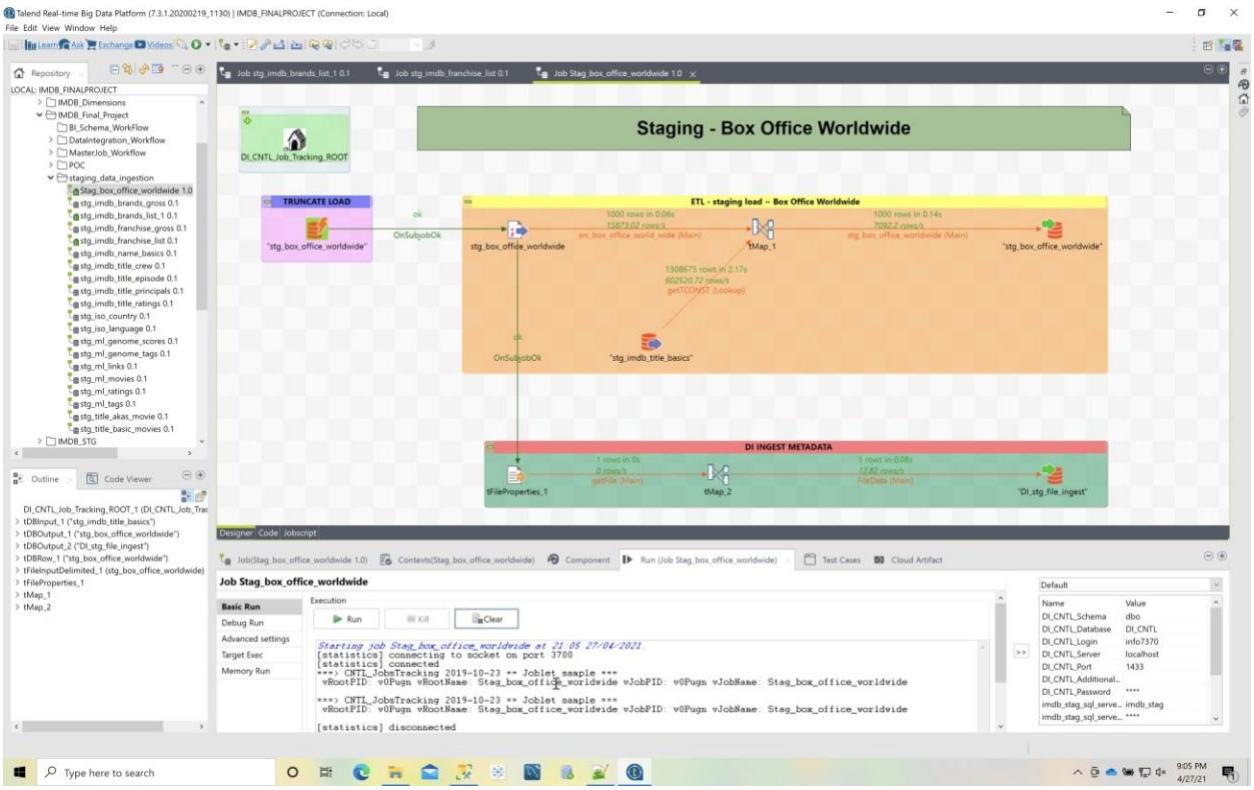
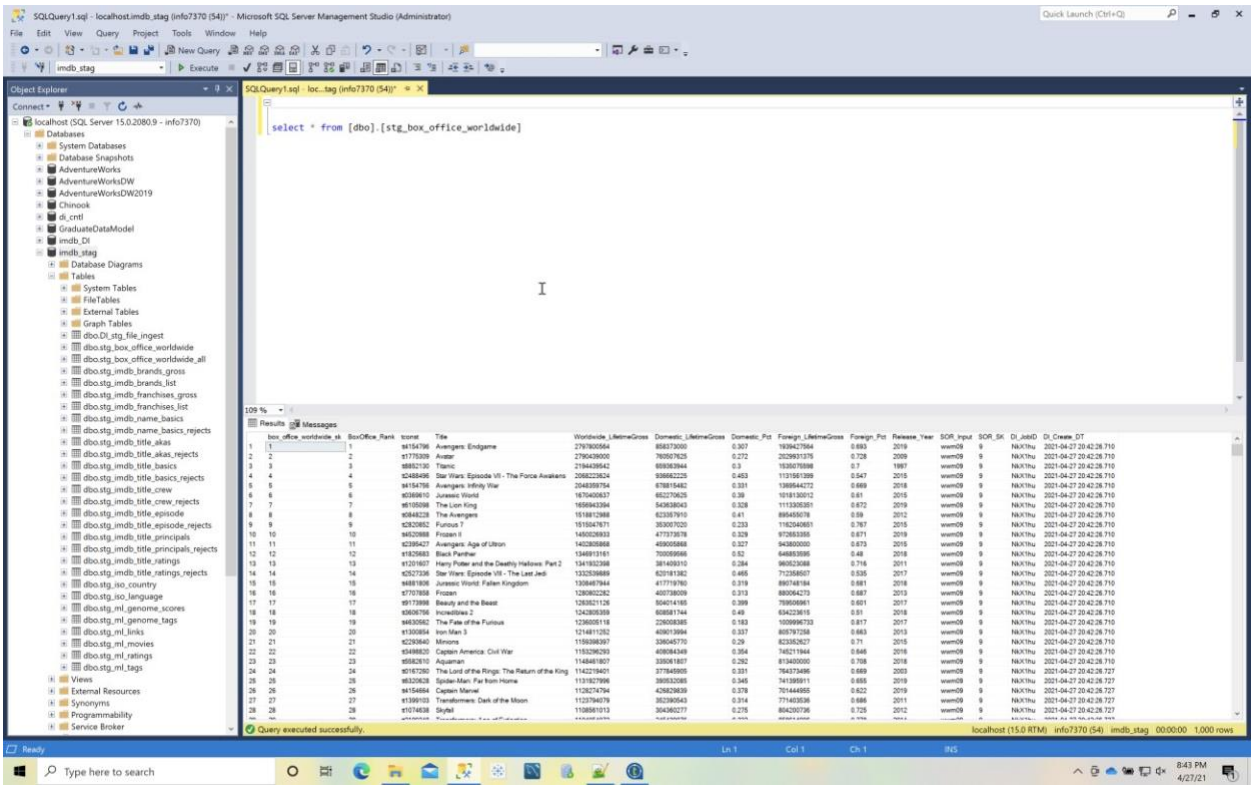
Staging Screen Shots- Talend

1. Stag_box_office_worldwide

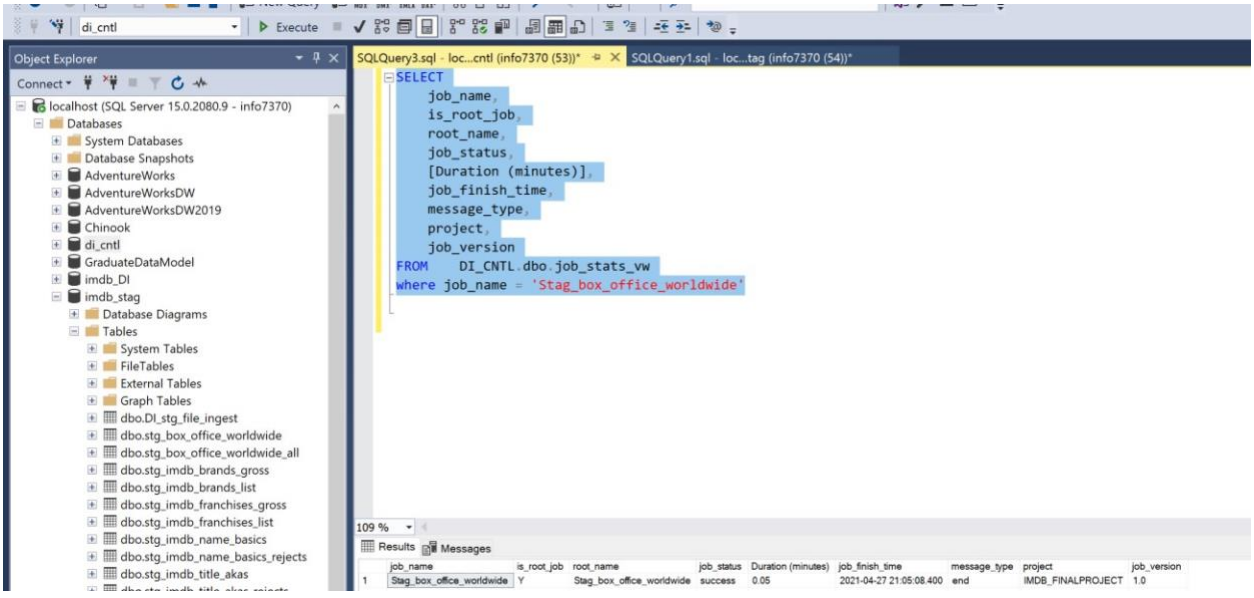


Submitted By: Shweta Gupta

Staging Screen Shots- Talend



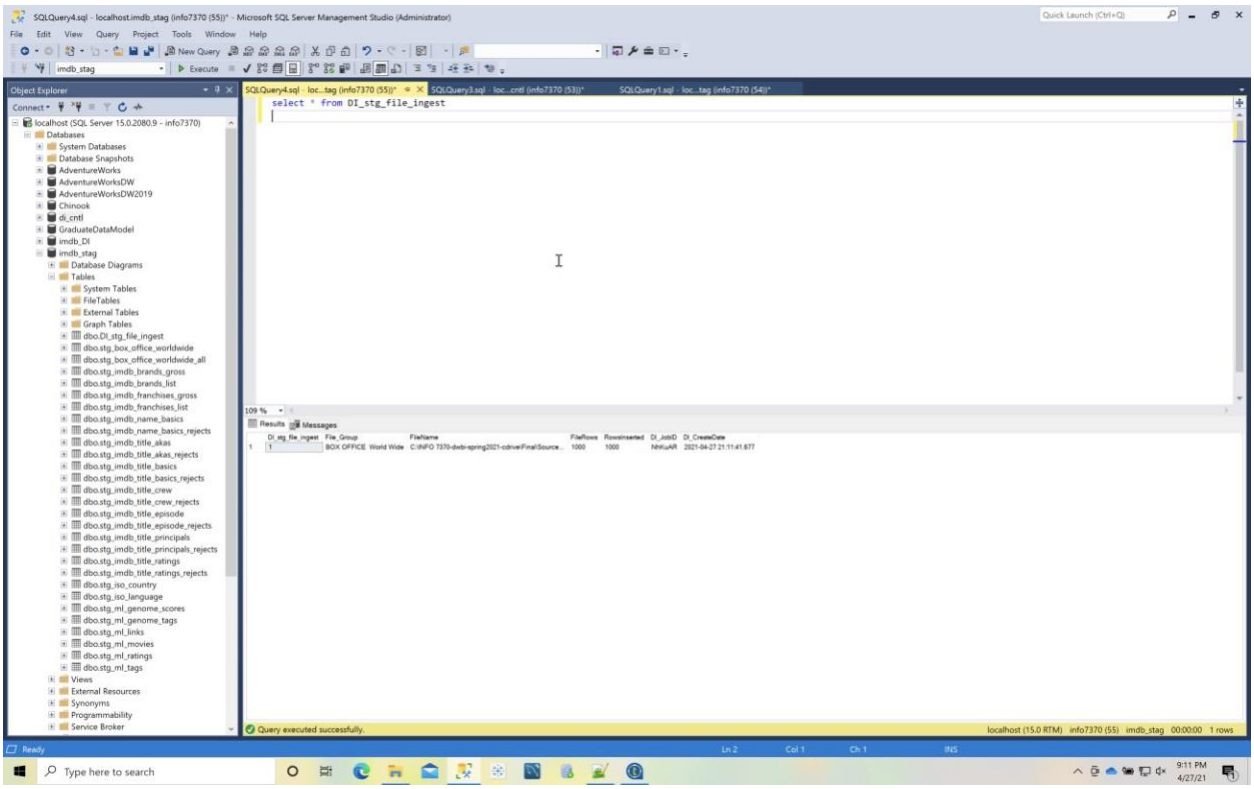
Time:



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

DI INJEST:



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

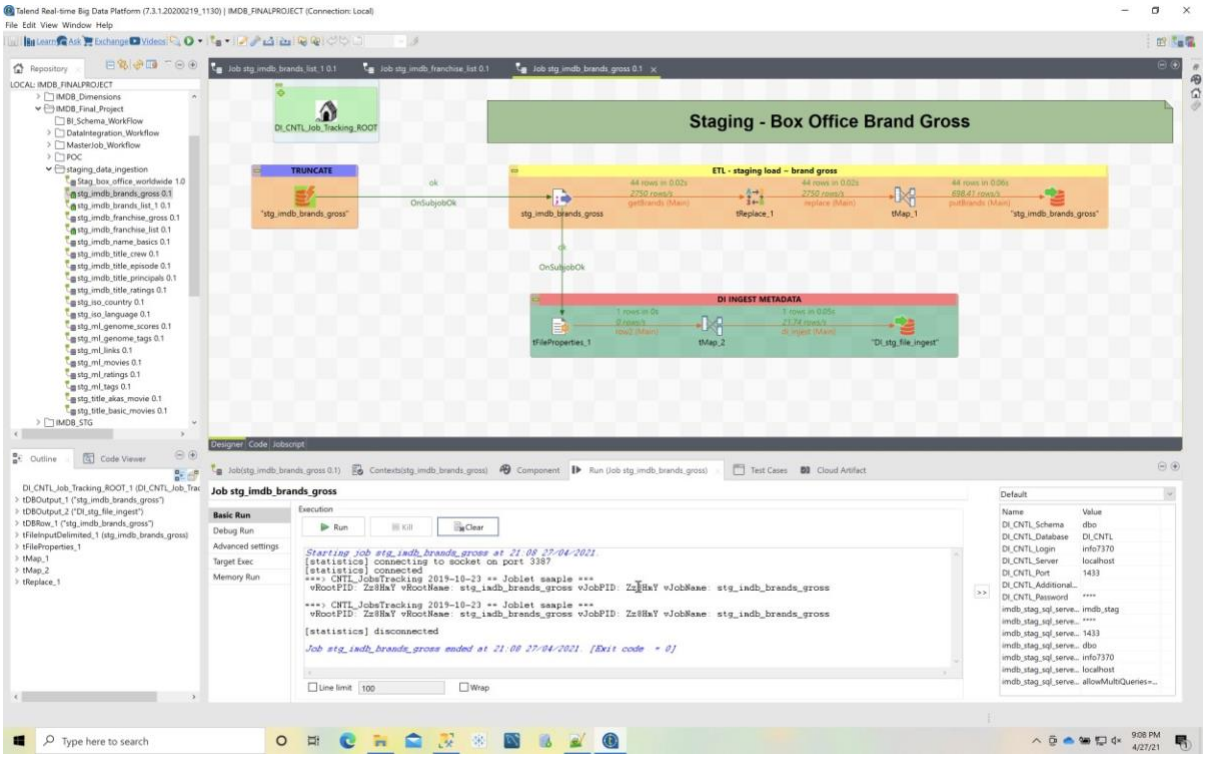
reject: for this table we are not supposed to do any rejects.

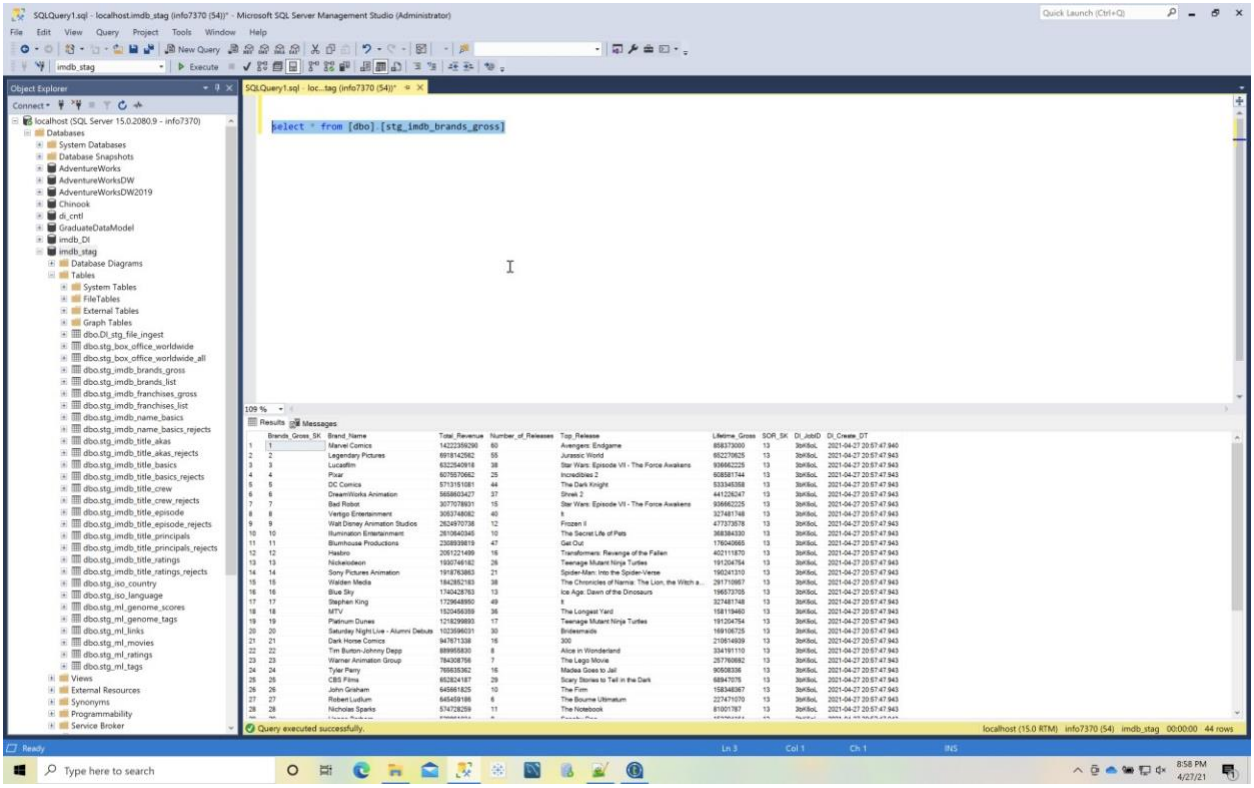
structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component and adding cols using tmap for create_date and DI_ETL_ID which is process id, other than that we have also completed bonus part where in we are populating DI_stg_injest table.

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

2. stg_imdb_brands_gross

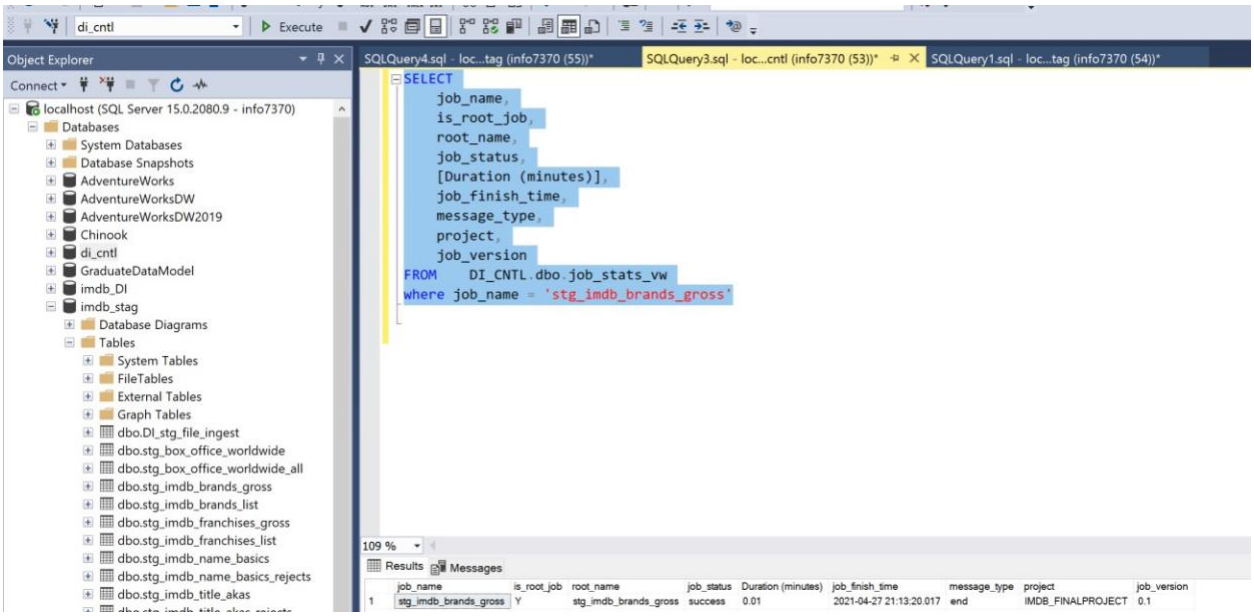




Submitted By: Shweta Gupta

Staging Screen Shots- Talend

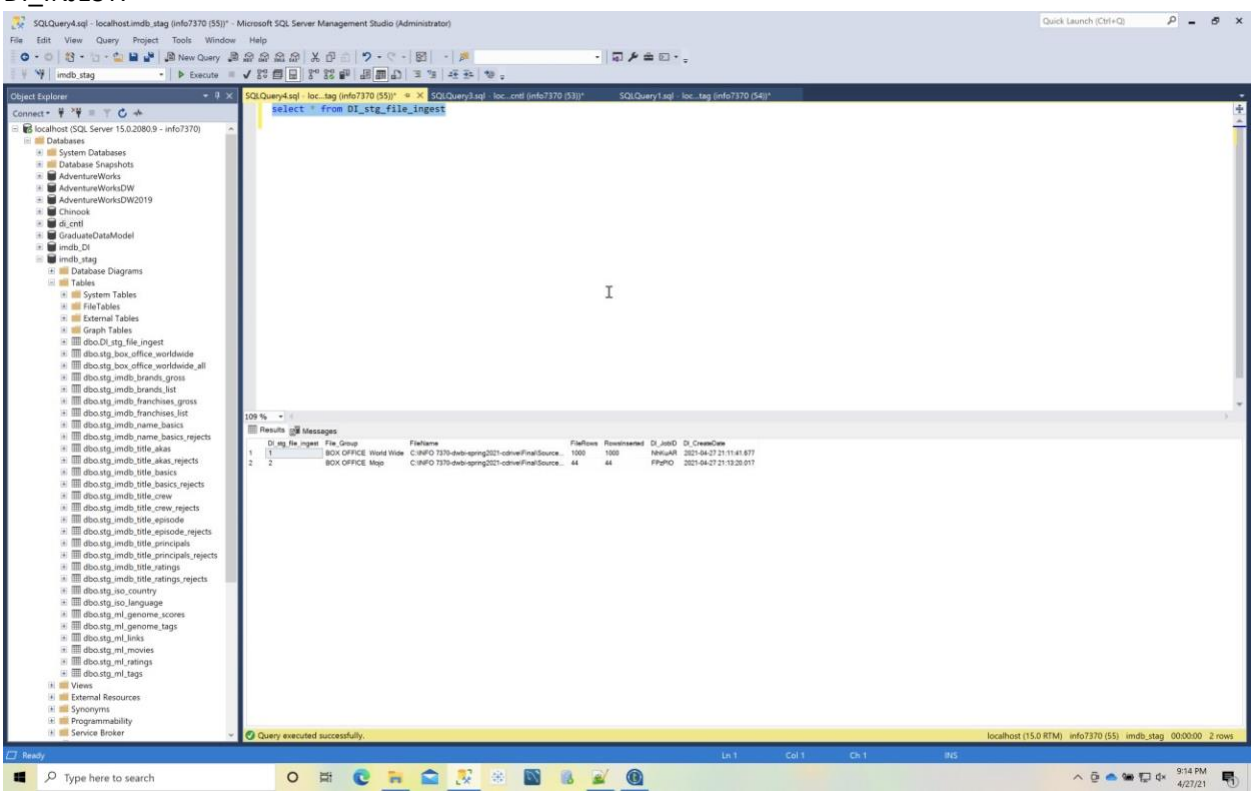
TIME



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

DI_INJEST:



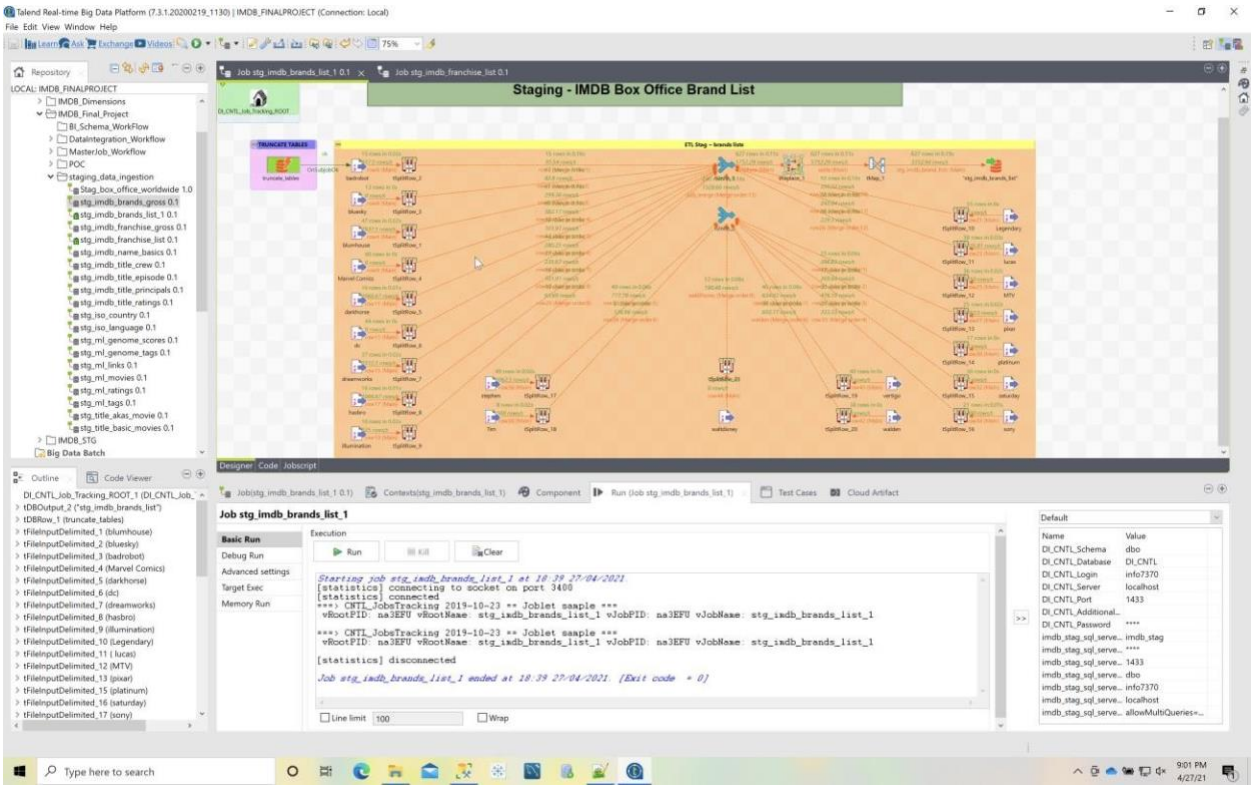
data consistency: data is loaded from file to stag table, data remains consistent with sk values and referential integrity.

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.

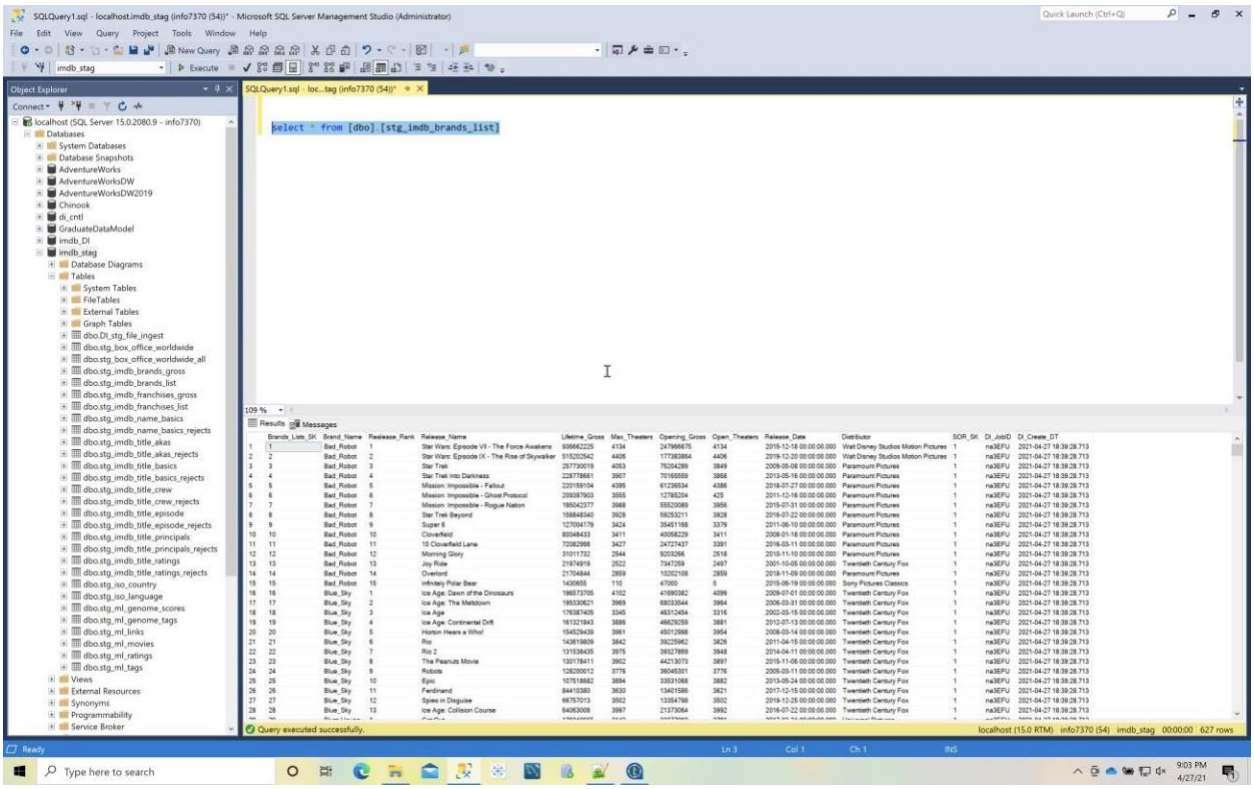
3. stg_imdb_brands_list

Staging Screen Shots- Talend



Submitted By: Shweta Gupta

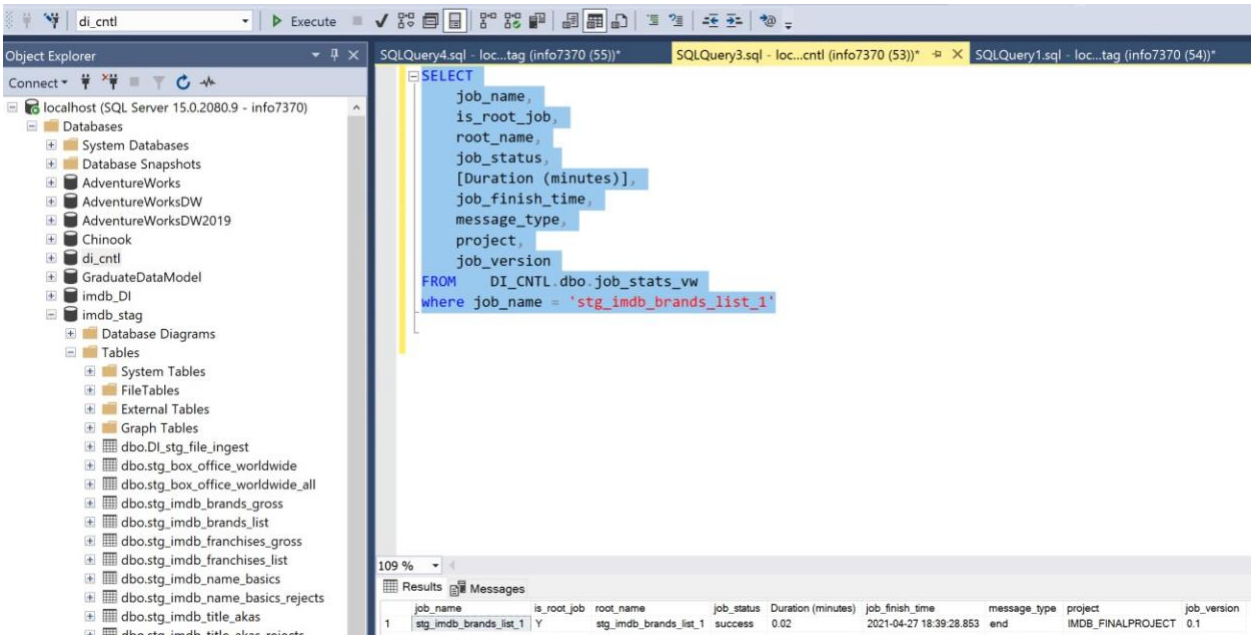
Staging Screen Shots- Talend



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

TIME:



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

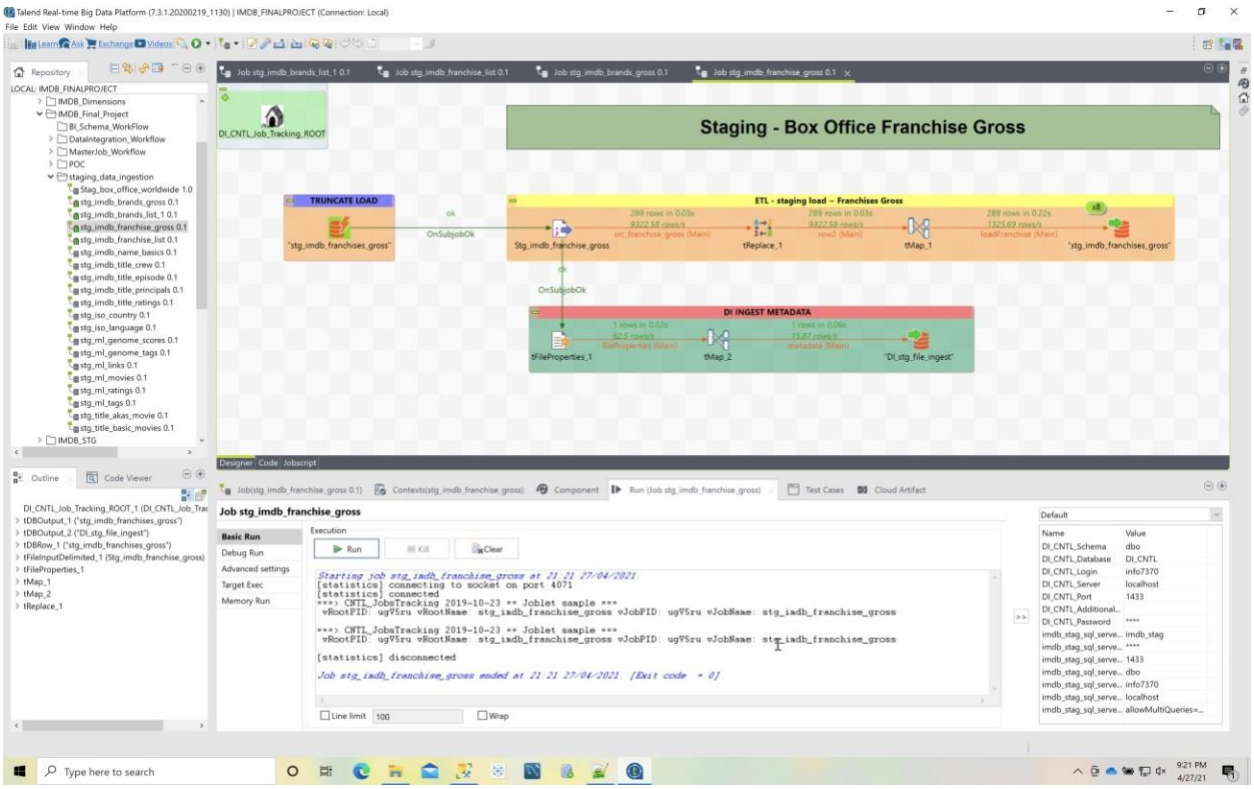
reject: for this table we are not supposed to do any rejects.

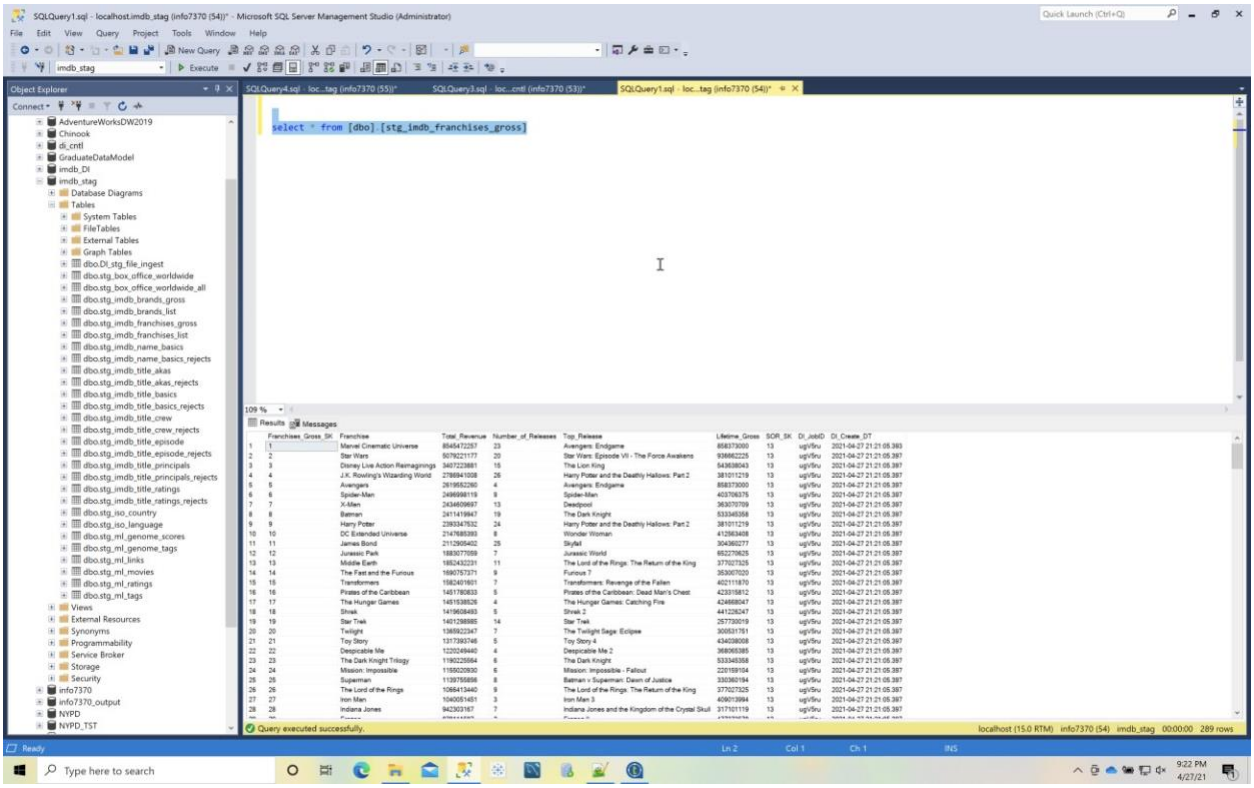
structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component, tUnite to merge all the files into a single result set table also, added cols using tmap for create_date and DI_ETL_ID which is process id, other than that we have also completed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

4. stg_imdb_franchise_gross

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

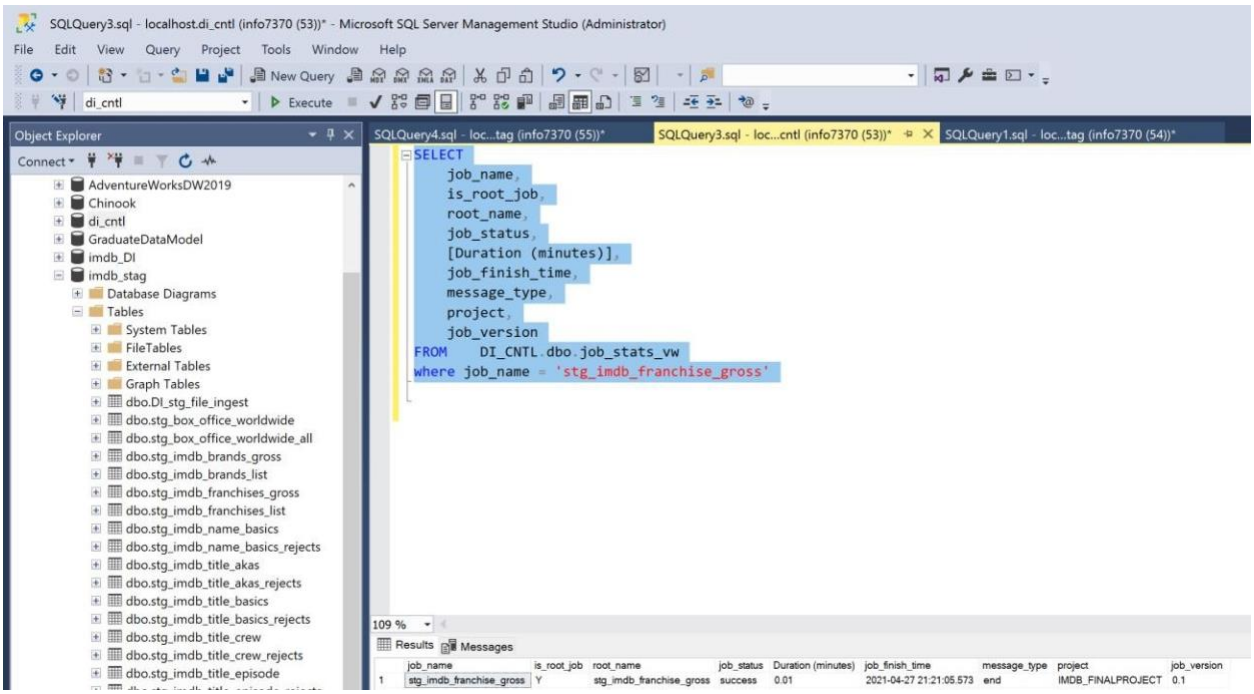




Submitted By: Shweta Gupta

Staging Screen Shots- Talend

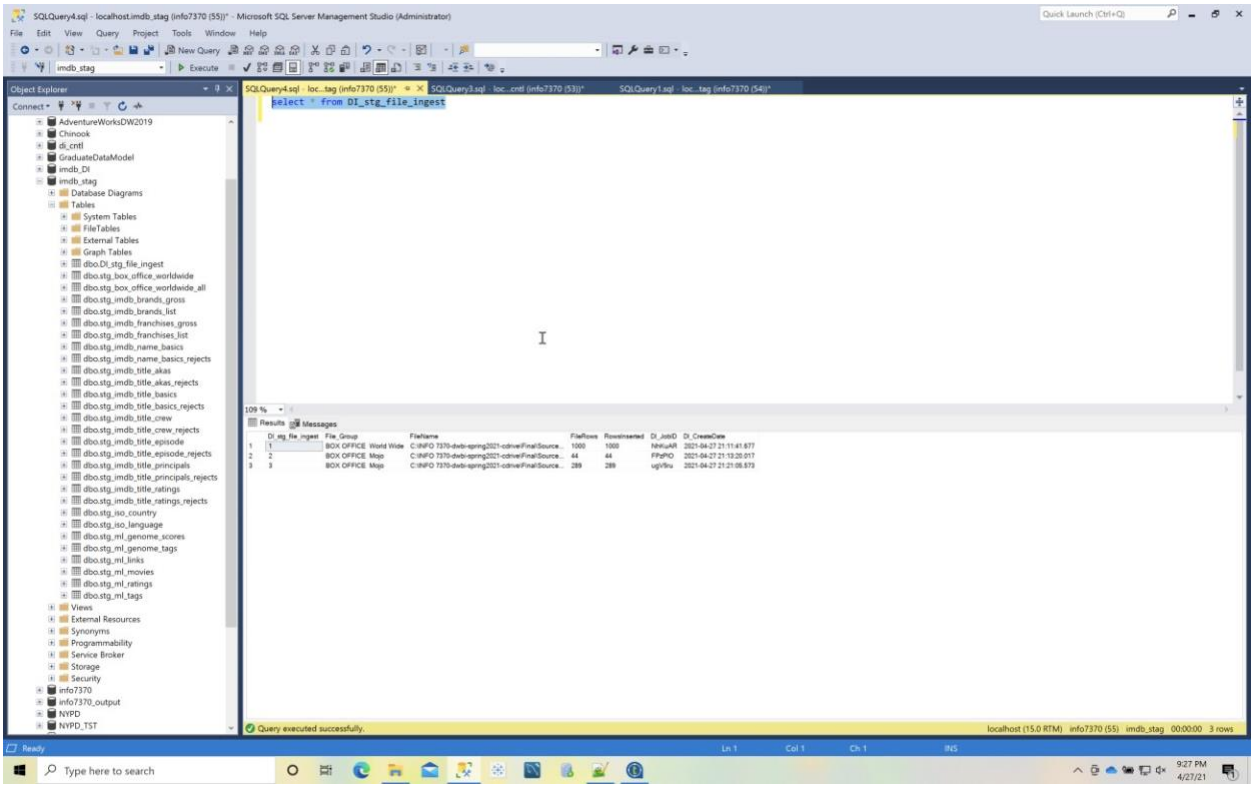
TIME:



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

DI INJEST:



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

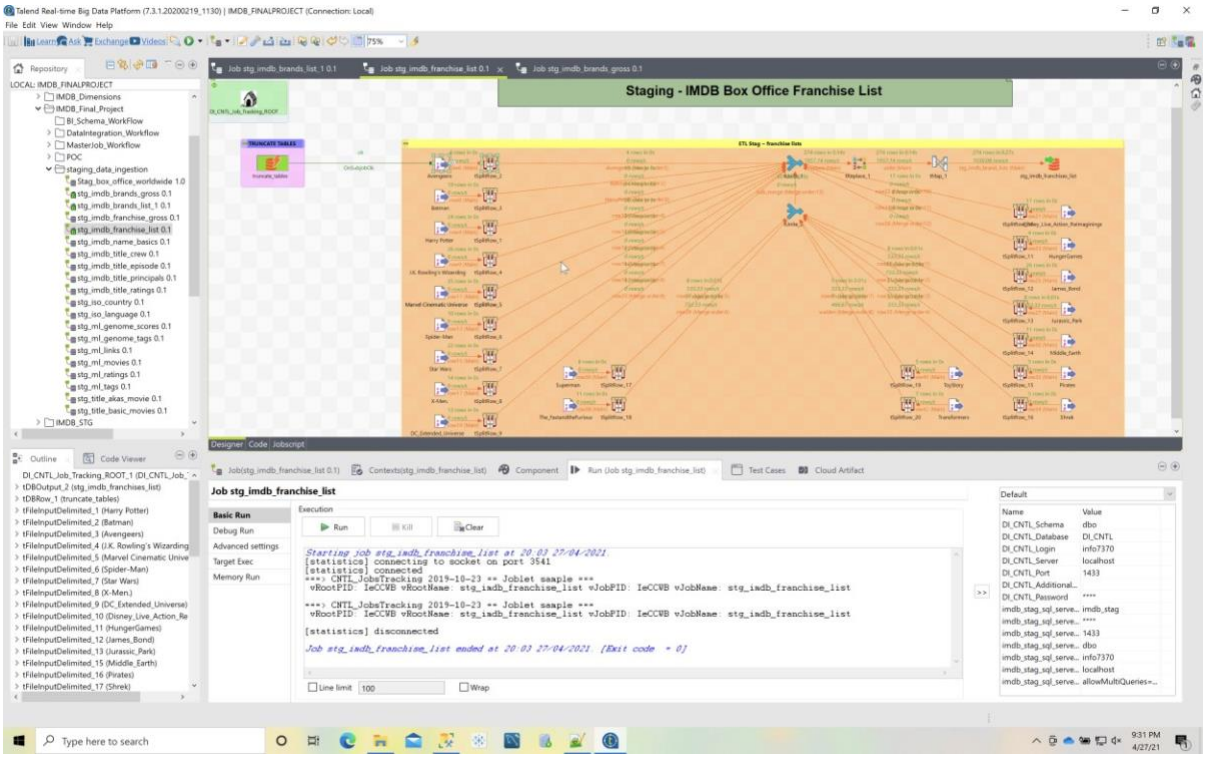
reject: for this table we are not supposed to do any rejects.

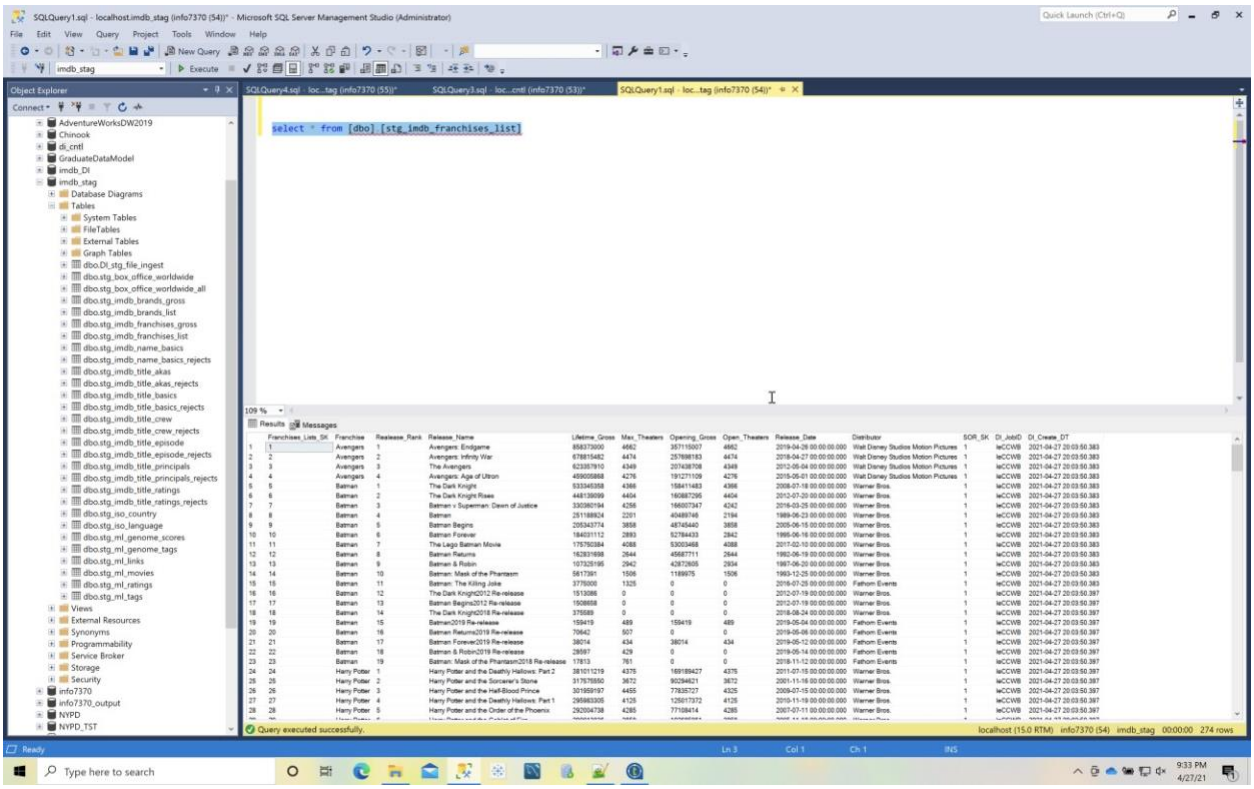
structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component, we have added cols using tmap for create_date and DI_ETL_ID which is process id, other than that we have also completed bonus part where in we are populating DI_stg_injest table.

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

5. stg_imdb_franchise_list

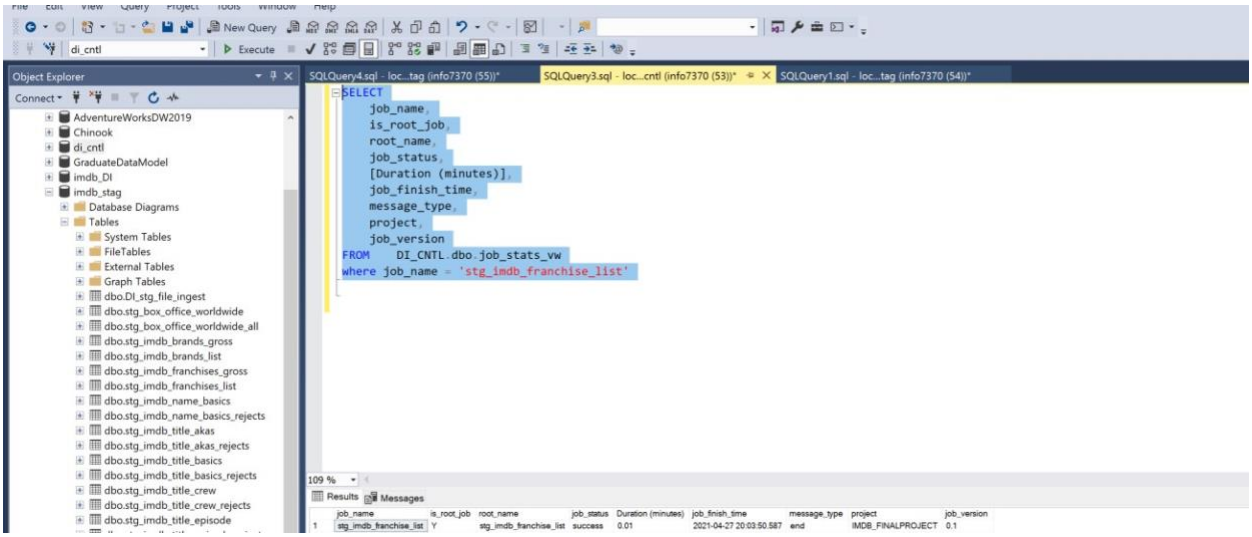




Submitted By: Shweta Gupta

Staging Screen Shots- Talend

TIME:



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

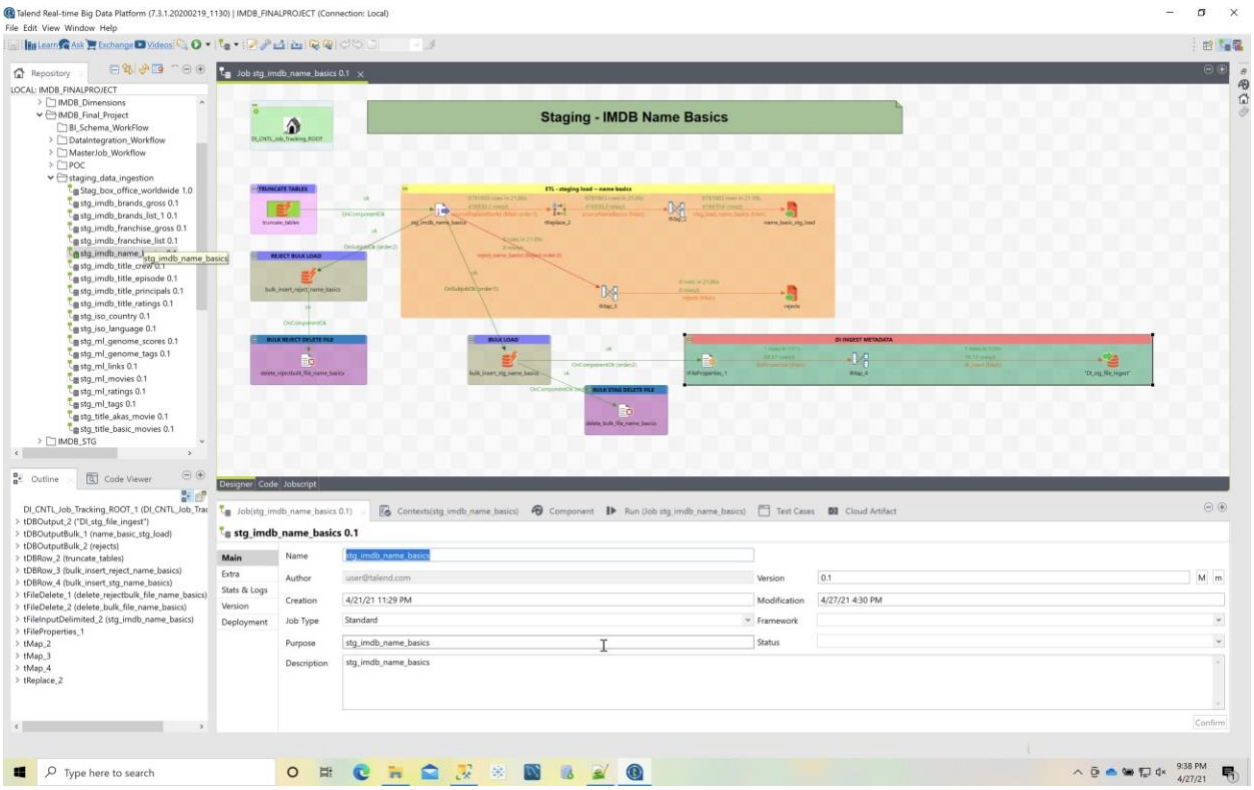
reject: for this table we are not supposed to do any rejects.

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.

Submitted By: Shweta Gupta

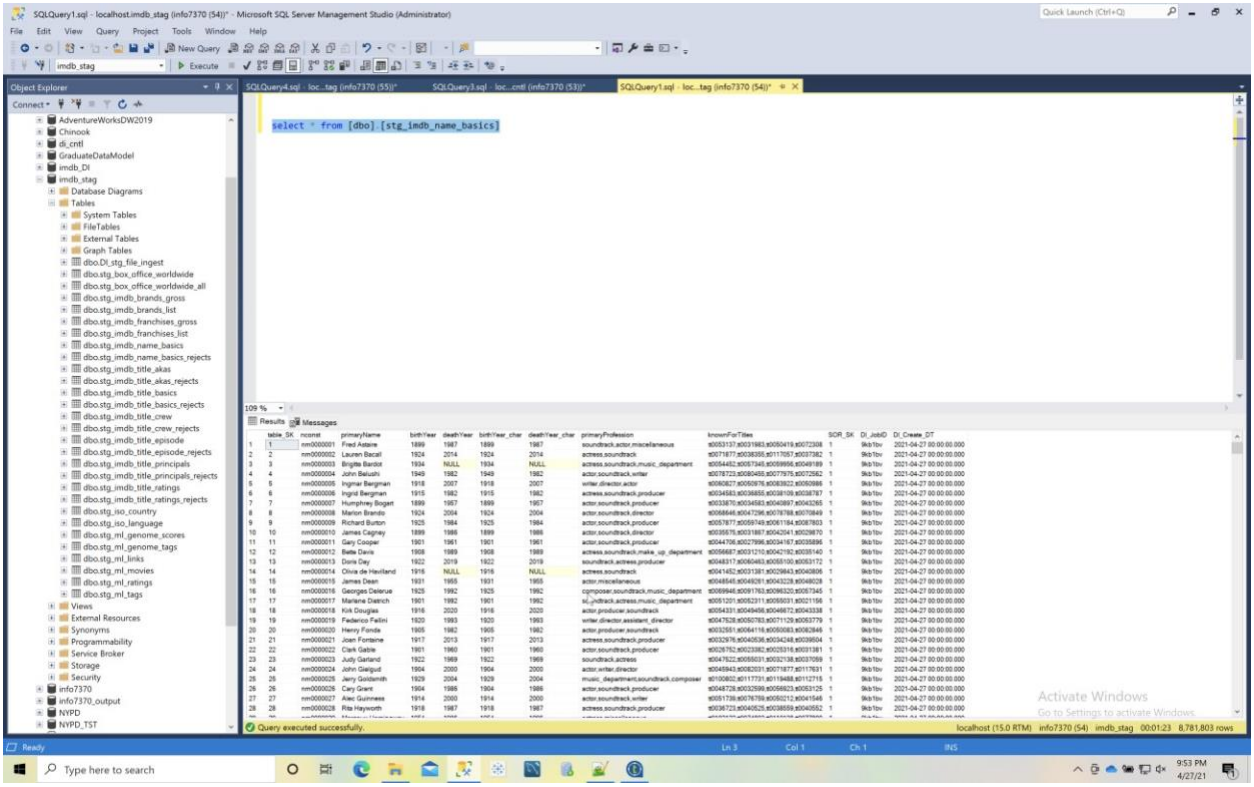
Staging Screen Shots- Talend

6. stg_imdb_name_basics & Rejects

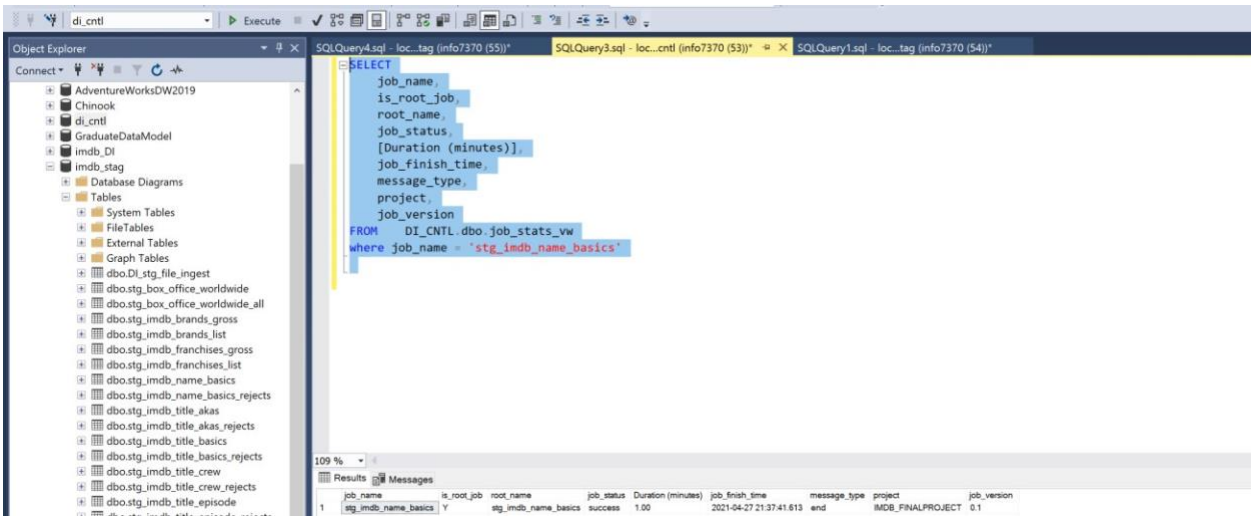


Submitted By: Shweta Gupta

Staging Screen Shots- Talend



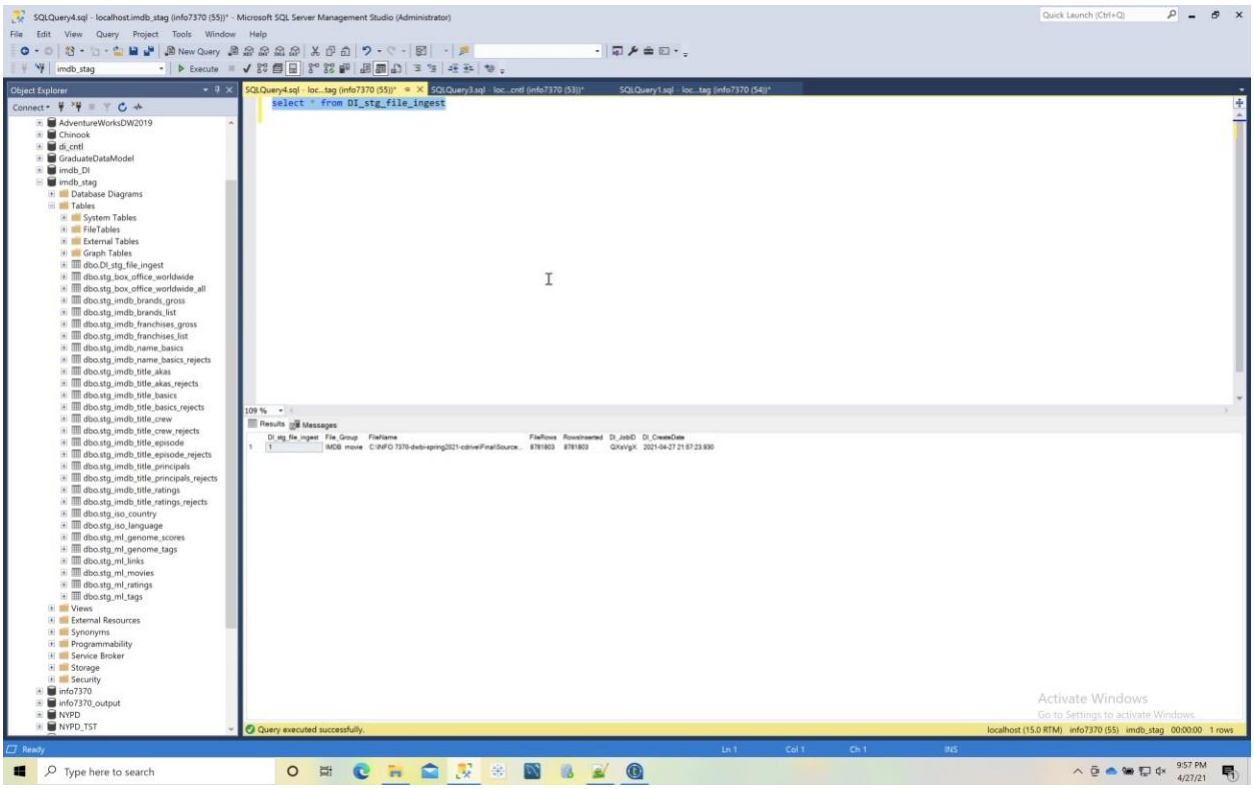
TIME:



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

DI INJEST:



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,

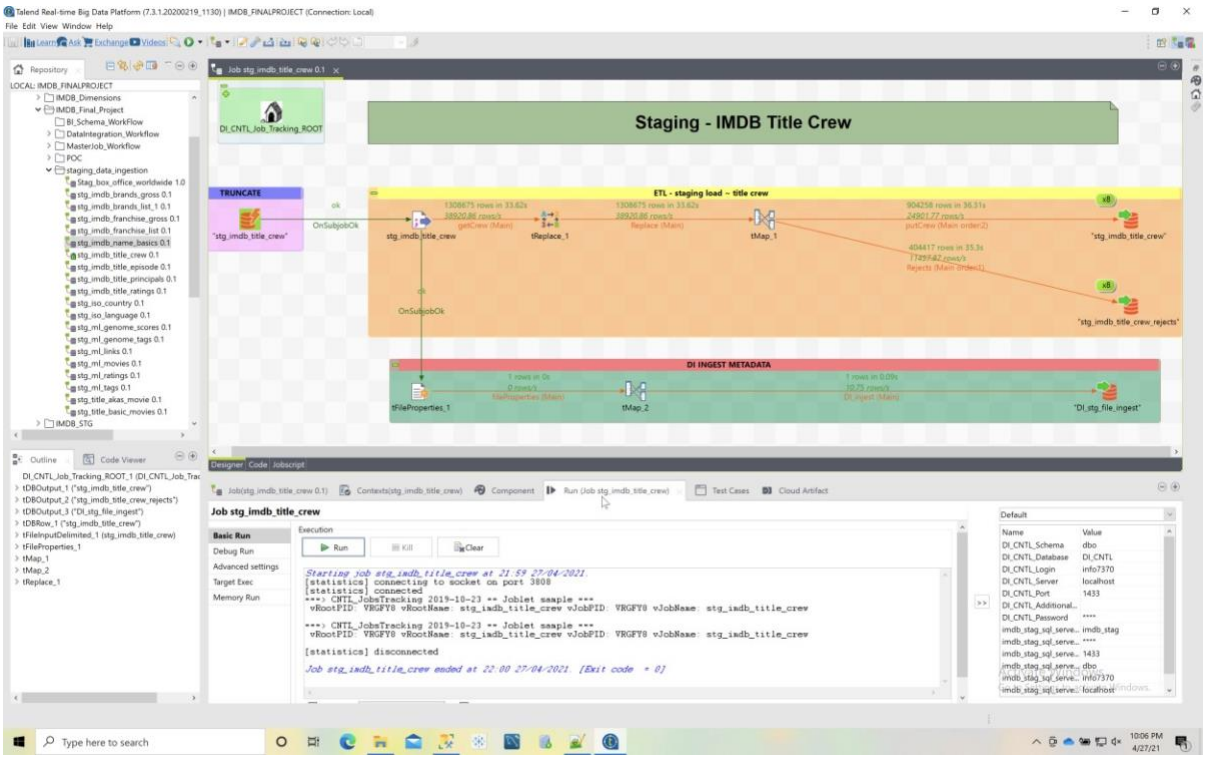
For DI standard we are deleting bulk file which is created in the process

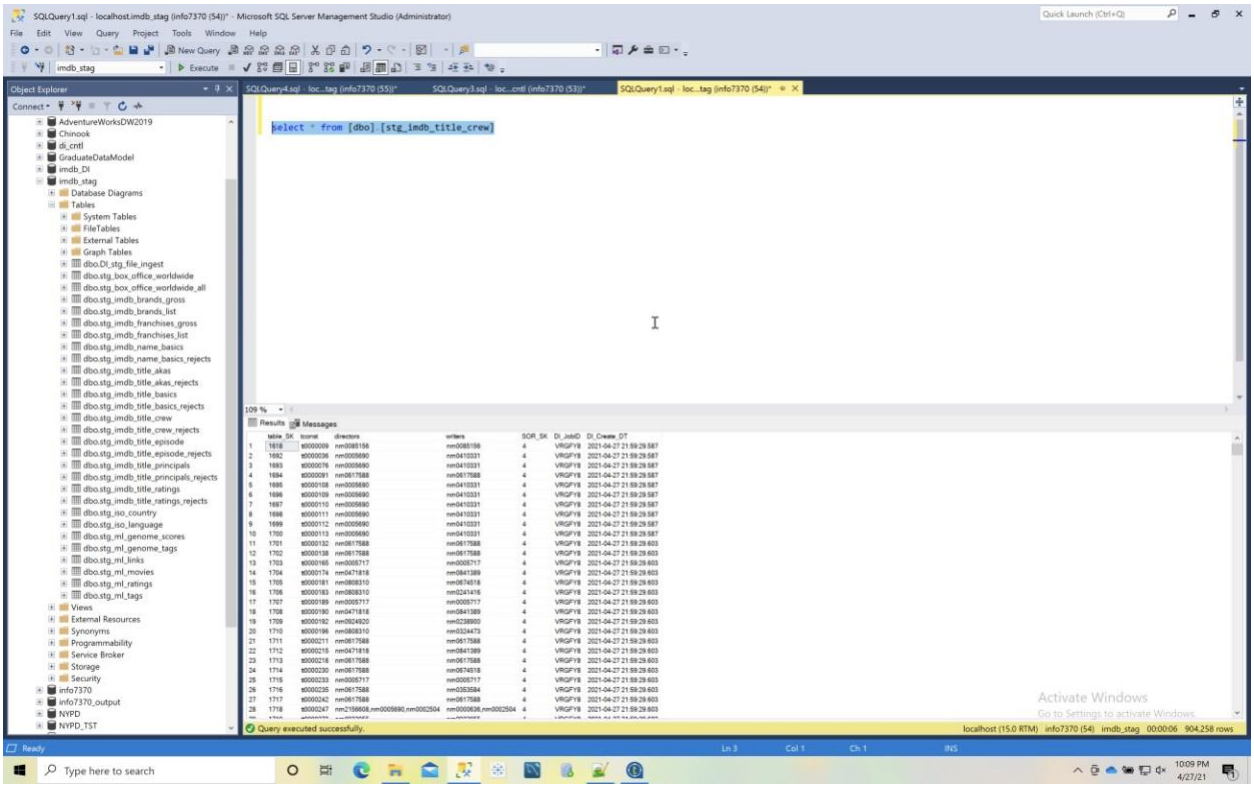
We are also capturing DI injest metadata and storing in the table which is bonus point.

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

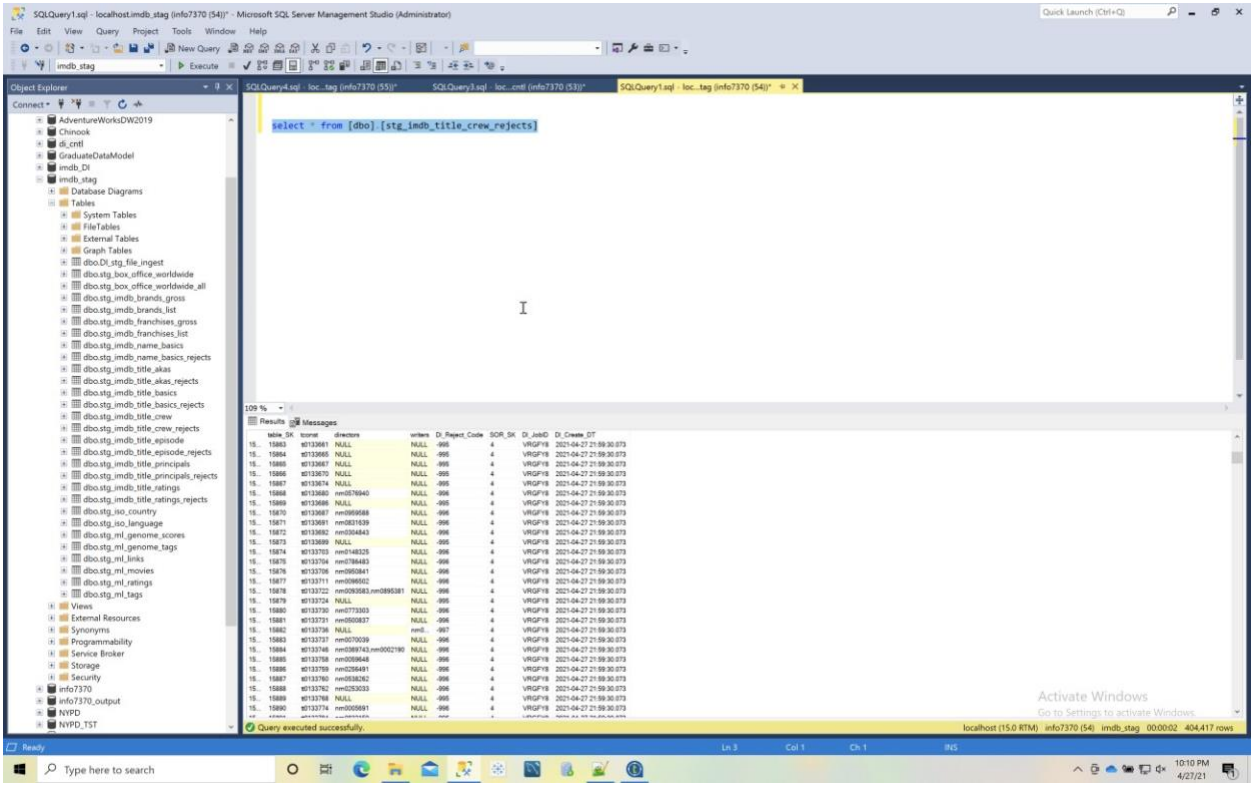
7. stg_imdb_title_crew & Rejects



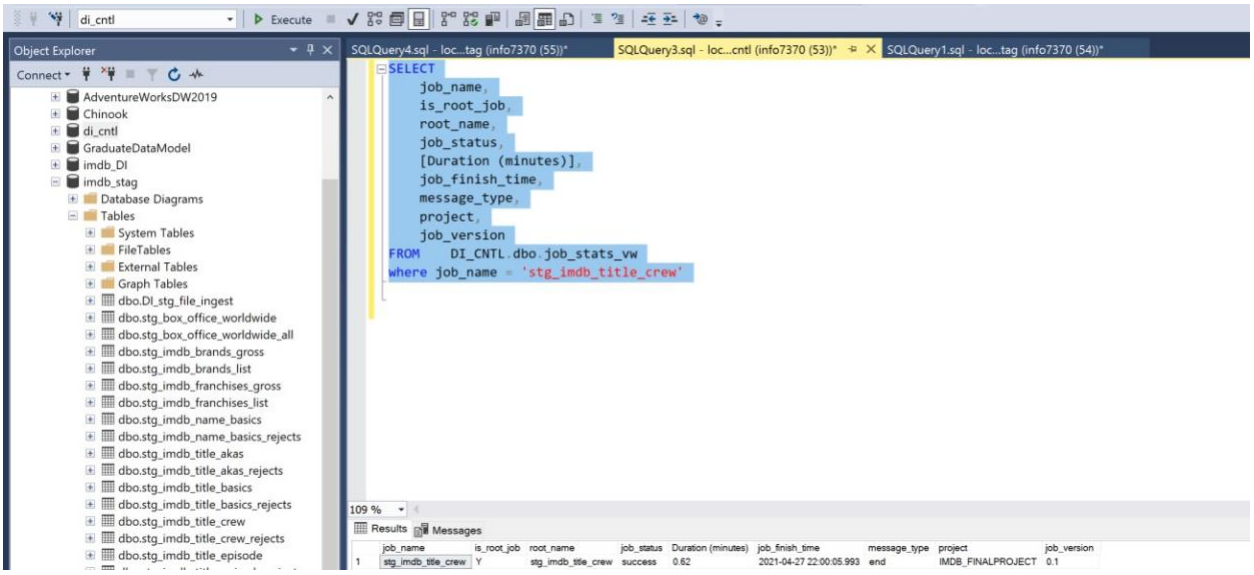


Submitted By: Shweta Gupta

Staging Screen Shots- Talend



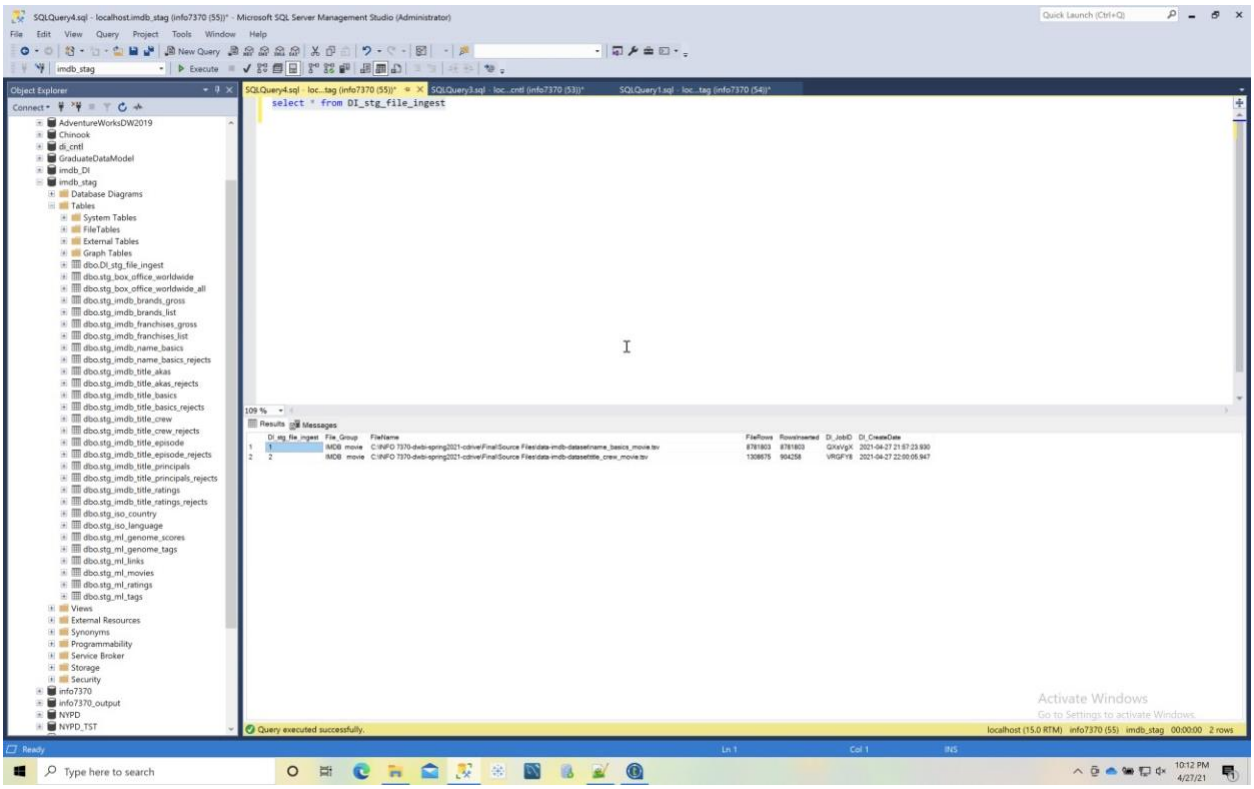
TIME:



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

DI INJEST:



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 :

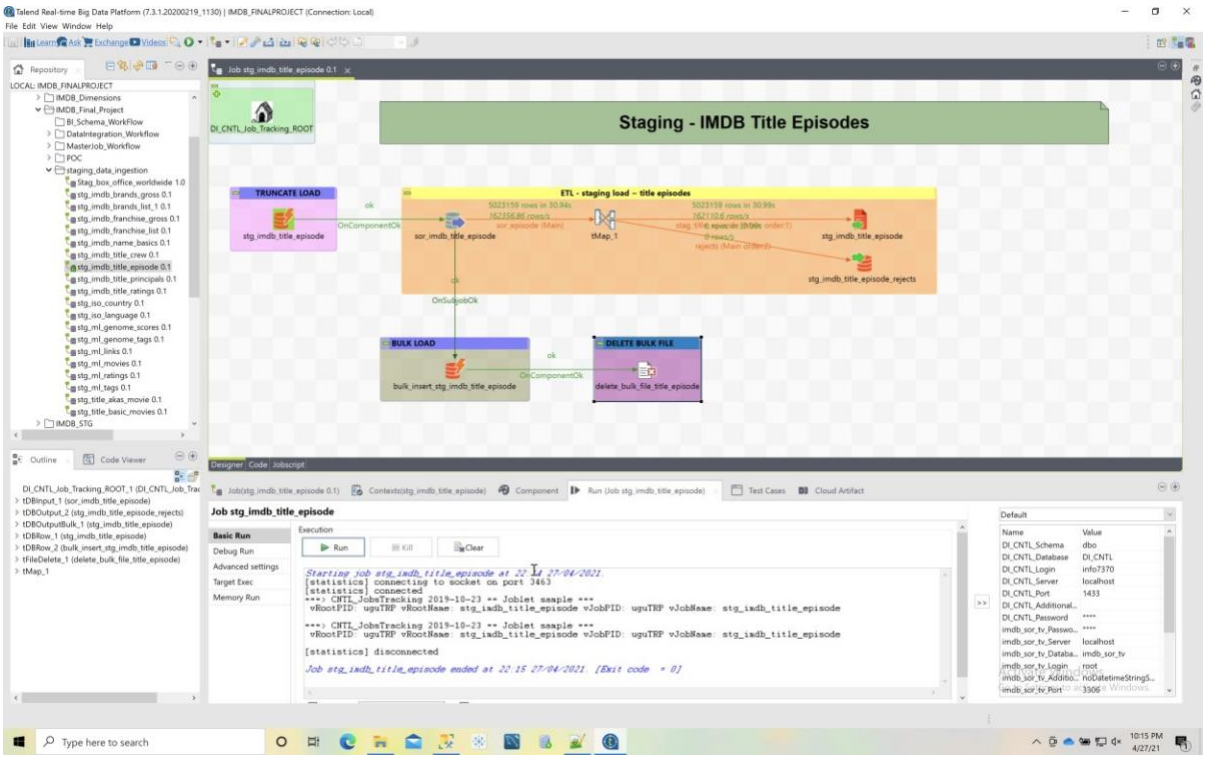
((Relational.ISNULL(Replace.directors)) && (Relational.ISNULL(Replace.writers))) ? -995 : 0

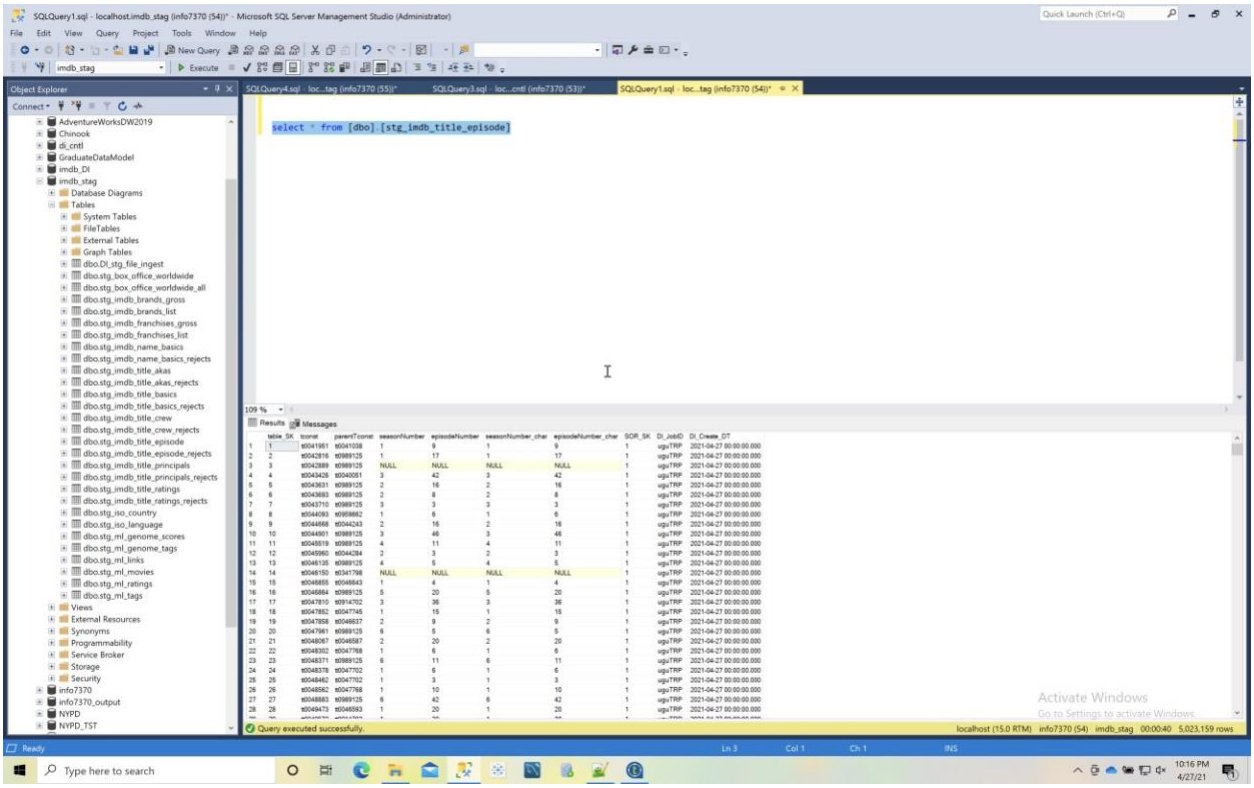
structural integration processes: initially we are truncating the table, we are doing data cleansing using tReplace component.
For performance we are using bulk components,
For DI standard we are deleting bulk file which is created in the process
We are also capturing DI injest metadata and storing in the table which is bonus point.

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

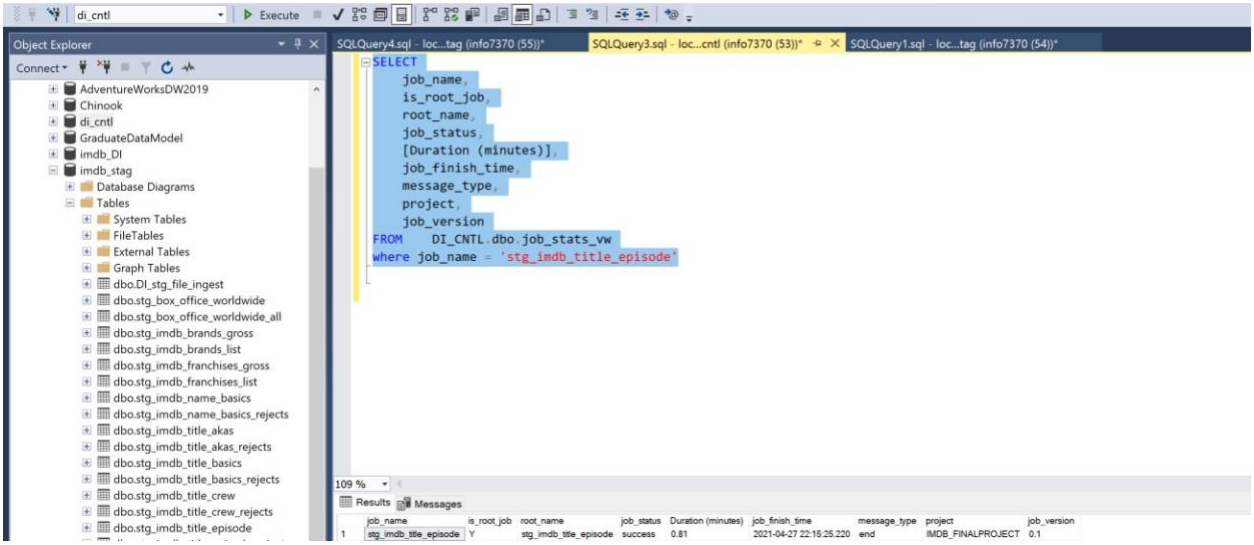
8. stg_imdb_title_episode & Rejects





Submitted By: Shweta Gupta

Staging Screen Shots- Talend



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,

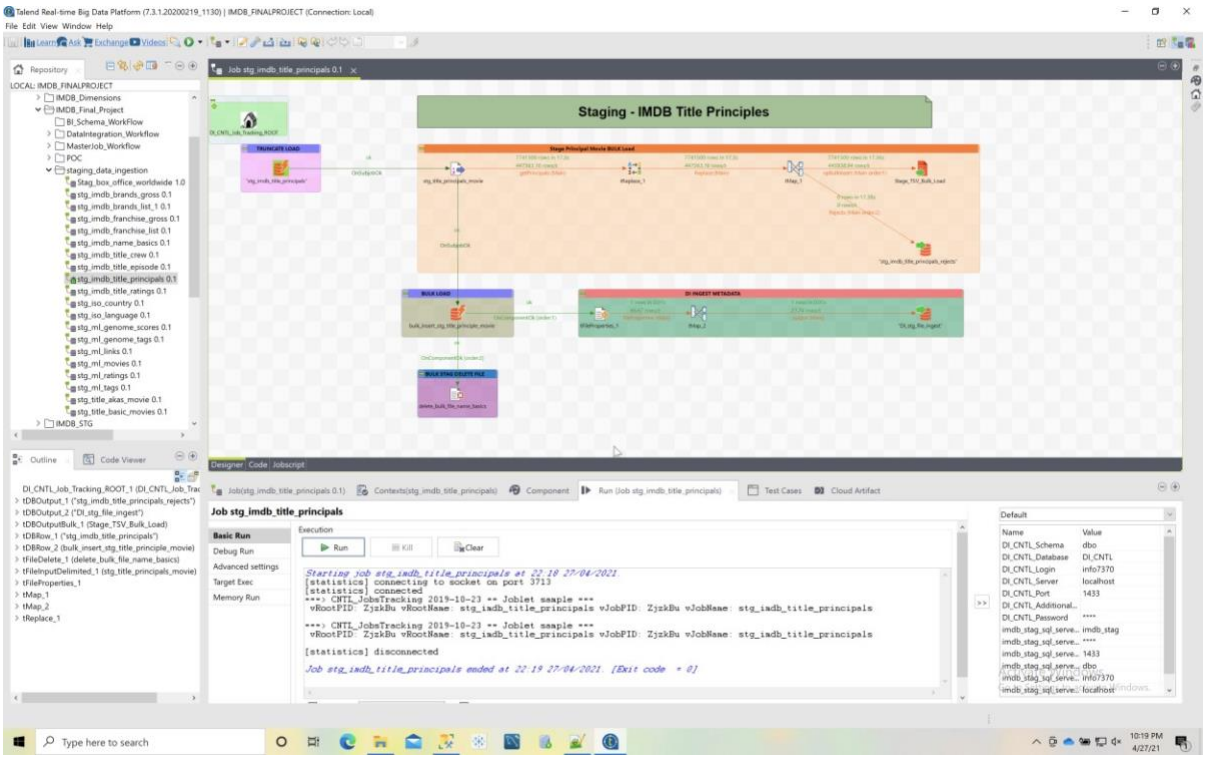
For DI standard we are deleting bulk file which is created in the process

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

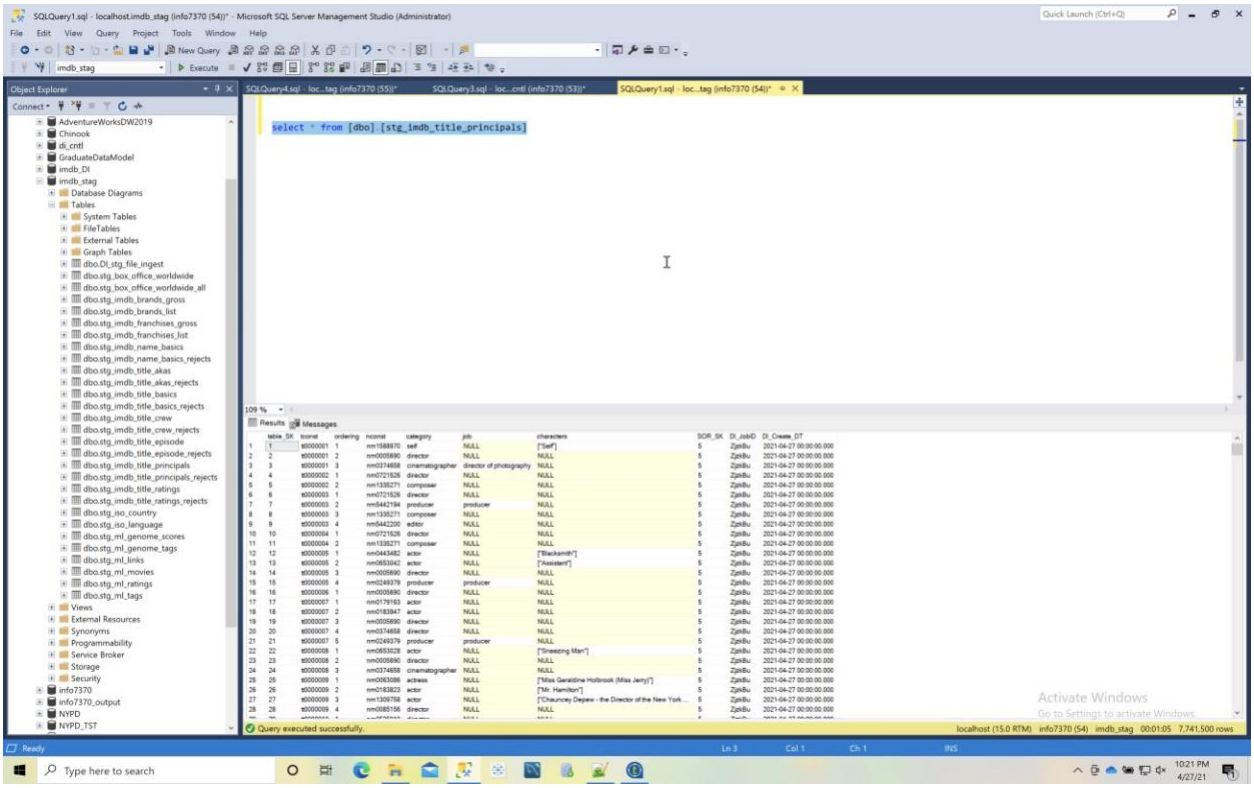
Submitted By: Shweta Gupta

Staging Screen Shots- Talend

9. stg_imdb_title_principals & Rejects



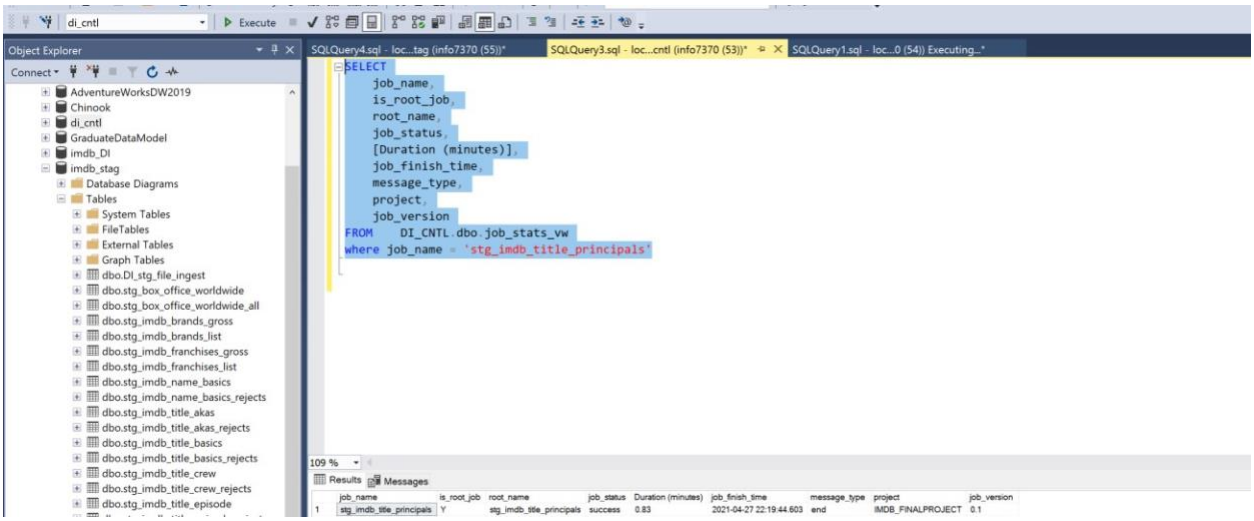
S



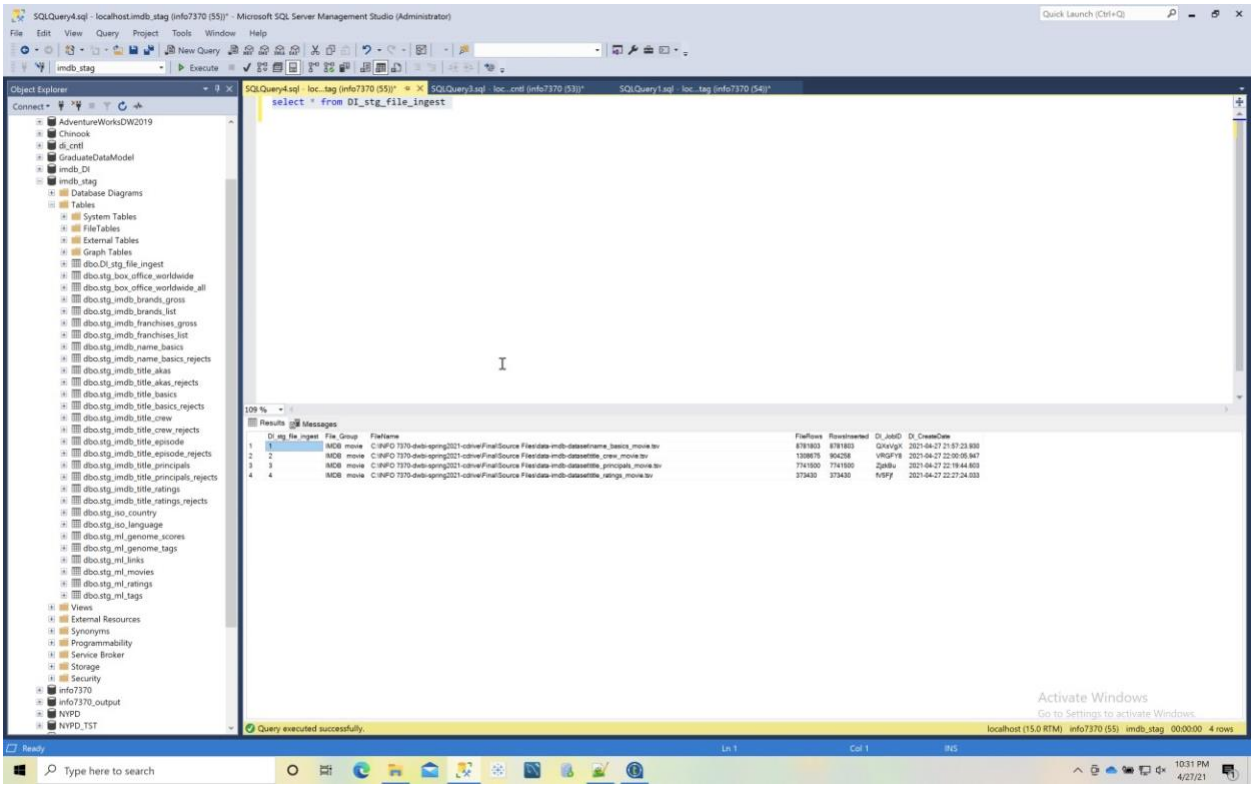
Submitted By: Shweta Gupta

Staging Screen Shots- Talend

TIME:



DI INJEST:



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.

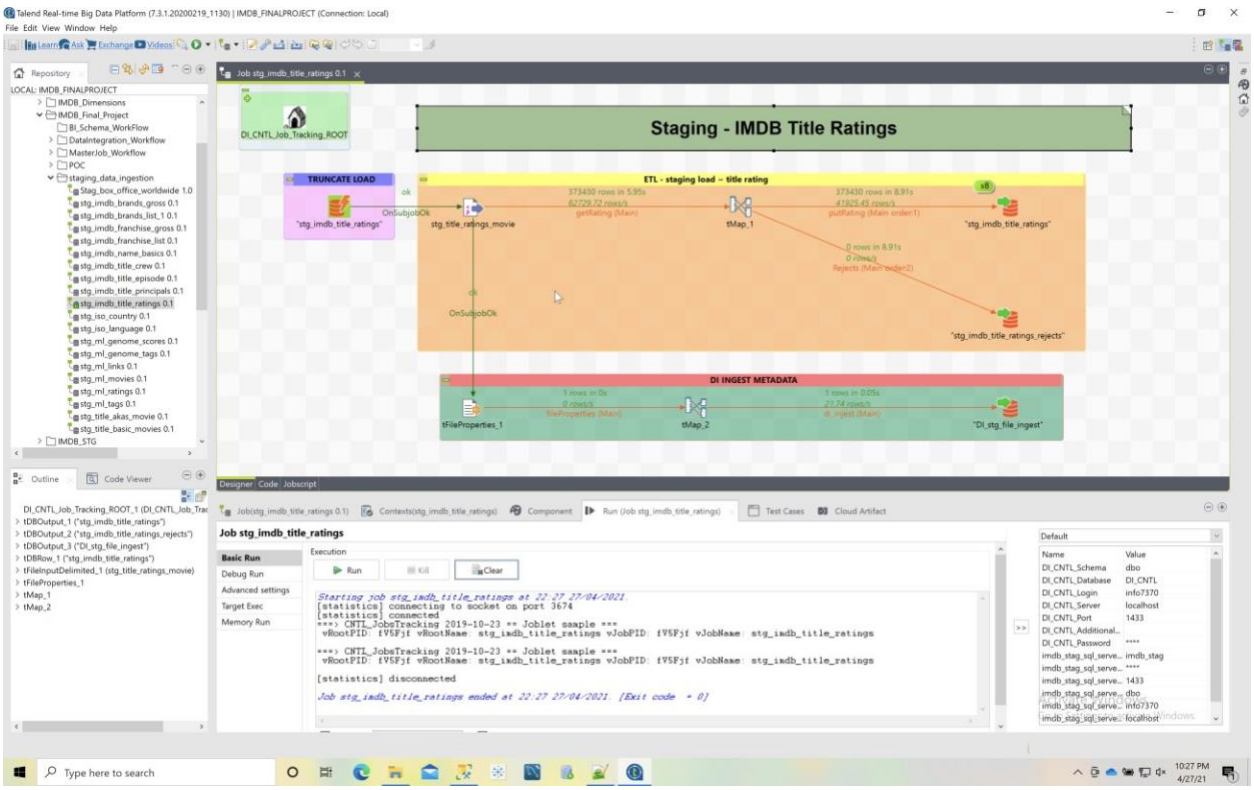
structural integration processes: initially we are truncating the table, we are doing

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

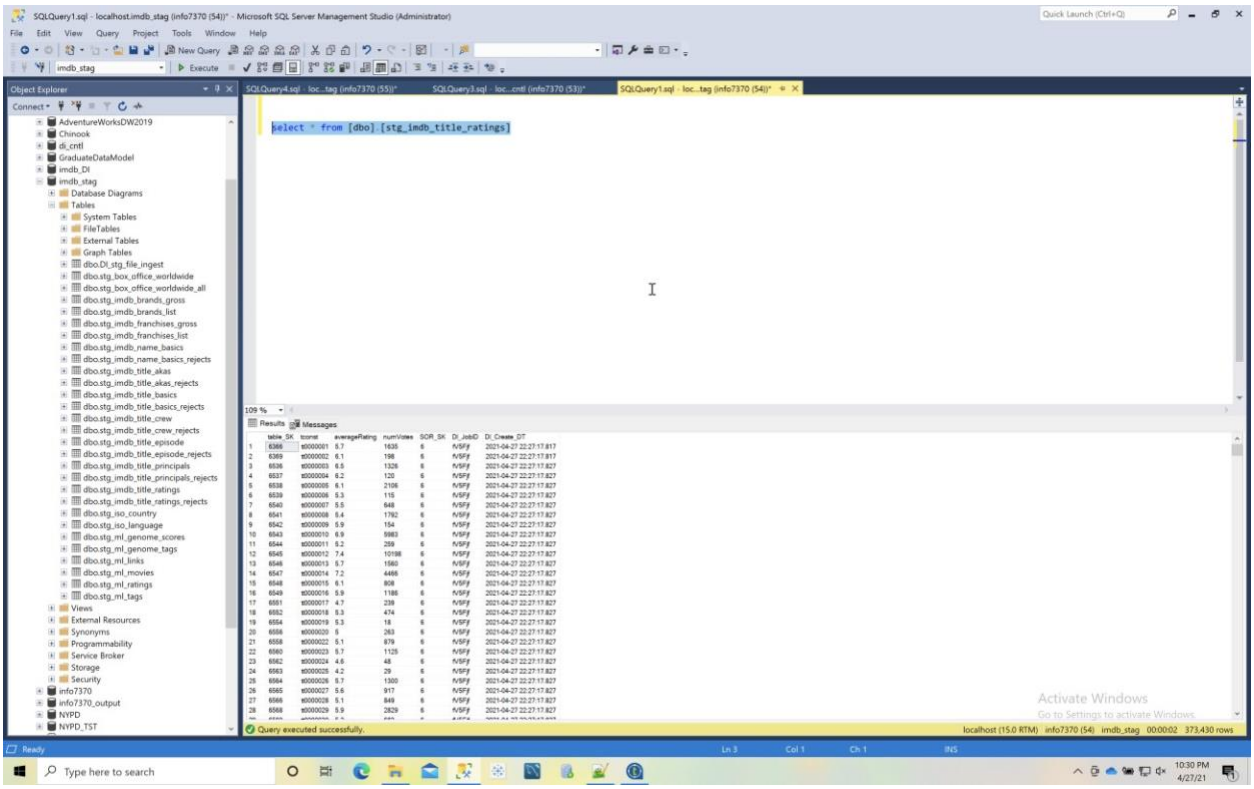
data cleansing using tReplace component.
For performance we are using bulk components,
For DI standard we are deleting bulk file which is created in the process
We are also capturing DI injest metadata and storing in the table which is bonus point.

10. stg_imdb_title_ratings & Rejects

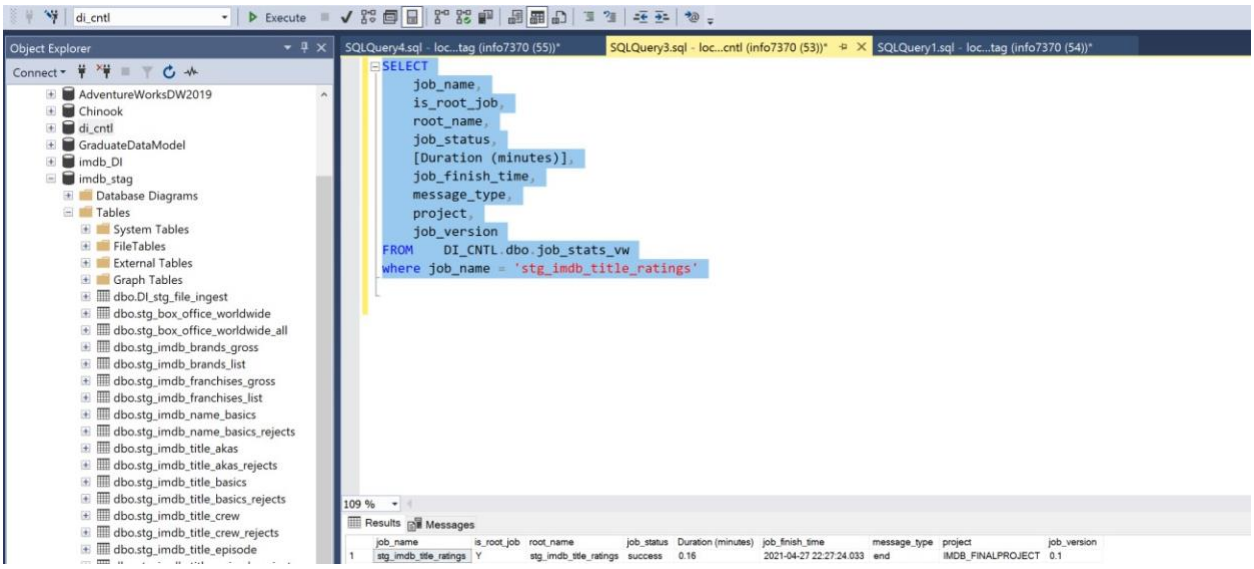


Submitted By: Shweta Gupta

Staging Screen Shots- Talend



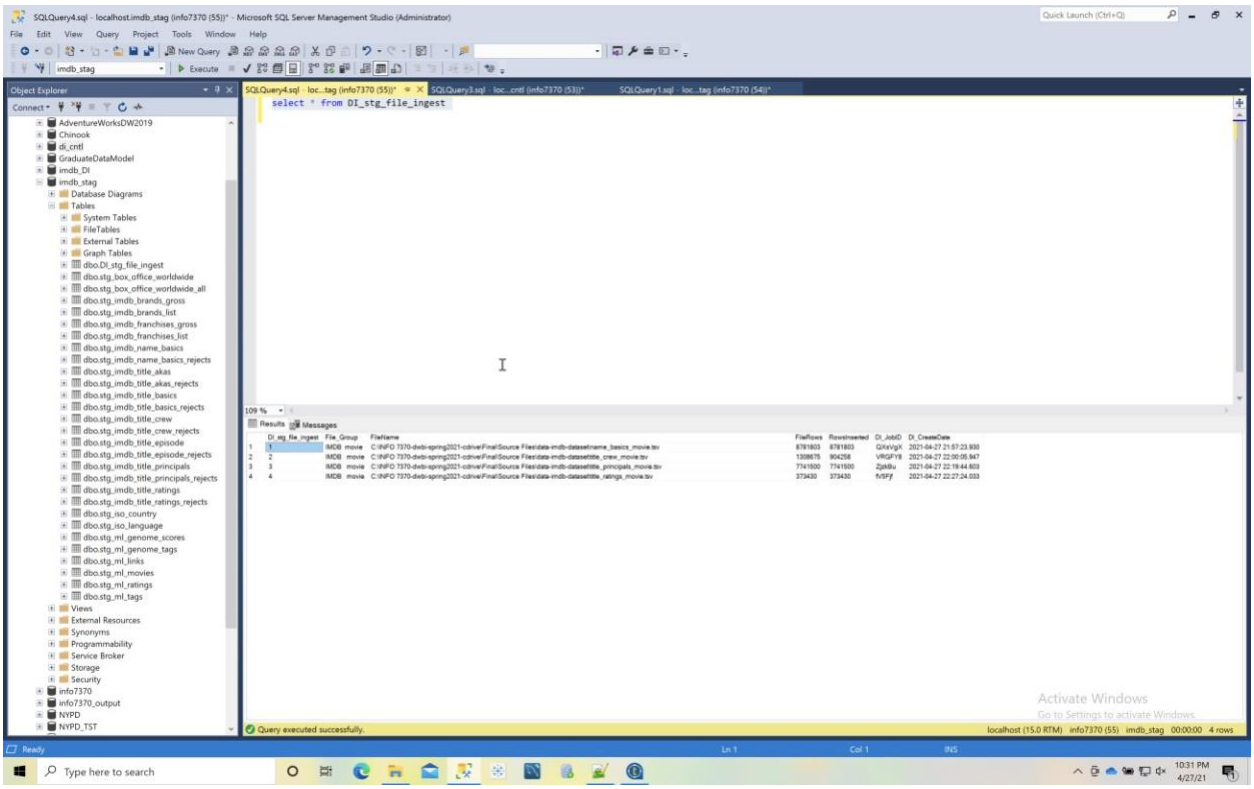
TIME



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

DI INJEST:



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.

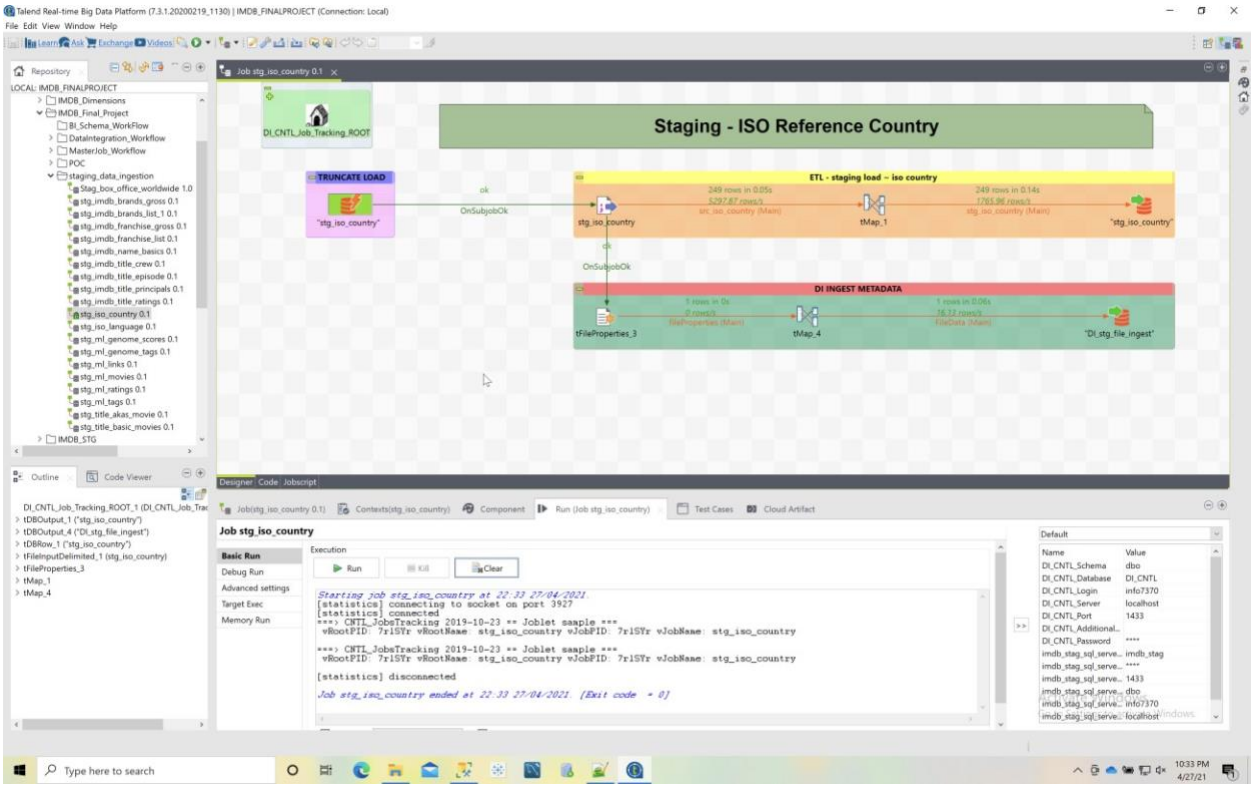
For Performance we have enabled paralleization to speed up the load

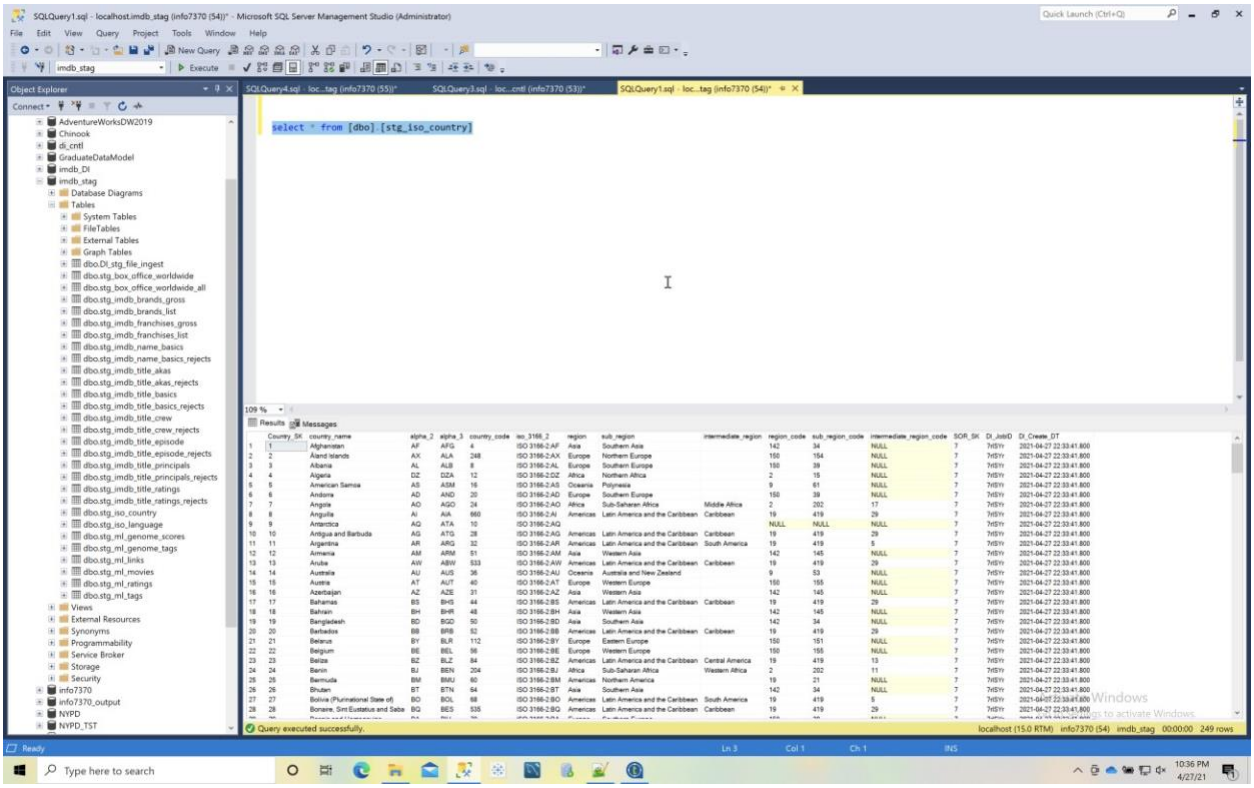
We are also capturing DI injest metadata and storing in the table which is bonus point.

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

11. stg_iso_country

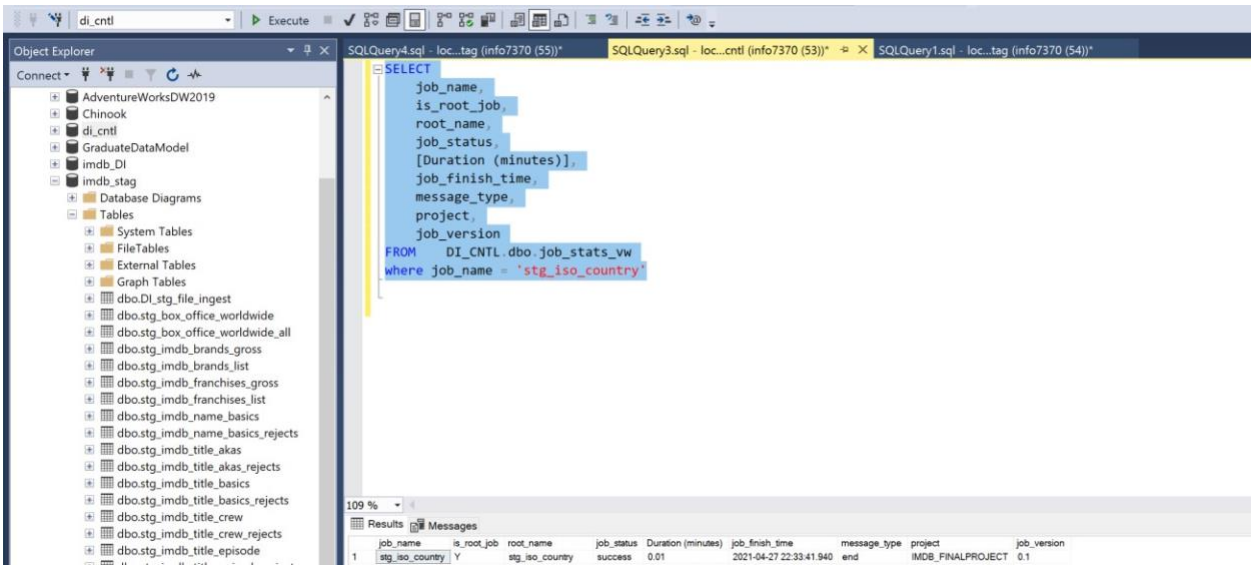




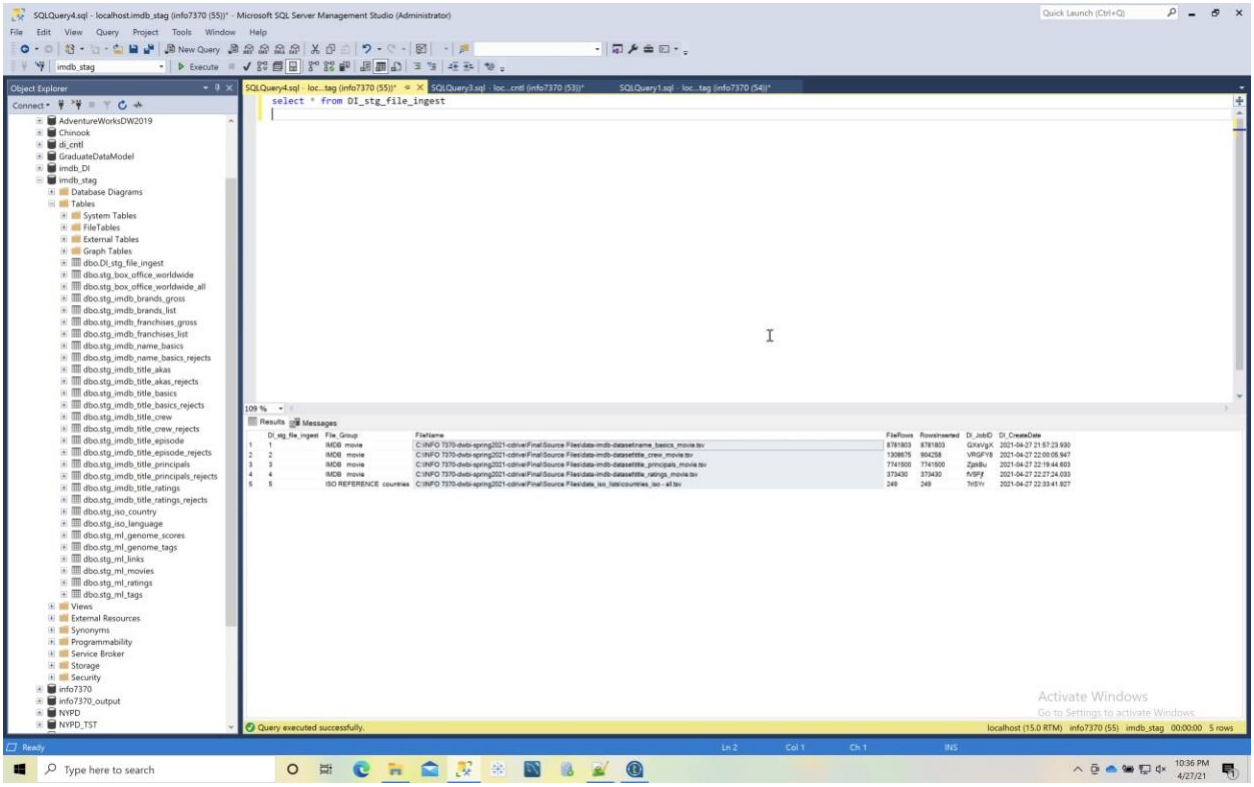
Submitted By: Shweta Gupta

Staging Screen Shots- Talend

TIME:



DI INJEST:



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

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

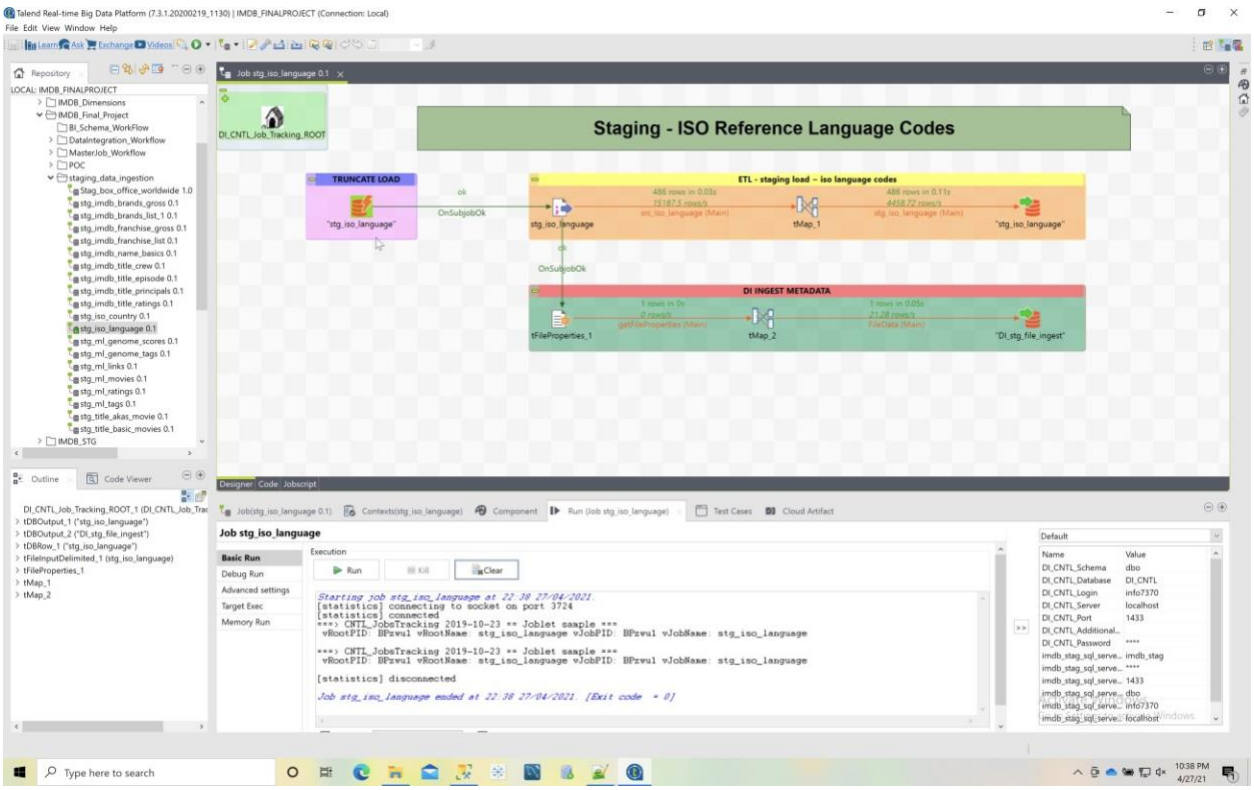
reject: No rejects required for this table.

structural integration processes: initially we are truncating the table.

For Performance we have enabled paralleization to speed up the load

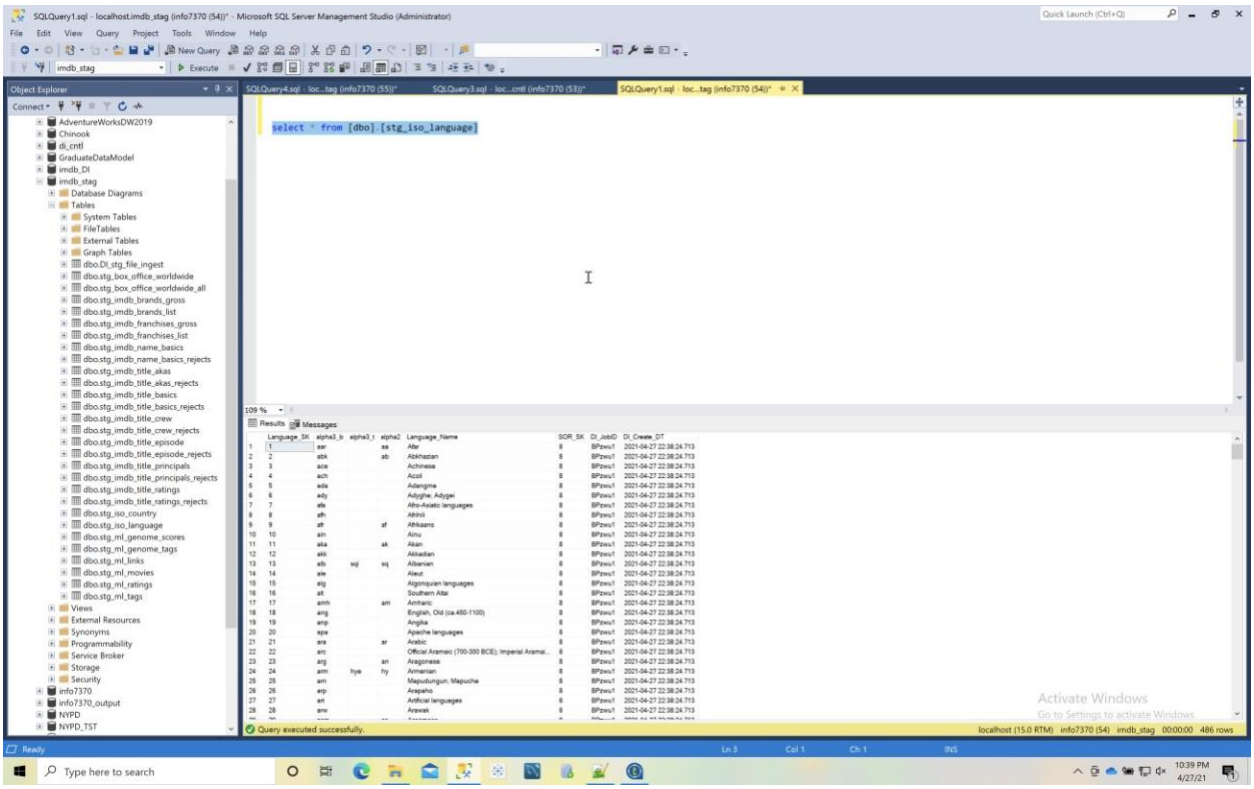
We are also capturing DI injest metadata and storing in the table which is bonus point.

12. stg_iso_language

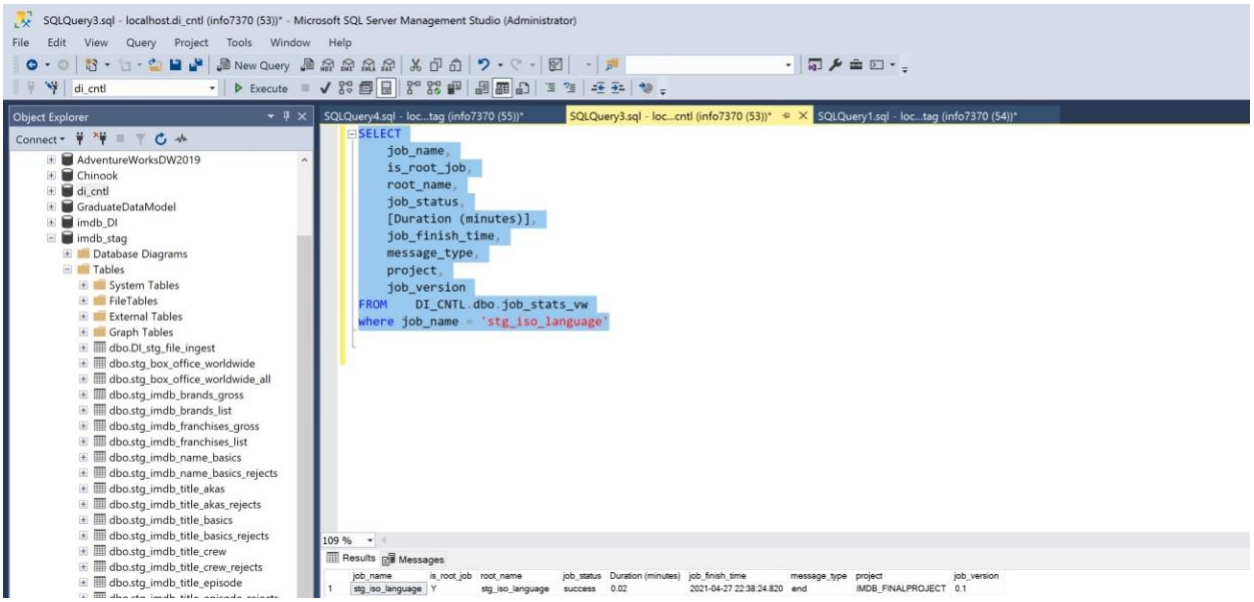


Submitted By: Shweta Gupta

Staging Screen Shots- Talend



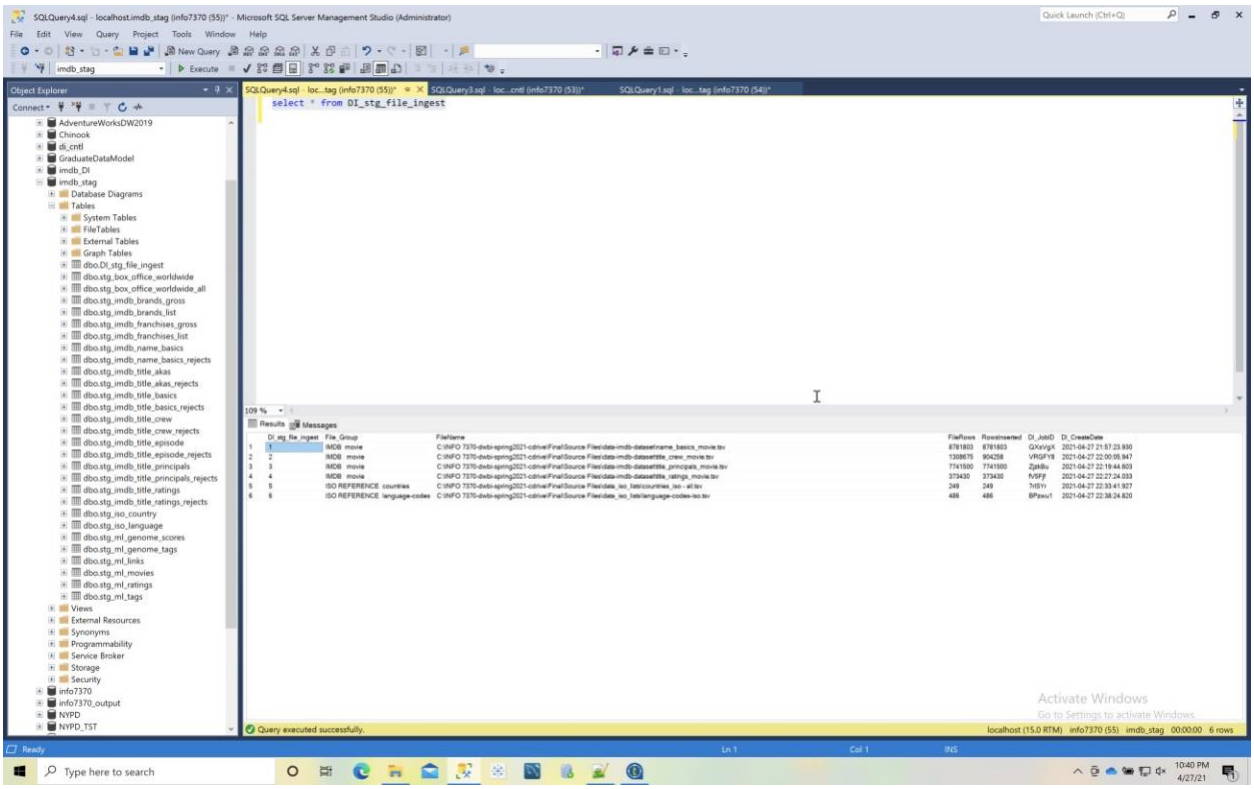
TIME:



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

DI INJEST:



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

reject: No rejects required for this table.

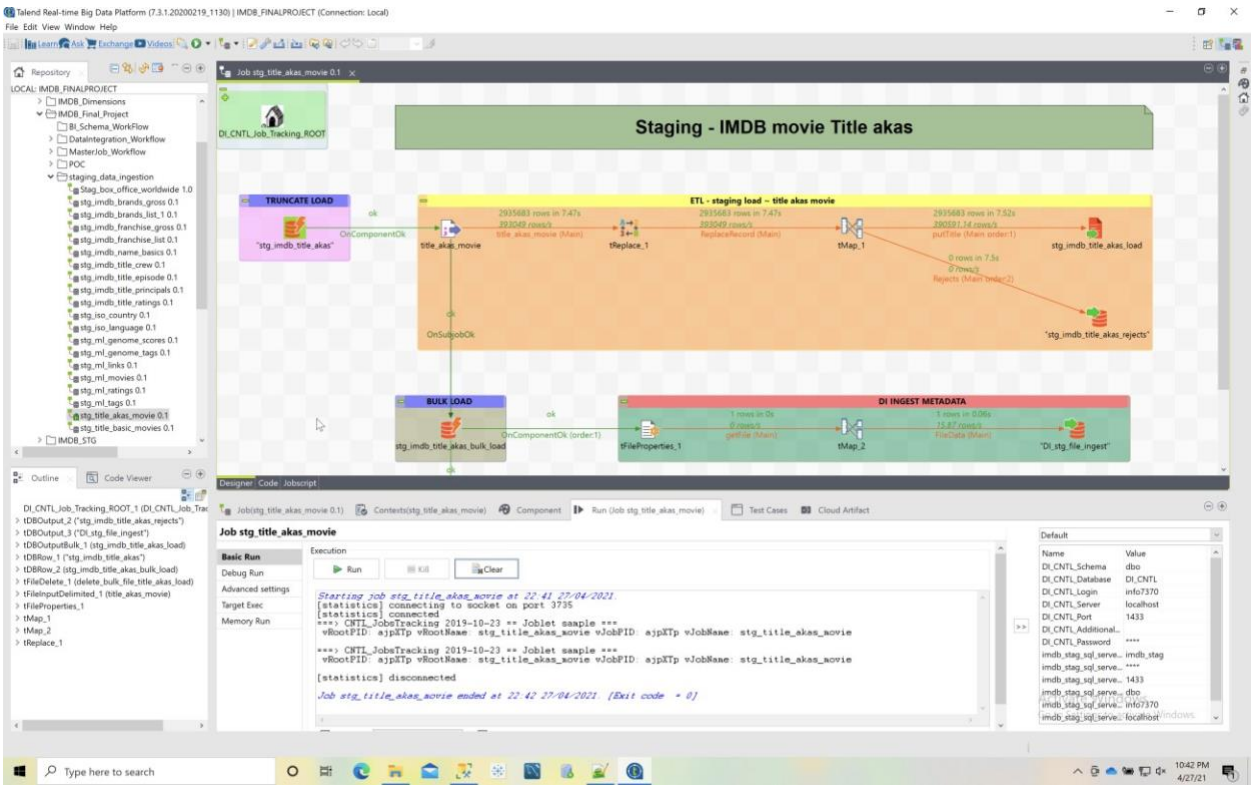
structural integration processes: initially we are truncating the table.

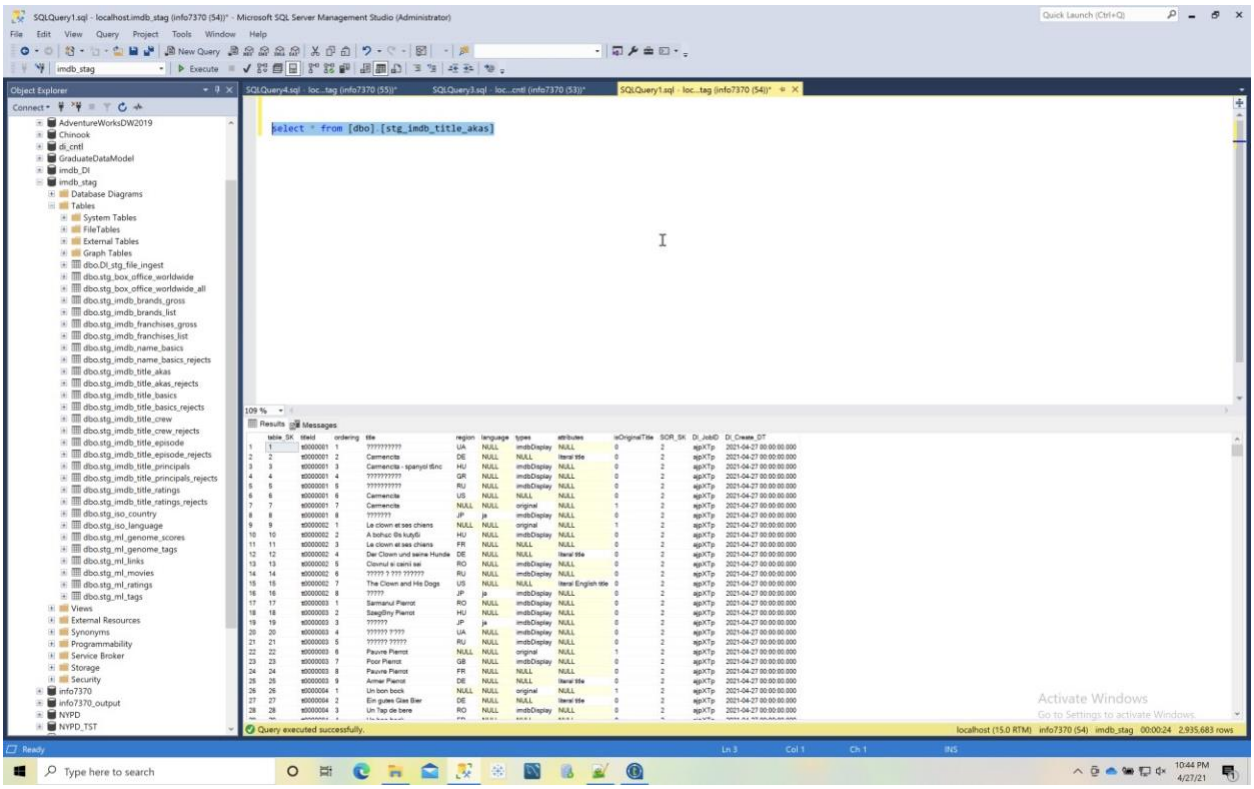
For Performance we have enabled paralleization to speed up the load

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

13. stg_title_akas_movie & Rejects

Staging Screen Shots- Talend

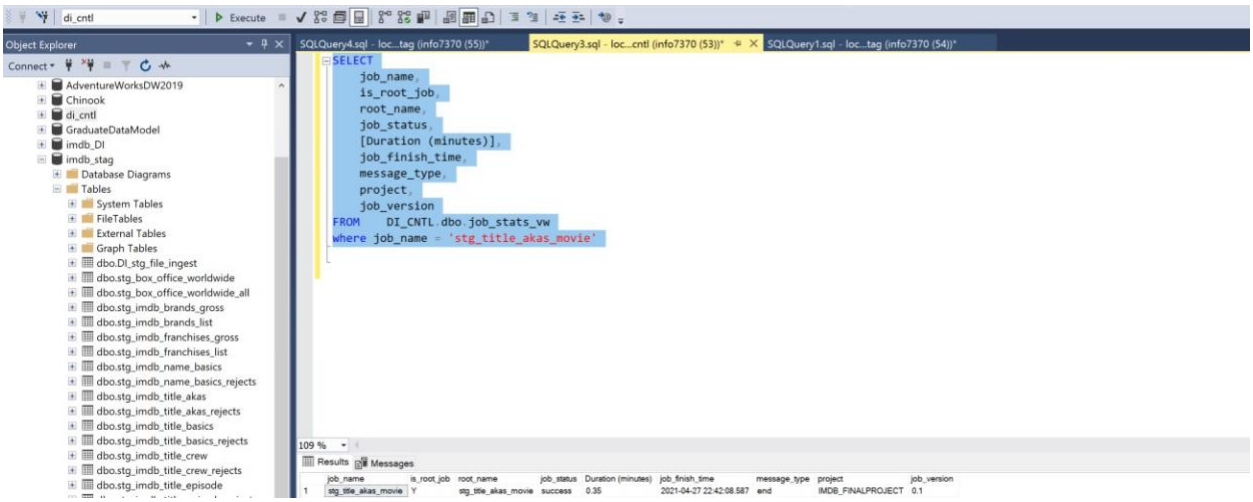




Submitted By: Shweta Gupta

Staging Screen Shots- Talend

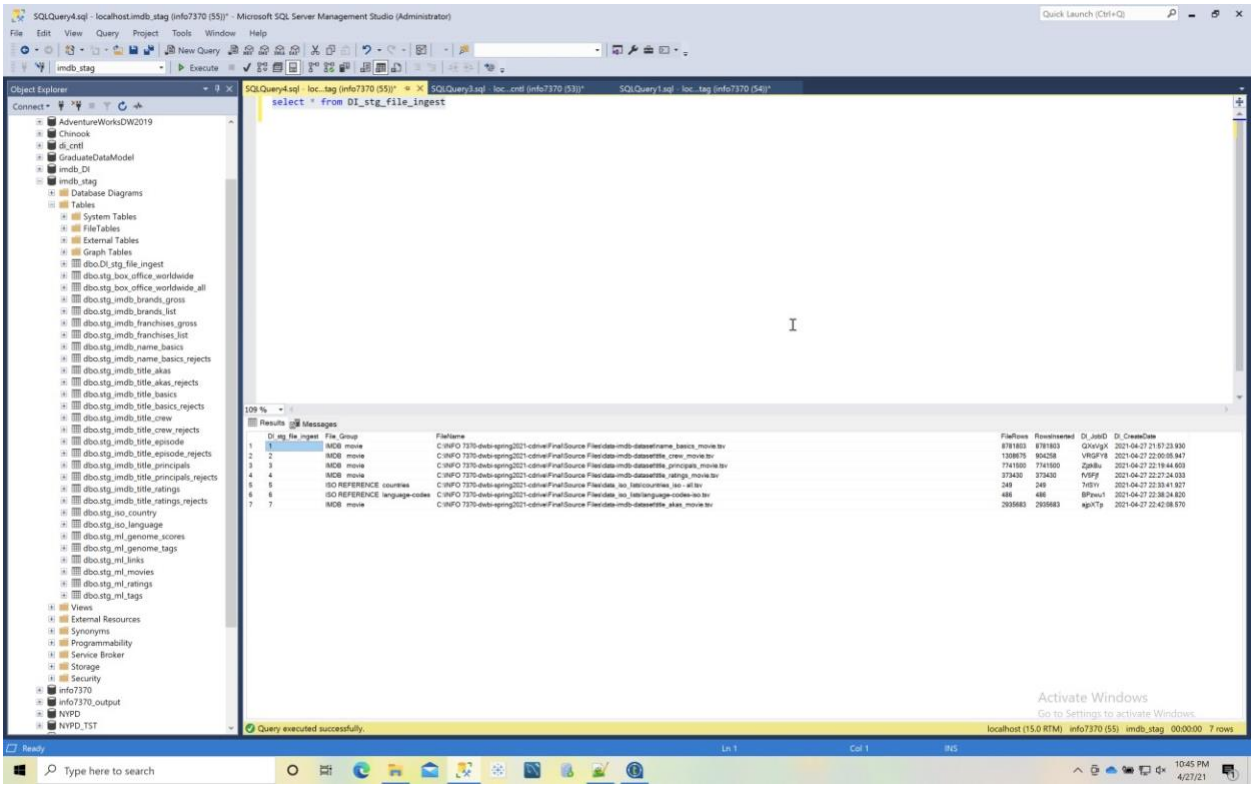
TIME:



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

DI INJEST:



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,

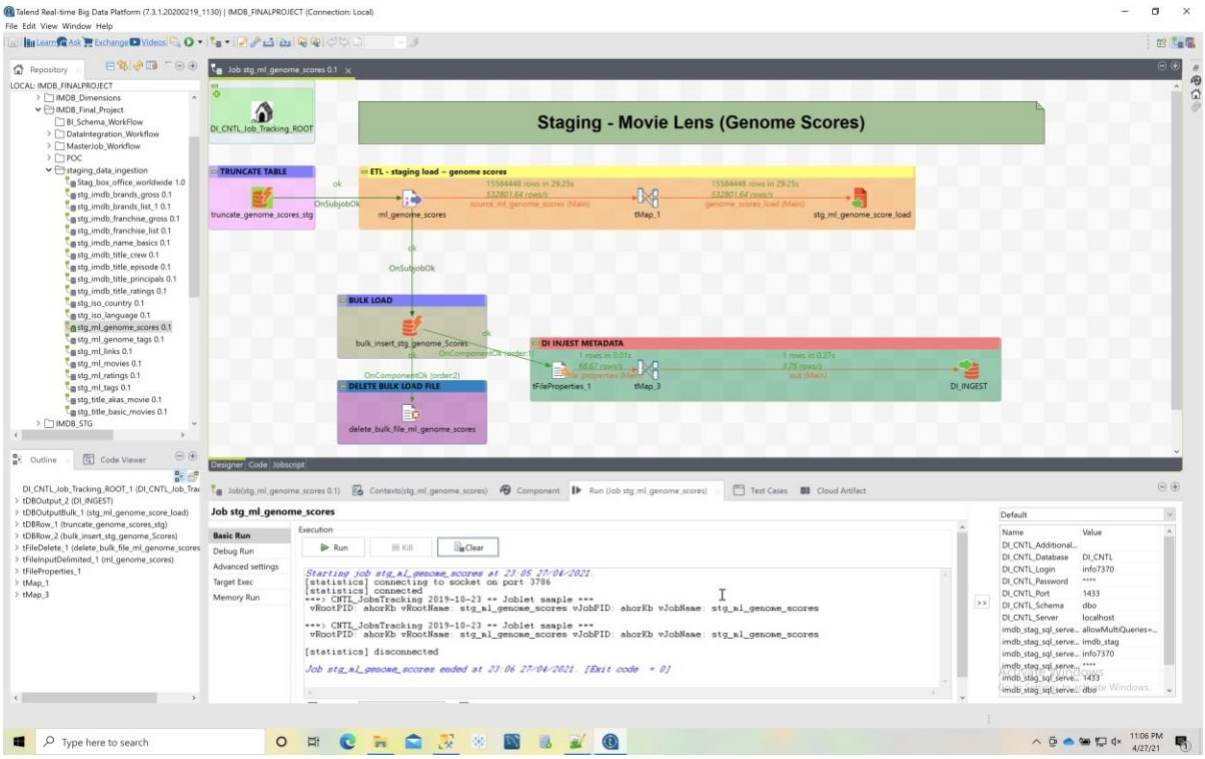
For DI standard we are deleting bulk file which is created in the process

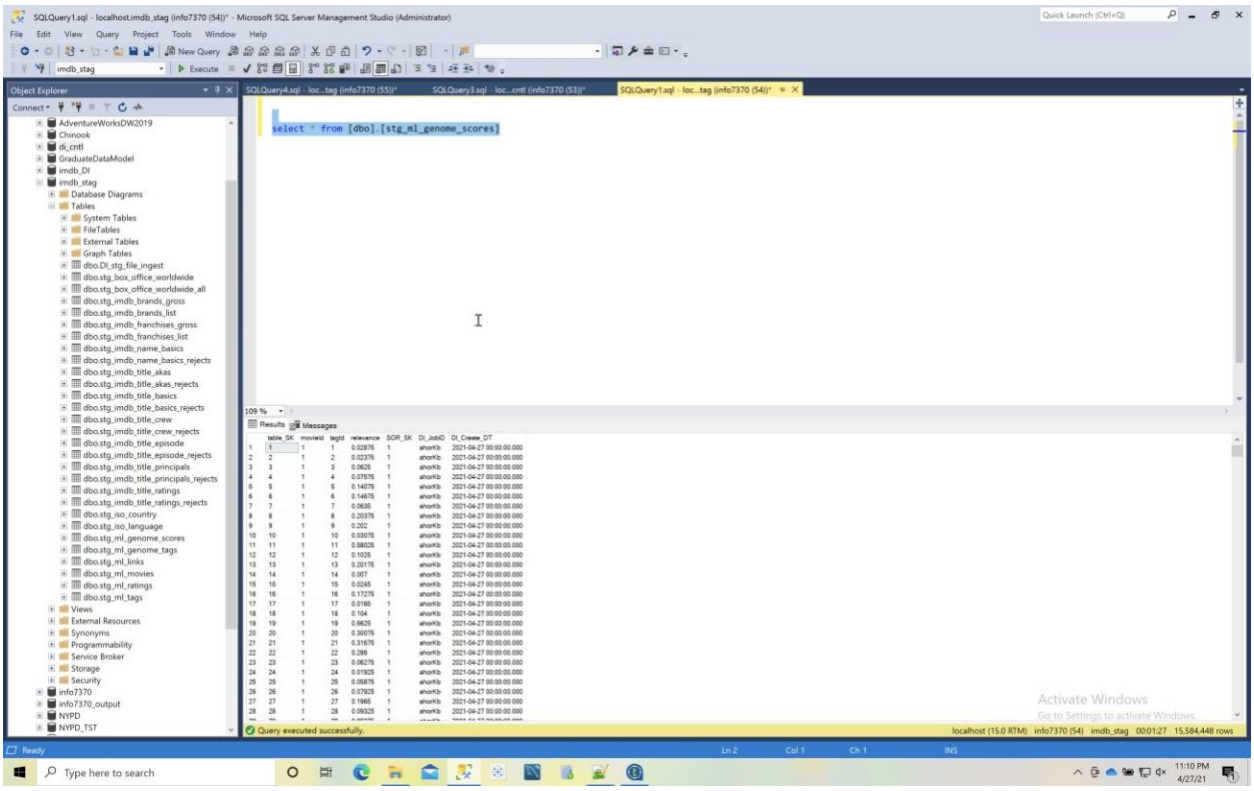
We are also capturing DI injest metadata and storing in the table which is bonus point.

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

14. stg_ml_genome_scores

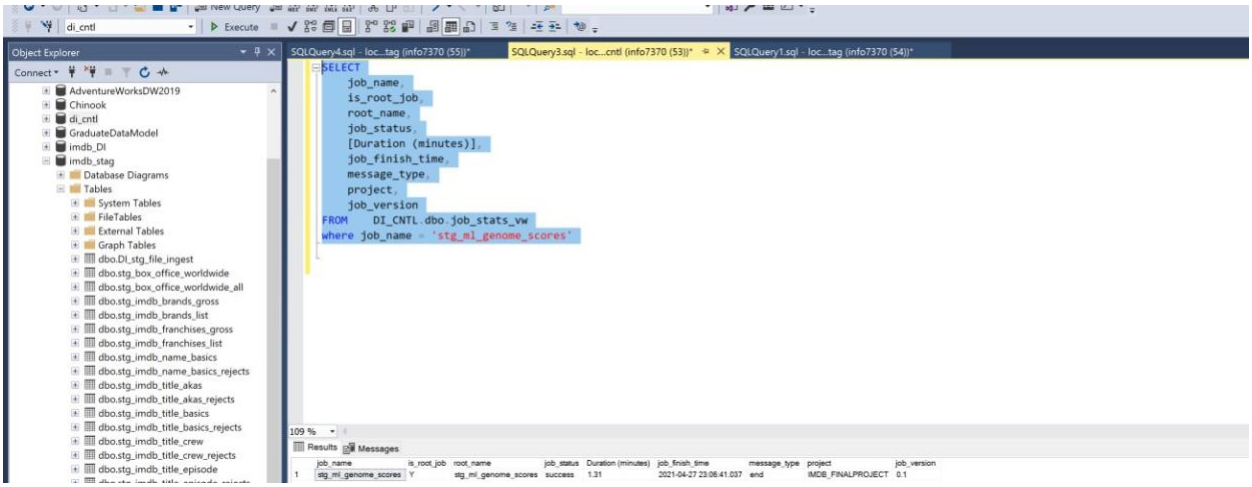




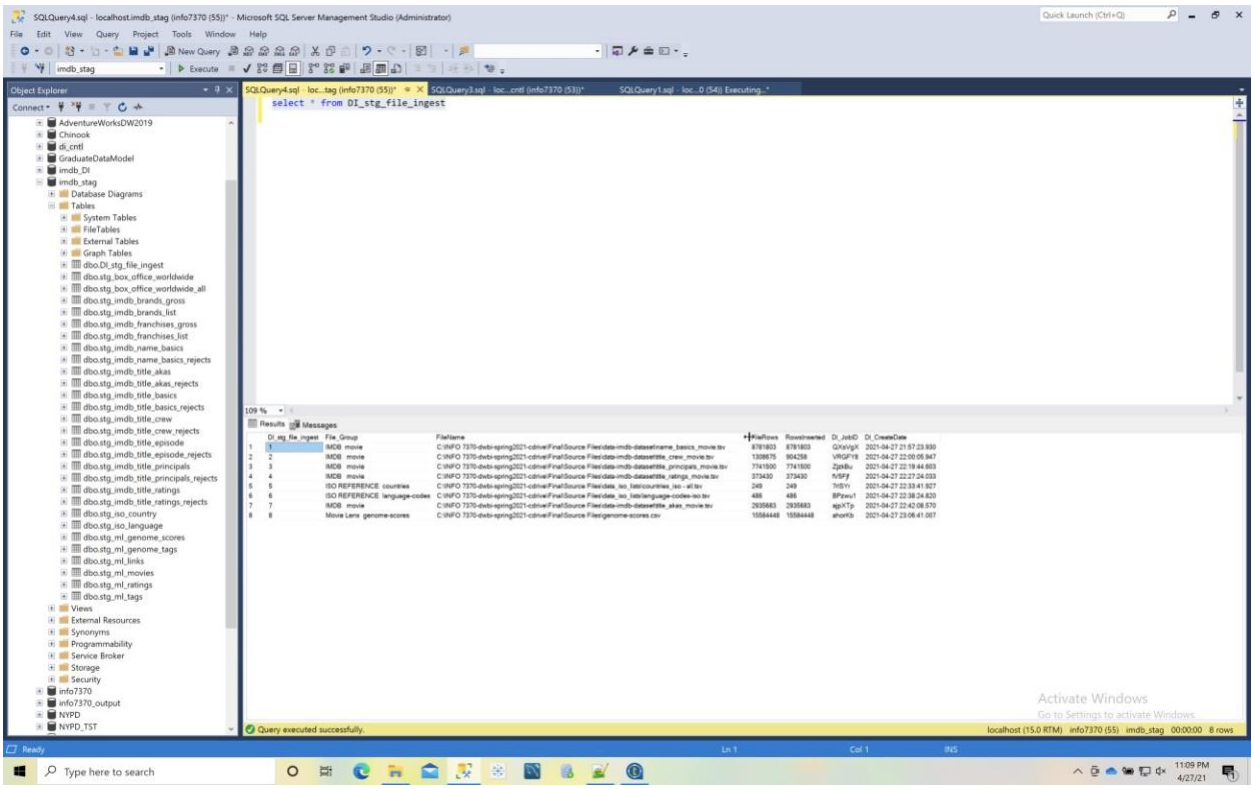
Submitted By: Shweta Gupta

Staging Screen Shots- Talend

TIME:



DI INJEST:



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

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

reject: No reject required for this table.

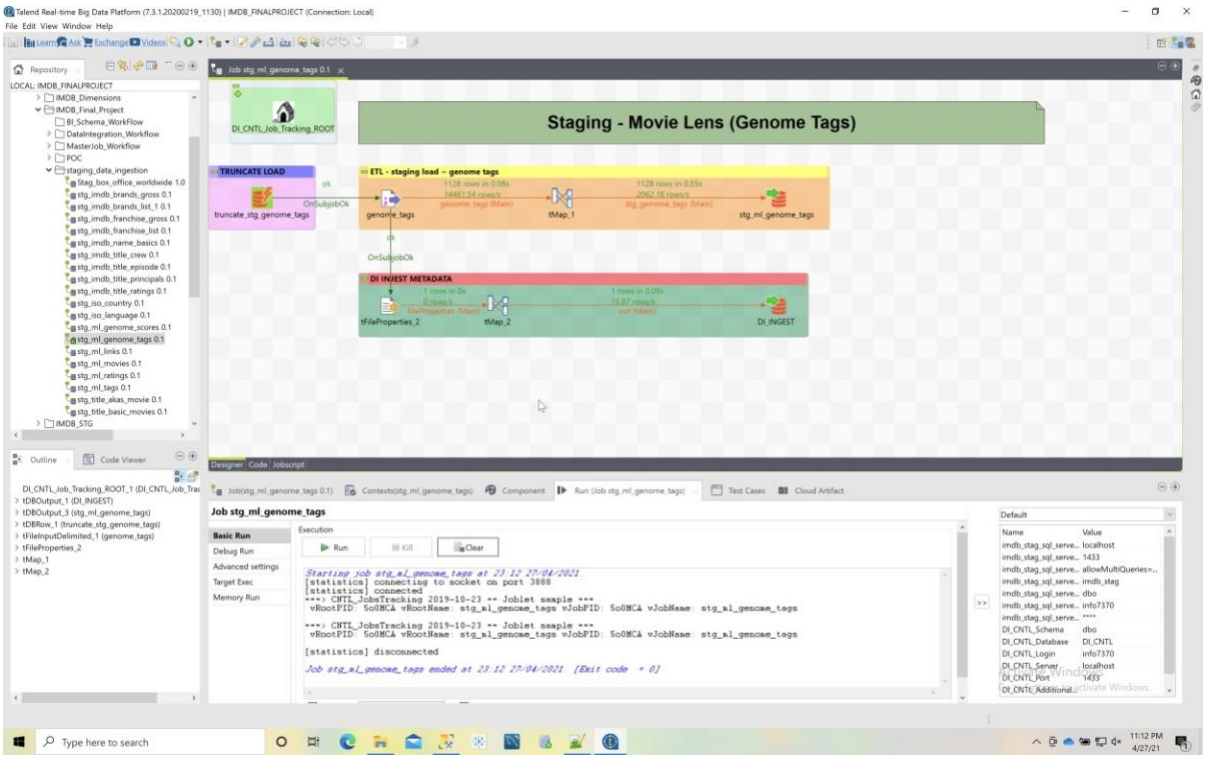
structural integration processes: initially we are truncating the table

For performance we are using bulk components,

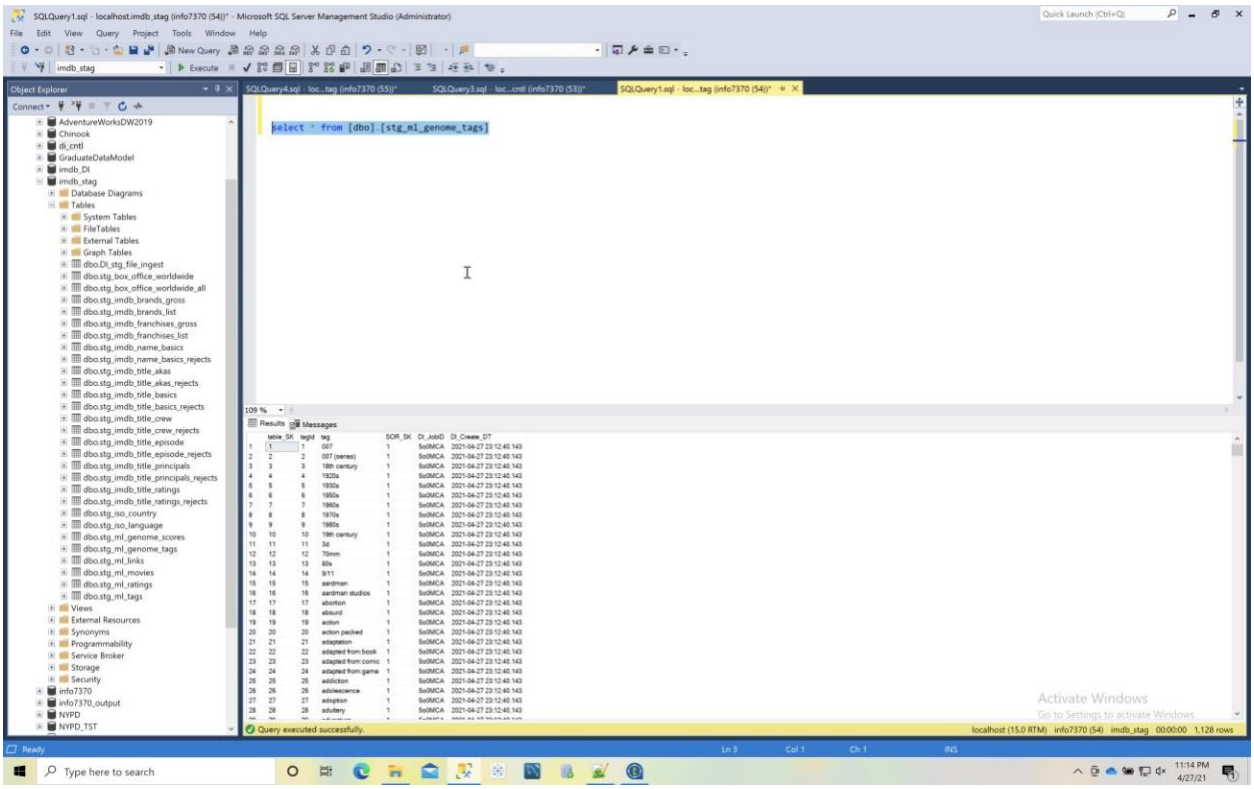
For DI standard we are deleting bulk file which is created in the process

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

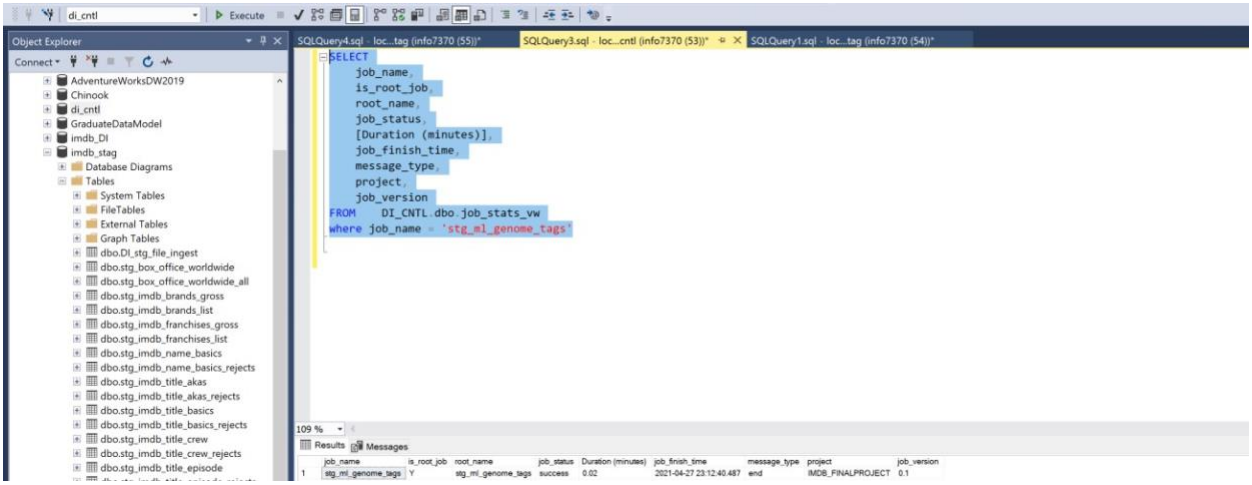
15. stg_ml_genome_tags



Staging Screen Shots- Talend



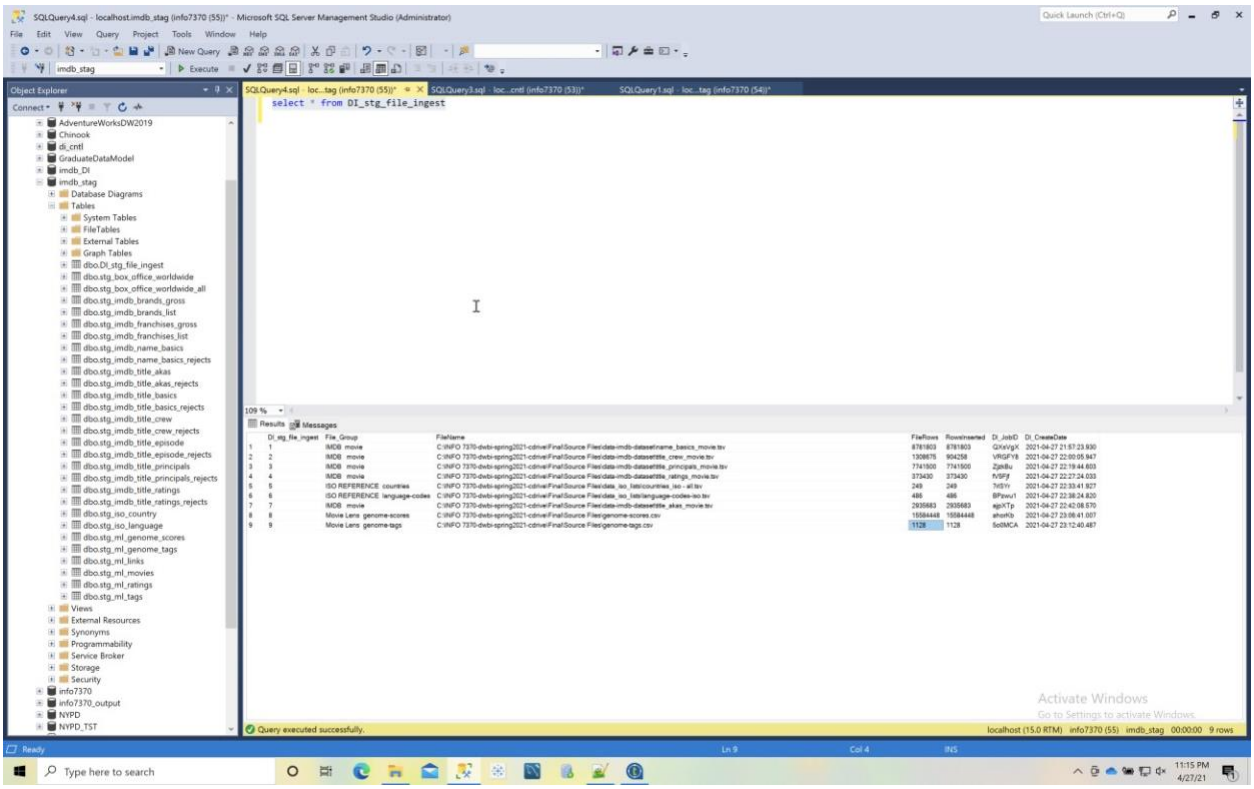
TIME:



Submitted By: Shweta Gupta

Staging Screen Shots- Talend

DI INJEST:



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.

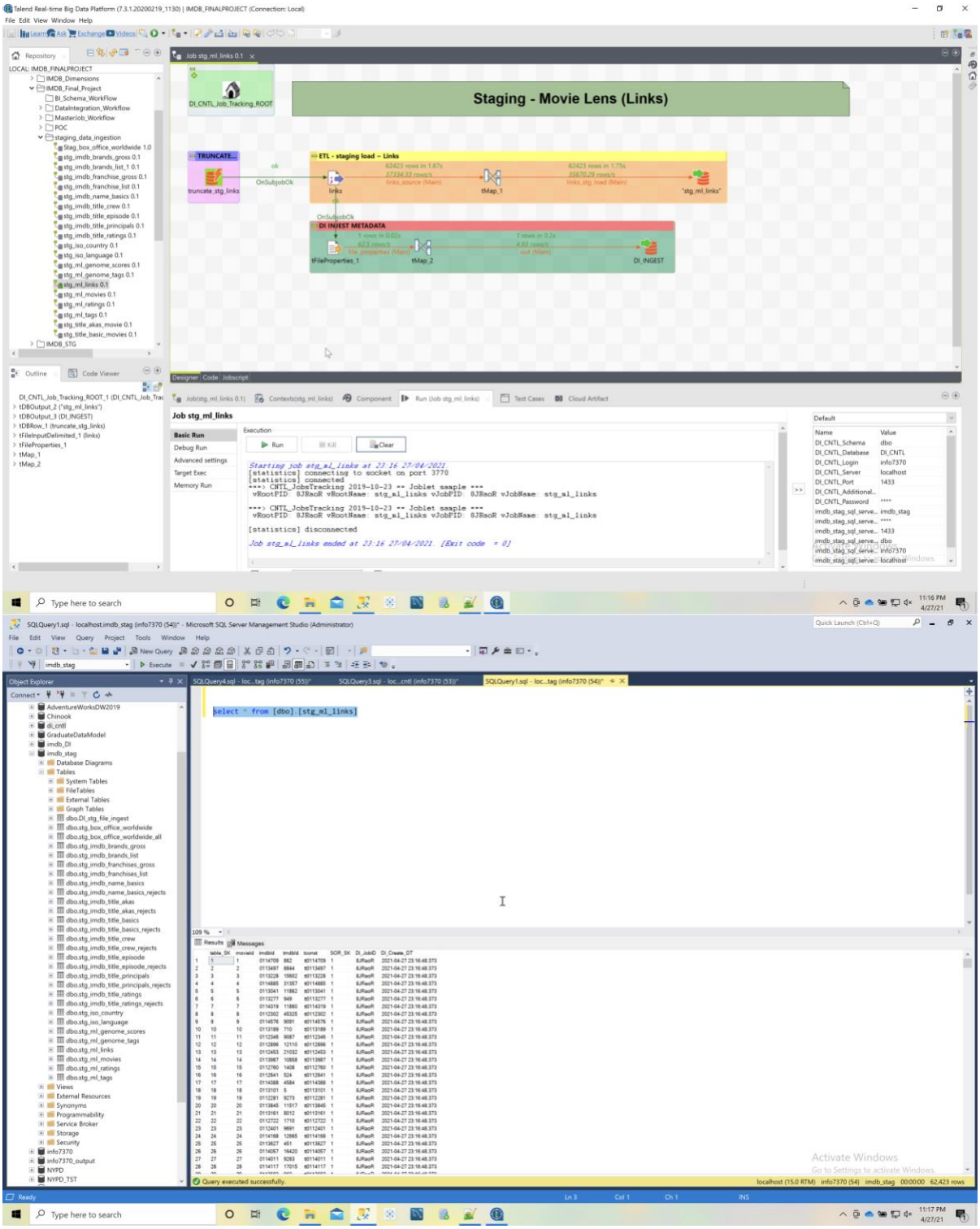
For Performance we have enabled parallelization to speed up the load

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

Submitted By: Shweta Gupta

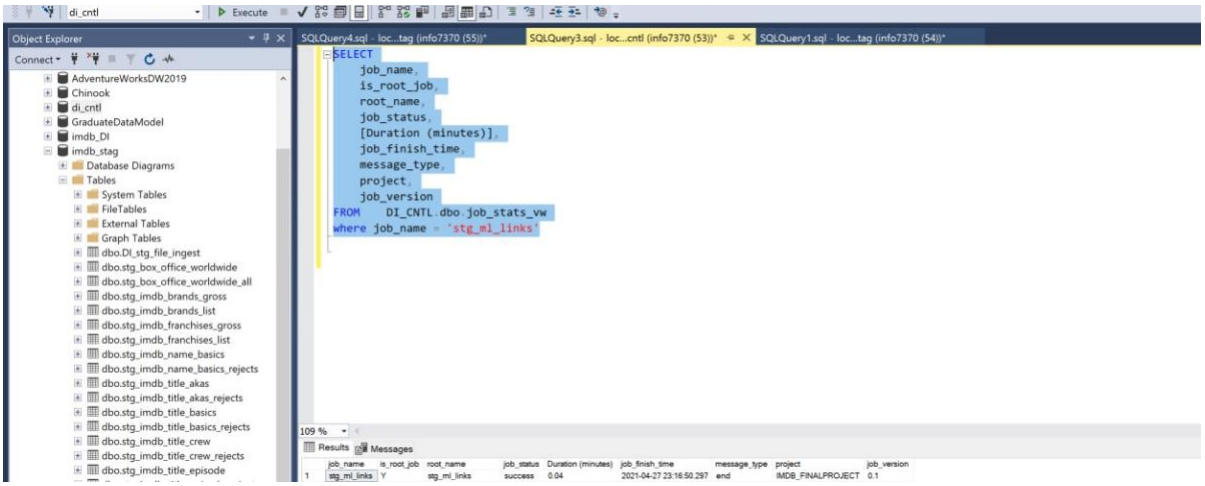
Staging Screen Shots- Talend

16. stg_ml_links

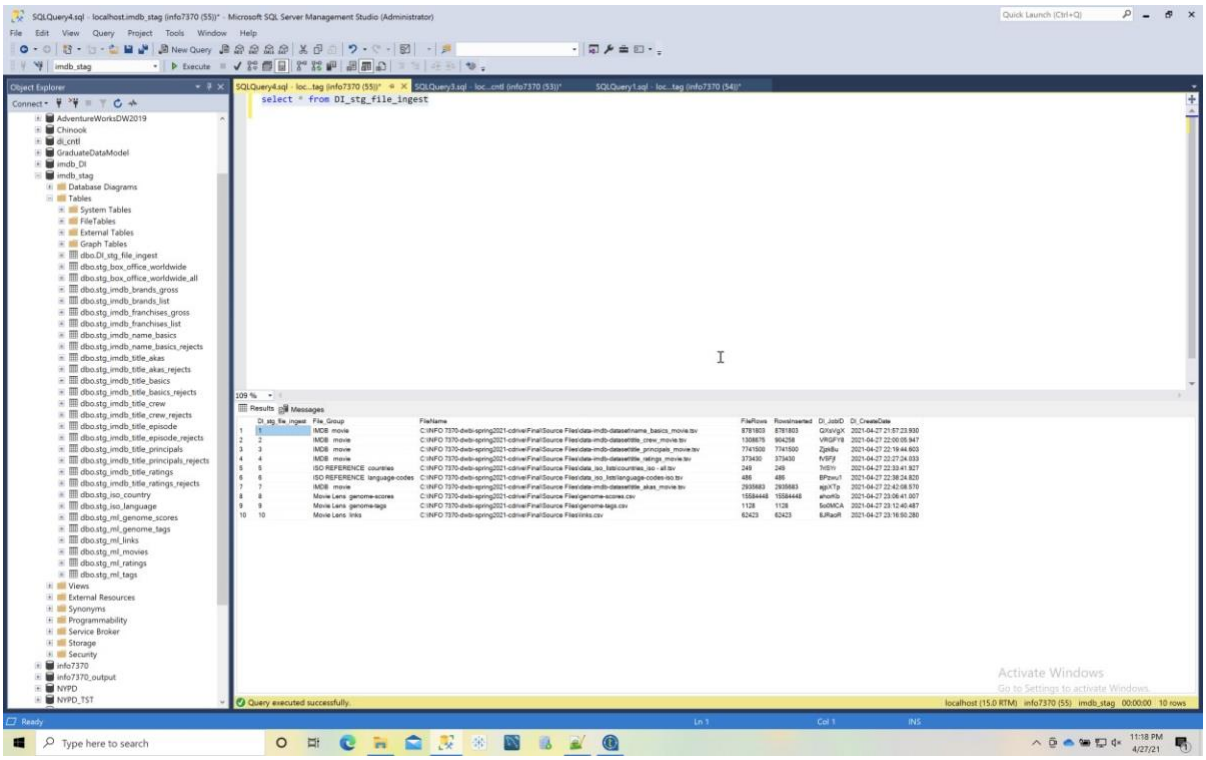


TIME:

Staging Screen Shots- Talend



DI INJEST:



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.

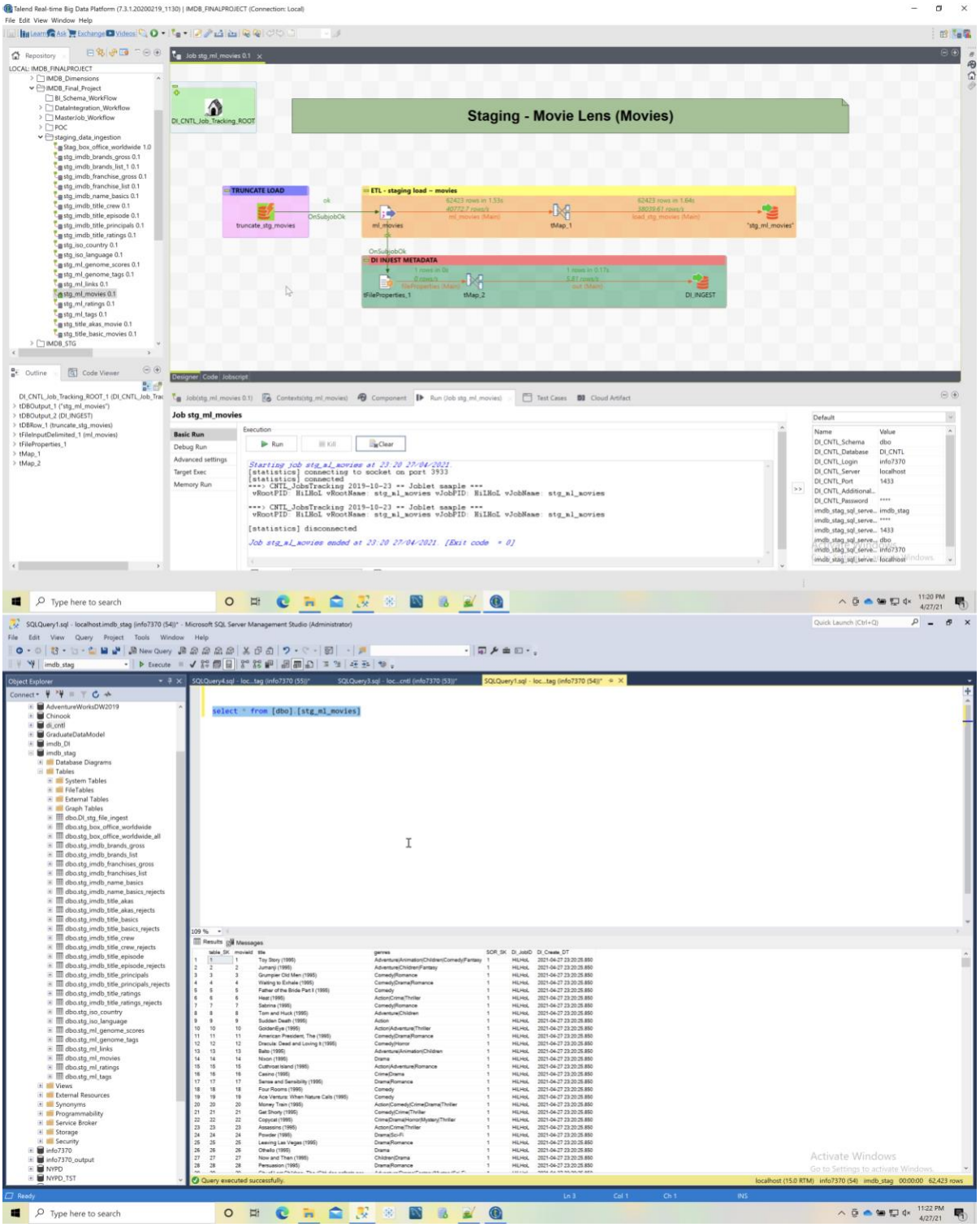
For Performance we have enabled parallelization to speed up the load

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

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

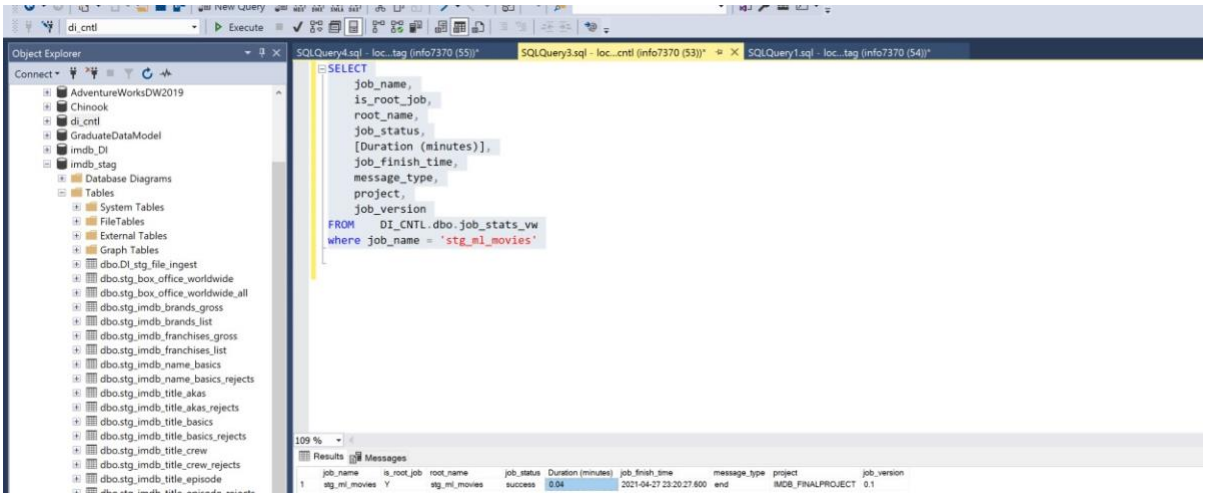
17. stg_ml_movies



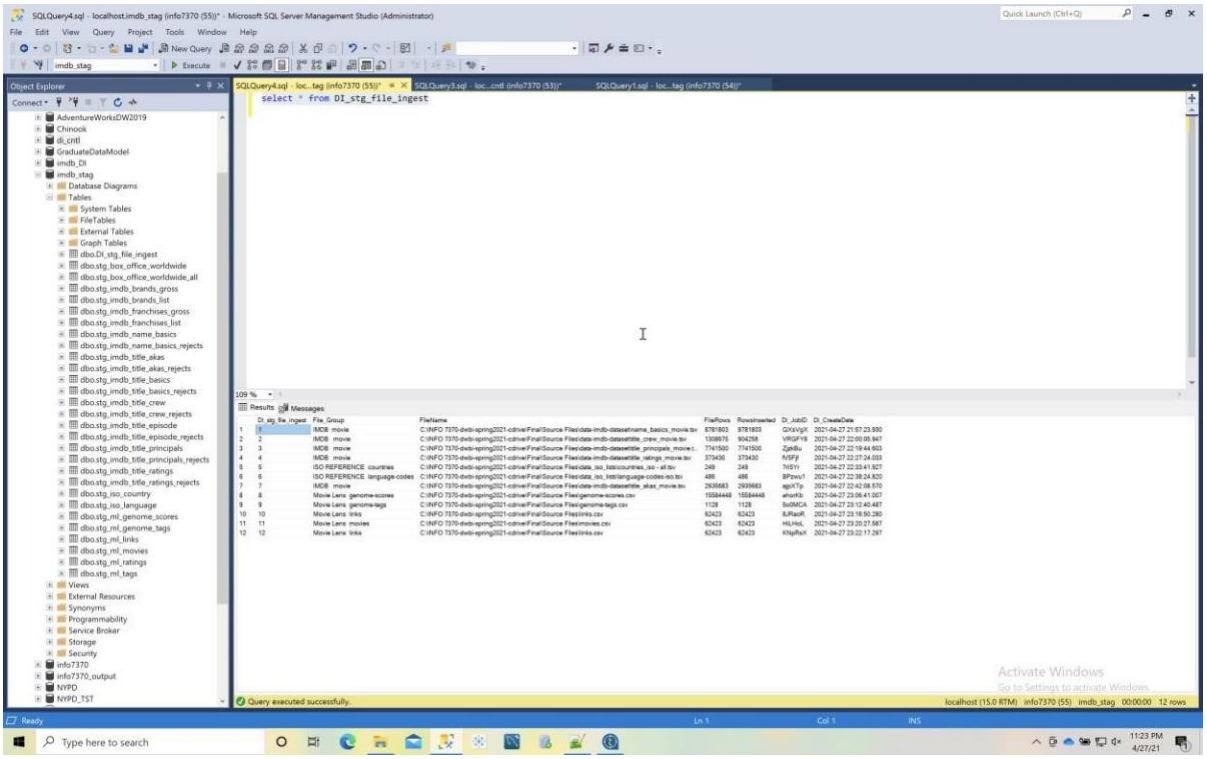
TIME:

Submitted By: Shweta Gupta

Staging Screen Shots- Talend



DI INJEST:



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.

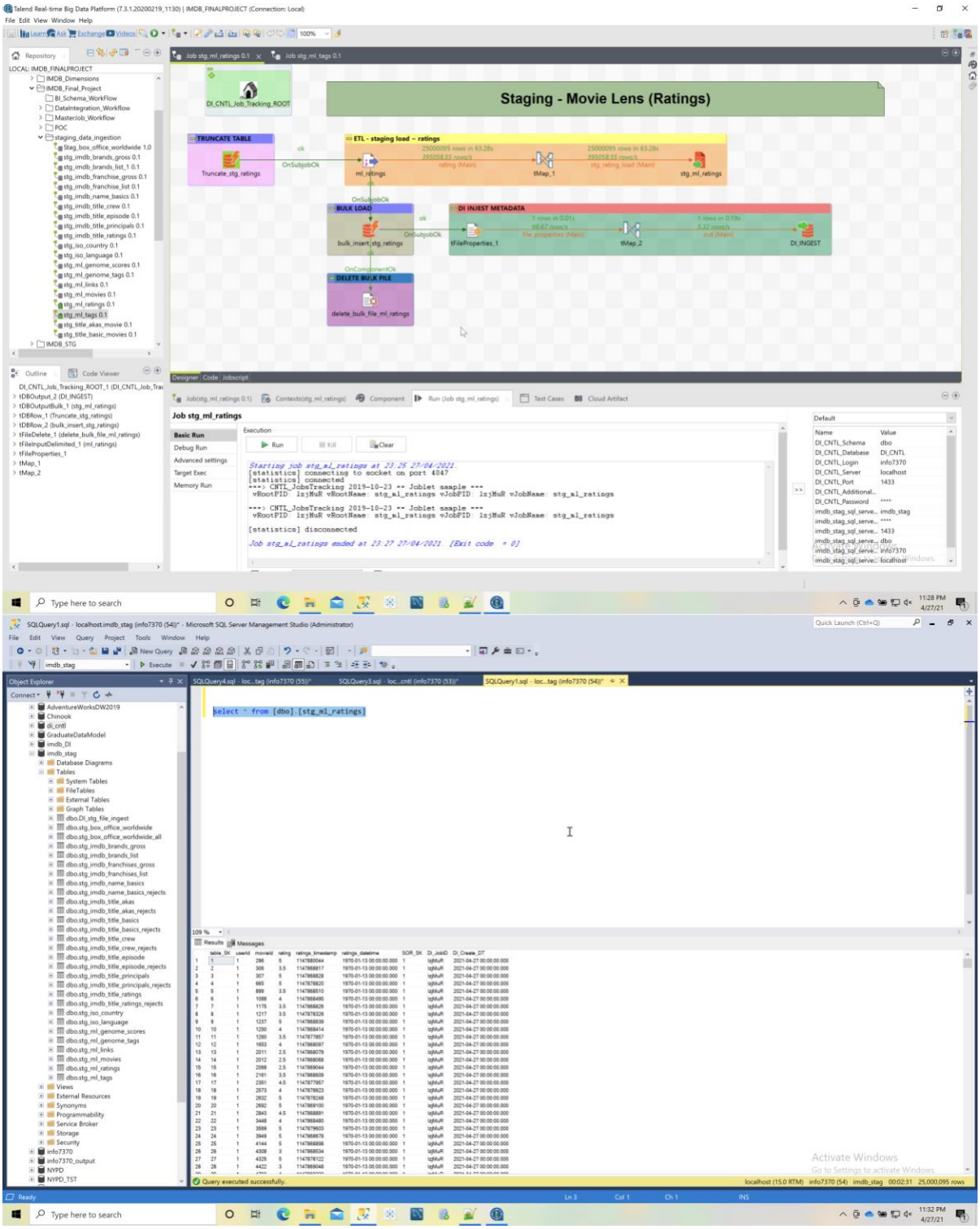
For Performance we have enabled parallelization to speed up the load

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

Submitted By: Shweta Gupta

Staging Screen Shots- Talend

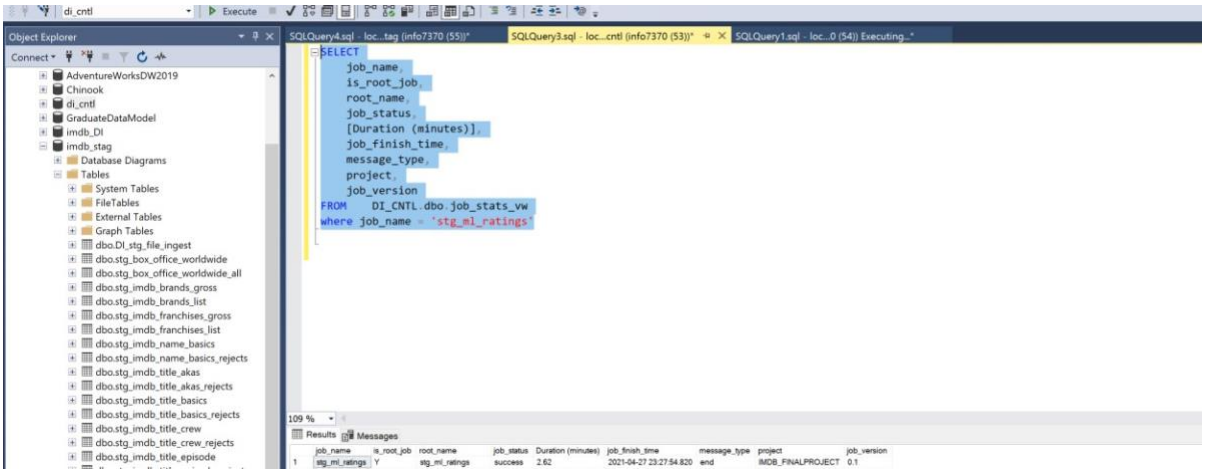
18. stg_ml_ratings



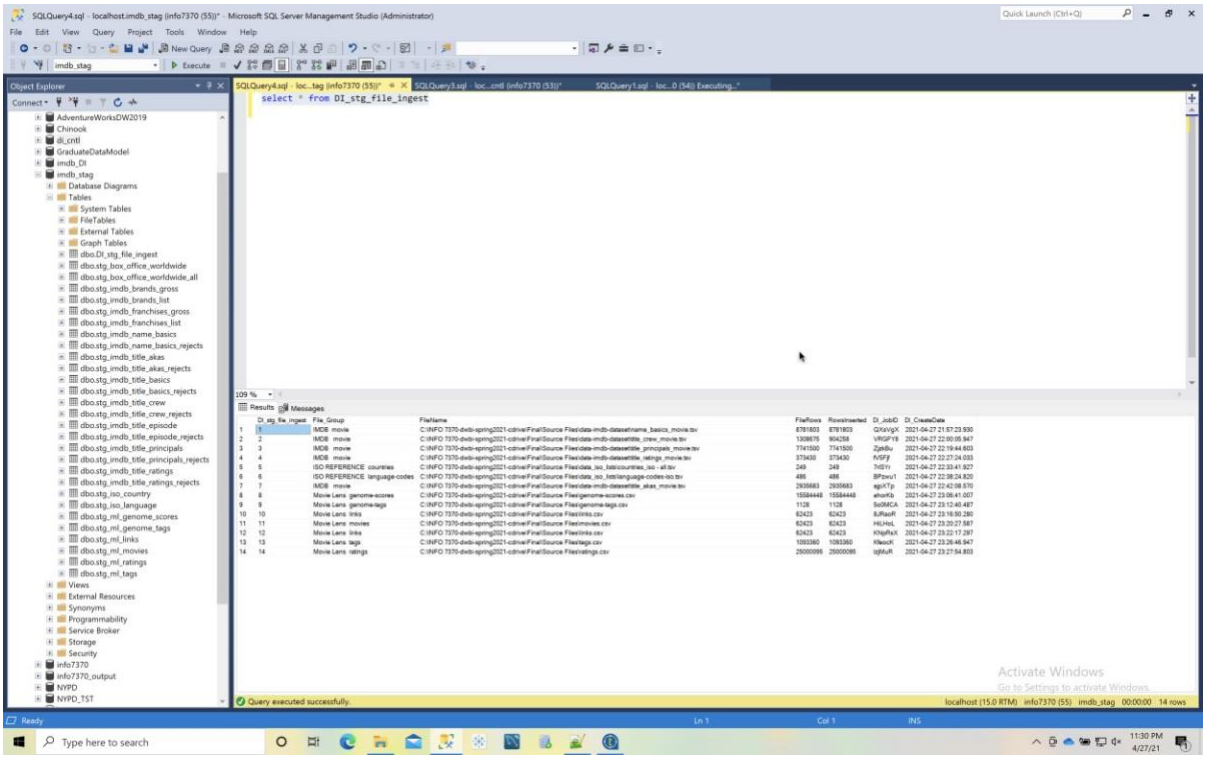
TIME:

Submitted By: Shweta Gupta

Staging Screen Shots- Talend



DI INJEST:



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

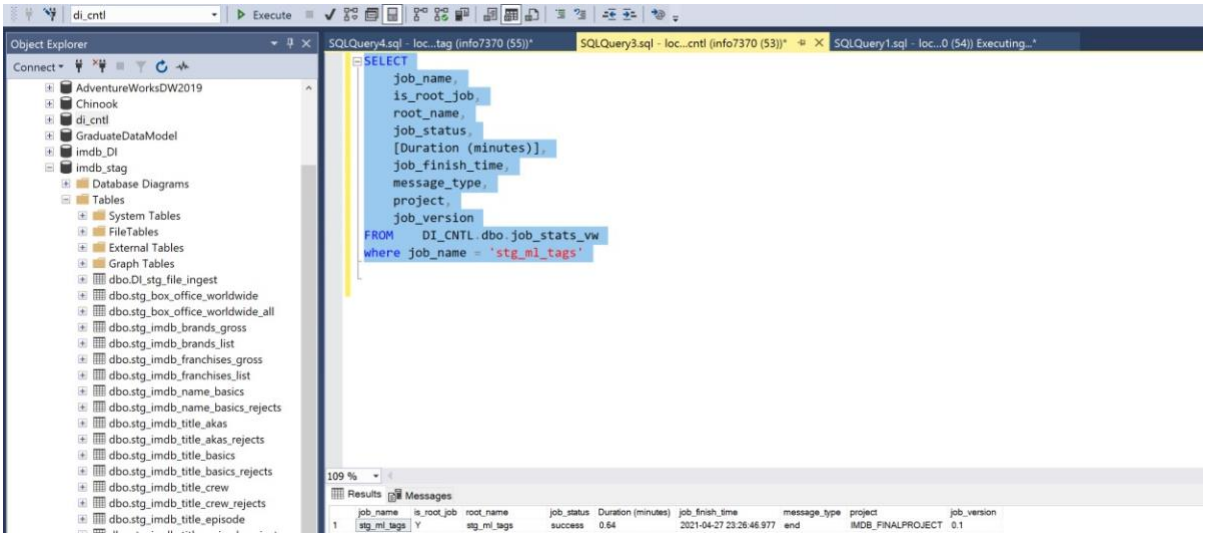
We are also capturing DI injest metadata and storing in the table which is bonus point.

Submitted By: Shweta Gupta

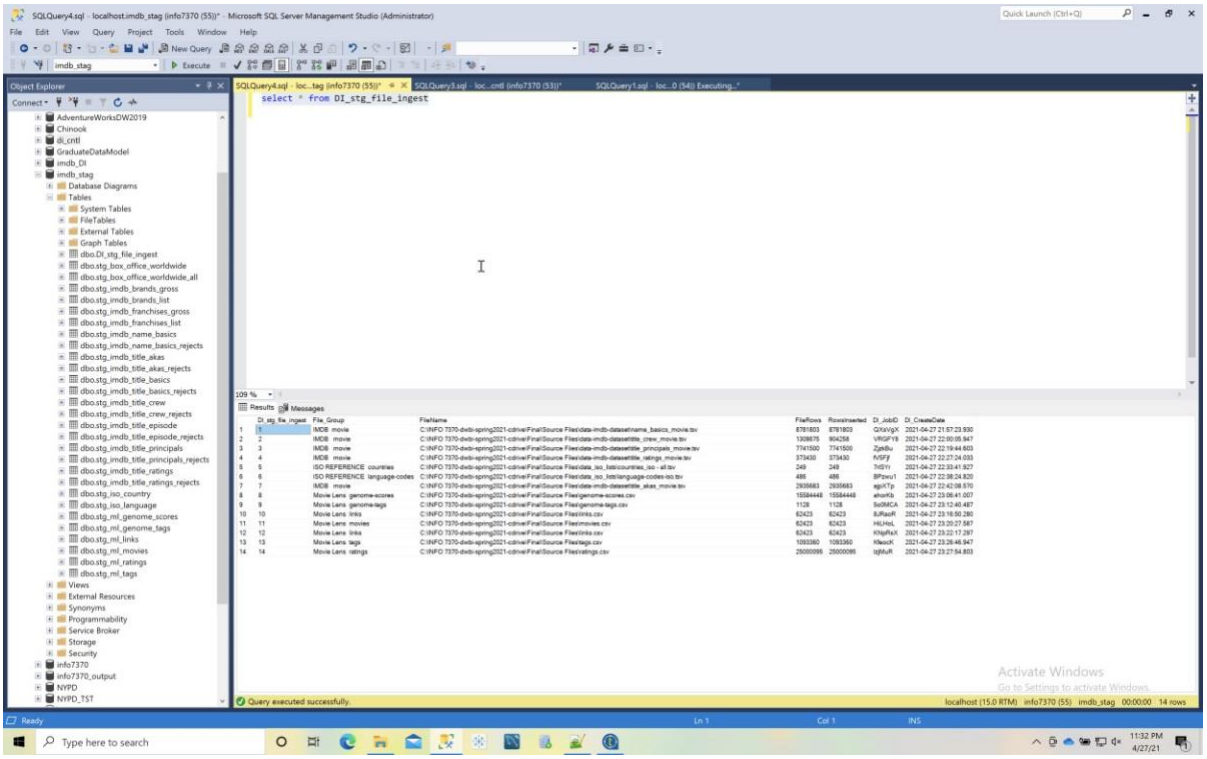
Staging Screen Shots- Talend

19. stg_ml_tags

TIME:



DI INJEST:



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

reject: No rejects required for this table.

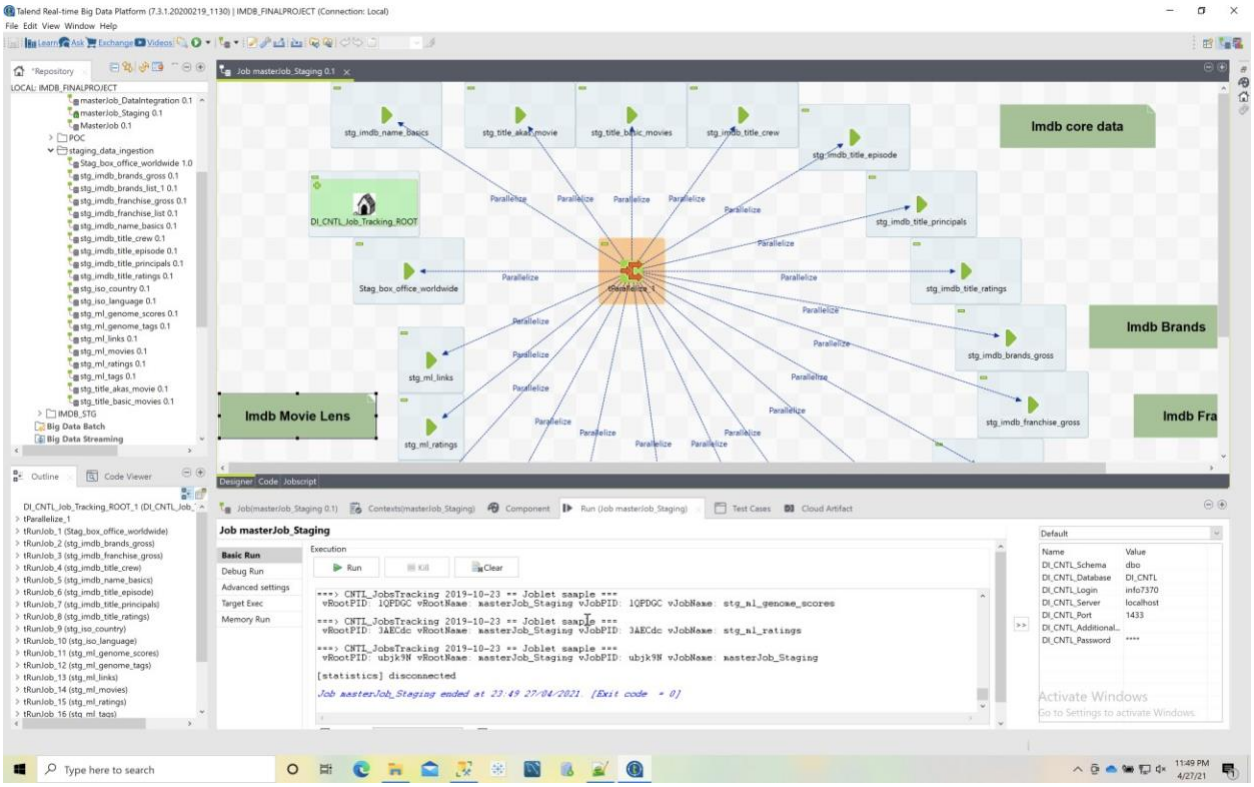
structural integration processes: initially we are truncating the table.

Submitted By: Shweta Gupta

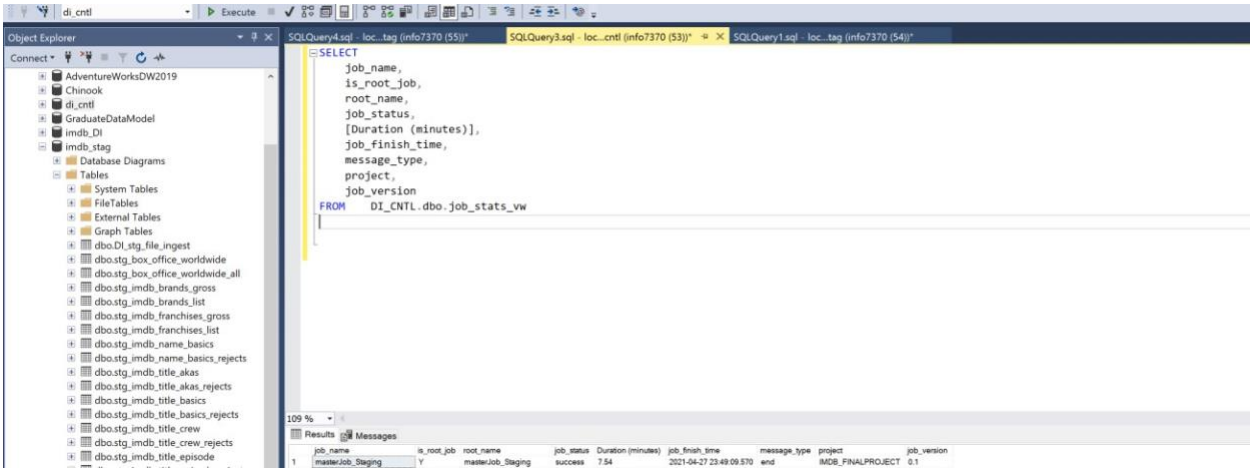
Staging Screen Shots- Talend

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

STAGING MASTER:



TIME: 7.54min



Submitted By: Shweta Gupta