



IMDb Data & Analysis Project: BI Schema & BI Dashboards

IMDb Menu All Search IMDb IMDbPro Watchlist Sign In

WATCH INSTANTLY! VIDEOS CHARACTERS SYNOPSIS GALLERY

IT BEGINS

DUNE

SEE IT IN THEATERS Own Now WATCH INSTANTLY!

Dune

Original title: Dune: Part One
2021 · PG-13 · 2h 35m

IMDb RATING 8.2/10 376K YOUR RATING ☆ Rate POPULARITY 6

Cast & crew · User reviews · Trivia · IMDbPro All topics

+ See Showtimes and tickets

More watch options

+ Add to Watchlist

Action Adventure Drama

Feature adaptation of Frank Herbert's science fiction novel about the son of a noble family entrusted with the protection of the most valuable asset and most vital element in the galaxy.

Director Denis Villeneuve

Writers Jon Spaihts (screenplay by) · Denis Villeneuve (screenplay by) · Eric Roth (screenplay by)

Stars Timothée Chalamet · Rebecca Ferguson · Zendaya

4.5K User reviews 445 Critic reviews 74 Metascore

IMDbPro See production, box office & company info

IMDb Project: BI

BI Deliverables

- Create & load data for 3 BI Schemas
 - Talend
- Create BI Dashboards answering questions
 - Tableau (Movies)
 - Power BI (TV & People)



Deliverables (Uploaded Files)

What to upload:

- Data Integration
 - List table row counts
 - Time to load each schema
- Talend
 - Export all your jobs with dependencies
 - Screenshots of your jobs' workflows
 - Explain your data consistency, reject and structural integration processes
- BI
 - Screenshots of dashboards & explain purpose of each
 - PowerBI pbix files - export your data
 - Tableau twb files - export your data



Deliverables (during project review)

Two online sessions:

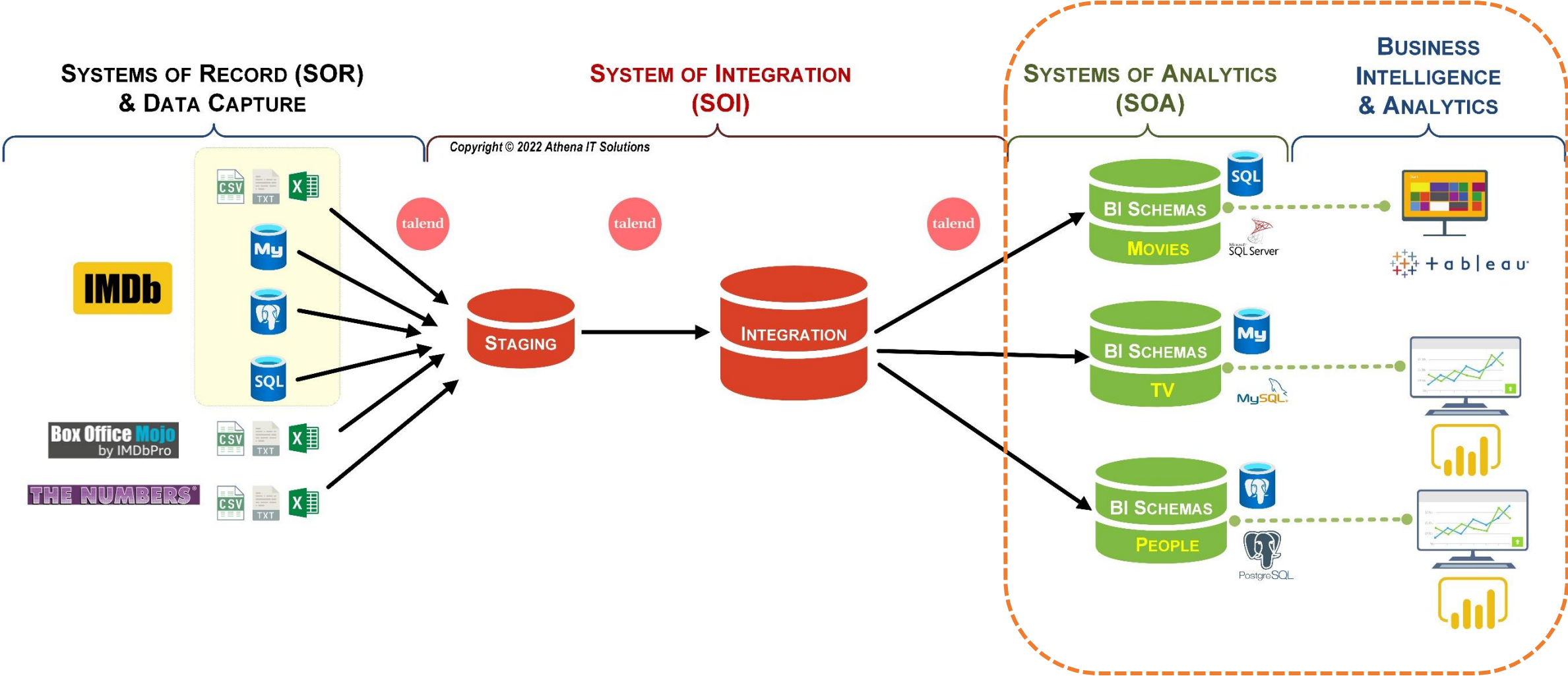
- Session with TA where you will run the complete load from source files to dimensional schema
 - Completeness of data integration
 - Total time to run
 - Table row counts per table
- Team Presentation
 - Review of data integration
 - Workflow, Transformations & Rejects
 - Review of BI
 - Answering ?s in Power BI
 - Displaying visualizations in Tableau
 - Any business analysis you feel tells a story



IMDb Project

BI Schema

IMDb Project - Architecture



IMDb Project - BI

- Create dashboards to be able to track entities in dimensional model such as movies, TV episodes and other titles with the people involved with associated revenue and ratings
- Use Tableau for Movies
 - Analysis of Box Office data
- Use Microsoft Power BI for TV & People
- Load data from IMDb INT schema
- SQL Script
 - Project imdb - BI Movies schema tables - sql server - 2022-04-28.sql
 - Project imdb - BI TV schema tables - MySQL- 2022-04-28.sql
 - Project imdb - BI People schema tables - PostgreSQL - 2022-04-28.sql

IMDb Project - BI

- 3 BI schemas:
 - Movies
 - TV Episodes and Series
 - People
 - Actors, Actresses, Writers (Screenwriters), Directors
- Each schema will be filtered by BI lists assigned to teams
 - TCONST - rows
 - NCONST - rows
 - Tables and Columns not needed for analysis
- Use the list to initially filter the data selected for the BI schema but may have to extend list to answer questions on dashboard.

Three BI Schemas

1. IMDb_BI_Movies
2. IMDb_BI_TV
3. IMDb_BI_People

Movies	TV	People
imdb_bi_movies	imdb_bi_tv	imdb_bi_people
SQL Server	MySQL	PostgreSQL

TableName	Movies	TV	People
BI_list_BRG_titles_names			
BI_list_names			
BI_list_titles			
DI_stg_file_ingest	N/A	N/A	N/A
dim_imdb_genres			
dim_imdb_job_category			
dim_imdb_name_basics			
dim_imdb_name_basics_knownForTitles			
dim_imdb_name_basics_primaryProfession			
dim_imdb_primaryProfession			
dim_imdb_title_akas	N/A	N/A	N/A
dim_imdb_title_basics			
dim_imdb_title_basics_genres			
dim_imdb_title_crew_directors			
dim_imdb_title_crew_writers			
dim_imdb_title_episode	N/A		
dim_imdb_title_principals			
dim_imdb_brand		N/A	N/A
dim_imdb_Franchise		N/A	N/A
fct_imdb_brands_gross		N/A	N/A
fct_imdb_brands_list		N/A	N/A
fct_imdb_franchises_gross		N/A	N/A
fct_imdb_franchises_list		N/A	N/A
fct_imdb_title_ratings		N/A	N/A
fct_movies_box_office_worldwide		N/A	N/A
fct_numbers_daily_box_office		N/A	N/A
fct_numbers_franchise_all_box_office		N/A	N/A
fct_numbers_franchise_movies_box_office		N/A	N/A
dim_imdb_title_basics_episodes_vw	N/A		N/A

Three BI Schemas

1. IMDb_BI_Movies
2. IMDb_BI_TV
3. IMDb_BI_People

F	Full Load
L	Filtered by List
E	Extended List

Order	TableName	Category	tconst	nconst	Movies	TV	People
1	BI_list_titles				1, E	1, E	2
2	BI_list_names				2	2	1
3	BI_list_BRG_titles_names				2	2	2
4	dim_imdb_genres		N/A	N/A	F	F	F
5	dim_imdb_job_category		N/A	N/A	F	F	F
6	dim_imdb_primaryProfession		N/A	N/A	F	F	F
7	dim_imdb_titleType		N/A	N/A	F	F	F
10	dim_imdb_title_basics		Y		E	E	L
11	dim_imdb_name_basics			Y	L	L	L
12	dim_imdb_name_basics_knownForTitles		Y	Y	L	L	L
13	dim_imdb_name_basics_primaryProfession			Y	L	L	L
15	dim_imdb_title_basics_genres		Y		E	L	L
16	dim_imdb_title_crew_directors		Y	Y	E	L	L
17	dim_imdb_title_crew_writers		Y	Y	E	L	L
18	dim_imdb_title_principals		Y	Y	E	L	L
19	fct_imdb_title_ratings		Y		E	L	L
20	dim_imdb_title_episode	tv	Y		N/A	L	L
21	dim_imdb_brand	movies			F	N/A	F
22	dim_imdb_Franchise	movies			F	N/A	F
23	fct_imdb_brands_gross	movies			F	N/A	F
24	fct_imdb_brands_list	movies			F	N/A	F
25	fct_imdb_franchises_gross	movies			F	N/A	F
26	fct_imdb_franchises_list	movies			F	N/A	F
28	fct_movies_box_office_worldwide	movies			F	N/A	F
29	fct_numbers_daily_box_office	movies			F	N/A	F
30	fct_numbers_franchise_all_box_office	movies			F	N/A	F
31	fct_numbers_franchise_movies_box_office	movies			F	N/A	F

BI Lists – Use for BI analysis & dashboards

Project Team
assigned files to filter
BI schema

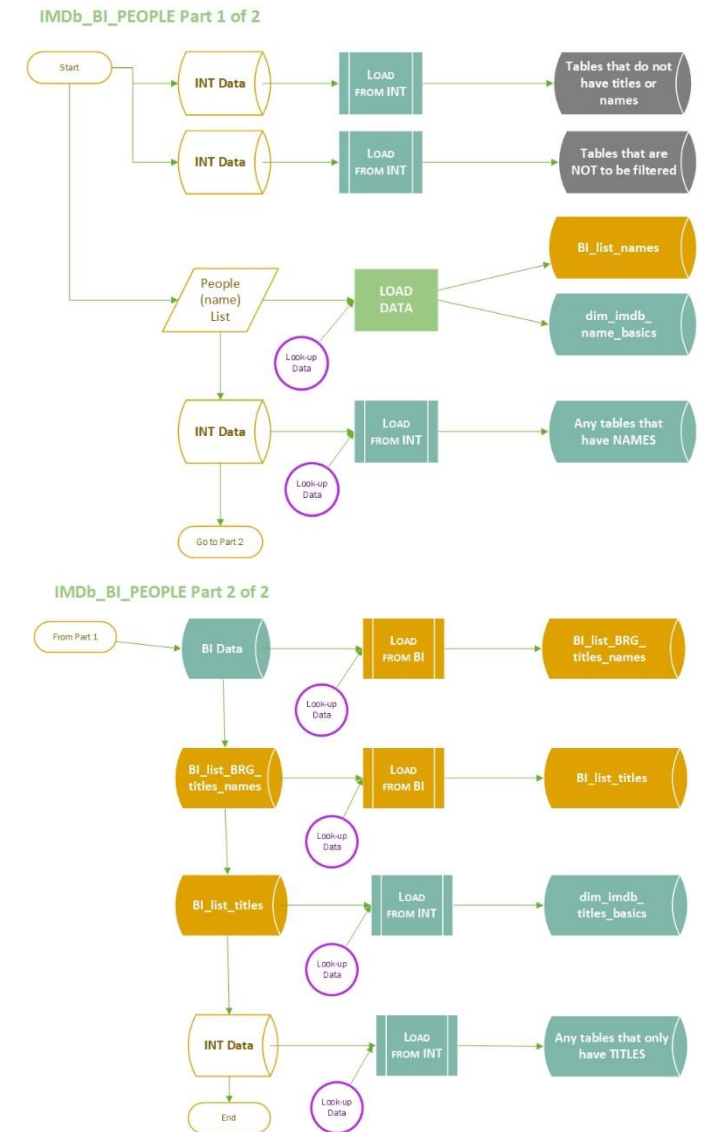
People List (OneDrive \Project - imdb\data_imdb_lists)	Team	Team	Team
BI List - People - Best Screenwriters.csv	4	15	25
BI List - People - The Best Actors Ever - 2010s.csv	6	16	26
BI List - People - The Best Actresses and Actors - Born in the 1960s.csv	7	17	27
BI List - People - The Best Directors Ever!.csv	8	18	28
BI List - People - The Best Female Directors.csv	9	19	29
BI List - People - The Most Popular Actresses.csv	10	20	31
BI List - People - The World's 1000 Greatest Actress.csv	11	21	
BI List - People - Top 1000 Actors and Actresses.csv	12	22	
BI List - People - Top 250 directors by IMDB rating.csv	13	23	
BI List - People - Top 500 Villain Actors.csv	14	24	
TV List OneDrive \Project - imdb\data_imdb_lists)	Team	Team	Team
BI List - TV - 100 Greatest Children TV Shows .csv	4	14	23
BI List - TV - 100 Greatest Modern TV Series .csv	6	15	24
BI List - TV - Best crime series.csv	7	16	25
BI List - TV - Best Drama TV Shows.csv	8	17	26
BI List - TV - IMDb Highest Rated TV Episodes.csv	9	18	27
BI List - TV - The Best Murder Mystery Thriller & Crime TV Shows.csv	10	19	28
BI List - TV - Top 100 tv comedy shows of all time.csv	11	20	29
BI List - TV - Top 250 TV Series.csv	12	21	31
BI List - TV - Top Sci-Fi Shows.csv	13	22	

BI Schema – Filtering by Lists

- Three tables used for gathering filtered data
 - BI_list_titles
 - BI_list_names
 - BI_list_BRG_titles_names

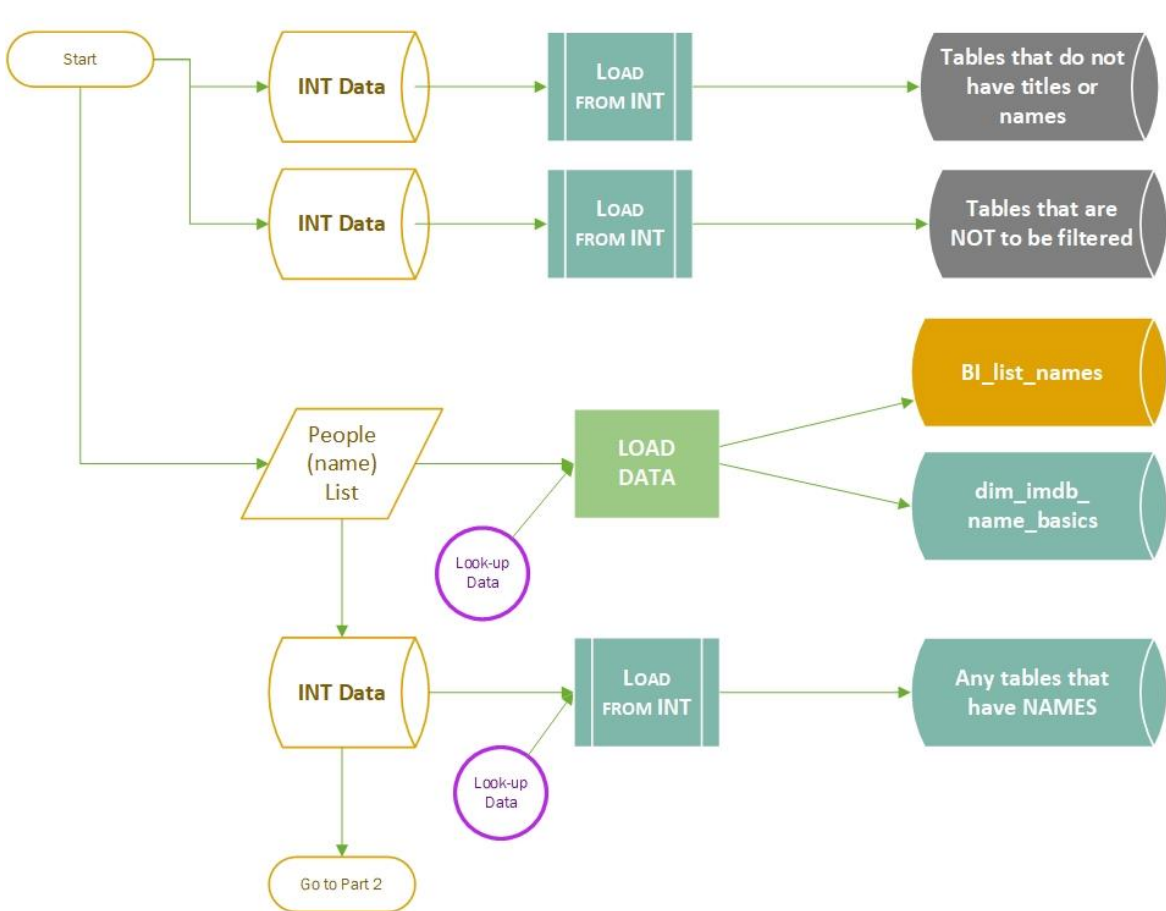
BI Schema – Filtering process

- For People oriented lists – imdb_bi_people
 - Read in list of names into BI_list_names
 - Ingest all data from INT schema where nconsts (names_sk) is the sole criteria
 - Populate BI_list_BRG_titles_names from all the BI tables that have NCONST and TCONST (titles_sv and names_sk)
 - Populate BI_list_titles from BI_list_BRG_titles_names
 - Ingest all data from INT schema where tconsts (titles_sk) is the sole criteria
 - Load all data applicable to BI_schema that is not filtered by NCONST or TCONST
 - (note: this can be done before any of the above)

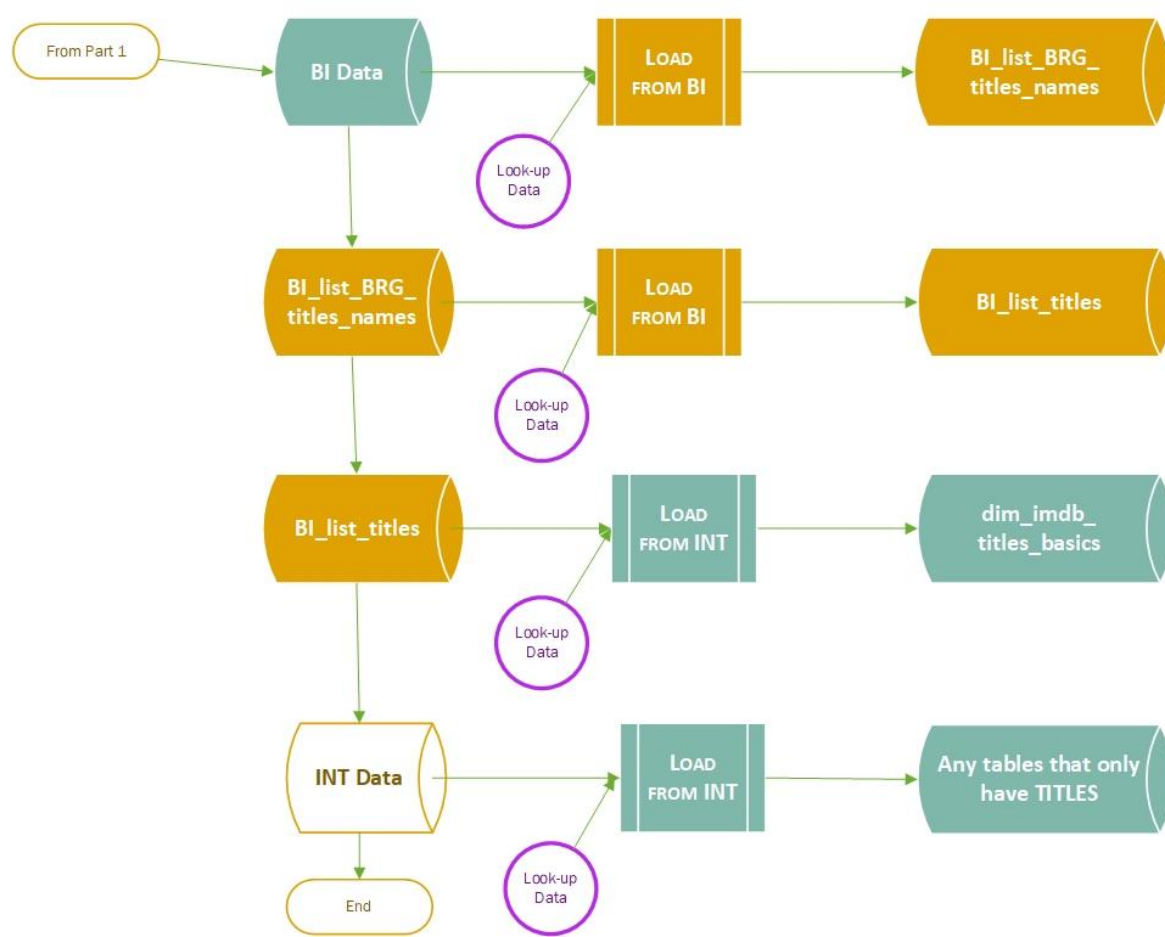


BI Schema – Filtering process

IMDb_BI_PEOPLE Part 1 of 2



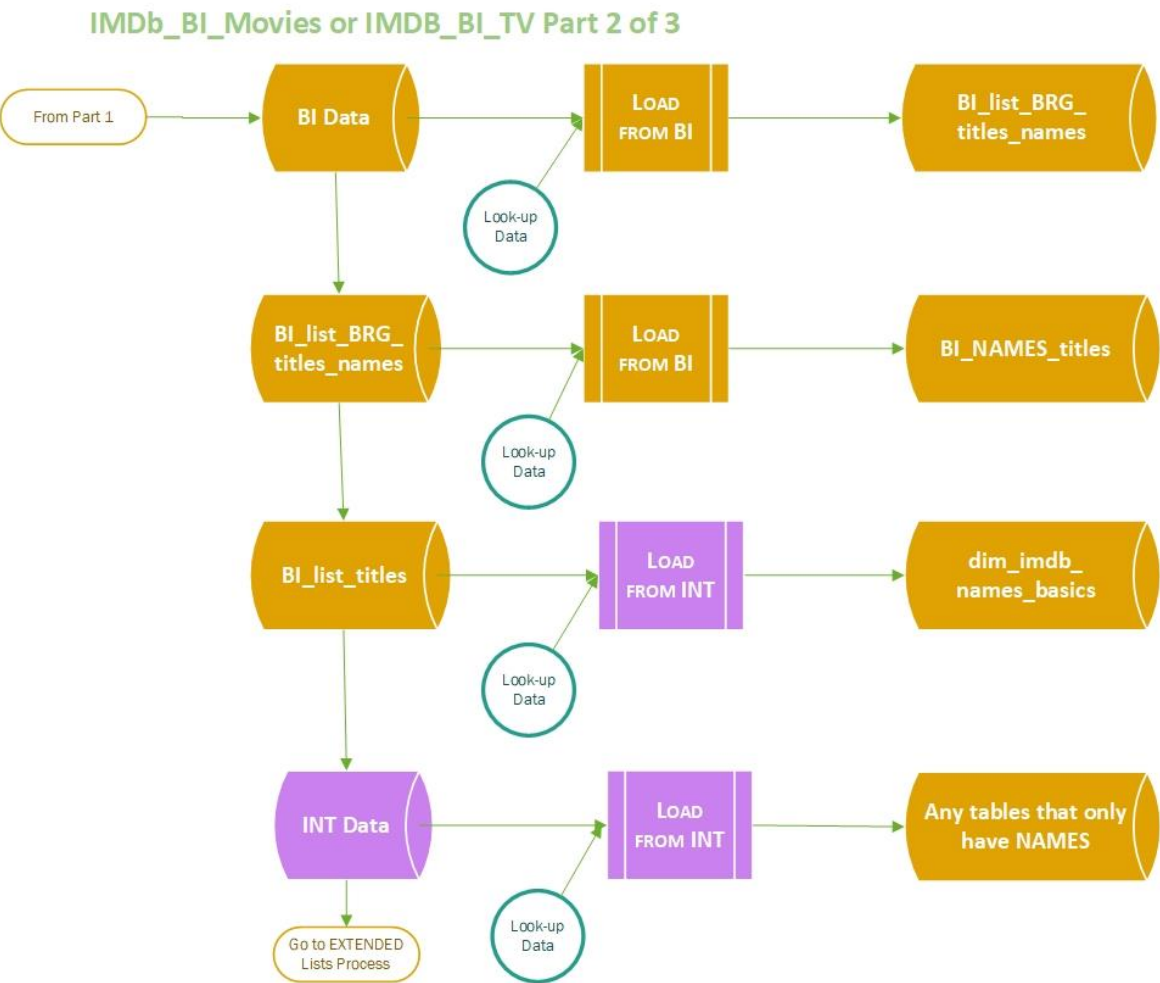
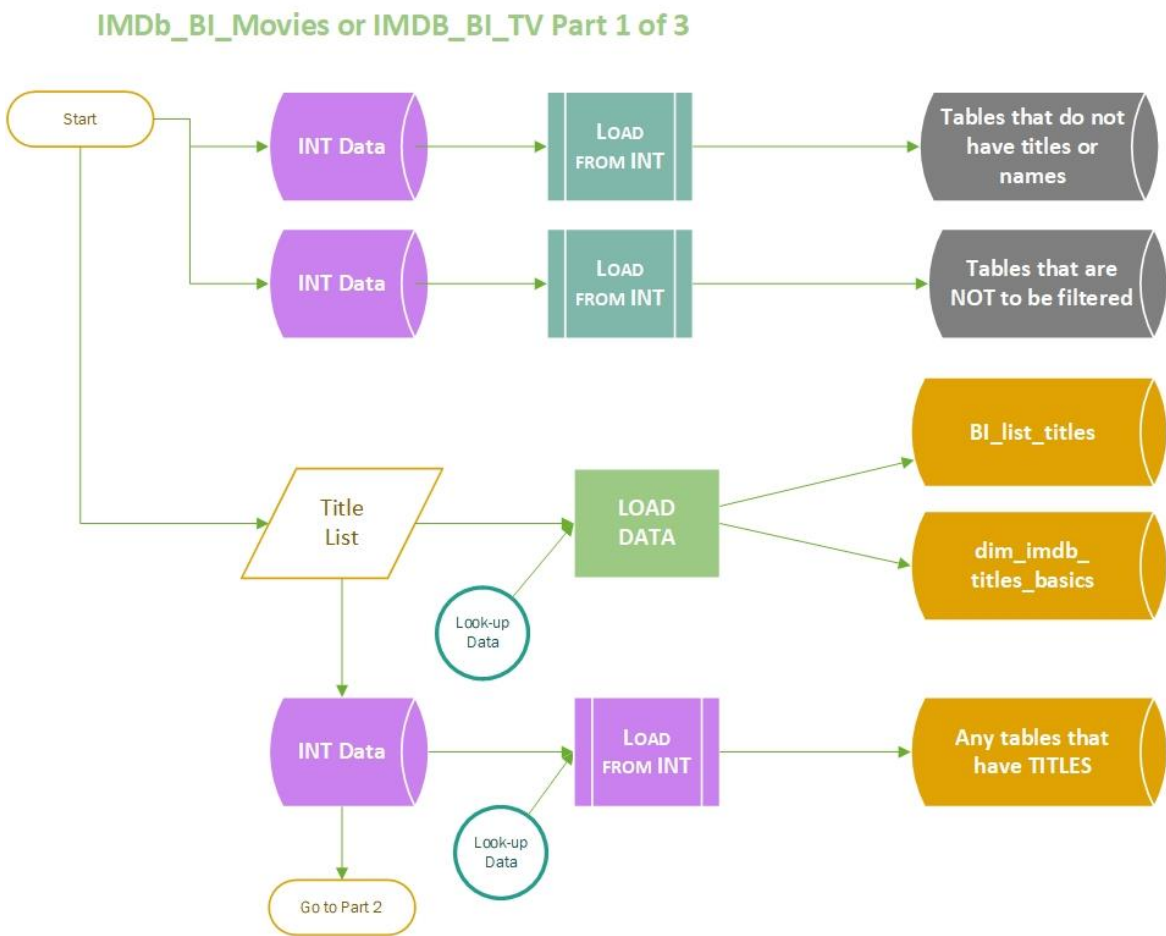
IMDb_BI_PEOPLE Part 2 of 2



BI Schema – Filtering process

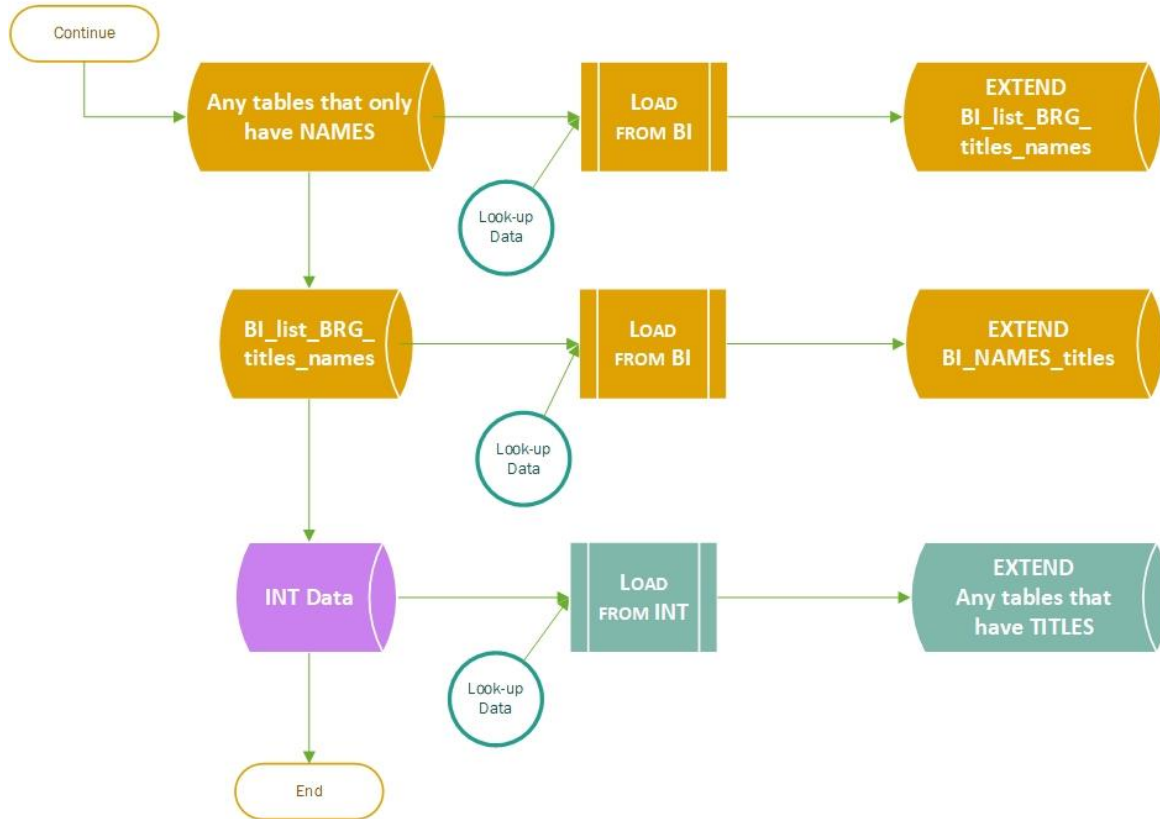
- For Title oriented lists
 - Read in list of titles into BI_list_titles
 - Ingest all data from INT schema where tconsts (titles_sk) is the sole criteria
 - Populate BI_list_BRG_titles_names from all the BI tables that have NCONST and TCONST (titles_sv and names_sk)
 - This would be done while loading BI schema tables that have NCONST and TCONST
 - Populate BI_list_names from BI_list_BRG_titles_names
 - Add data into BI tables that have NCONST and TCONST (titles_sv and names_sk) where
 - names_sk in BI_list_BRG_titles_names but titles_sv not in BI_list_titles
 - Expand BI_list_BRG_titles_names and titles_sv not in BI_list_titles based on above
 - Add data to dim_imdb_title_basics based on expanded BI_list_titles based on above
 - Ingest all data from INT schema where nconsts (names_sk) is the sole criteria
 - Load all data applicable to BI_schema that is not filtered by NCONST or TCONST
 - (note: this can be done before any of the above)

BI Schema – Filtering process



BI Schema – Filtering process

IMDb_BI_Movies or IMDB_BI_TV Part 2 of 2



- Initial pass is the titles with the people that were in those titles
- Extended means adding in the OTHER titles that those people appeared in

BI Schema – Data not impacted by filtering

- For IMDb_BI_Movies ingest all data from:
 - dim_imdb_brand
 - dim_imdb_Franchise
 - fct_imdb_brands_gross
 - fct_imdb_brands_list
 - fct_imdb_franchises_gross
 - fct_imdb_franchises_list
 - fct_movies_box_office_worldwide
 - fct_numbers_daily_box_office
 - fct_numbers_franchise_all_box_office
 - fct_numbers_franchise_movies_box_office
- NOTE: Do NOT filter the above by titles

BI Schema – Table excluded

- For IMDb_BI_Movies exclude:
 - dim_imdb_title_episode
- For IMDb_BI_TV & IMDb_BI_People exclude:
 - dim_imdb_brand
 - dim_imdb_Franchise
 - fct_imdb_brands_gross
 - fct_imdb_brands_list
 - fct_imdb_franchises_gross
 - fct_imdb_franchises_list
 - fct_movies_box_office_worldwide
 - fct_numbers_daily_box_office
 - fct_numbers_franchise_all_box_office
 - fct_numbers_franchise_movies_box_office

IMDb BI: Movies

- Dashboard: Top 1000 movies by all time worldwide gross revenue
 - Release date, running time, rankings, revenue related data, etc.
 - Genre(s)
 - Ratings
 - When a top 1000 movie is selected information is provided about the following:
 - Actors & actresses
 - Writers
 - Directors
 - A url is available to link titles and names to IMDb web page
- Use fct_movies_box_office_worldwide (1,000 movies) as the filter list for this BI Schema
- Compare daily box office for selected films

IMDb BI: TV

- Dashboard: TV Shows
 - Release date
 - Genre(s)
 - Ratings
 - When a tv title is selected information is provided about the following:
 - Actors & actresses
 - Writers
 - Directors
 - A url is available to link titles and names to IMDb web page
 - Select TV series & list seasons & episodes
- Each team is assigned an IMDb list of TV series and episodes

IMDb BI: People

- Dashboard: People involved with titles (movies & TV)
 - Actors & actresses
 - Writers
 - Directors
 - List movie and TV titles they worked in along with attributes
 - List job-related attributes for movie or TV titles they worked in
 - Be able to link to IMDb web page for these people
-
- Each team is assigned an IMDb list of TV series and episodes