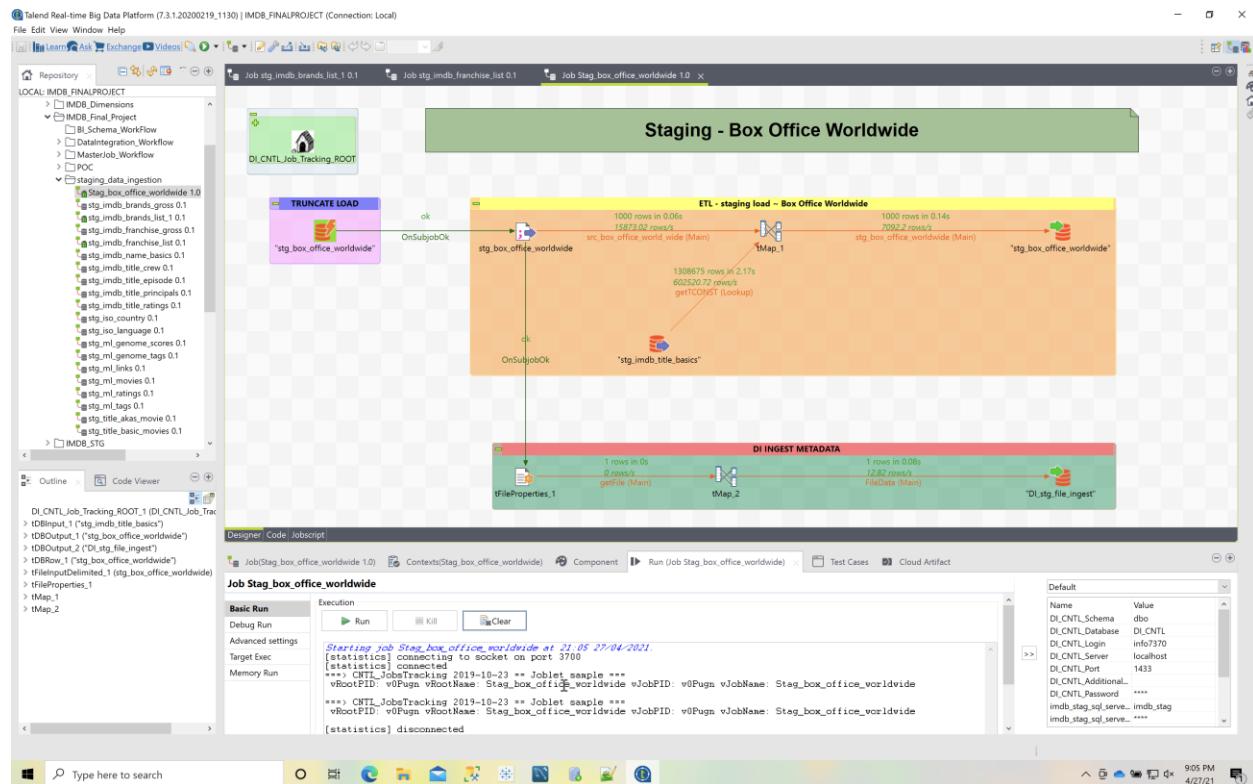


Staging Screen Shots- Talend

1. Stag_box_office_worldwide



SQLQuery1.sql - localhost\imdb_stag (info7370 (54)) - Microsoft SQL Server Management Studio (Administrator)

```
select * from [dbo].[stg_box_office_worldwide]
```

Results Messages

box_office_worldwide_sk	BoxOffice_Rank	Score	Title	Worldwide_LifetimeGross	Domestic_LifetimeGross	Domestic_Pct	Foreign_LifetimeGross	Foreign_Pct	Release_Year	SDS_Input	SDS_SK	Qd_Cache_ID	Qd_Cache_Dt
1	1	2	Avengers: Endgame	2797000000000	8887200000	0.307	1930627044	0.693	2019	www09	9	N4XHu	2021-04-27 20:42:26.710
2	2	1775309	Avatar	2794300000000	76057625	0.272	2029631736	0.728	2009	www09	9	N4XHu	2021-04-27 20:42:26.710
3	3	2	Titanic	69593944	0.3	1535075988	0.7	1997	www09	9	N4XHu	2021-04-27 20:42:26.710	
4	4	4	Star Wars: Episode VI - The Force Awakens	26438542	39986204	0.453	1336544272	0.547	2015	www09	9	N4XHu	2021-04-27 20:42:26.710
5	5	6	Star Wars: Episode VII - The Last Jedi	204835974	678915482	0.331	1336544272	0.669	2016	www09	9	N4XHu	2021-04-27 20:42:26.710
6	6	6	Star Wars: Episode VIII - The Last Jedi	10589610	65279655	0.39	1018130011	0.61	2015	www09	9	N4XHu	2021-04-27 20:42:26.710
7	7	7	Jurassic World	168443394	84800282	0.328	1100000001	0.672	2019	www09	9	N4XHu	2021-04-27 20:42:26.710
8	8	8	The Avengers	151812398	82357910	0.41	85845078	0.59	2012	www09	9	N4XHu	2021-04-27 20:42:26.710
9	9	9	Furious 7	151041871	36307020	0.233	116234065	0.767	2015	www09	9	N4XHu	2021-04-27 20:42:26.710
10	10	10	Frozen	148020833	47737878	0.329	97205335	0.671	2019	www09	9	N4XHu	2021-04-27 20:42:26.710
11	11	11	Avengers: Age of Ultron	142000000	45992588	0.329	1000000000	0.672	2015	www09	9	N4XHu	2021-04-27 20:42:26.710
12	12	12	Black Panther	134691161	70006966	0.52	64488395	0.48	2018	www09	9	N4XHu	2021-04-27 20:42:26.710
13	13	13	Harry Potter and the Deathly Hallows - Part 2	134193238	38140910	0.284	96052308	0.716	2011	www09	9	N4XHu	2021-04-27 20:42:26.710
14	14	14	Star Wars: Episode V - The Empire Strikes Back	125210000	62118089	0.487	1000000000	0.688	2017	www09	9	N4XHu	2021-04-27 20:42:26.710
15	15	15	Jurassic World: Fallen Kingdom	10481806	417719760	0.319	89074814	0.661	2018	www09	9	N4XHu	2021-04-27 20:42:26.710
16	16	16	Frozen	4070788	12808232	0.313	88006427	0.687	2013	www09	9	N4XHu	2021-04-27 20:42:26.710
17	17	17	Spider-Man: Far From Home	103910000	26421712	0.293	999999999	0.597	2019	www09	9	N4XHu	2021-04-27 20:42:26.710
18	18	18	Incredibles 2	124280539	60581744	0.49	634223815	0.51	2018	www09	9	N4XHu	2021-04-27 20:42:26.710
19	19	19	The Fate of the Furious	120800518	22608385	0.183	10099673	0.817	2017	www09	9	N4XHu	2021-04-27 20:42:26.710
20	20	20	Avengers: Endgame	121441281	4000000000	0.301	1000000000	0.693	2019	www09	9	N4XHu	2021-04-27 20:42:26.710
21	21	21	Men in Black: International	115936204	336045770	0.29	82336297	0.71	2019	www09	9	N4XHu	2021-04-27 20:42:26.710
22	22	22	Captain America: Civil War	115226293	40864349	0.354	745211944	0.646	2016	www09	9	N4XHu	2021-04-27 20:42:26.710
23	23	23	Aquaman	114481707	33001807	0.292	813400000	0.708	2018	www09	9	N4XHu	2021-04-27 20:42:26.710
24	24	24	The Lord of the Rings: The Return of the King	102874405	37380269	0.328	700000000	0.669	2003	www09	9	N4XHu	2021-04-27 20:42:26.727
25	25	25	Spider-Man: Far From Home	1131927996	39053208	0.348	741395911	0.655	2019	www09	9	N4XHu	2021-04-27 20:42:26.727
26	26	26	Captain Marvel	11516464	112374794	0.378	701444955	0.622	2019	www09	9	N4XHu	2021-04-27 20:42:26.727
27	27	27	Star Wars: Episode IX - The Rise of Skywalker	113073979	52328393	0.374	741395911	0.661	2019	www09	9	N4XHu	2021-04-27 20:42:26.727
28	28	28	Skyfall	10704583	304860277	0.278	804200738	0.725	2012	www09	9	N4XHu	2021-04-27 20:42:26.727

Query executed successfully.

Time:

SQLQuery3.sql - loc.cntl (info7370 (53)) - SQLQuery1.sql - loc.tag (info7370 (54))*

```
SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM    DI_CNTL.dbo.job_stats_vw
where job_name = 'Stag_box_office_worldwide'
```

Results Messages

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
Stag_box_office_worldwide	Y	Stag_box_office_worldwide	success	0.06	2021-04-27 21:05:08.400	end	IMDB_FINALPROJECT	1.0

DI INJECT:

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the 'imdb_stag' database which contains the 'DI_stg_file_Ingest' table. The central pane displays a query window with the following SQL code:

```
select * from DI_stg_file_Ingest
```

The results pane shows one row of data from the 'DI_stg_file_Ingest' table:

FileGroup	FileName	FileTowm	RowInserted	DI_JobID	DI_CreateDate
1	BOX OFFICE World_Wide_C:\INFO 7370\dwh-spring2021\odbc\FinalSource...	1000	1000	NHKuAR	2021-04-27 21:11:41.677

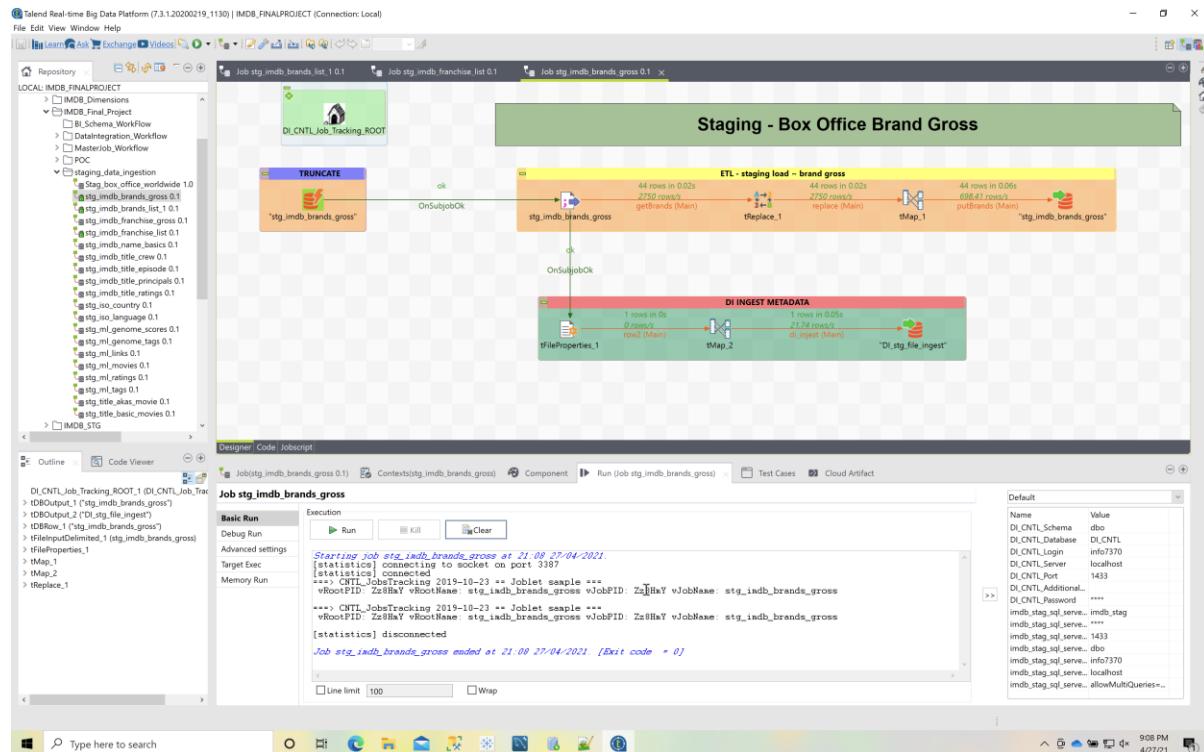
At the bottom of the screen, the status bar indicates "localhost (15.0 RTM) info7370 (55) imdb_stag 00:00:00 1 rows".

data consistency: data is loaded from file to stag table, data remains consistent with sk values and referential integrity

reject: for this table we are not supposed to do any rejects.

structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component and adding cols using tmap for create_date and DI_ETL_ID which is process id, other than that we have also completed bonus part where in we are populating DI_stg_inject table.

2. stg_imdb_brands_gross



The screenshot shows the SQL Server Management Studio interface with the following details:

- File:** SQLQuery1.sql - localhost.imdb_stag (info7370 (54)) - Microsoft SQL Server Management Studio (Administrator)
- Object Explorer:** Shows the database structure including tables like 'stg_imdb.brands.gross'.
- SQL Query Editor:**

```
select * from [dbo].[stg_imdb_brands_gross]
```
- Results:**

Brand	Gross_Mkt	Brand_Name	Total_Revenue	Number_of_Releases	Top_Release	Lifetime_Gross	SRK_SK	PL_Ahead	DI_Create_DT
1	Marvel Comics	Marvel Comics	142233926	60	Avengers Endgame	888739000	13	3hr60m	2021-04-27 20:57:47 943
2	Legendary Pictures	Legendary Pictures	691814262	55	Jurassic World	65270825	13	3hr60m	2021-04-27 20:57:47 943
3	Lucfilm	Lucfilm	862311262	55	Star Wars: Episode VII - The Force Awakens	65270825	13	3hr60m	2021-04-27 20:57:47 943
4	Pixar	Pixar	607957062	38	Incredibles 2	60889144	13	3hr60m	2021-04-27 20:57:47 943
5	DC Comics	DC Comics	571315101	44	The Dark Knight	533543568	13	3hr60m	2021-04-27 20:57:47 943
6	DreamWorks Animation	DreamWorks Animation	568803427	37	Sherk 2	441252647	13	3hr60m	2021-04-27 20:57:47 943
7	Buena Vista	Buena Vista	307295021	58	Star Wars: Episode VII - The Force Awakens	441252647	13	3hr60m	2021-04-27 20:57:47 943
8	Virgo Entertainment	Virgo Entertainment	305734802	40	k	327481748	13	3hr60m	2021-04-27 20:57:47 943
9	Walt Disney Animation Studios	Walt Disney Animation Studios	26497078	12	Frozen	47737378	13	3hr60m	2021-04-27 20:57:47 943
10	Blumhouse Productions	Blumhouse Productions	20480000	10	The Secret Life of Pets	32964060	13	3hr60m	2021-04-27 20:57:47 943
11	Hasbro	Hasbro	203083981	47	Get Out	17604066	13	3hr60m	2021-04-27 20:57:47 943
12	Habibi	Habibi	205121499	16	Transformers: Revenge of the Fallen	402111870	13	3hr60m	2021-04-27 20:57:47 943
13	Nickelodeon	Nickelodeon	192037000	15	Transformers: Dark of the Moon	395111870	13	3hr60m	2021-04-27 20:57:47 943
14	Sony Pictures Animation	Sony Pictures Animation	191638063	21	Spider-Man: Into the Spider-Verse	180243110	13	3hr60m	2021-04-27 20:57:47 943
15	Walt Disney Media	Walt Disney Media	184282163	38	The Chronicles of Narnia: The Lion, the Witch & the Wardrobe	291709867	13	3hr60m	2021-04-27 20:57:47 943
16	Blue Sky	Blue Sky	179428000	13	Ice Age: Dawn of the Dinosaurs	291709867	13	3hr60m	2021-04-27 20:57:47 943
17	Stephan King	Stephan King	172964890	69	I	32841748	13	3hr60m	2021-04-27 20:57:47 943
18	MTV	MTV	152046530	36	The Longest Yard	18119464	13	3hr60m	2021-04-27 20:57:47 943
19	Platinum Dunes	Platinum Dunes	121639893	17	Teenage Mutant Ninja Turtles	181204754	13	3hr60m	2021-04-27 20:57:47 943
20	Saturday Night Live - Alumni Debut	Saturday Night Live - Alumni Debut	10000000	30	Brooklyn Nine-Nine	181204754	13	3hr60m	2021-04-27 20:57:47 943
21	Dark Horse Comics	Dark Horse Comics	94761338	16	300	210944939	13	3hr60m	2021-04-27 20:57:47 943
22	Tim Burton/Johnny Depp	Tim Burton/Johnny Depp	88995830	8	Alice in Wonderland	33479111	13	3hr60m	2021-04-27 20:57:47 943
23	Warner Bros. Animation Group	Warner Bros. Animation Group	79563795	7	The LEGO Movie	32964060	13	3hr60m	2021-04-27 20:57:47 943
24	Tyler Perry	Tyler Perry	79563362	16	Madagascar	90583338	13	3hr60m	2021-04-27 20:57:47 943
25	CBS Films	CBS Films	652624187	29	Scary Stories to Tell in the Dark	68447075	13	3hr60m	2021-04-27 20:57:47 943
26	John Goodman	John Goodman	64889142	10	The Polar Express	68447075	13	3hr60m	2021-04-27 20:57:47 943
27	Robert Ludlum	Robert Ludlum	645491918	6	The Bourne Ultimatum	227471070	13	3hr60m	2021-04-27 20:57:47 943
28	Nicholas Sparks	Nicholas Sparks	57472629	11	The Notebook	8100787	13	3hr60m	2021-04-27 20:57:47 943

TIME

The screenshot shows the SSMS interface with the following details:

- Object Explorer:** Shows the database structure under "localhost (SQL Server 15.0.2080.9 - info7370)".
- SQL Query Editor:** Contains the following T-SQL query:

```
SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM    DI_CNTL.dbo.job_stats_vw
where   job_name = 'stg_imdb_brands_gross'
```
- Results Pane:** Displays the output of the query:

	job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
1	stg_imdb_brands_gross	Y	stg_imdb_brands_gross	success	0.01	2021-04-27 21:13:20.017	end	IMDB_FINALPROJECT	0.1

DI_INJECT:

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the 'imdb_stag' database and its tables. The central pane displays a query results grid for the 'DI_stg_file_Ingest' table. The results show two rows of data, both corresponding to 'BOX OFFICE World Wide'. The columns include File Group, FileName, FileRows, RowInserted, DI_JobID, and DI_CreateDate. The first row has File Group '1', FileName 'C:\INFO 7370\dwh-spring2021-cdive\FinalSource...', FileRows '1000', RowInserted '1000', DI_JobID 'NNKuAR', and DI_CreateDate '2021-04-27 21:11:41.677'. The second row has File Group '2', FileName 'C:\INFO 7370\dwh-spring2021-cdive\FinalSource...', FileRows '44', RowInserted '44', DI_JobID 'FPzHO', and DI_CreateDate '2021-04-27 21:13:20.017'. The status bar at the bottom indicates 'localhost (15.0 RTM) info7370 (55) imdb_stag 00:00:00 2 rows'.

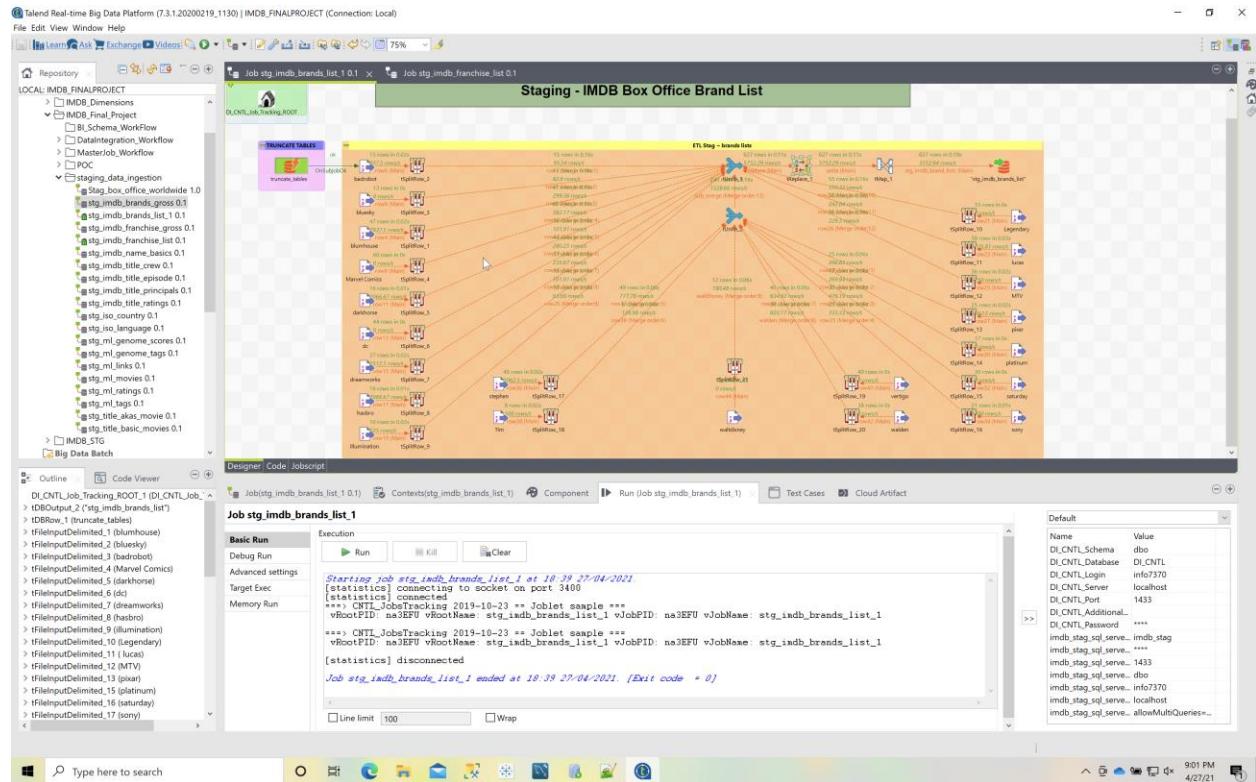
File Group	FileName	FileRows	RowInserted	DI_JobID	DI_CreateDate
1	C:\INFO 7370\dwh-spring2021-cdive\FinalSource...	1000	1000	NNKuAR	2021-04-27 21:11:41.677
2	C:\INFO 7370\dwh-spring2021-cdive\FinalSource...	44	44	FPzHO	2021-04-27 21:13:20.017

data consistency: data is loaded from file to stag table, data remains consistent with sk values and referential integrity.

reject: for this table we are not supposed to do any rejects.

structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component and adding cols using tmap for create_date and DI_ETL_ID which is process id, other than that we have also completed bonus part where in we are populating DI_stg_injest table.

3. stg_imdb_brands_list



SQLQuery1.sql - localhost.imdb.stag (info7370 (54)) - Microsoft SQL Server Management Studio (Administrator)

File Edit View Query Project Tools Window Help

Object Explorer SQLQuery1.sql - loc.tag (info7370 (54)) *

Connect to... localhost (SQL Server 15.0.2080.9 - info7370) New Query Execute Find Refresh Home Help Exit

select * from [dbo].[stg_imdb_brands_list]

Results Messages

Brands_List_SK	Brand_Name	Release_Rank	Release_Name	Udeme_Gross	Max_Theaters	Opening_Gross	Open_Theaters	Release_Date	GenreList	SOR_SK	Di_Jakob	Di_Chow	DT
1	Bad Robot	1	Star Wars: Episode VII - The Force Awakens	9496225	2114	24796075	4136	2015-12-18 00:00:00.000	Walt Disney Studios Motion Pictures	na3FV	2021-04-27 18:39:28.713		
2	Bad Robot	2	Star Wars: Episode IX - The Rise of Skywalker	51520342	4408	17738384	4056	2019-12-20 00:00:00.000	Walt Disney Studios Motion Pictures	na3FV	2021-04-27 18:39:28.713		
3	Bad Robot	3	Star Trek	25730019	4053	7524239	3849	2009-05-08 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
4	Bad Robot	4	Star Trek: Darkness	25019701	7007	7007009	4386	2009-05-08 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
5	Bad Robot	5	Mission Impossible - Fallout	220189104	4395	61236534	4386	2018-07-27 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
6	Bad Robot	6	Mission Impossible - Ghost Protocol	20897003	3655	12785204	425	2011-12-16 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
7	Bad Robot	7	Mission Impossible - Rogue Nation	19833200	3988	8039899	3979	2015-07-10 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
8	Bad Robot	8	Star Trek Beyond	158484240	3628	54253211	3528	2016-07-22 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
9	Bad Robot	9	Super 8	12700479	3424	35411616	3378	2011-06-10 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
10	Bad Robot	10	Cloverfield	80548433	3411	40068229	3411	2008-01-18 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
11	Bad Robot	11	10 Cloverfield Lane	7238200	2447	24772373	2393	2016-03-18 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
12	Bad Robot	12	Morning Glory	31011732	2364	9203266	2816	2010-11-10 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
13	Bad Robot	13	Jay Ride	2194749	2522	7347259	2467	2001-10-05 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
14	Bad Robot	14	Overlord	21704843	2389	12029268	2389	2018-08-10 00:00:00.000	Paramount Pictures	1	na3FV	2021-04-27 18:39:28.713	
15	Bad Robot	15	Infinay Polar Bear	1420658	110	47000	5	2015-09-19 00:00:00.000	Sony Pictures Classics	1	na3FV	2021-04-27 18:39:28.713	
16	Blue_Sky	1	Ice Age: Dawn of the Dinosaurs	19857205	4102	4198052	4098	2009-07-01 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
17	Blue_Sky	2	Ice Age: The Meltdown	19853200	3999	6822000	4098	2008-07-01 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
18	Blue_Sky	3	Ice Age	17537405	3345	46312454	3318	2002-03-19 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
19	Blue_Sky	4	Ice Age: Continental Drift	161321843	3888	46629259	3881	2013-07-10 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
20	Blue_Sky	5	Horton Hears a Who!	16423400	3981	40082000	3981	2008-07-01 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
21	Blue_Sky	6	Rio	14591969	3542	39275942	3426	2011-04-18 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
22	Blue_Sky	7	Rio 2	131938435	3975	3927889	3948	2014-04-11 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
23	Blue_Sky	8	The Peanuts Movie	130781411	3902	44213073	3887	2015-11-06 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
24	Blue_Sky	9	Rio 3	132000000	3778	36007171	3583	2018-07-10 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
25	Blue_Sky	10	Epic	10751882	3884	33531056	3882	2013-05-24 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
26	Blue_Sky	11	Fernleaf	84410580	9830	13461956	5671	2017-12-15 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
27	Blue_Sky	12	Spiral in Disguise	47393700	13022	13888888	13022	2019-07-10 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	
28	Blue_Sky	13	Ice Age: Collision Course	64063000	3997	21370944	3952	2016-07-22 00:00:00.000	Twentieth Century Fox	1	na3FV	2021-04-27 18:39:28.713	

Query executed successfully.

localhost (15.0 RKM) | info7370 (54) | imdb_stag | 0000000 | 627 rows

TIME:

```

SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM    DI_CNTL.dbo.job_stats_vw
where job_name = 'stg_imdb_brands_list_1'
  
```

	job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
1	stg_imdb_brands_list_1	Y	stg_imdb_brands_list_1	success	0.02	2021-04-27 18:39:28.853	end	IMDB_FINALPROJECT	0.1

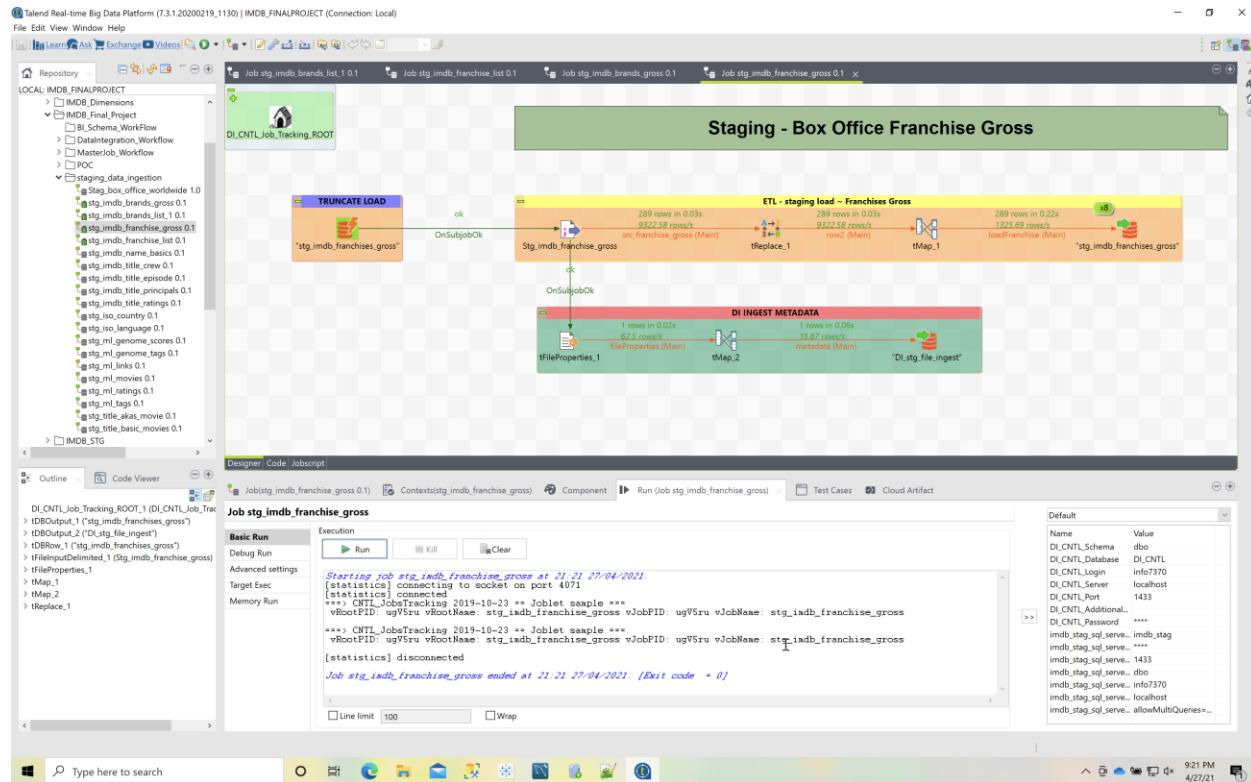
data consistency: data is loaded from multiple files, to make consistent data we used

tReplace

reject: for this table we are not supposed to do any rejects.

structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component, tUnite to merge all the files into a single result set table also, added cols using tmap for create_date and DI_ETL_ID which is process id, other than that we have also completed bonus part where in we are populating DI_stg_injest table.

4. stg_imdb_franchise_gross



Object Explorer

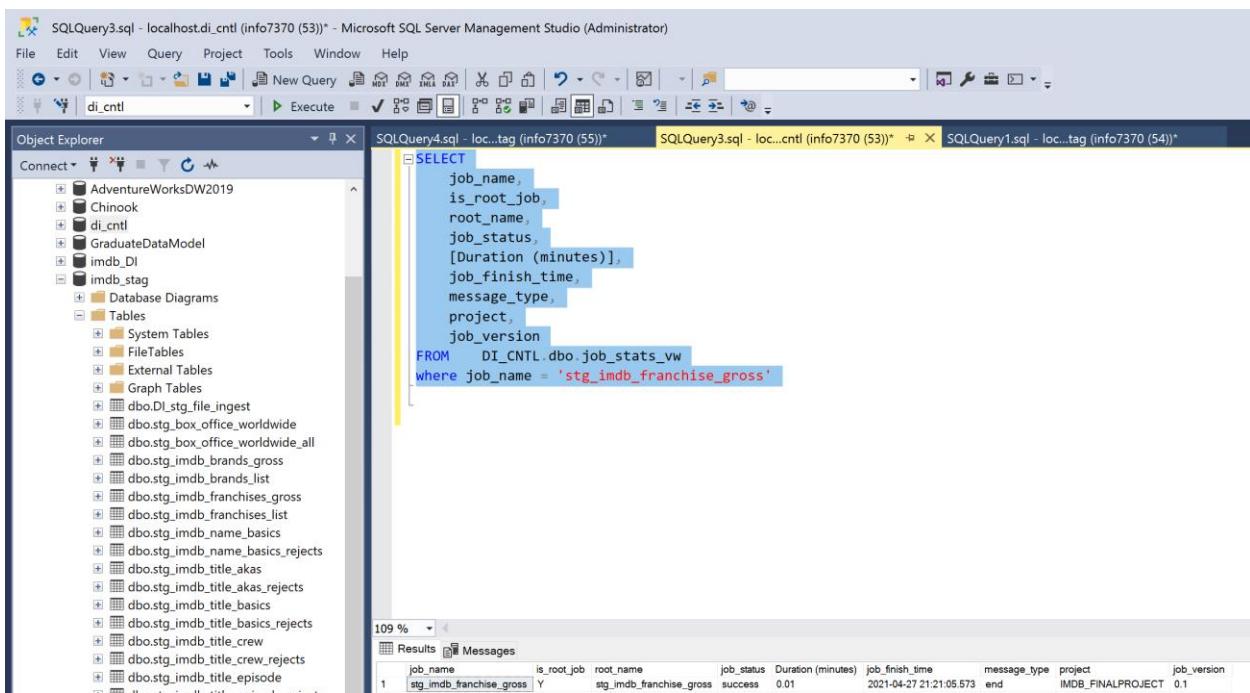
Results

Messages

Franchise	Gross	Total Revenue	Number_of_Releases	Top Releases	Length_Gross	SR_RK	D_Rank	D_Cause
Warner Bros. Entertainment	86479220	12	1	Star Wars: Episode I - The Phantom Menace	16873000	13	ug95ru	2021-04-27 21:21:05.397
Star Wars	6079221177	20	1	Star Wars: Episode VII - The Force Awakens	53866225	13	ug95ru	2021-04-27 21:21:05.397
Disney Live Action Remakes	5407223881	15	1	The Lion King	54388043	13	ug95ru	2021-04-27 21:21:05.397
J.K. Rowling's Wizarding World	27885008	26	1	Harry Potter and the Deathly Hallows: Part 2	387011219	13	ug95ru	2021-04-27 21:21:05.397
Average	249988220	2	1	Avengers: Endgame	403706375	13	ug95ru	2021-04-27 21:21:05.397
Spider-Man	249988191	9	1	Spider-Man: Homecoming	363070709	13	ug95ru	2021-04-27 21:21:05.397
X-Men	243469607	13	1	Deadpool	412956349	13	ug95ru	2021-04-27 21:21:05.397
Batman	241191500	19	1	The Dark Knight	387011219	13	ug95ru	2021-04-27 21:21:05.397
Harry Potter	239334705	24	1	Harry Potter and the Deathly Hallows: Part 2	412956349	13	ug95ru	2021-04-27 21:21:05.397
DC Extended Universe	214785350	8	1	Wonder Woman	412956349	13	ug95ru	2021-04-27 21:21:05.397
Shrek	17232000	29	1	Spiral	652270625	13	ug95ru	2021-04-27 21:21:05.397
Jurassic Park	183377059	7	1	Jurassic World	377027023	13	ug95ru	2021-04-27 21:21:05.397
Middle Earth	185342231	11	1	The Lord of the Rings: The Return of the King	377027023	13	ug95ru	2021-04-27 21:21:05.397
The Fast and the Furious	18693200	9	1	Furious 7	403706375	13	ug95ru	2021-04-27 21:21:05.397
Titanic	182461801	2	1	Titanic	422118070	13	ug95ru	2021-04-27 21:21:05.397
Pirates of the Caribbean	148178053	5	1	Titanic: Revenge of the Fallen	42311812	13	ug95ru	2021-04-27 21:21:05.397
The Hunger Games	148138826	4	1	Pirates of the Caribbean: Dead Man's Chest	42311812	13	ug95ru	2021-04-27 21:21:05.397
Shrek	14809000	1	1	The Hunger Games: Catching Fire	426686247	13	ug95ru	2021-04-27 21:21:05.397
Star Trek	14013980	14	1	Shrek 2	257730019	13	ug95ru	2021-04-27 21:21:05.397
Twilight	136522347	7	1	Star Trek	300531751	13	ug95ru	2021-04-27 21:21:05.397
Toy Story	133717000	6	1	The Twilight Saga: Eclipse	300531751	13	ug95ru	2021-04-27 21:21:05.397
Despicable Me	122028440	4	1	Toy Story 3	368605385	13	ug95ru	2021-04-27 21:21:05.397
The Dark Knight Trilogy	119225654	6	1	Despicable Me 2	368605385	13	ug95ru	2021-04-27 21:21:05.397
Mission: Impossible	109202200	1	1	The Dark Knight	53234503	13	ug95ru	2021-04-27 21:21:05.397
Superman	113755856	8	1	Mission: Impossible: Fallout	53234503	13	ug95ru	2021-04-27 21:21:05.397
The Lord of the Rings	106643440	9	1	Batman v Superman: Dawn of Justice	332360194	13	ug95ru	2021-04-27 21:21:05.397
Indiana Jones	10400000	3	1	The Lord of the Rings: The Return of the King	377027023	13	ug95ru	2021-04-27 21:21:05.397
Indiana Jones and the Kingdom of the Crystal Skull	94236167	2	1	Iron Man 3	317701119	13	ug95ru	2021-04-27 21:21:05.397

Query executed successfully.

TIME:



The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists several databases and objects, including AdventureWorksDW2019, Chinook, di_cntl, GraduateDataModel, imdb_DL, and imdb_stg. The SQL Query window on the right contains a SELECT statement:

```
SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM   DI_CNTL.dbo.job_stats_vw
where job_name = 'stg_imdb_franchise_gross'
```

The results pane at the bottom shows one row of data:

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
stg_imdb_franchise_gross	Y	stg_imdb_franchise_gross	success	0.01	2021-04-27 21:21:05.573	end	IMDB_FINALPROJECT	0.1

DI INJECT:

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists several databases and tables under the 'imdb_stag' database, including 'di_stg_file_ingest'. The 'di_stg_file_ingest' table has three rows of data:

DI_STG_File_Ingest	File_Group	FileName	FileRows	RowInserted	DI_JobID	DI_CreateDate
1	BOX OFFICE World Wide	C:\INFO 7370\dele-spring2021\cine\FinalSource...	1000	1000	NHkUAR	2021-04-27 21:11:41.677
2	BOX OFFICE Map	C:\INFO 7370\dele-spring2021\cine\FinalSource...	44	44	PPHO	2021-04-27 21:13:20.017
3	BOX OFFICE Map	C:\INFO 7370\dele-spring2021\cine\FinalSource...	389	389	ugVeu	2021-04-27 21:09:37.973

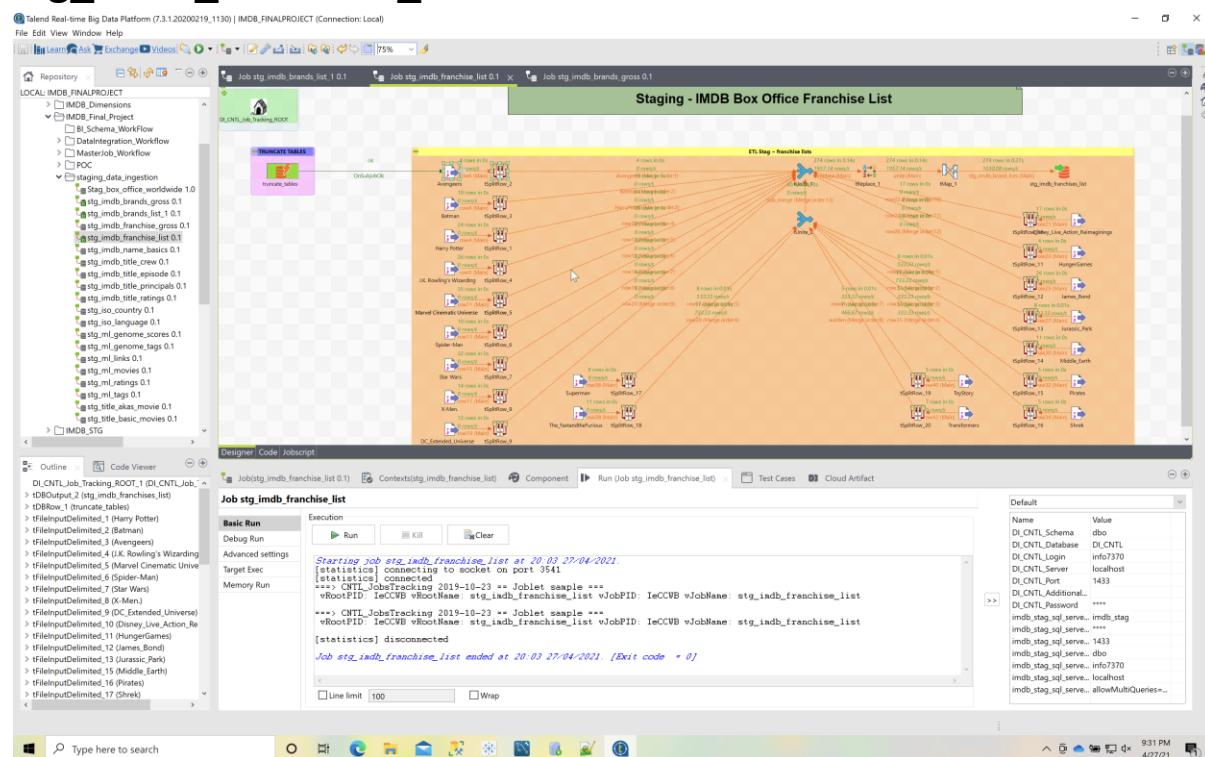
Below the table, a message states: "Query executed successfully."

data consistency: data is loaded from the file, we are using tReplace to make data clean and consistent, identity column making the stg consistent.

reject: for this table we are not supposed to do any rejects.

structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component, we have added cols using tmap for create_date and DI_ETL_ID which is process id, other than that we have also completed bonus part where in we are populating DI_STG_INJECT table.

5. stg_imdb_franchise_list



The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. The top menu bar includes 'File', 'Edit', 'View', 'Query', 'Project', 'Tools', 'Window', and 'Help'. The left pane shows the 'Object Explorer' with the 'imdb_stg' database selected. The central pane displays the results of a query:

```

select * from [dbo].[stg_imdb_franchise_list]

```

The results grid shows the following data (approximate values):

Franchise	Franchise_SK	Franchise	Release_Rank	Release_Name	Lifetime_Gross	Max_Theaters	Opening_Gross	Open_Theaters	Release_Date	Distributor	SOR_SK	DI_JobID	DI_Creation_DT
Avengers	1	Avengers	1	Avengers: Endgame	4862	35719507	4474	4474	2019-04-26 00:00:00	Walt Disney Studios Motion Pictures	1	IeCCWB	2021-04-27 20:00:00
Avengers	2	Avengers	2	Avengers: Infinity War	4474	25798183	4474	4474	2018-04-27 00:00:00	Walt Disney Studios Motion Pictures	1	IeCCWB	2021-04-27 20:00:00
Avengers	3	Avengers	3	Thor: Ragnarok	1235	27043919	1235	1235	2017-10-20 00:00:00	Walt Disney Studios Motion Pictures	1	IeCCWB	2021-04-27 20:00:00
Avengers	4	Avengers	4	Avengers: Age of Ultron	4276	19721109	4276	4276	2016-05-01 00:00:00	Walt Disney Studios Motion Pictures	1	IeCCWB	2021-04-27 20:00:00
Batman	1	Batman	1	The Dark Knight	4366	15841143	4366	4366	2008-07-18 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	2	Batman	2	The Dark Knight Rises	4089	14893903	4089	4089	2012-07-20 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	3	Batman	3	Batman v Superman: Dawn of Justice	4356	16607047	4342	4342	2016-03-25 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	4	Batman	4	Batman v Superman: Dawn of Justice	2201	20409746	2194	2194	2016-03-23 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	5	Batman	5	Batman Begins	4258	20349371	4258	4258	2005-06-17 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	6	Batman	6	Batman Begins	2993	57294433	2842	2842	1995-06-16 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	7	Batman	7	The LEGO Batman Movie	4088	3003486	4088	4088	2017-02-10 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	8	Batman	8	Batman Returns	2844	19257002	2844	2844	2003-07-12 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	9	Batman	9	Batman & Robin	2942	42873605	2934	2934	1997-06-20 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	10	Batman	10	Batman: Mask of the Phantasm	1508	1189975	1508	1508	1993-12-25 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	11	Batman	11	Batman: The Killing Joke	375000	1325	0	0	2016-07-25 00:00:00	Fathom Events	1	IeCCWB	2021-04-27 20:00:00
Batman	12	Batman	12	The Dark Knight Re-take	131033	0	0	0	2012-07-25 00:00:00	Fathom Events	1	IeCCWB	2021-04-27 20:00:00
Batman	13	Batman	13	Batman Begins/2012 Re-release	108658	0	0	0	2012-07-19 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	14	Batman	14	The Dark Knight/2012 Re-release	37589	0	0	0	2018-04-24 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Batman	15	Batman	15	Batman Returns/2012 Re-release	10819	489	10819	489	2018-04-24 00:00:00	Fathom Events	1	IeCCWB	2021-04-27 20:00:00
Batman	16	Batman	16	Batman Returns/2019 Re-release	70642	507	0	0	2019-06-06 00:00:00	Fathom Events	1	IeCCWB	2021-04-27 20:00:00
Batman	17	Batman	17	Batman Returns/2019 Re-release	38014	434	38014	434	2019-06-12 00:00:00	Fathom Events	1	IeCCWB	2021-04-27 20:00:00
Batman	18	Batman	18	Batman Returns/2019 Re-release	23051	429	0	0	2019-06-12 00:00:00	Fathom Events	1	IeCCWB	2021-04-27 20:00:00
Batman	19	Batman	19	Batman: Mask of the Phantasm/2018 Re-release	17813	761	0	0	2018-11-12 00:00:00	Fathom Events	1	IeCCWB	2021-04-27 20:00:00
Harry Potter	24	Harry Potter	1	Harry Potter and the Deathly Hallows: Part 2	4375	16919407	4375	4375	2011-07-18 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Harry Potter	25	Harry Potter	2	Harry Potter and the Chamber of Secrets	31767859	4472	30204674	4472	2001-11-16 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Harry Potter	26	Harry Potter	3	Harry Potter and the Prisoner of Azkaban	30159917	4465	77385727	4325	2001-07-15 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Harry Potter	27	Harry Potter	4	Harry Potter and the Half-Blood Prince	26983356	4125	12601732	4125	2010-10-19 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00
Harry Potter	28	Harry Potter	5	Harry Potter and the Order of the Phoenix	26200478	4285	7718414	4285	2007-07-11 00:00:00	Warner Bros.	1	IeCCWB	2021-04-27 20:00:00

TIME:

The screenshot shows the SSMS interface with three tabs at the top: SQLQuery4.sql - loc...tag (info7370 (55)), SQLQuery3.sql - loc...cntl (info7370 (53)), and SQLQuery1.sql - loc...tag (info7370 (54)). The main area displays a query window with the following code:

```

SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM    DI_CNTL.dbo.job_stats_vw
where job_name = 'stg_imdb_franchise_list'
  
```

The results grid shows one row of data:

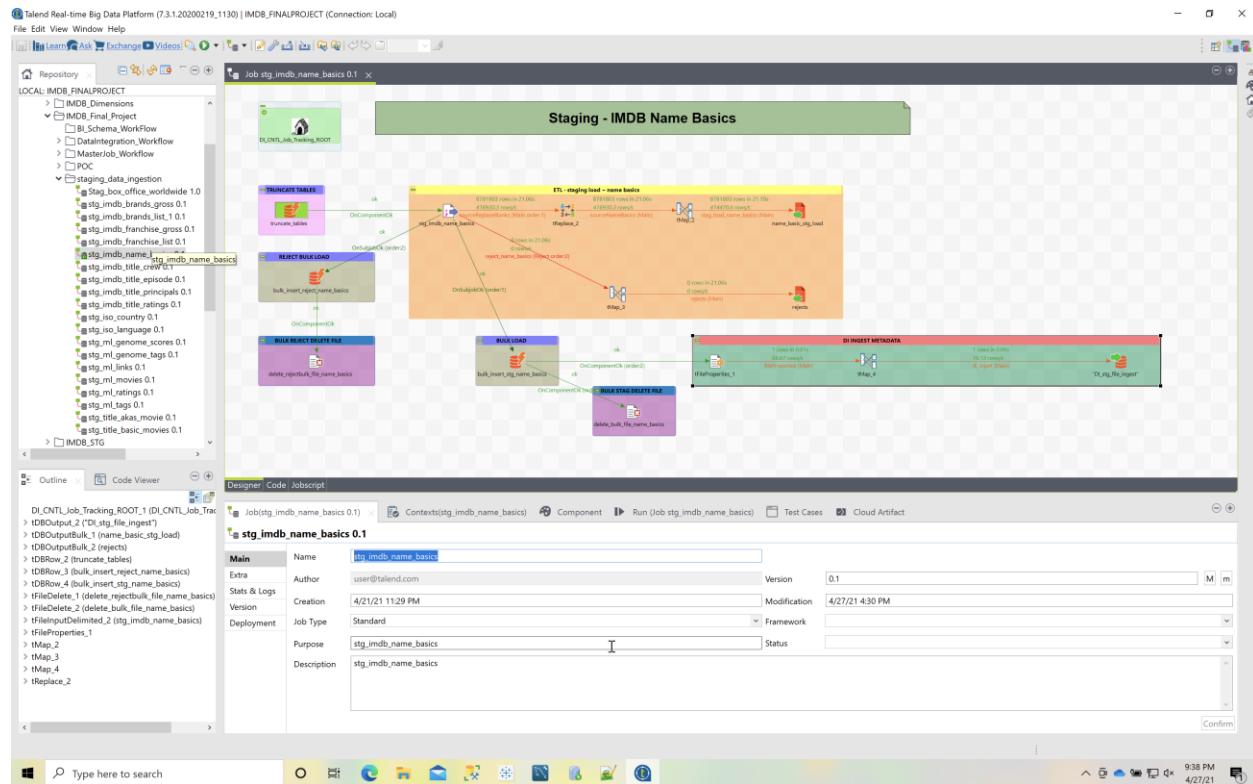
job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
stg_imdb_franchise_list	Y	stg_imdb_franchise_list	success	0.01	2021-04-27 20:03:50.587	end	IMDB_FINALPROJECT	0.1

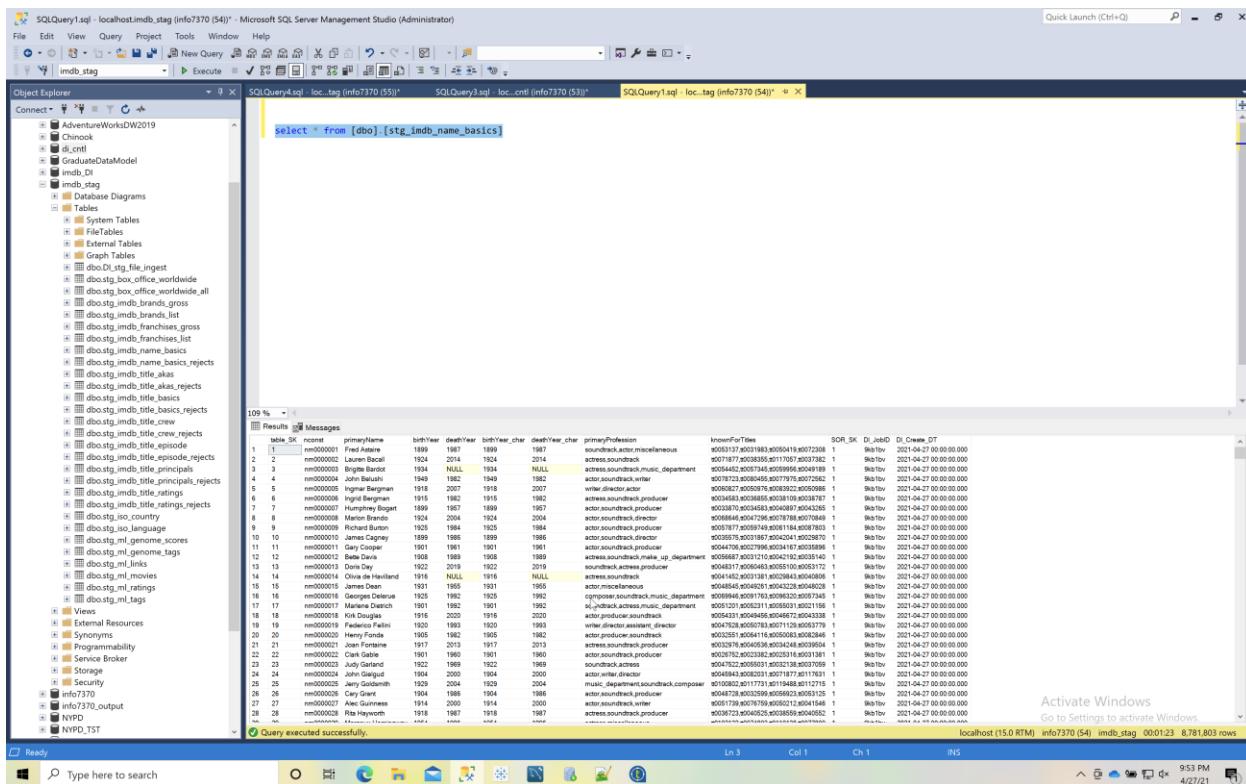
data consistency: data is loaded from multiple files, to make consistent data we have used tReplace component to clean the data.

reject: for this table we are not supposed to do any rejects.

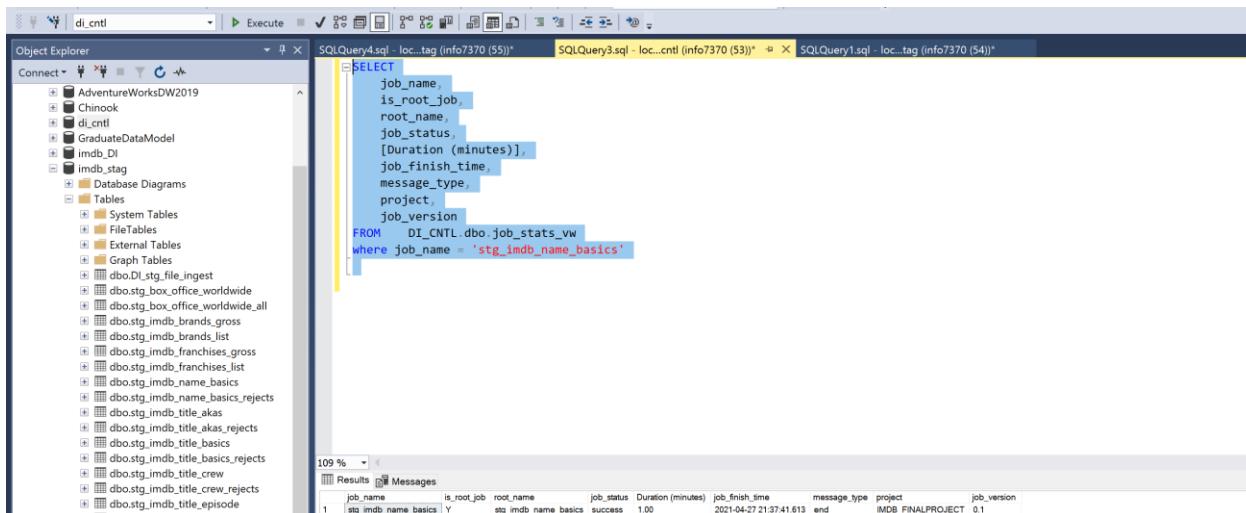
structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component, tUnite to merge all the files into a single result set table also, added cols using tmap for create_date and DI_ETL_ID which is process id, other than that we have also completed bonus part where in we are populating DI_stg_injest table.

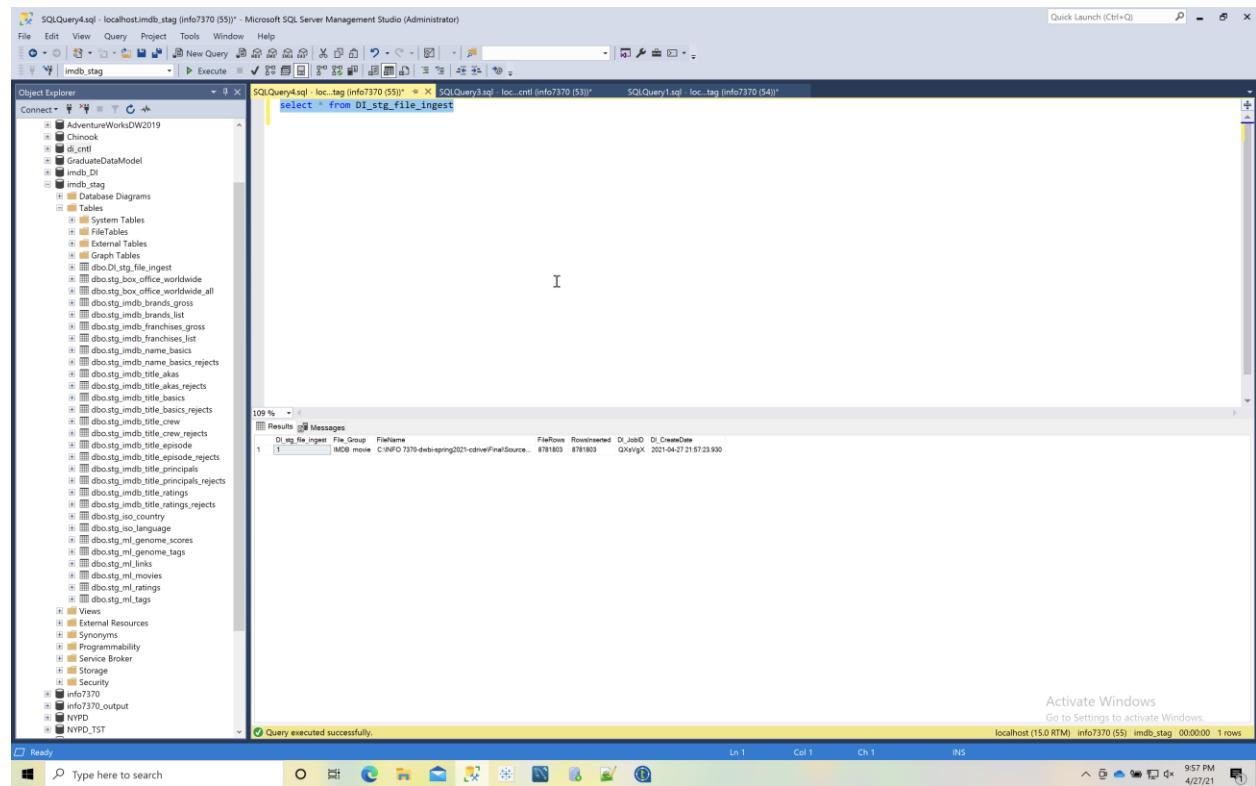
6. stg_imdb_name_basics & Rejects





TIME:



DI INJECT:

data consistency: data is loaded from tsv files, to make consistent data we have used tReplace component to clean the data.

reject: for rejects we are checking trim columns and check each row against schema structure, we have not received any rejects for this table.

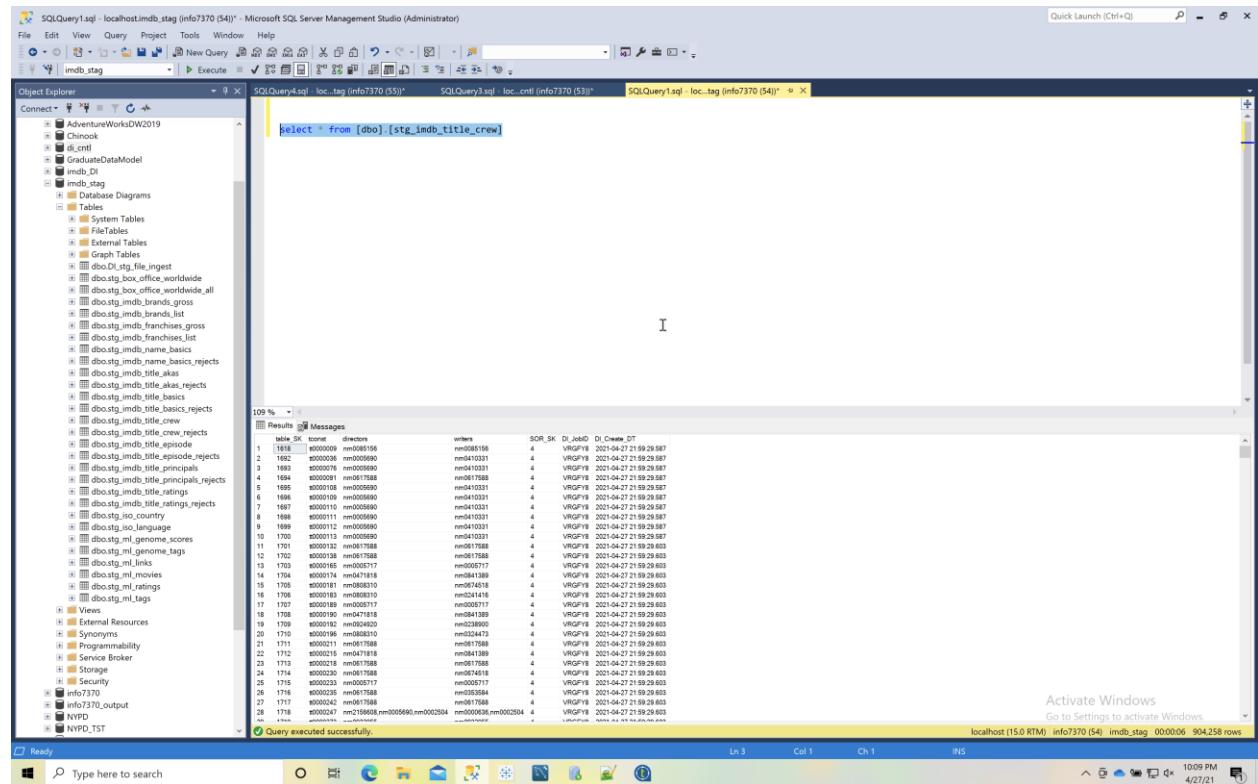
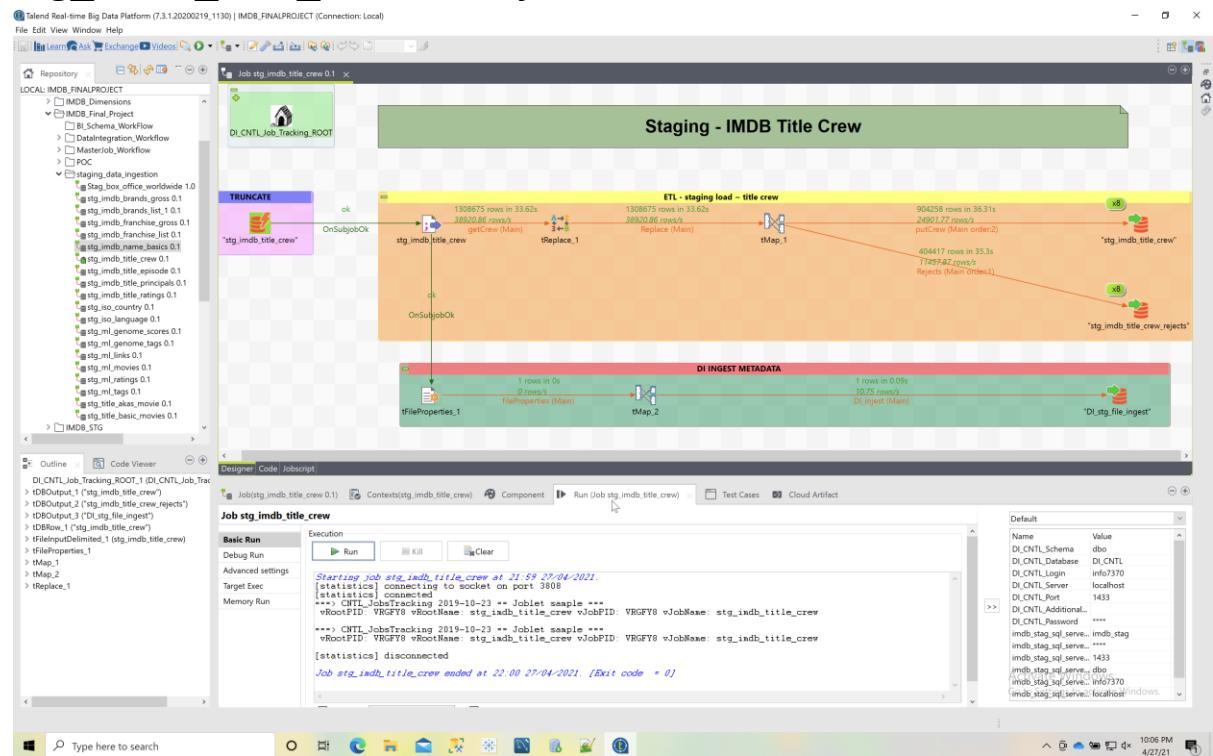
structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component.

For performance we are using bulk components,

For DI standard we are deleting bulk file which is created in the process

We are also capturing DI inject metadata and storing in the table which is bonus point.

7. stg_imdb_title_crew & Rejects



SQLQuery1.sql - localhost.imdb_stag (Info7370 (54)) - Microsoft SQL Server Management Studio (Administrator)

```
select * from [dbo].[stg_imdb_title_crew_rejects]
```

Object Explorer Results Messages

109 % Results Messages

table_Sk	score	directors	writers	Di_Reject_Code	SOR_SK	Di_JobId	Di_Create
16_18683	0	sf133641	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18685	0	sf133655	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18686	0	sf133667	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18688	0	sf133670	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18689	0	sf133671	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18690	0	sf133680	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18694	0	sf133690	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18695	0	sf133698	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18696	0	sf133700	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18697	0	sf133704	nm078641	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18698	0	sf133706	nm090641	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18699	0	sf133707	nm090642	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18700	0	sf133722	nm095983,rm030381	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18701	0	sf133724	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18702	0	sf133725	nm095983	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18704	0	sf133730	nm148325	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18705	0	sf133730	nm148325	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18706	0	sf133730	nm148325	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18707	0	sf133730	nm148325	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18708	0	sf133730	nm148325	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18709	0	sf133730	nm148325	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18710	0	sf133731	nm050837	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18711	0	sf133736	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18712	0	sf133740	nm070739	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18713	0	sf133740	nm095743,rm032198	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18714	0	sf133740	nm095743,rm032198	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18715	0	sf133758	nm095948	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18716	0	sf133759	nm256432	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18717	0	sf133762	nm256432	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18718	0	sf133762	nm256432	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18719	0	sf133762	nm256432	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18720	0	sf133764	NULL	996	4	VRSFV8	2021-04-27 21:99:30.073
16_18721	0	sf133764	nm000591	996	4	VRSFV8	2021-04-27 21:99:30.073

Results Messages

109 % Results Messages

Query executed successfully.

localhost (15.0 RTM) info7370 (54) imdb_stag 0000002 404,417 rows

Activate Windows
Go to Settings to activate Windows.

Ready Type here to search

10:10 PM 4/27/21

TIME:

SQLQuery4.sql - loc...tag (Info7370 (55))

```
SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM    DI_CNTL.dbo.job_stats_vw
where   job_name = 'stg_imdb_title_crew'
```

Object Explorer Results Messages

109 % Results Messages

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
stg_imdb_title_crew	Y	stg_imdb_title_crew	success	0.62	2021-04-27 22:00:05.993	end	IMDB_FINALPROJECT	0.1

DI INJECT:

FileRow	RowInserted	Di.JobID	Di.CreateDate	FileGroup	FileName
1	8781803	QXvVgX	2021-04-27 21:57:23.890	IMDB	movie C:\INFO 7370-dwlr-spring2021\dwlr\Final\Source Files\imdb-dataset\name_basics_movie.csv
2	1308678	VRGFfB	2021-04-27 22:00:05.847	IMDB	movie C:\INFO 7370-dwlr-spring2021\dwlr\Final\Source Files\imdb-dataset\title_crew_movie.csv

data consistency: data is loaded from tsv files, to make consistent data we have used tReplace component to clean the data.

reject: for rejects we are checking trim columns and check each row against schema structure, we are also using referential integrity reject for col di_reject_code

```
(Relational.ISNULL(Replace.directors) && !(Relational.ISNULL(Replace.writers))) ? -997 :  
(Relational.ISNULL(Replace.writers) && !(Relational.ISNULL(Replace.directors))) ? -996 :
```

```
((Relational.ISNULL(Replace.directors)) && (Relational.ISNULL(Replace.writers))) ? -995 : 0
```

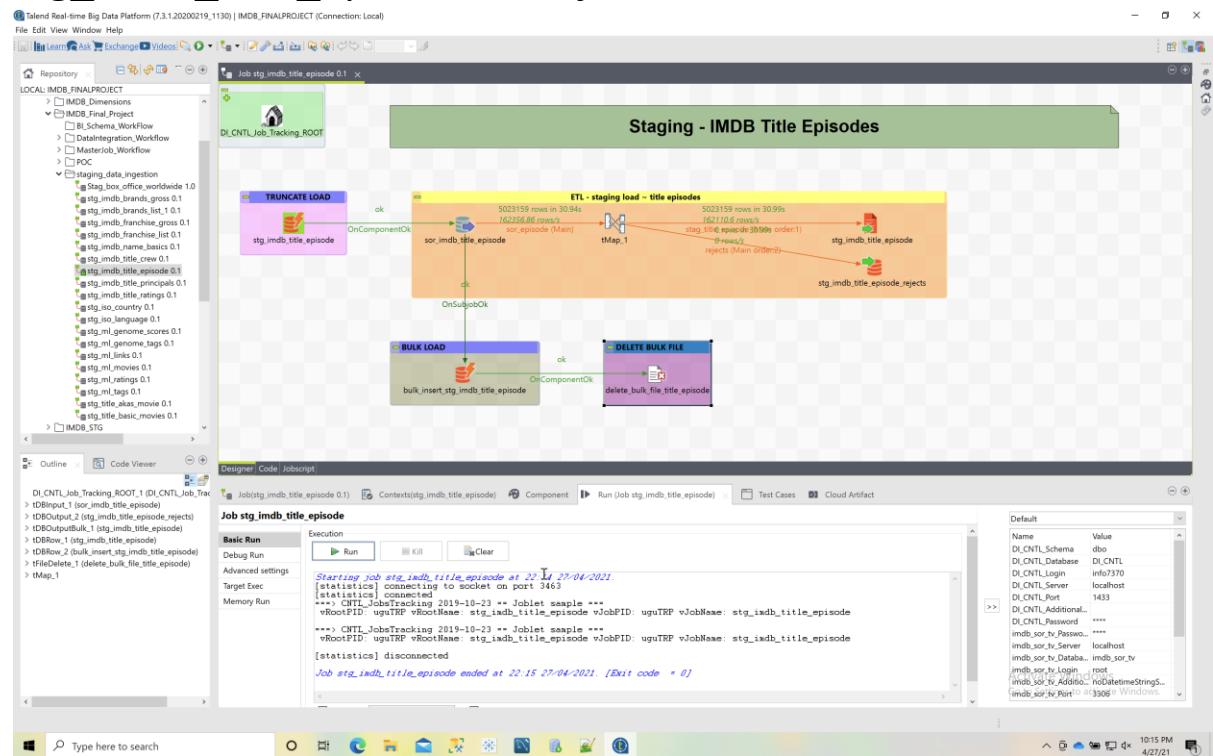
structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component.

For performance we are using bulk components,

For DI standard we are deleting bulk file which is created in the process

We are also capturing DI inject metadata and storing in the table which is bonus point.

8. stg_imdb_title_episode & Rejects



Column	Value
DL_CNTL.Schema	dbo
DL_CNTL.Database	DI_CNTL
DL_CNTL.Login	info7370
DL_CNTL.Server	localhost
DL_CNTL.Port	1433
DL_CNTL.Additional...	
DL_CNTL.Password	****
imbd_sor_tv.Password	****
imbd_sor_tv.Server	localhost
imbd_sor_tv.Database	imbd_sor_tv
imbd_sor_tv.Login	root
imbd_sor_tv.Addition...	noDatetimeString5..
imbd_sor_tv.Port	3306
Windows	

```

SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM   DI_CNTL.dbo.job_stats_vw
where job_name = 'stg_imdb_title_episode'
  
```

	job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
1	stg_imdb_title_episode	Y	stg_imdb_title_episode	success	0.81	2021-04-27 22:15:25.220	end	IMDB_FINALPROJECT	0.1

data consistency: data is loaded from sor tv mysql db data is consistent using sk column

reject: for rejects we are checking trim columns and check each row against schema structure, we have not received any rejects for this table.

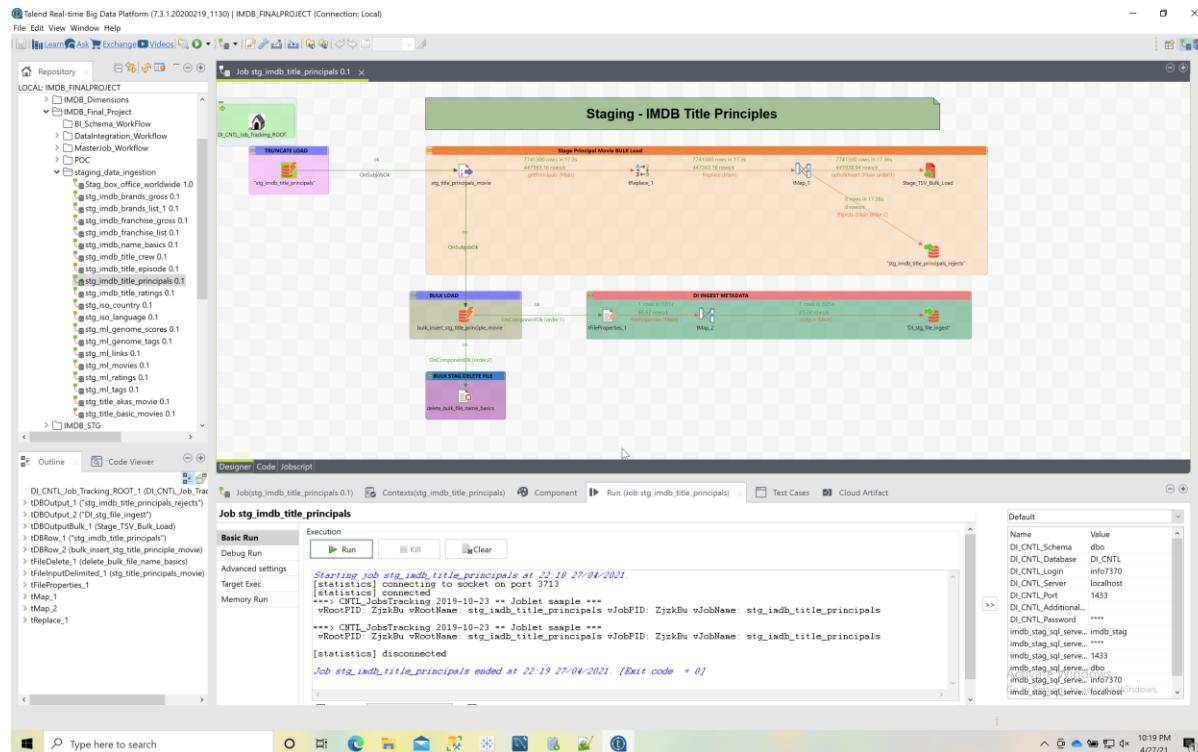
structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component.

For performance we are using bulk components,

For DI standard we are deleting bulk file which is created in the process

We are also capturing DI ingest metadata and storing in the table which is bonus point.

9. stg_imdb_title_principals & Rejects



S

table_sk	tonce	ordering	name	category	job	characters	SOR_SK	Q_JobID	D_Cause	D_Start
1	2	1	nm108870	self	NULL	["Self"]	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
2	2	1	s000001	2	director	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
3	3	1	s000001	3	nm074658	cinematographer	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
4	4	2	nm000002	1	composer	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
5	5	2	nm133271	composer	NULL	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
6	6	1	nm071626	director	NULL	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
7	7	1	nm000003	1	producer	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
8	8	3	nm000003	2	producer	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
9	9	4	nm044220	editor	NULL	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
10	10	1	nm072192	director	NULL	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
11	11	1	nm000004	1	producer	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
12	12	1	nm004482	actor	NULL	["Backstab"]	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
13	13	2	nm005042	actor	NULL	["Assassin"]	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
14	14	1	nm000005	1	producer	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
15	15	4	nm024979	producer	NULL	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
16	16	1	nm000090	director	NULL	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
17	17	1	nm000006	1	producer	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
18	18	2	nm019347	actor	NULL	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
19	19	3	nm000090	director	NULL	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
20	20	1	nm000007	1	producer	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
21	21	5	nm004879	producer	producer	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
22	22	1	nm005028	actor	NULL	["Sneaking Man"]	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
23	23	2	nm000090	director	NULL	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
24	24	1	nm000008	1	cinematographer	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
25	25	1	nm005086	actress	NULL	["Miss Geraldine Hobrook (Miss Jerry)"]	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
26	26	2	nm019823	actor	NULL	["Tommy Dorsey - The Director of the New York..."]	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
27	27	1	nm000009	1	producer	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00
28	28	4	nm008516	director	NULL	NULL	5	ZjskBu	2021-04-27 00:00:00	2021-04-27 00:00:00

TIME:

```

SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM    DI_CNTL.dbo.job_stats_vw
where   job_name = 'stg_imdb_title_principals'

```

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
stg_imdb_title_principals	Y	stg_imdb_title_principals	success	0.83	2021-04-27 22:19:44.603	end	IMDB_FINALPROJECT	0.1

DI INJECT:

```

select * from DI_stg_file_ingest

```

FileGroup	Filename	FileRow	Registered	DiJobID	DiChecksum
	C:\INFO_7370-dvbs-spring2021-cdive\FinalSource\Field-data-imdb-datasetname_basic_movie.tsv	8791803	8791803	QXxxIx	2021-04-27 21:57:23.930
	C:\INFO_7370-dvbs-spring2021-cdive\FinalSource\Field-data-imdb-datasettitle_crew_movie.tsv	1038675	904258	VRGfY8	2021-04-27 22:00:05.947
	C:\INFO_7370-dvbs-spring2021-cdive\FinalSource\Field-data-imdb-datasettitle_principals_movie.tsv	7741600	7741600	ZjdJlu	2021-04-27 22:19:44.603
	C:\INFO_7370-dvbs-spring2021-cdive\FinalSource\Field-data-imdb-datasettitle_string_movie.tsv	373400	373400	AxFF	2021-04-27 22:27:24.693

data consistency: data is loaded from tsv files, to make consistent data we have used tReplace component to clean the data.

reject: for rejects we are checking trim columns and check each row against schema structure, we have not received any rejects for this table.

structural integration processes: initially we are truncating the table, we are doing

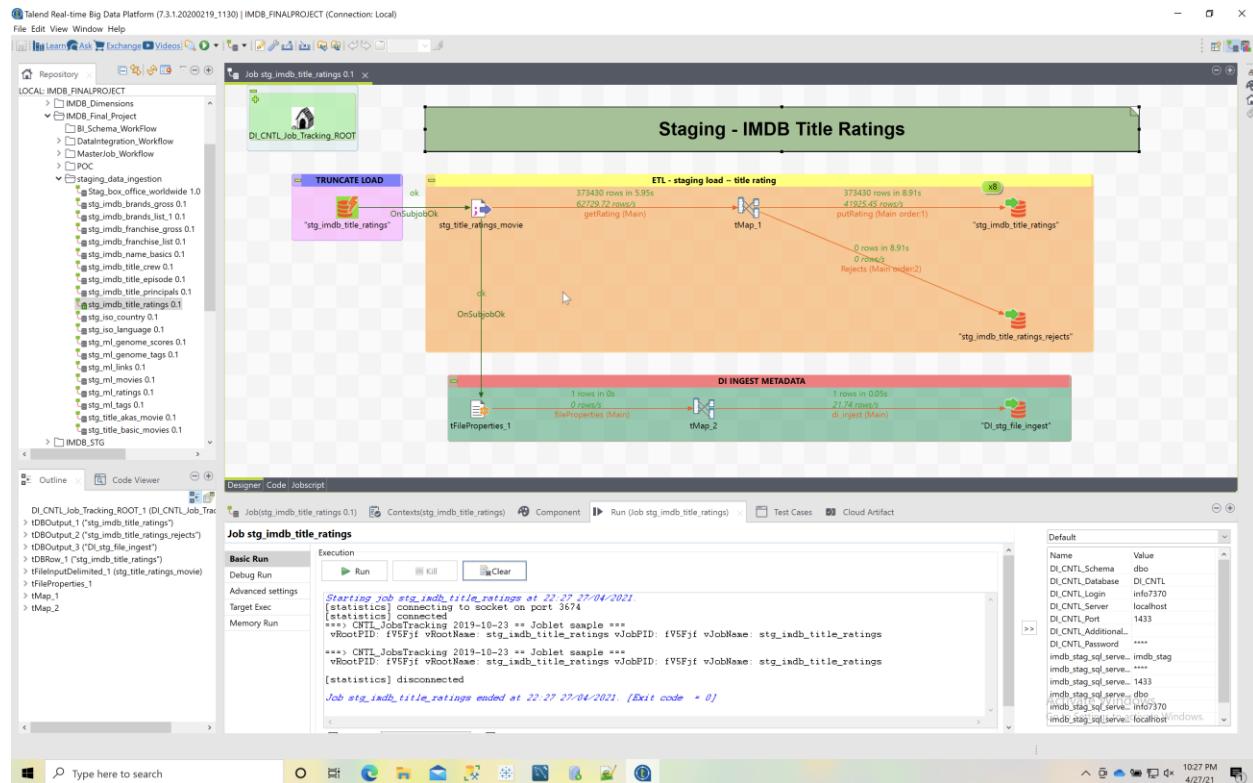
data cleansing using tReplace component.

For performance we are using bulk components,

For DI standard we are deleting bulk file which is created in the process

We are also capturing DI ingest metadata and storing in the table which is bonus point.

10. stg_imdb_title_ratings & Rejects



SQLQuery1.sql - localhost.imdb.stag (info7370 (54)) - Microsoft SQL Server Management Studio (Administrator)

File Edit View Query Project Tools Window Help

Object Explorer SQLQuery4.sql - loc...tag (info7370 (55)) SQLQuery3.sql - loc...cntl (info7370 (53)) SQLQuery1.sql - loc...tag (info7370 (54))

```
select * from [dbo].[stg_imdb_title_ratings]
```

I

Results Messages

table_SK	tonic	averageRating	numVotes	SCR_SK	DI_JOBID	DI_Created_DT	
1		1.89	198	6	N/SF	2021-04-27 22:27:17.87	
2	6369	0000002	6.1	198	6	N/SF	2021-04-27 22:27:17.87
3	6536	0000003	6.5	1326	6	N/SF	2021-04-27 22:27:17.87
4	6538	0000004	4.2	20	6	N/SF	2021-04-27 22:27:17.87
5	6538	0000005	6.1	2106	6	N/SF	2021-04-27 22:27:17.87
6	6539	0000006	5.3	115	6	N/SF	2021-04-27 22:27:17.87
7	6541	0000007	5.4	44	6	N/SF	2021-04-27 22:27:17.87
8	6541	0000008	5.4	1792	6	N/SF	2021-04-27 22:27:17.87
9	6542	0000009	5.9	198	6	N/SF	2021-04-27 22:27:17.87
10	6543	0000010	6.9	983	6	N/SF	2021-04-27 22:27:17.87
11	6545	0000011	5.2	205	6	N/SF	2021-04-27 22:27:17.87
12	6545	0000012	7.4	10108	6	N/SF	2021-04-27 22:27:17.87
13	6546	0000013	5.7	1560	6	N/SF	2021-04-27 22:27:17.87
14	6546	0000014	7.2	2449	6	N/SF	2021-04-27 22:27:17.87
15	6548	0000015	6.1	808	6	N/SF	2021-04-27 22:27:17.87
16	6549	0000016	5.9	1186	6	N/SF	2021-04-27 22:27:17.87
17	6550	0000017	4.7	29	6	N/SF	2021-04-27 22:27:17.87
18	6552	0000018	5.3	474	6	N/SF	2021-04-27 22:27:17.87
19	6554	0000019	5.3	18	6	N/SF	2021-04-27 22:27:17.87
20	6555	0000020	5	43	6	N/SF	2021-04-27 22:27:17.87
21	6558	0000022	5.1	379	6	N/SF	2021-04-27 22:27:17.87
22	6560	0000023	5.7	1125	6	N/SF	2021-04-27 22:27:17.87
23	6562	0000024	4.6	48	6	N/SF	2021-04-27 22:27:17.87
24	6563	0000025	4.2	9	6	N/SF	2021-04-27 22:27:17.87
25	6564	0000026	5.7	1300	6	N/SF	2021-04-27 22:27:17.87
26	6565	0000027	5.6	917	6	N/SF	2021-04-27 22:27:17.87
27	6566	0000028	5.1	49	6	N/SF	2021-04-27 22:27:17.87
28	6568	0000029	5.9	2829	6	N/SF	2021-04-27 22:27:17.87

Activate Windows
Go to Settings to activate Windows

localhost (15.0 RTM) info7370 (54) imdb_stag 000002 373,430 rows

Query executed successfully.

Ready Type here to search

Ln 3 Col 1 Ch 1 INS

10:30 PM 4/27/21 10:30 PM 4/27/21

TIME

SQLQuery4.sql - loc...tag (info7370 (55)) SQLQuery3.sql - loc...cntl (info7370 (53)) SQLQuery1.sql - loc...tag (info7370 (54))

Object Explorer di_cntl

```
SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM DI_CNTL.dbo.job_stats_vw
WHERE job_name = 'stg_imdb_title_ratings'
```

Results Messages

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
stg_imdb_title_ratings	Y	stg_imdb_title_ratings	success	0.16	2021-04-27 22:24.033	end	IMDB_FINALPROJECT	0.1

DI INJECT:

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists several databases and tables, including 'AdventureWorksDW2019', 'Chinook', 'di_cntl', 'GlobalDataModel', 'imdb_Dt', and 'imdb_stag'. The 'imdb_stag' database is expanded, showing tables like 'Database Diagrams', 'Tables', 'System Tables', 'External Tables', 'Graph Tables', and numerous 'dbo.DI_stg...' tables. The 'Results' tab in the center displays the output of the query 'select * from DI_stg_file_Ingest'. The results grid has columns: FileRowID, RowInserted, DI_JobID, DI_CreateDate, FileGroup, FileName, and FileRow. There are four rows of data, each corresponding to a file named 'C:\INFO 7370-dwbspring2021\dwbs\source\Filed\data\imdb-dataset\name_basics_movie.tsv'. The 'Activate Windows' watermark is visible at the bottom right.

FileRowID	RowInserted	DI_JobID	DI_CreateDate	FileGroup	FileName
1	8781803	QXxVgX	2021-04-27 21:57:23.930		C:\INFO 7370-dwbspring2021\dwbs\source\Filed\data\imdb-dataset\name_basics_movie.tsv
2	1308675	904258	2021-04-27 22:00:06.947		C:\INFO 7370-dwbspring2021\dwbs\source\Filed\data\imdb-dataset\crew_movie.tsv
3	7741600	ZjBhu	2021-04-27 22:19:44.803		C:\INFO 7370-dwbspring2021\dwbs\source\Filed\data\imdb-dataset\principals_movie.tsv
4	873430	NPFy	2021-04-27 22:27:24.633		C:\INFO 7370-dwbspring2021\dwbs\source\Filed\data\imdb-dataset\string_movie.tsv

data consistency: data is loaded from tsv files, to make consistent data we have used sk column.

reject: for rejects we are checking trim columns and check each row against schema structure, we have not received any rejects for this table.

structural integration processes: initially we are truncating the table.

For Performance we have enabled parallelization to speed up the load

We are also capturing DI inject metadata and storing in the table which is bonus point.

11. stg_iso_country

Staging - ISO Reference Country

Job stg_iso_country

Basic Run

```

Starting job etl_stg_isocountryr_stg_23_27_04_2021
[statistics] connecting to socket on port 3927
[statistics] connected
*** Joblet sample *** vRootPID: 7zISYr vRootName: stg_iso_country vJobName: stg_iso_country
*** CNTL_JobTeaching 2019-10-23 ** Joblet sample ***
vRootPID: 7zISYr vRootName: stg_iso_country vJobName: stg_iso_country
[statistics] disconnected
Job stg_iso_country ended at 23:33:27 04/2021. [Exit code = 0]

```

Default

Name	Value
DI_CNTL_Schema	dbo
DI_CNTL_Database	DI_CNTL
DI_CNTL_Login	info7370
DI_CNTL_Server	localhost
DI_CNTL_Port	1433
DI_CNTL_Password	****
imdb_stg_sql_server	imdb_stag
imdb_stag_sql_server	imdb_stag
imdb_stag_port	1433
imdb_stag_sql_server	imdb_stag
imdb_stag_sql_server	info7370
imdb_stag_sql_server	localhost

SQL Server Management Studio (Administrator)

File Edit View Project Tools Window Help

Object Explorer

SQLQuery1.sql - localhost.imdb_stag (Info7370 (54)) - Microsoft SQL Server Management Studio (Administrator)

SQLQuery4.sql - loc_tag (Info7370 (55)) - SQLQuery3.sql - loc_cntl (Info7370 (53)) - SQLQuery1.sql - loc_tag (Info7370 (54)) -

select * from [dbo].[stg_iso_country]

Results

Country_Sk	country_name	alpha_2	alpha_3	country_code	iso_3166_2	region	sub_region	intermediate_region	region_code	sub_region_code	intermediate_region_code	SQR_SK	Di_JobID	Di_Create_DT
1	Afghanistan	AF	AFG	4	ISO 3166-2:Af	Asia	Southern Asia	142	34	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
2	Aland Islands	AX	ALA	248	ISO 3166-2:AX	Europe	Northern Europe	150	154	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
3	Albania	AL	ALB	249	ISO 3166-2:AL	Europe	South Europe	150	39	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
4	Algeria	DZ	DZA	12	ISO 3166-2:DZ	Africa	Northern Africa	2	15	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
5	American Samoa	AS	ASM	16	ISO 3166-2:AS	Oceania	Polynesia	9	61	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
6	Angola	AO	AGO	24	ISO 3166-2:AO	Africa	Sub-Saharan Africa	150	39	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
7	Angola	AO	AGO	24	ISO 3166-2:AO	Africa	Middle Africa	2	202	17	NULL	7	7zISYy	2021-04-27 22:33:41.800
8	Anguilla	AI	AIA	660	ISO 3166-2:AI	Americas	Latin America and the Caribbean	19	419	29	7	7zISYy	2021-04-27 22:33:41.800	
9	Antarctica	AT	ATA	28	ISO 3166-2:AT	Americas	Caribbean	NULL	NULL	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
10	Antigua and Barbuda	AG	ATG	28	ISO 3166-2:AG	Americas	Latin America and the Caribbean	15	29	7	7zISYy	2021-04-27 22:33:41.800		
11	Argentina	AR	ARG	32	ISO 3166-2:AR	Americas	Latin America and the Caribbean	19	419	5	7	7zISYy	2021-04-27 22:33:41.800	
12	Armenia	AM	ARM	81	ISO 3166-2:AM	Asia	Western Asia	142	145	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
13	Aruba	AW	ABW	36	ISO 3166-2:AW	Americas	Caribbean	15	29	7	7zISYy	2021-04-27 22:33:41.800		
14	Australia	AU	AUS	36	ISO 3166-2:AU	Oceania	Australia and New Zealand	9	53	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
15	Austria	AT	AUT	40	ISO 3166-2:AT	Europe	Western Europe	150	155	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
16	Azerbaijan	AZ	AZE	282	ISO 3166-2:AZ	Asia	Central Asia	142	145	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
17	Bahamas	BS	BHS	44	ISO 3166-2:BS	Americas	Latin America and the Caribbean	19	419	29	7	7zISYy	2021-04-27 22:33:41.800	
18	Bahrain	BH	BHR	48	ISO 3166-2:BH	Asia	Western Asia	142	145	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
19	Bangladesh	BD	BGD	50	ISO 3166-2:BD	Americas	South Asia	142	34	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
20	Barbados	BB	BRB	52	ISO 3166-2:BB	Americas	Latin America and the Caribbean	19	419	29	7	7zISYy	2021-04-27 22:33:41.800	
21	Belarus	BY	BLR	112	ISO 3166-2:BY	Europe	Eastern Europe	150	151	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
22	Belgium	BE	BEL	190	ISO 3166-2:BE	Europe	Western Europe	150	155	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
23	Belize	BZ	BZL	64	ISO 3166-2:BZ	Americas	Central America and the Caribbean	19	419	13	7	7zISYy	2021-04-27 22:33:41.800	
24	Benin	BJ	BEN	204	ISO 3166-2:BJ	Africa	Sub-Saharan Africa	2	202	11	7	7zISYy	2021-04-27 22:33:41.800	
25	Bermuda	BM	BMU	60	ISO 3166-2:BM	Americas	Northern America	19	61	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
26	Bhutan	BT	BTH	87	ISO 3166-2:BT	Asia	Central Asia	142	34	NULL	NULL	7	7zISYy	2021-04-27 22:33:41.800
27	Bolivia (Plurinational State of)	BO	BOL	68	ISO 3166-2:BO	Americas	Latin America and the Caribbean	19	419	5	7	7zISYy	2021-04-27 22:33:41.800	
28	Bonaire, Sint Eustatius and Saba	BQ	BES	535	ISO 3166-2:BQ	Americas	Latin America and the Caribbean	19	419	29	7	7zISYy	2021-04-27 22:33:41.800	

Query executed successfully.

TIME:

```

SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM DI_CNTL.dbo.job_stats_vw
where job_name = 'stg_iso_country'
  
```

	job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
1	stg_iso_country	Y	stg_iso_country	success	0.01	2021-04-27 22:33:41.940	end	IMDB_FINALPROJECT	0.1

DI INJECT:

```

select * from DI_stg_file_ingest
  
```

	File_Group	FileName	FileRow	RowInserted	DiJobID	DiCreateDate
1	IMDB.basics	C:\INFO 7370\dwb\spring2021\dwbe\FinalSource\Files\data\imdb-datasetnetflix\basics_movies.csv	8781603	8781603	104VgD	2021-04-27 21:57:23.800
2	IMDB.movie	C:\INFO 7370\dwb\spring2021\dwbe\FinalSource\Files\data\imdb-datasetnetflix\movies_movies.csv	12286701	12286701	VQdJ	2021-04-27 22:19:44.603
3	IMDB.movie	C:\INFO 7370\dwb\spring2021\dwbe\FinalSource\Files\data\imdb-datasetnetflix\principals_movies_principals.csv	7741500	7741500	ZpBu	2021-04-27 22:19:44.603
4	IMDB.movie	C:\INFO 7370\dwb\spring2021\dwbe\FinalSource\Files\data\imdb-datasetnetflix\ratings_movies_ratings.csv	373430	373430	NfSf	2021-04-27 22:17:24.033
5	ISO REFERENCE counts	C:\INFO 7370\dwb\spring2021\dwbe\FinalSource\Files\data\iso_isocountry_iso_all.csv	249	249	ThSt	2021-04-27 22:33:41.927

Activate Windows
Go to Settings to activate Windows.

localhost (15.0 RTM) info7370 (55) imdb_stag 00:00:00 5 rows

data consistency: data is loaded from tsv files, to make consistent data we have used sk column.

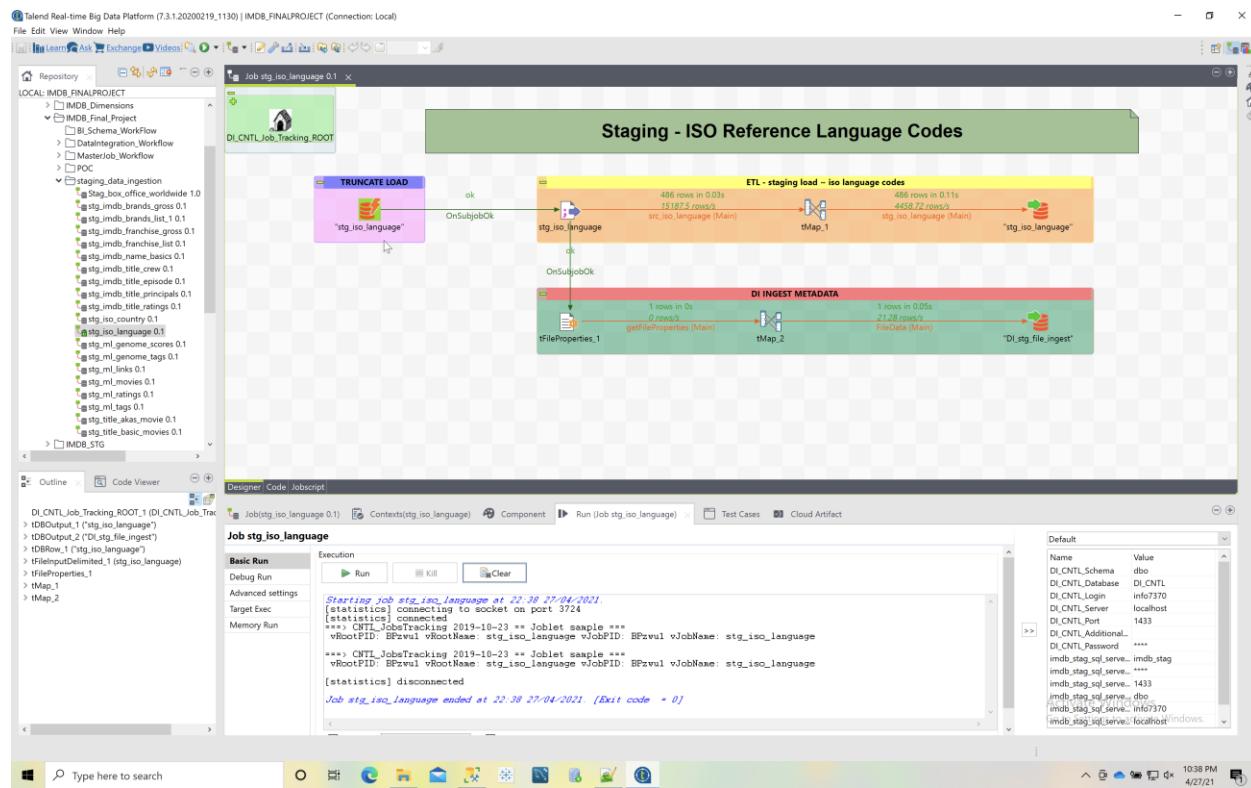
reject: No rejects required for this table.

structural integration processes: initially we are truncating the table.

For Performance we have enabled parallelization to speed up the load

We are also capturing DI ingest metadata and storing in the table which is bonus point.

12. stg_iso_language



SQLQuery1.sql - localhost.imdb_stag (info7370 (54)) - Microsoft SQL Server Management Studio (Administrator)

```
select * from [dbo].[stg_iso_language]
```

Results Messages

Language_SK	alpha1_b	alpha1_r	alpha2_b	Language_Name	SOR_SK	DI_ProjID	DI_Created_DT
1	azb	ئەزبەق	az	Azerbaijani	8	BPEnv1	2021-04-27 22:38:24.713
2	ab	አማርኛ	am	Afar	8	BPEnv1	2021-04-27 22:38:24.713
3	ace	አማርኛ	am	Achinese	8	BPEnv1	2021-04-27 22:38:24.713
4	ach	አማርኛ	am	Acoli	8	BPEnv1	2021-04-27 22:38:24.713
5	eda	አማርኛ	am	Adangme	8	BPEnv1	2021-04-27 22:38:24.713
6	ady	አማርኛ	am	Adyghe, Adygei	8	BPEnv1	2021-04-27 22:38:24.713
7	af	አማርኛ	am	Amharic, Amharic (Ethiopia)	8	BPEnv1	2021-04-27 22:38:24.713
8	arh	አማርኛ	am	Amharic	8	BPEnv1	2021-04-27 22:38:24.713
9	af	አማርኛ	am	Afrikaans	8	BPEnv1	2021-04-27 22:38:24.713
10	an	አማርኛ	am	Anu	8	BPEnv1	2021-04-27 22:38:24.713
11	akk	አማርኛ	am	Akkadian	8	BPEnv1	2021-04-27 22:38:24.713
12	sq	አማርኛ	sq	Albanian	8	BPEnv1	2021-04-27 22:38:24.713
13	ab	አማርኛ	sq	Albanian	8	BPEnv1	2021-04-27 22:38:24.713
14	ak	አማርኛ	sq	Albanian	8	BPEnv1	2021-04-27 22:38:24.713
15	alg	አማርኛ	sq	Algonquian languages	8	BPEnv1	2021-04-27 22:38:24.713
16	at	አማርኛ	sq	Algonquin	8	BPEnv1	2021-04-27 22:38:24.713
17	amn	አማርኛ	sq	Altonian	8	BPEnv1	2021-04-27 22:38:24.713
18	ang	አማርኛ	sq	English, Old (ca.400-1100)	8	BPEnv1	2021-04-27 22:38:24.713
19	arp	አማርኛ	sq	Angika	8	BPEnv1	2021-04-27 22:38:24.713
20	ape	አማርኛ	sq	Apache languages	8	BPEnv1	2021-04-27 22:38:24.713
21	ara	አማርኛ	sq	Arabic	8	BPEnv1	2021-04-27 22:38:24.713
22	arc	አማርኛ	sq	Official Aramaic (700-300 BCE); Imperial Aramaic	8	BPEnv1	2021-04-27 22:38:24.713
23	arg	አማርኛ	sq	Aragonese	8	BPEnv1	2021-04-27 22:38:24.713
24	amr	አማርኛ	sq	Armenian	8	BPEnv1	2021-04-27 22:38:24.713
25	hye	አማርኛ	sq	Mapudungun, Maputche	8	BPEnv1	2021-04-27 22:38:24.713
26	arp	አማርኛ	sq	Apache	8	BPEnv1	2021-04-27 22:38:24.713
27	arh	አማርኛ	sq	Arabic languages	8	BPEnv1	2021-04-27 22:38:24.713
28	avr	አማርኛ	sq	Awakat	8	BPEnv1	2021-04-27 22:38:24.713

Query executed successfully.

TIME:

SQLQuery3.sql - localhost.di_cntl (info7370 (53)) - Microsoft SQL Server Management Studio (Administrator)

```
SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM
    DI_CNTL.dbo.job_stats_vw
where job_name = 'stg_iso_language'
```

Results Messages

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
stg_iso_language	Y	stg_iso_language	success	0.02	2021-04-27 22:38:24.820	end	IMDB_FINALPROJECT	0.1

DI INJECT:

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the `imdb_stag` database and its tables. The main window displays a query results grid for the `DI_stg_file_Ingest` table. The results show 6 rows of data with columns: FileRow, RowNumber, File Group, FileName, FileRows, RowRejected, Di.JobID, and Di.CreateDate. The data includes various file names like `IMDB.movie`, `IMDB.movie`, `IMDB.movie`, `IMDB.movie`, `ISO REFERENCE.countries`, and `ISO REFERENCE.language-codes`. The last column, `Di.CreateDate`, shows dates from 2021-04-27 to 2021-04-28.

FileRow	RowNumber	File Group	FileName	FileRows	RowRejected	Di.JobID	Di.CreateDate
1	1	IMDB movie	C:\INFO 7370-debs-spring2021\cdtive\FinalSource\Flatdata\imdb-dataset\name_basics.movie.tsv	8781803	8781803	QXWVgX	2021-04-27 21:57:23.930
2	2	IMDB movie	C:\INFO 7370-debs-spring2021\cdtive\FinalSource\Flatdata\imdb-dataset\name_crew.movie.tsv	1308675	904258	VRGfY8	2021-04-27 22:00:08.947
3	3	IMDB movie	C:\INFO 7370-debs-spring2021\cdtive\FinalSource\Flatdata\imdb-dataset\name_principals.movie.tsv	7000000	2011800	ZgPjFw	2021-04-27 22:00:09.000
4	4	IMDB movie	C:\INFO 7370-debs-spring2021\cdtive\FinalSource\Flatdata\imdb-dataset\name_rejects.movie.tsv	373450	279450	NyFfY	2021-04-27 22:37:24.033
5	5	ISO REFERENCE.countries	C:\INFO 7370-debs-spring2021\cdtive\FinalSource\Flatdata\iso_lscountries.iso+alt.csv	249	248	7iSY	2021-04-27 22:37:41.927
6	6	ISO REFERENCE.language-codes	C:\INFO 7370-debs-spring2021\cdtive\FinalSource\Flatdata\iso_lslanguage-codes-iso.csv	486	486	BPAwv1	2021-04-27 22:38:24.820

data consistency: data is loaded from tsv files, to make consistent data we have used `sk` column.

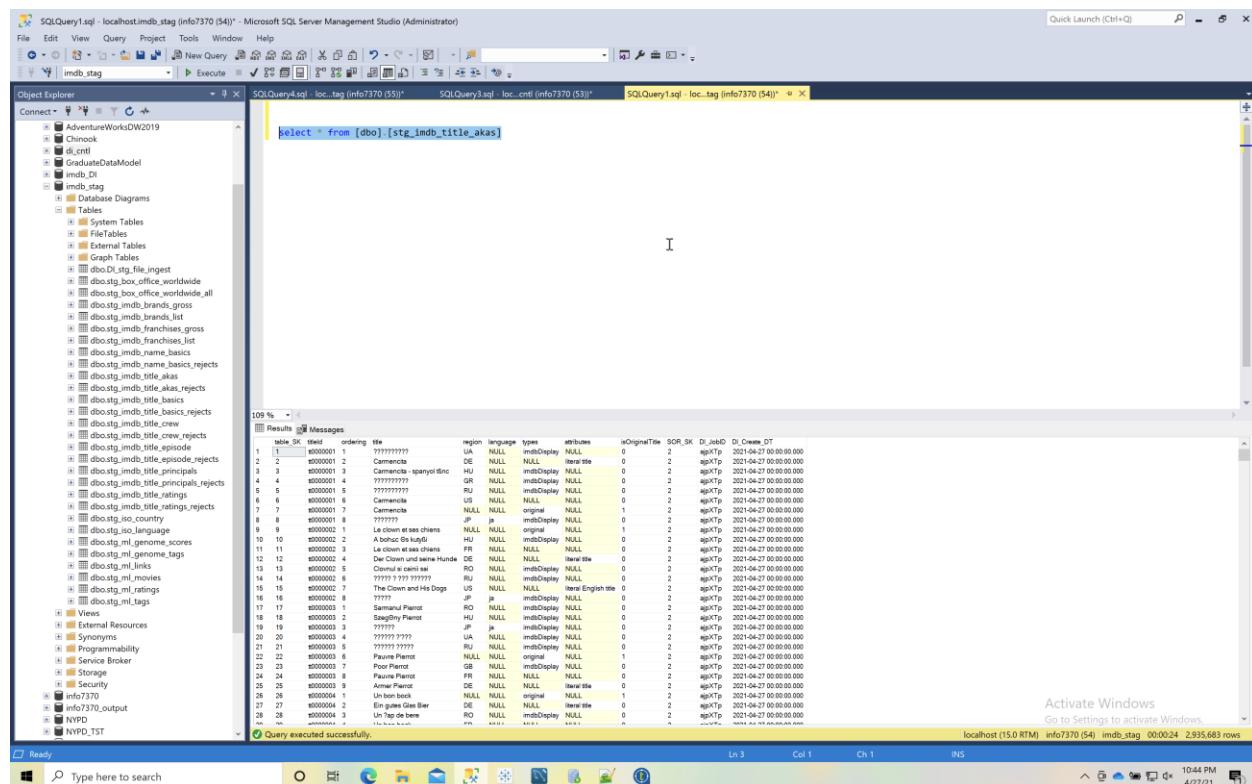
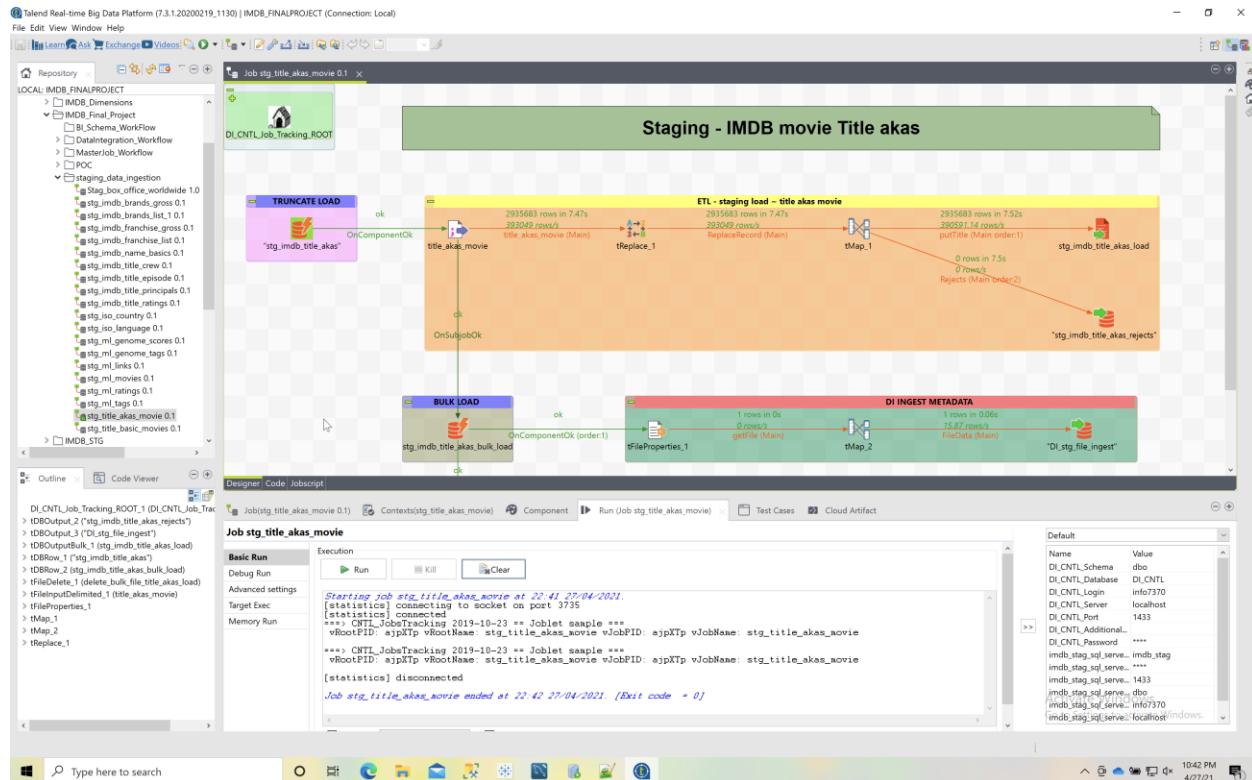
reject: No rejects required for this table.

structural integration processes: initially we are truncating the table.

For Performance we have enabled parallelization to speed up the load

We are also capturing DI inject metadata and storing in the table which is bonus point.

13. stg_title_akas_movie & Rejects



TIME:

```
SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM    DI_CNTL.dbo.job_stats_vw
WHERE   job_name = 'stg_title_akas_movie'
```

	job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
1	stg_title_akas_movie	Y	stg_title_akas_movie	success	0.35	2021-04-27 22:42:08.587	end	IMDB_FINALPROJECT	0.1

DI INJECT:

The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, a database named 'imdb_stag' is selected. Under the 'Tables' node, there is a single table named 'DI_stg_file_ingest'. In the main pane, a query is running: 'select * from DI_stg_file_ingest'. The results grid shows 109 rows of data, which includes file names and their corresponding file sizes and creation dates. The last few rows of the results grid are as follows:

FileRow	Reinserted	Di_JobID	Di_CreateDate
8719803	GxvVgX	2021-04-27 21:57:33.930	
1508975	VHnLqf	2021-04-27 22:19:44.647	
7741600	ZpBju	2021-04-27 22:19:44.603	
373430	3Tjw3D	2021-04-27 22:27:24.033	
249	TdP	2021-04-27 22:30:24.027	
486	BPewvt	2021-04-27 22:30:24.820	
2635683	egkTp	2021-04-27 22:42:08.570	

At the bottom of the results grid, it says '7 rows'. The status bar at the bottom right shows 'localhost (15.0 RTM) info7370 (S5) imdb_stag 00:00:00 10:45 PM 4/27/21'.

data consistency: data is loaded from tsv files, to make consistent data we have used tReplace component to clean the data and sk col

reject: for rejects we are checking trim columns and check each row against schema structure, we have not received any rejects for this table.

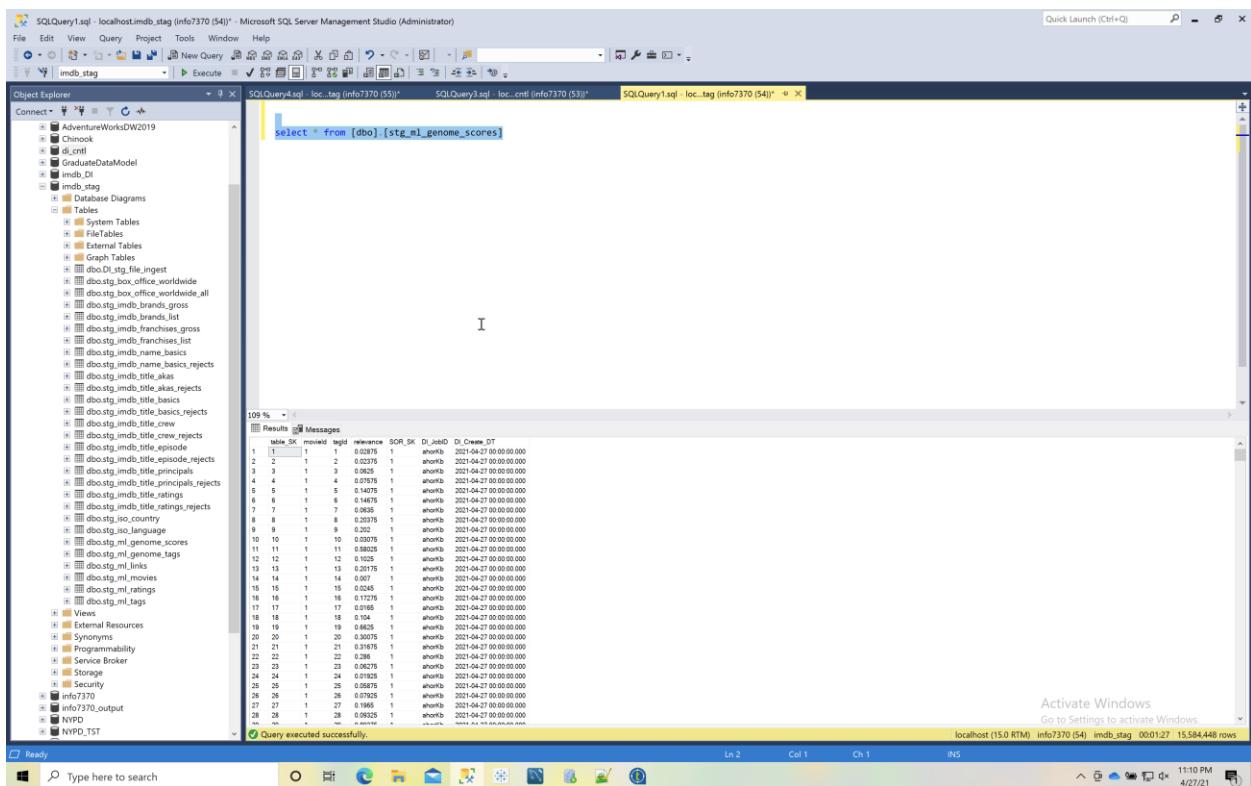
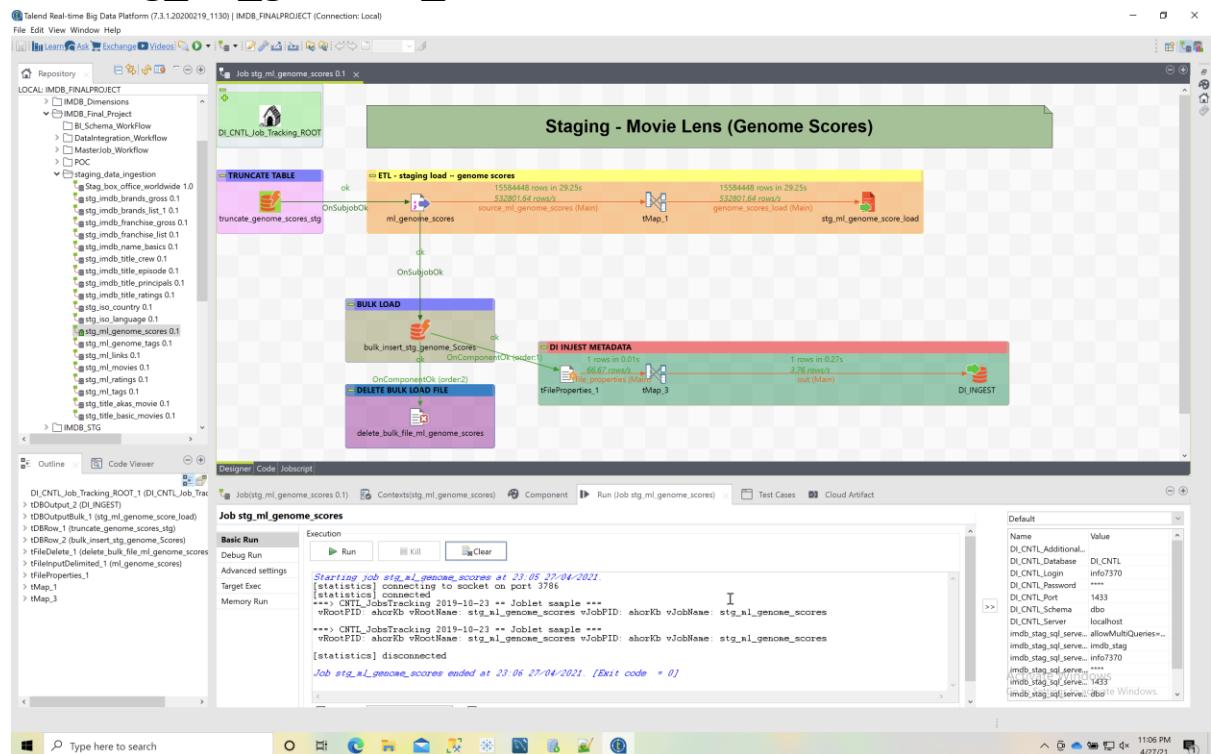
structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component.

For performance we are using bulk components,

For DI standard we are deleting bulk file which is created in the process

We are also capturing DI inject metadata and storing in the table which is bonus point.

14. stg_ml_genome_scores



TIME:

```

SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_start_time,
    message_type,
    project,
    job_version
FROM DI_CNTL.dbo.job_stats_vw
WHERE job_name = 'stg_ml_genome_scores'
  
```

The screenshot shows the SSMS interface with the Object Explorer on the left and a query editor window on the right. The query editor contains a SELECT statement from the DI_CNTL.dbo.job_stats_vw view, filtering for the job_name 'stg_ml_genome_scores'. The results grid shows one row with the following data:

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_start_time	message_type	project	job_version
stg_ml_genome_scores	Y		success	1.31	2021-04-27 23:06:41.037	end	IMDB_FINALPROJECT	0.1

DI INJECT:

The screenshot shows the SSMS interface with multiple queries running in separate tabs:

- Tab 1: SQLQuery4.sql - localhost.imdb_stag (info7370 (55)) - Microsoft SQL Server Management Studio (Administrator)


```
select * from DI_stg_file_ingest
```

 The results grid shows 8 rows of data from the DI_stg_file_ingest table. The columns include File_Group, File, Path, FileSize, RowCount, Duration, and CreationDate. One row is highlighted with a yellow background.
- Tab 2: SQLQuery3.sql - loc...cntl (info7370 (53))*


```
SELECT
        job_name,
        is_root_job,
        root_name,
        job_status,
        [Duration (minutes)],
        job_start_time,
        message_type,
        project,
        job_version
    FROM DI_CNTL.dbo.job_stats_vw
    WHERE job_name = 'stg_ml_genome_scores'
```

 The results grid shows one row with the same data as the first query.
- Tab 3: SQLQuery1.sql - loc...tag (info7370 (54))*


```
SELECT
        job_name,
        is_root_job,
        root_name,
        job_status,
        [Duration (minutes)],
        job_start_time,
        message_type,
        project,
        job_version
    FROM DI_CNTL.dbo.job_stats_vw
    WHERE job_name = 'stg_ml_genome_scores'
```

 The results grid shows one row with the same data as the first query.

The status bar at the bottom indicates "Query executed successfully." and "localhost (15.0 RTM) info7370 (55) imdb_stag 00:00:00 8 rows".

data consistency: data is loaded from tsv files, to make consistent data we have used skip column.

reject: No reject required for this table.

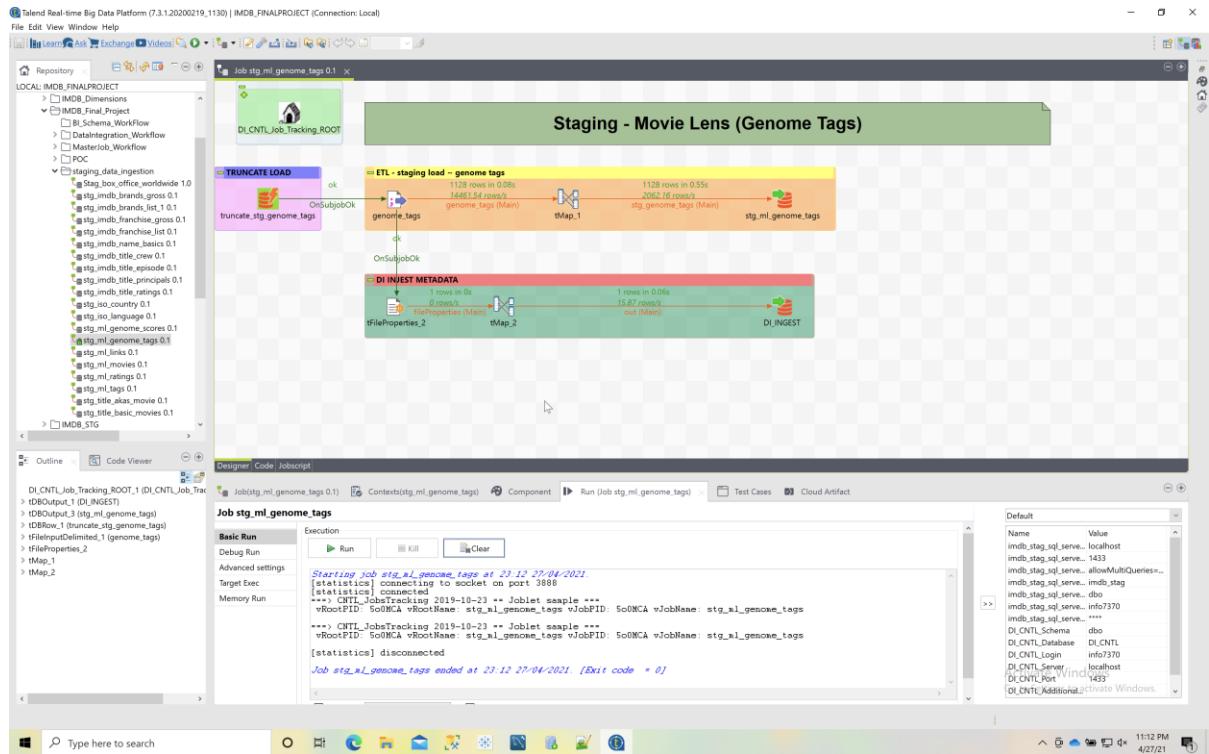
structural integration processes: initially we are truncating the table

For performance we are using bulk components,

For DI standard we are deleting bulk file which is created in the process

We are also capturing DI ingest metadata and storing in the table which is bonus point.

15. stg_ml_genome_tags



SQLQuery1.sql - localhost\imdb_stag (info7370 (54)) - Microsoft SQL Server Management Studio (Administrator)

```
select * from [dbo].[stg_mi_genome_tags]
```

Results Messages

table_SK	ligid	tag	SOR_SK	Di_JobID	Di_Created
1	2	007 (series)	1	90dMCA	2021-04-27 23:12:40.143
2	3	3 18th century	1	90dMCA	2021-04-27 23:12:40.143
3	4	4 19th century	1	90dMCA	2021-04-27 23:12:40.143
4	5	5 1950s	1	90dMCA	2021-04-27 23:12:40.143
5	6	6 1950s	1	90dMCA	2021-04-27 23:12:40.143
6	7	7 1960s	1	90dMCA	2021-04-27 23:12:40.143
7	8	8 1970s	1	90dMCA	2021-04-27 23:12:40.143
8	9	9 1980s	1	90dMCA	2021-04-27 23:12:40.143
9	10	10 19th century	1	90dMCA	2021-04-27 23:12:40.143
10	11	11 20th century	1	90dMCA	2021-04-27 23:12:40.143
11	12	12 70mm	1	90dMCA	2021-04-27 23:12:40.143
12	13	13 80s	1	90dMCA	2021-04-27 23:12:40.143
13	14	14 90s	1	90dMCA	2021-04-27 23:12:40.143
14	15	15 aardman	1	90dMCA	2021-04-27 23:12:40.143
15	16	16 aardman studios	1	90dMCA	2021-04-27 23:12:40.143
16	17	17 absurd	1	90dMCA	2021-04-27 23:12:40.143
17	18	18 absurd	1	90dMCA	2021-04-27 23:12:40.143
18	19	19 action	1	90dMCA	2021-04-27 23:12:40.143
19	20	20 action packed	1	90dMCA	2021-04-27 23:12:40.143
20	21	21 aeronauton	1	90dMCA	2021-04-27 23:12:40.143
21	22	22 adapted from book	1	90dMCA	2021-04-27 23:12:40.143
22	23	23 adapted from comic	1	90dMCA	2021-04-27 23:12:40.143
23	24	24 adapted from game	1	90dMCA	2021-04-27 23:12:40.143
24	25	25 addiction	1	90dMCA	2021-04-27 23:12:40.143
25	26	26 addisence	1	90dMCA	2021-04-27 23:12:40.143
26	27	27 addisence	1	90dMCA	2021-04-27 23:12:40.143
27	28	28 addury	1	90dMCA	2021-04-27 23:12:40.143
28	29	29 addury	1	90dMCA	2021-04-27 23:12:40.143

Activate Windows
Go to Settings to activate Windows

localhost (15.0 RTM) - info7370 (54) | imdb_stag | 00:00:00 | 1,128 rows

Query executed successfully.

TIME:

SQLQuery4.sql - loc.tag (info7370 (55)) - Microsoft SQL Server Management Studio (Administrator)

```
SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM DI_CNTL.dbo.job_stats_vw
where job_name = 'stg_mi_genome_tags'
```

Results Messages

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
stg_mi_genome_tags	Y	stg_mi_genome_tags	success	0.02	2021-04-27 23:12:40.487	end	IMDB_FINALPROJECT	0.1

DI INJECT:

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists the database structure, including the 'imdb_stag' database and its tables. The central pane displays a T-SQL query:

```
select * from DI_stg_file_Ingest
```

The results pane shows the output of the query, which is a table with 9 rows of data. The columns are:

FileRowID	RowIngested	DI.JobID	DI.CreateDate
8781803	8781803	GxVqX5	2021-04-27 21:57:23.930
1308675	904258	VRGvY9	2021-04-27 22:00:05.947
7911600	904259	ZgHkP9	2021-04-27 22:00:06.943
373420	373430	KV9fY	2021-04-27 22:27:24.033
249	249	TlSYy	2021-04-27 22:33:41.307
486	486	IPWwU	2021-04-27 22:38:24.850
2065883	2065883	NPgJL	2021-04-27 22:45:57.770
1558444	1558444	shoKb	2021-04-27 23:06:41.870
1128	1128	8d0MCA	2021-04-27 23:12:40.487

Below the results, a message says "Query executed successfully." The status bar at the bottom right shows "localhost (15.0 RTM) info7370 (55) imdb_stag 00:00:00 9 rows".

data consistency: data is loaded from tsv files, to make consistent data we have used sk column.

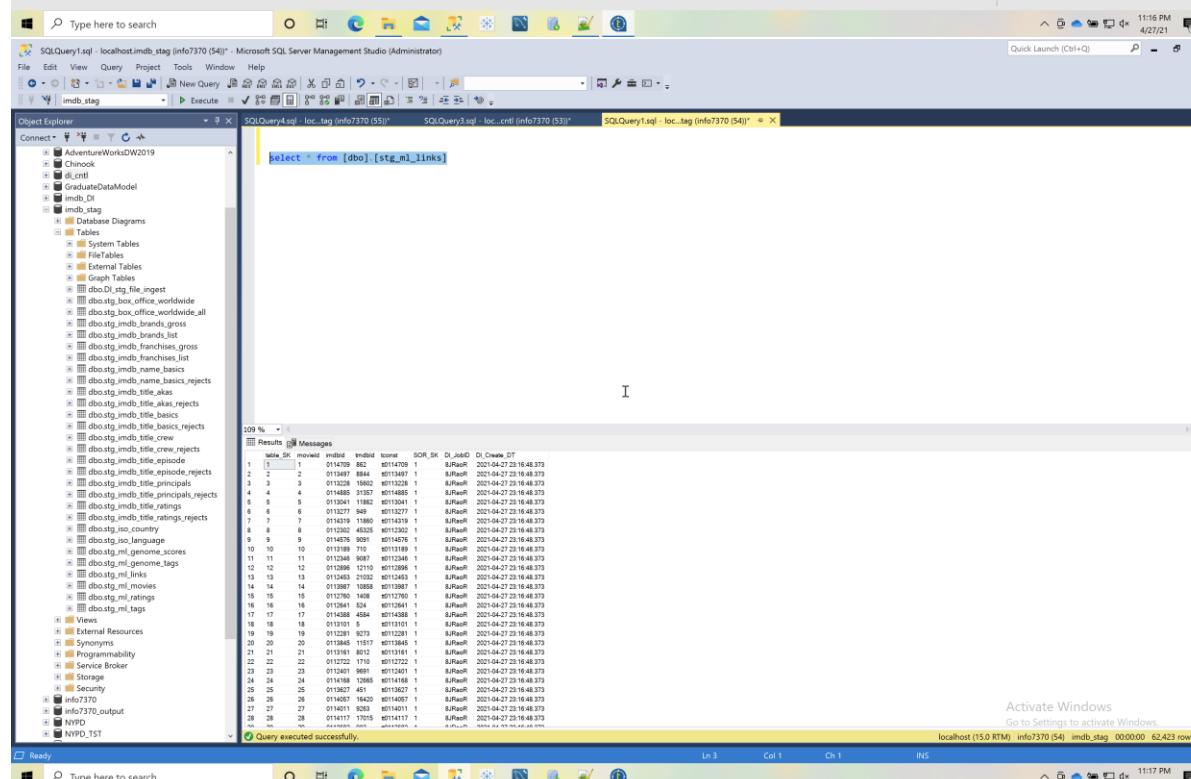
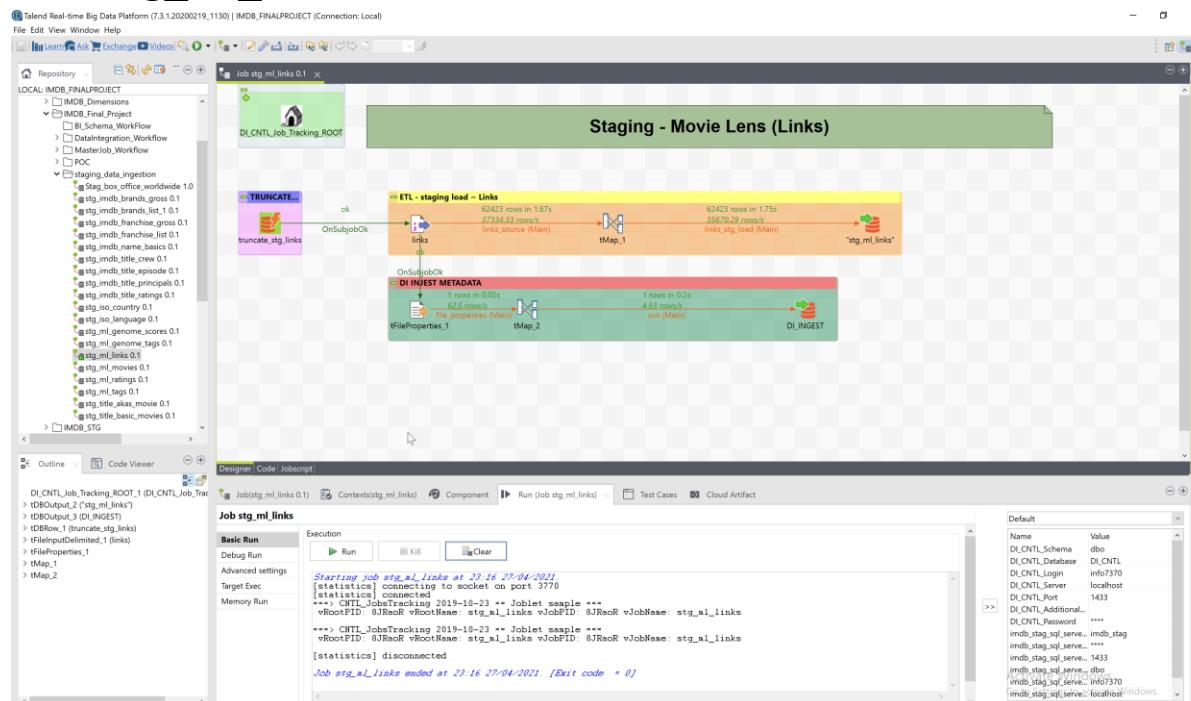
reject: No rejects required for this table.

structural integration processes: initially we are truncating the table.

For Performance we have enabled parallelization to speed up the load

We are also capturing DI ingest metadata and storing in the table which is bonus point.

16. stg_ml_links



TIME •

```

SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM DI_CNTL.dbo.job_stats_vw
WHERE job_name = 'stg_ml_links'

```

	job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
1	sg_ml_links	Y		success	0.04	2021-04-27 23:16:50.297	end	IMDB_FINALPROJECT	0.1


```

select * from DI_STG_file_ingest

```

Di_ag_file_ingest	File Group	Filename	FileRows	RowUpdated	Di_JobID	Di_CreateDate
1	1	C:\INFO\7370\dbspring2021\cdrive\finalSource\Fileslinks\imdb-dataservice_basics_movies.tsv	6781903	6781803	QXvGjg	2021-04-27 21:57:33.00
2	2	C:\INFO\7370\dbspring2021\cdrive\finalSource\Fileslinks\imdb-dataservice_genres_movies.tsv	1989126	1989126	qyDgk	2021-04-27 21:57:33.047
3	3	C:\INFO\7370\dbspring2021\cdrive\finalSource\Fileslinks\imdb-dataservice_principals_movies.tsv	774160	774160	Zgdxlu	2021-04-27 23:19:44.603
4	4	C:\INFO\7370\dbspring2021\cdrive\finalSource\Fileslinks\imdb-dataservice_ratings_movies.tsv	373450	373450	fVfPj	2021-04-27 22:37:24.033
5	5	C:\INFO\7370\dbspring2021\cdrive\finalSource\Fileslinks\iso_languages_codes_iso_all_tv_episodes.tsv	249	249	hTmJ	2021-04-27 23:38:24.827
6	6	C:\INFO\7370\dbspring2021\cdrive\finalSource\Fileslinks\iso_languages_codes_tv_episodes.tsv	486	486	BPrzvU	2021-04-27 23:38:24.823
7	7	C:\INFO\7370\dbspring2021\cdrive\finalSource\Fileslinks\imdb-dataservice_tv_movies.tsv	295983	2935683	apnTp	2021-04-27 22:42:08.670
8	8	C:\INFO\7370\dbspring2021\cdrive\finalSource\Fileslinks\imdb-dataservice_tv_tvseries.tsv	1683448	1683448	WZLc	2021-04-27 23:12:49.407
9	9	MovieLens-genome-scores	1128	1128	SofMC	2021-04-27 23:12:49.487
10	10	MovieLens-links	62423	62423	8JfRqR	2021-04-27 23:16:50.280

data consistency: data is loaded from tsv files, to make consistent data we have used sk column.

reject: No rejects required for this table.

structural integration processes: initially we are truncating the table.

For Performance we have enabled parallelization to speed up the load

We are also capturing DI inject metadata and storing in the table which is bonus point.

17. stg_ml_movies

Talend Real-time Big Data Platform (7.3.1.20200219_1130) [IMDB_FINALPROJECT (Connection: Local)]

Job stg_ml_movies 0.1

Staging - Movie Lens (Movies)

Job stg_ml_movies

Execution

Starting job stg_ml_movies at 23:20 27/04/2021
[statistics] connecting to socket on port 3393
[statistics] connected
*** CNTL_JobTracking 2019-10-23 -- Joblet sample ***
vRootID: HilHel vJobName: stg_ml_movies
*** CNTL_JobTracking 2019-10-23 -- Joblet sample ***
vRootID: HilHel vJobName: stg_ml_movies
[statistics] disconnected
Job stg_ml_movies ended at 23:20 27/04/2021. [Exit code = 0]

Default

Name	Value
DL_CNTL_Schema	dbo
DL_CNTL_Database	DL_CNTL
DL_CNTL_Login	info7370
DL_CNTL_Port	localhost
DL_CNTL_Port	1433
DL_CNTL_Additional...	
DL_CNTL_Password	****
imdb_stag_sql_seve...	imdb_stag
imdb_stag_sql_seve...	****
imdb_stag_sql_seve...	1433
imdb_stag_sql_seve...	info7370
imdb_stag_sql_seve...	localhost

SQL Server Management Studio (Administrator)

Object Explorer

Results

```
select * from [dbo].[stg_ml_movies]
```

id	imdb_sk	movie_id	title	genres	score	sk	di_jobid	di_create_dt
1	1	2	Toy Story (1995)	Adventure/Animation/Children/Comedy/Fantasy	1	H1Hck	2021-04-27 23:20:55.890	
2	2	2	Jumanji (1995)	Adventure/Children/Fantasy	1	H1Hck	2021-04-27 23:20:55.890	
3	3	3	Citizen Ruth (1996)	Comedy/Drama/Thriller	1	H1Hck	2021-04-27 23:20:55.890	
4	4	4	Waiting to Exhale (1995)	Comedy/Drama/Romance	1	H1Hck	2021-04-27 23:20:55.890	
5	5	5	Father of the Bride Part I (1995)	Comedy	1	H1Hck	2021-04-27 23:20:55.890	
6	6	6	Heat (1995)	Action/Crime/Mystery/Thriller	1	H1Hck	2021-04-27 23:20:55.890	
7	7	7	Tem and Huck (1995)	Comedy/Fantasy	1	H1Hck	2021-04-27 23:20:55.890	
8	8	8	The Birdcage (1996)	Comedy	1	H1Hck	2021-04-27 23:20:55.890	
9	9	9	Home Alone (1990)	Comedy	1	H1Hck	2021-04-27 23:20:55.890	
10	10	10	Goldfinger (1964)	Action/Adventure/Thriller	1	H1Hck	2021-04-27 23:20:55.890	
11	11	11	American President; The (1995)	Comedy/Drama/Romance	1	H1Hck	2021-04-27 23:20:55.890	
12	12	12	Dracula; Dead and Living It (1995)	Comedy/Drama/Horror	1	H1Hck	2021-04-27 23:20:55.890	
13	13	13	Born (1996)	Drama	1	H1Hck	2021-04-27 23:20:55.890	
14	14	14	Nixon (1995)	Action/Adventure/Romance	1	H1Hck	2021-04-27 23:20:55.890	
15	15	15	Cuthbert (1996)	Crime/Drama	1	H1Hck	2021-04-27 23:20:55.890	
16	16	16	Crash (1996)	Crime/Drama	1	H1Hck	2021-04-27 23:20:55.890	
17	17	17	Sense and Sensibility (1995)	Drama/Romance	1	H1Hck	2021-04-27 23:20:55.890	
18	18	18	Notting Hill (1999)	Comedy	1	H1Hck	2021-04-27 23:20:55.890	
19	19	19	Ace Ventura: When Nature Calls (1998)	Comedy	1	H1Hck	2021-04-27 23:20:55.890	
20	20	20	Money Train (1985)	Action/Comedy/Com-Drama/Thriller	1	H1Hck	2021-04-27 23:20:55.890	
21	21	21	Get Shorty (1995)	Comedy/Thriller	1	H1Hck	2021-04-27 23:20:55.890	
22	22	22	GoodFellas (1990)	Crime/Drama/Mystery/Thriller	1	H1Hck	2021-04-27 23:20:55.890	
23	23	23	Assassins (1995)	Action/Crime/Thriller	1	H1Hck	2021-04-27 23:20:55.890	
24	24	24	Philadelphia (1993)	Drama/Juris	1	H1Hck	2021-04-27 23:20:55.890	
25	25	25	Leaving Las Vegas (1995)	Drama/Romance	1	H1Hck	2021-04-27 23:20:55.890	
26	26	26	Orphelia (1995)	Drama	1	H1Hck	2021-04-27 23:20:55.890	
27	27	27	Death Becomes Her (1993)	Comedy/Drama	1	H1Hck	2021-04-27 23:20:55.890	
28	28	28	Persuasion (1995)	Drama/Romance	1	H1Hck	2021-04-27 23:20:55.890	

Activate Windows
Go to Settings to activate Windows.

TIME:

```

SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM   DI_CNTL.dbo.job_stats_vw
WHERE job_name = 'stg_ml_movies'
  
```

Results

	job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
1	stg_ml_movies	Y	stg_ml_movies	success	0.04	2021-04-27 23:20:27.600	end	IMDB_FINALPROJECT	0.1

DI INJECT:

```

select * from DI_stg_file_ingest
  
```

Results

File Group	File	FileID	Requested	D_JobID	D_CreatedOn
1	IMDB movie	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\imdb-datasummaries_basics_movie.tsv	1718153	979103	2021-04-27 23:12:57.97 23:50
2	IMDB movie	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\imdb-datasummaries_crew_movie.tsv	1308716	904236	2021-04-27 22:50:05.947
3	IMDB movie	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\imdb-datasummaries_principals_movie.t	7741500	7741500	ZyFfBv 2021-04-27 22:19:44.803
4	IMDB movie	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\imdb-datasummaries_revenue_movie.t	375436	379430	NvPF 2021-04-27 22:27.44.583
5	ISO REFERENCE courses	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\iso_isoreferencecourses_at_imdb.tsv	1307171	1307171	zJZG 2021-04-27 22:33:24.877
6	IMDB language-codes	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\iso_IsoLanguageCodesIso_tsv	486	486	BPzWtU 2021-04-27 22:38:24.820
7	IMDB movie	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\imdb-datasummaries_akimovetv_movie.tsv	2939508	2939508	sgP7Pw 2021-04-27 22:42:00.570
8	Movie genome-scores	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\movie_genomescores_imdb_movie.tsv	1393444	1393444	oJQD 2021-04-27 22:42:00.570
9	Movie genome-tags	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\movie_genometags_imdb_movie.csv	1128	1128	So0MCA 2021-04-27 23:12:40.487
10	Movie Lens links	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\movielens_links.csv	62423	62423	8JnRfR 2021-04-27 23:16:50.287
11	Movie Lens movies	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\movielens_movies.csv	62423	62423	qJLjHx 2021-04-27 23:16:50.287
12	Movie Lens tags	C:\INFO\7370\di\spring2021\di\final\Source\Filereads\movielens_tags.csv	62423	62423	KqjpxX 2021-04-27 23:22:17.297

Activate Windows
Go to Settings to activate Windows
localhost (15.0 RTM) info7370 (55) imbd_stag 00:00:00 12 rows

data consistency: data is loaded from tsv files, to make consistent data we have used sk column.

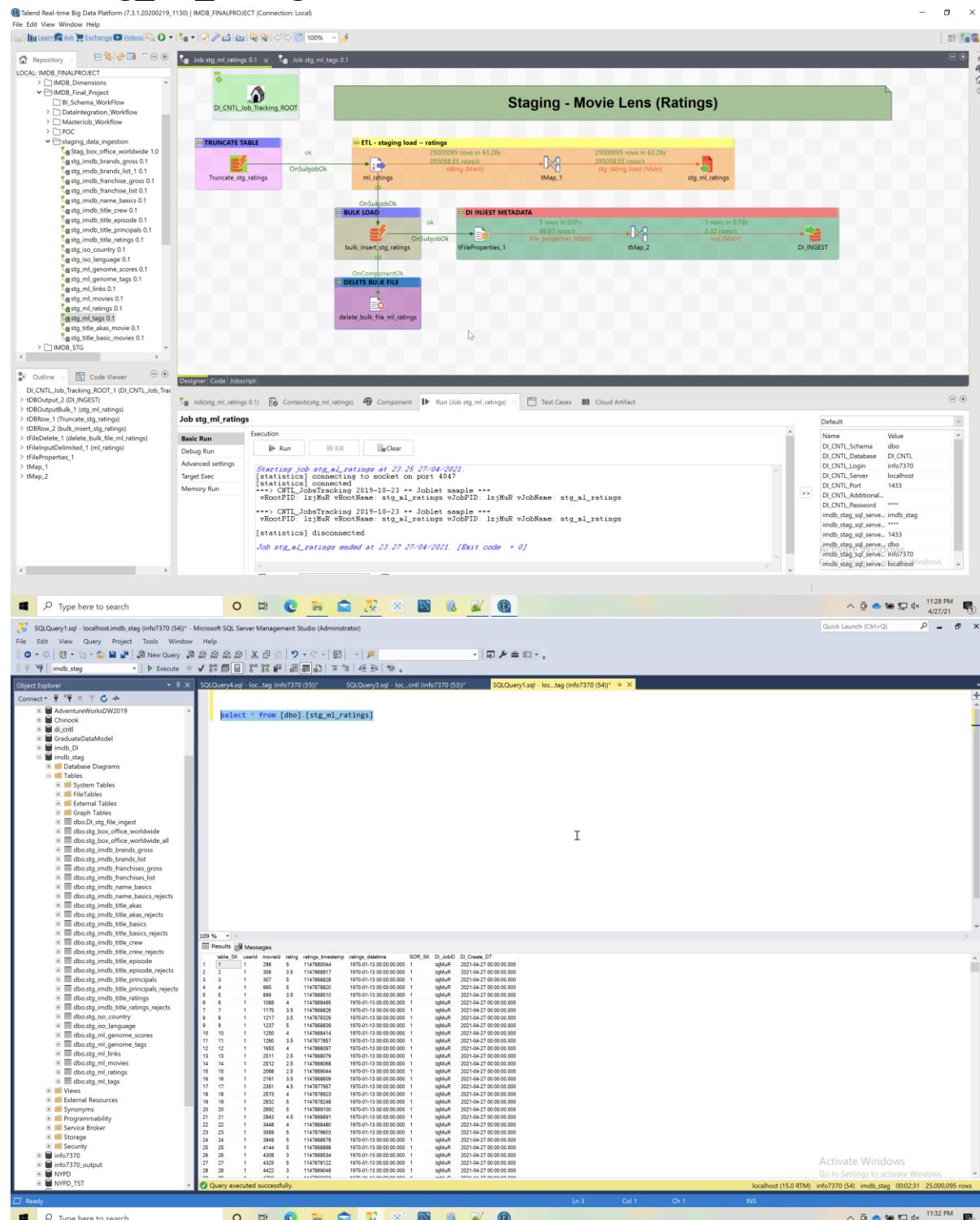
reject: No rejects required for this table.

structural integration processes: initially we are truncating the table.

For Performance we have enabled parallelization to speed up the load

We are also capturing DI inject metadata and storing in the table which is bonus point.

18. stg_ml_ratings



TIME:

```

SELECT 
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM DI_CNTL.dbo.job_stats_vw
where job_name = 'stg_ml_ratings'

```

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
stg_ml_ratings	Y	stg_ml_ratings	success	2.62	2021-04-27 23:27:54.820	end	IMDB_FINALPROJECT	0.1

DI INJECT:

```

select * from DI_STG_FILE_INGEST

```

File Name	RowInserted	Di_Audit	Di_CreatedOn
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_basics.movie.tsv	8781603	8781603	QX0VgK 2021-04-27 23:17:33.930
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_crew.movie.tsv	1000079	1000079	VqG9Pf 2021-04-27 23:17:44.947
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_genome.movie.tsv	774150	774150	ZgjF 2021-04-27 23:17:44.803
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_retag.movie.tsv	373430	373430	rVfFy 2021-04-27 23:27:24.033
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\iso_at_imdb.movie.tsv	249	249	hMzC 2021-04-27 23:27:24.037
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\iso_imdb.movie.tsv	488	488	BPPwvI 2021-04-27 23:27:24.820
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_shes.movie.tsv	2835683	2835683	apXTP 2021-04-27 23:42:08.870
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_shes.genome.movie.tsv	10000448	10000448	whDkW 2021-04-27 23:42:08.877
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_shes.genome.movie.tsv	1128	1128	SdMCs 2021-04-27 23:42:40.487
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_shes.genome.movie.tsv	62423	62423	UfRasf 2021-04-27 23:46:50.280
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_shes.genome.movie.tsv	62423	62423	KNgPwK 2021-04-27 23:22:17.287
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_shes.genome.movie.tsv	1093360	1093360	KlwoK 2021-04-27 23:26:46.947
C:\INFO\7370\dbspring2021\di\final\source\FileLinks\imdb-datasourcename_shes.genome.movie.tsv	28000096	28000096	tzfjuR 2021-04-27 23:27:54.803

data consistency: data is loaded from tsv files, to make consistent data we have used sk column.

reject: No reject required for this table.

structural integration processes: initially we are truncating the table

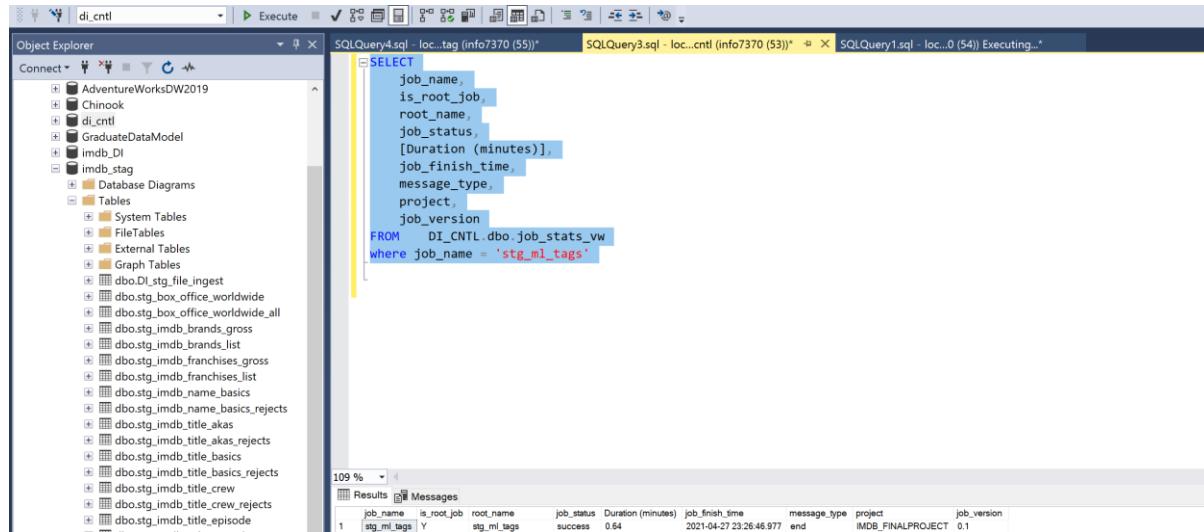
For performance we are using bulk components,

For DI standard we are deleting bulk file which is created in the process

We are also capturing DI inject metadata and storing in the table which is bonus point.

19. stg_ml_tags

TIME:



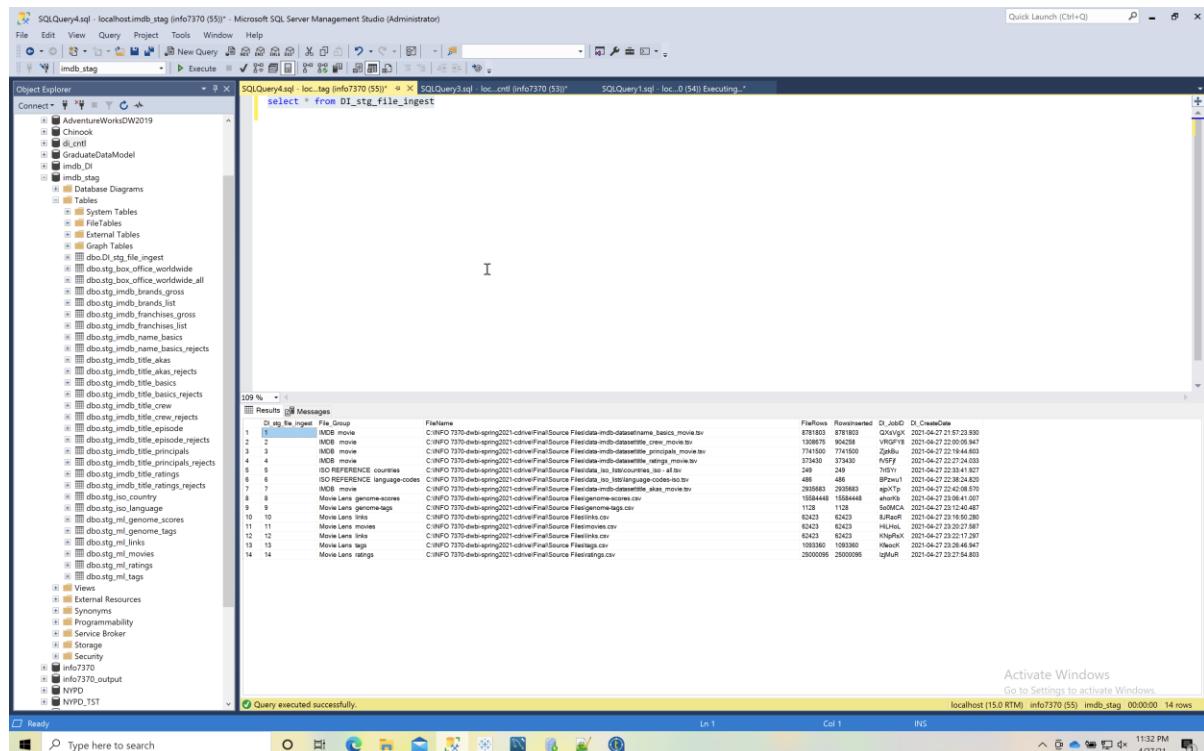
```

SELECT
    job_name,
    is_root_job,
    root_name,
    job_status,
    [Duration (minutes)],
    job_finish_time,
    message_type,
    project,
    job_version
FROM   DI_CNTL.dbo.job_stats_vw
where job_name = 'stg_ml_tags'
  
```

The screenshot shows the Object Explorer on the left with the 'di_CNTL' database selected. In the center, a query window displays the above SQL code. Below it, the results pane shows a single row of data:

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_finish_time	message_type	project	job_version
stg_ml_tags	Y	stg_ml_tags	success	0.64	2021-04-27 23:26:46.977	end	IMDB_FINALPROJECT	0.1

DI INJECT:



```

select * from DI_stg_file_ingest
  
```

The screenshot shows the Object Explorer on the left with the 'di_CNTL' database selected. In the center, a query window displays the above SQL code. Below it, the results pane shows a large dataset with 14 rows of file information:

File Group	Filename	FileSize	Reclaimed	Di_Auto	Di_Created
IMDB movie	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\imdb-datasource_name_basics_movie.tsv	8781603	8781603	QxVgXf	2021-04-27 23:17:23.930
IMDB movie	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\imdb-datasource_crew_movie.tsv	1309875	904305	VRFqFm	2021-04-27 23:20:09.847
IMDB movie	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\imdb-datasource_genre_movie.tsv	1741513	1042000	2qGJkA	2021-04-27 23:20:44.683
IMDB movie	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\imdb-datasource_rating_movie.tsv	573430	379430	hFfF	2021-04-27 22:27:24.033
ISO REFERENCE countries	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\iso_intlcountries_iso-all.tsv	249	249	THYY	2021-04-27 23:31:41.827
ISO REFERENCE language-codes	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\iso_intllanguage-codes-iso.tsv	489	489	BpHfP	2021-04-27 23:31:42.249
IMDB movie	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\imdb-datasource_title_rejects_rejects	2935683	2935683	egKTP	2021-04-27 22:42:08.570
IMDB movie	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\imdb-datasource_title_rejects_rejects	15584448	15584448	shwRk	2021-04-27 23:38:41.507
Movie Lens genre-tags	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\ml-100k\ml-100k-genre-tags.csv	1123	1123	SGCQ	2021-04-27 23:16:50.487
Movie Lens lens	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\ml-100k\ml-100k-lens.csv	62423	62423	RRfRsf	2021-04-27 23:16:50.280
Movie Lens movies	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\ml-100k\ml-100k-movies.csv	62423	62423	HLHsl	2021-04-27 23:20:27.587
Movie Lens ratings	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\ml-100k\ml-100k-ratings.csv	62423	62423	WzJZ	2021-04-27 23:16:50.787
Movie Lens tags	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\ml-100k\ml-100k-tags.csv	1093360	1093360	KwKwK	2021-04-27 23:26:46.947
Movie Lens ratings	C:\INFO\7370\dev\spring2021\cdbe\FinalSource\Filesdata\ml-100k\ml-100k-ratings.csv	28000096	28000096	tzfMuR	2021-04-27 23:27:54.803

At the bottom of the results pane, it says "Query executed successfully."

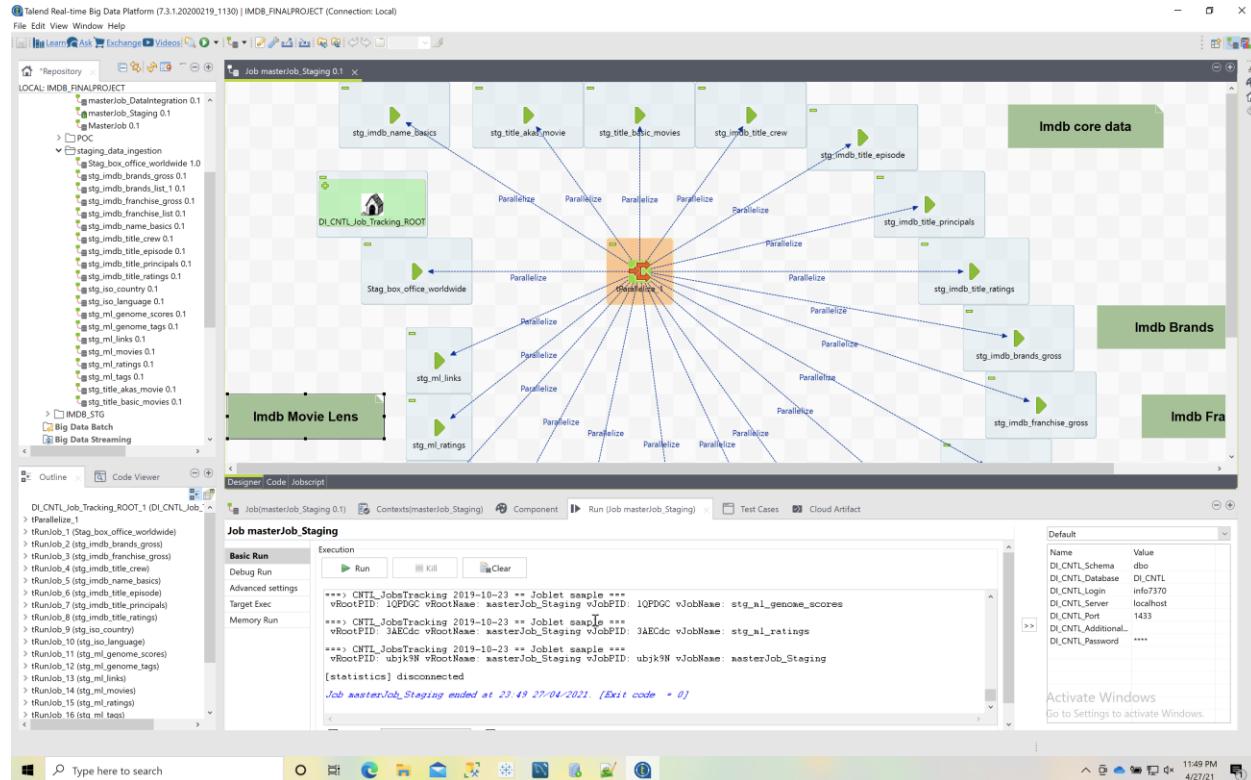
data consistency: data is loaded from tsv files, to make consistent data we have used sk column.

reject: No rejects required for this table.

structural integration processes: initially we are truncating the table.

For Performance we have enabled parallelization to speed up the load
 We are also capturing DI ingest metadata and storing in the table which is bonus point.

STAGING MASTER:



TIME: 7.54min

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. The Object Explorer on the left shows a database structure with multiple schemas and tables. In the center, there are four tabs: "SQLQuery4.sql - loc_tag (info7370 (55))", "SQLQuery3.sql - loc_cntl (info7370 (53))", "SQLQuery1.sql - loc_tag (info7370 (54))", and "Messages". The "Messages" tab is active, displaying a single row of log information. The "Results" tab shows a query result set from the "job_stats_vw" view, listing various job statistics. The status bar at the bottom indicates the query took 7.54 minutes to execute.

job_name	is_root_job	root_name	job_status	Duration (minutes)	job_start_time	message_type	project	job_version
masterJob_Staging	Y	masterJob_Staging	success	7.54	2021-04-27 23:49:09.570	end	IMDB_FINALPROJECT	0.1