

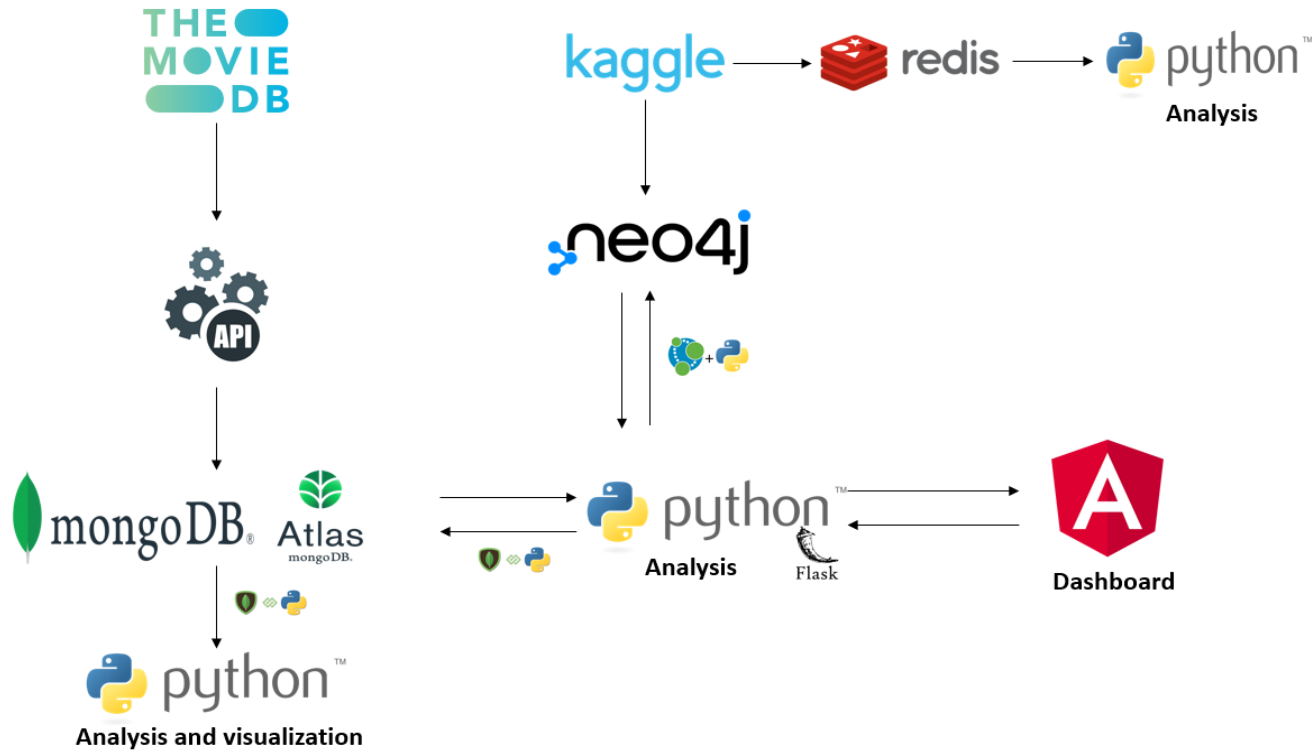
TV Shows & Movie Analysis

INTRODUCTION

- Data is of movies and TV shows scraped from the movie database (TMDB) and kaggle.
- TV shows scraped from TMDB has 94k records
- Movie data scraped from TMDB has 0.5M records
- Movie and TV shows data from Kaggle has 7.7k records with cast information.
- These data is fed into MongoDB, Neo4j and Redis

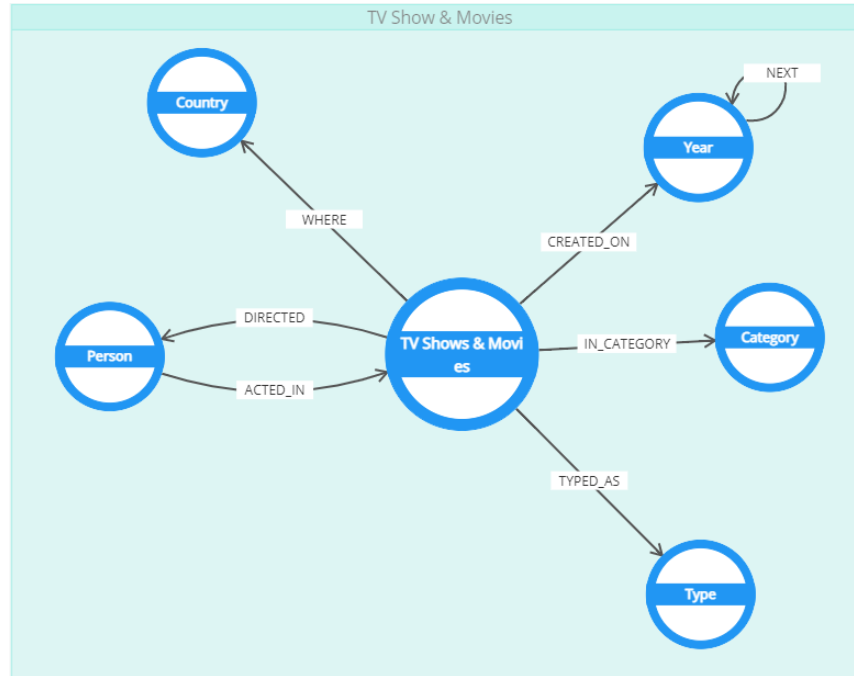
DATA PIPELINE OVERVIEW

STAATLICH
ANERKANNTE
HOCHSCHULE



NEO4J DATA MODEL – TV SHOWS AND MOVIES

STAATLICH
ANERKANNTE
HOCHSCHULE



neo4j

hackolade

MONGO DB DATA MODEL – TV SHOWS

STAATLICH
ANERKANNTE
HOCHSCHULE

TV_SERIES		
::	_id	old
::	backdrop_path	str
::	created_by	arr
::	[0]	doc
::	id	num
::	credit_id	str
::	name	str
::	gender	num
::	profile_path	str
::	episode_run_time	arr
::	[0]	num
::	first_air_date	date
::	genres	arr
::	homepage	str
::	id	num
::	in_production	bool
::	language	arr
::	[0]	str
::	last_air_date	date
::	last_episode_to_air	doc
::	air_date	date
::	episode_number	num
::	id	num
::	name	str
::	overview	str
::	production_code	str
::	season_number	num
::	show_id	num
::	still_path	str
::	vote_average	num
::	vote_count	num
::	name	str
::	next_episode_to_air	str

::	networks	arr
::	[0]	old
::	number_of_episodes	num
::	number_of_seasons	num
::	origin_country	arr
::	[0]	str
::	original_language	str
::	original_name	str
::	popularity	num
::	poster_path	str
::	production_companies	arr
::	[0]	doc
::	id	num
::	logo_path	str
::	name	str
::	origin_country	str
::	seasons	arr
::	[0]	doc
::	air_date	date
::	episode_count	num
::	id	num
::	name	str
::	overview	str
::	poster_path	str
::	season_number	num
::	status	str
::	type	str
::	vote_average	num
::	vote_count	num



MONGO DB DATA MODEL – MOVIES

STAATLICH
ANERKANNTE
HOCHSCHULE

New Collection		
:: backdrop_path	str	
:: belongs_to_collection	str	
:: adult	bool	
:: Genre	arr	#
:: [0] Object	doc	
:: id	num	
:: name	str	
:: budget	num	#
:: homepage	str	
:: id	num	
:: imdb_id	num	
:: original_language	str	#
:: original_title	str	
:: overview	str	
:: popularity	num	#
:: poster_path	str	
:: production_companies	arr	
:: [0] object	doc	
:: id	num	
:: logo_path	str	
:: name	str	
:: origin_country	str	

:: production_countries	arr	
:: [0] object	doc	
:: iso_3661_1	str	
:: name	str	
:: release_date	num	
:: revenue	num	
:: runtime	num	
:: spoken_languages	arr	
:: [0] object	doc	
:: iso_639_1	str	
:: name	str	
:: status	str	#
:: tagline	str	
:: title	str	
:: video	bool	
:: vote_average	num	
:: vote_count	num	



REDIS DATA MODEL – TV SHOWS AND MOVIES

Redis Hash

Show id	
type	Value
title	value
director	value
cast	value
country	value
date	value
Release year	value
rating	value
duration	value
listed in	Redis set
description	value
quantity	value
purchased	value



USER STORIES

	USER STORY	DATABASE
1.	Popularity comparison between multiple genres over years	MongoDB
2.	Auto completing TV shows when it is searched partially	MongoDB Atlas
3.	Releases during pandemic versus past 20 years	MongoDB
4.	Top 10 movies by revenue in a year (Filter using dropdown)	MongoDB
5.	Recommendation system for a user if he has already watched a TV Show	Neo4j
6.	Best cast to work with given the lead actors	Neo4j
7.	Similarity between genres – Which genres are usually combined	Neo4j
8.	Evolution of movies released by countries	Neo4j
9.	Seasons comparison India vs rest of the world	Neo4j
10.	Analysis on positive/negative reviews	Redis
11.	Users less than 18 cant watch UR rated shows	Redis
12.	Error logging demonstration	Redis
13.	Using autocomplete in mongo dB and find cast from neo4j	MongoDB & Neo4j
14.	Popularity of shows based on genre	MongoDB & Neo4j
15.	Animated International TV shows with 2 Seasons	MongoDB & Neo4j

FUTURE WORK

- Automate the data updation process.
- Fetch crew and cast data for all TV shows and Movies.
- Recommendations based on user search stored in cache.
- Realtime application of age restriction for movies and TV shows