



UNIVERSITÀ DEGLI STUDI DI SALERNO
DIPARTIMENTO DI INFORMATICA
CORSO DI BASI DI DATI II

DISNEY+

Rossi Rosaria 0522501340
Elia Teresa 0522501189



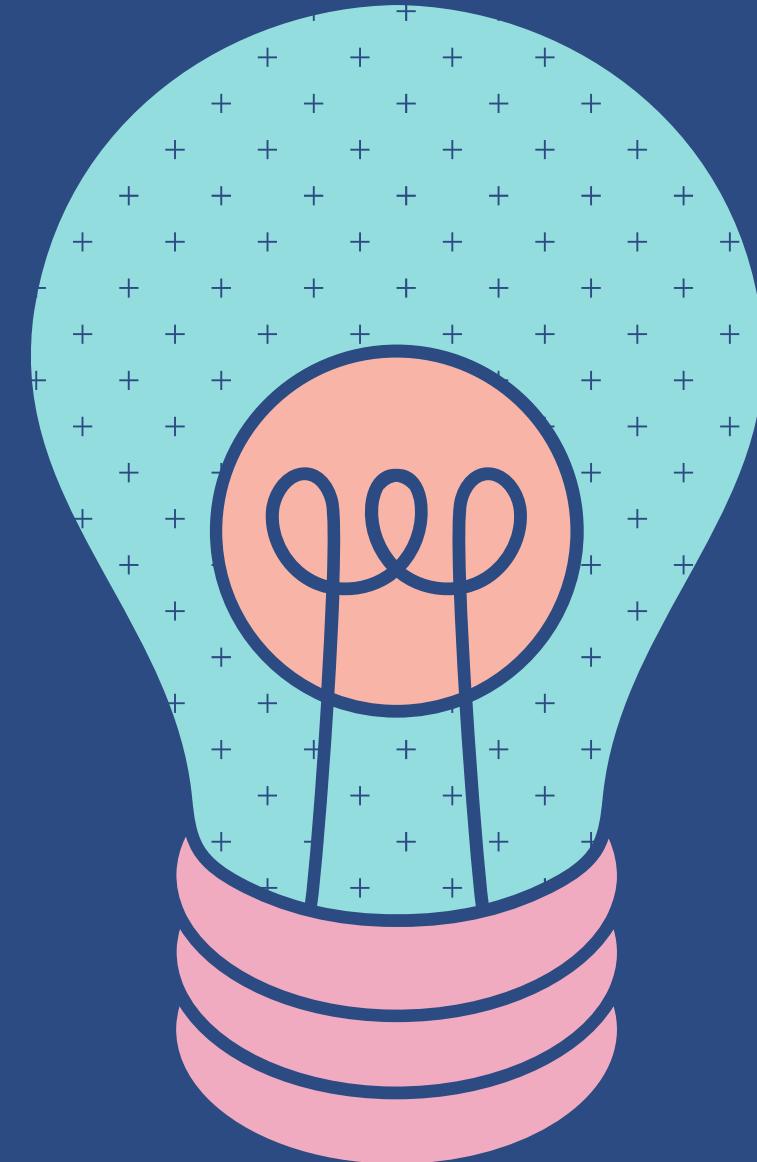
Il progetto ha lo scopo di sviluppare un database per effettuare la gestione dei dati della piattaforma streaming, rilasciata nel 2020: Disney+.



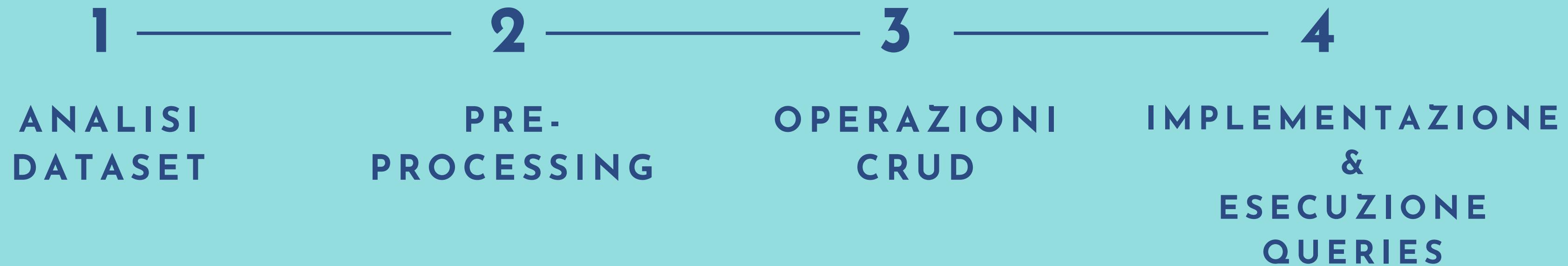
Le tecnologie utilizzate sono NoSQL e MongoDB.



E' stato sviluppato un software in linguaggio Python per l'interrogazione del DB.



Fasi di sviluppo del progetto



Fasi di sviluppo del progetto

1

ANALISI
DATASET

2

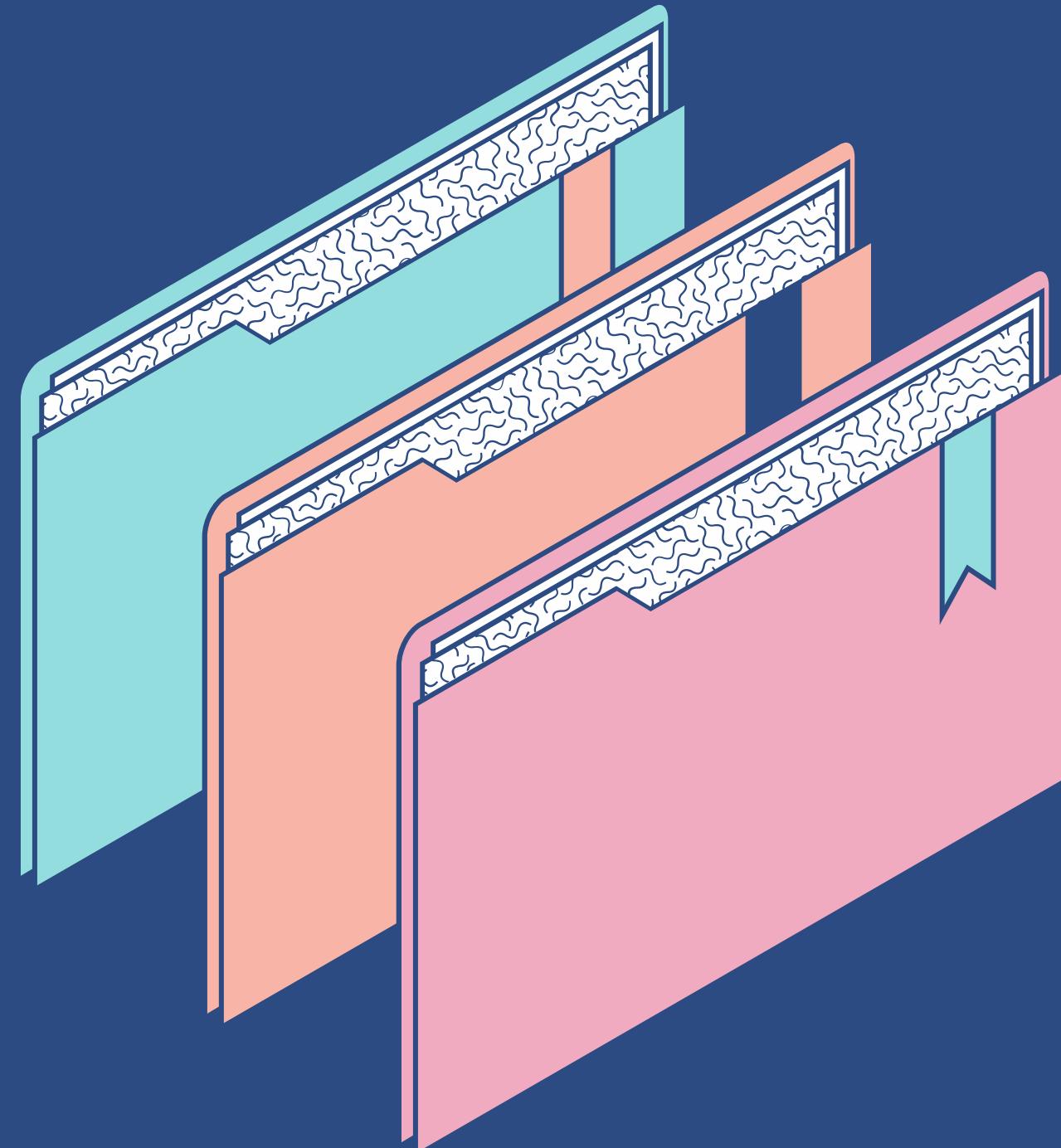
PRE-PROCESSING

3

OPERAZIONI
CRUD

4

IMPLEMENTAZIONE
&
ESECUZIONE
QUERIES



ANALISI DATASET

Il dataset riporta le informazioni circa serie tv e film presenti sulla piattaforma Disney+ e prodotti dall'anno 1900 ad oggi.

Fornisce in dettaglio informazioni circa i registi e gli attori che hanno preso parte alla loro produzione e gli score ottenuti da parte del pubblico.

credits.csv

Il dataset contiene più di 26K voci, circa attori e direttori di titoli di serie tv o film presenti su Disney+.

Il file presenta la struttura seguente

person_id	id	name	character	role
23433	tm74391	Deems Taylor	Narrator - Narrative Introductions	ACTOR
5910	tm74391	Walt Disney	Mickey Mouse (segment 'The Sorcerer's Apprentice') (voice)	ACTOR
23436	tm74391	Julietta Novis	Soloist (segment 'Ave Maria') (singing voice)	ACTOR
23434	tm74391	Leopold Stokowski	Himself - Conductor of The Philadelphia Orchestra	ACTOR
23441	tm74391	Paul Satterfield		DIRECTOR
23442	tm74391	Hamilton Luske		DIRECTOR
23443	tm74391	Jim Handley		DIRECTOR
6164	tm74391	Ford Beebe		DIRECTOR
23444	tm74391	T. Hee		DIRECTOR
23446	tm74391	Wilfred Jackson		DIRECTOR
23440	tm74391	William Roberts		DIRECTOR
13776	tm74391	Ben Sharpsteen		DIRECTOR
23438	tm74391	James Algar		DIRECTOR

titles.csv

Il dataset contiene +1500 titoli di serie tv o film presenti su Disney+ con tutte le informazioni anesse.

Il file presenta la struttura seguente

id	title	type	description	release_year	age_certification	runtime	genres	production_countries	seasons	imdb_id	imdb_score	imdb_votes	tmdb_popularity	tmdb_score
tm74391	Fantasia	MOVIE	Walt Disney's ti	1940	G	120	['animation', ['US']]			tt0032455	7.7	94681.0	57.751	7.4
tm67803	Snow White a	MOVIE	A beautiful girl,	1937	G	83	['fantasy', 'fa	['US']]		tt0029583	7.6	195321.0	107.137	7.1
tm82546	Pinocchio	MOVIE	Lonely toymake	1940	G	88	['animation', ['US']]			tt0032910	7.5	141937.0	71.16	7.1
tm79357	Bambi	MOVIE	Bambi's tale un	1942	G	70	['animation', ['US']]			tt0034492	7.3	140406.0	68.136	7.0
tm62671	Treasure Islan	MOVIE	Enchanted by th	1950	PG	96	['family', 'acti	['GB', 'US']]		tt0043067	6.9	8229.0	10.698	6.5
tm61729	The Adventur	MOVIE	The Wind in the	1949	G	68	['fantasy', 'ho	['US']]		tt0041094	6.9	15405.0	10.862	6.5
tm61052	Cinderella	MOVIE	Cinderella has f	1950	G	74	['fantasy', 'ar	['US']]		tt0042332	7.3	157099.0	144.284	7.0
tm87946	Dumbo	MOVIE	Dumbo is a bab	1941	G	63	['animation', ['US']]			tt0033563	7.2	129548.0	69.875	7.0
tm4623	The Three Ca	MOVIE	For Donald's bi	1944	G	71	['animation', ['US']]			tt0038166	6.3	13889.0	11.75	6.4
tm77826	The Reluctant	MOVIE	Humorist Robe	1941	G	74	['fantasy', 'ar	['US']]		tt0034091	6.8	3280.0	6.958	6.6
tm89831	Fun and Fanc	MOVIE	Jiminy Cricket h	1947	G	73	['animation', ['US']]			tt0039404	6.4	8228.0	10.058	6.3
tm3616	Steamboat W	MOVIE	Mickey Mouse,	1928	G	8	['animation']	['US']]					10.513	7.4
tm77407	Saludos Amig	MOVIE	Live-action seg	1942		42	['animation', ['US']]			tt0036326	6.1	6506.0	20.418	5.8
tm71205	How to Play E	MOVIE	Goofy shows us	1942		8	['comedy', 'a	['US']]					4.179	6.6
tm5822	Melody Time	MOVIE	In the grand tra	1948		75	['comedy', 'm	['US']]		tt0040580	6.2	5993.0	8.856	6.2
tm8579	Ferdinand the	MOVIE	This Oscar-winn	1938	G	8	['comedy']	['US']]					10.983	6.5
tm134279	Three Little Pi	MOVIE	The two pigs bu	1933		8	['comedy']	['US']]					8.883	6.8

Fasi di sviluppo del progetto

1

ANALISI
DATASET

2

PRE-
PROCESSING

3

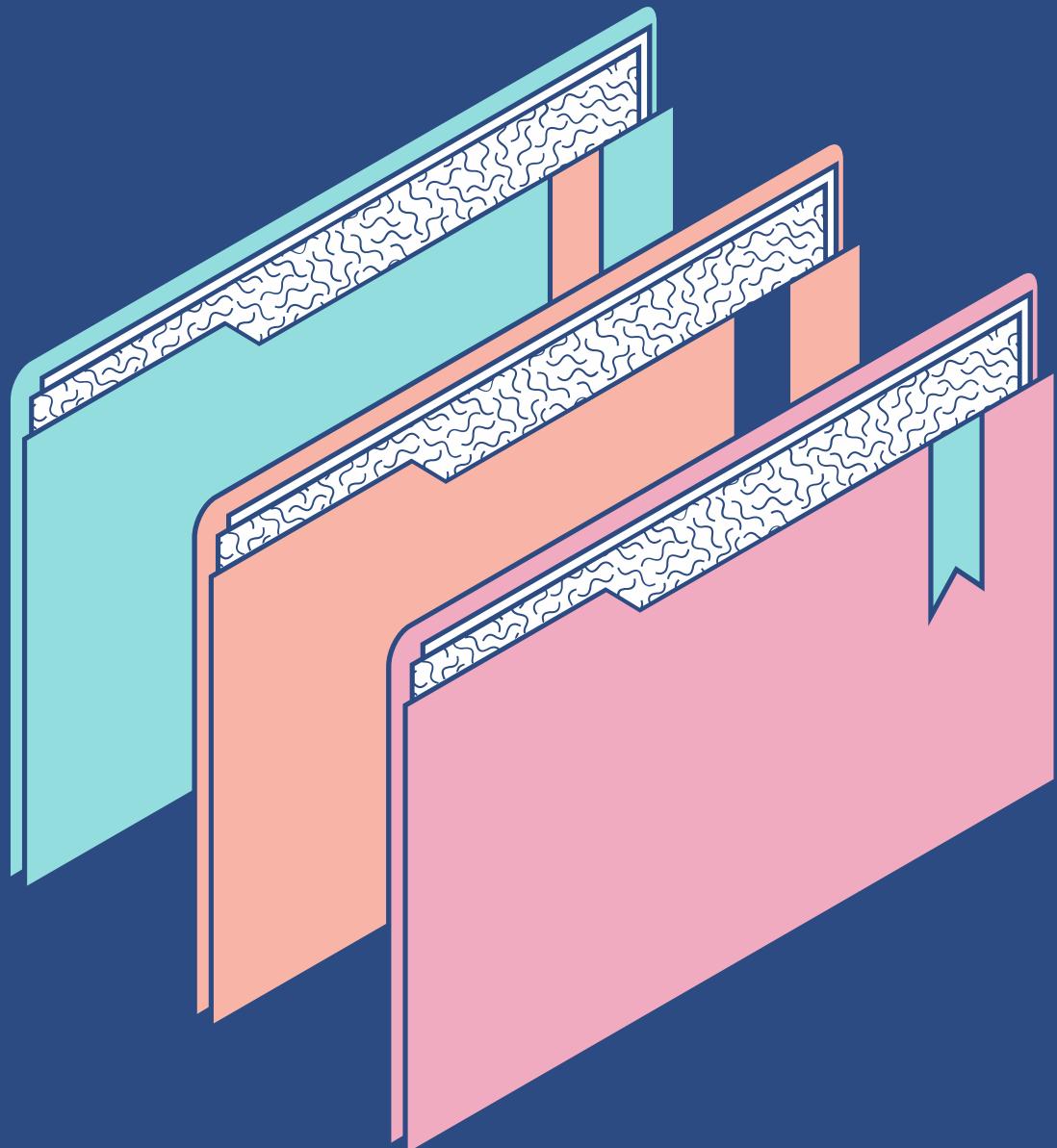
OPERAZIONI
CRUD

4

IMPLEMENTAZIONE
&
ESECUZIONE
QUERIES

PRE-PROCESSING

CREDITS.CSV



1. Presenza di valori null in corrispondenza delle voci inerenti ai registi per l'attributo character.
2. Scissione del file in: *director.csv* e *actor.csv*
3. Rimozione attributo ridondante role dai file creati.
4. In *director.csv*: presenza superflua dell'attributo character.

credits.csv --> director.csv & actor.csv

person_id	id	name	character	role
23433	tm74391	Deems Taylor	Narrator - Narrative Introductions	ACTOR
5910	tm74391	Walt Disney	Mickey Mouse (segment 'The Sorcerer's Apprentice') (voice)	ACTOR
23436	tm74391	Julietta Novis	Soloist (segment 'Ave Maria') (singing voice)	ACTOR
23434	tm74391	Leopold Stokowski	Himself - Conductor of The Philadelphia Orchestra	ACTOR
23441	tm74391	Paul Satterfield		DIRECTOR

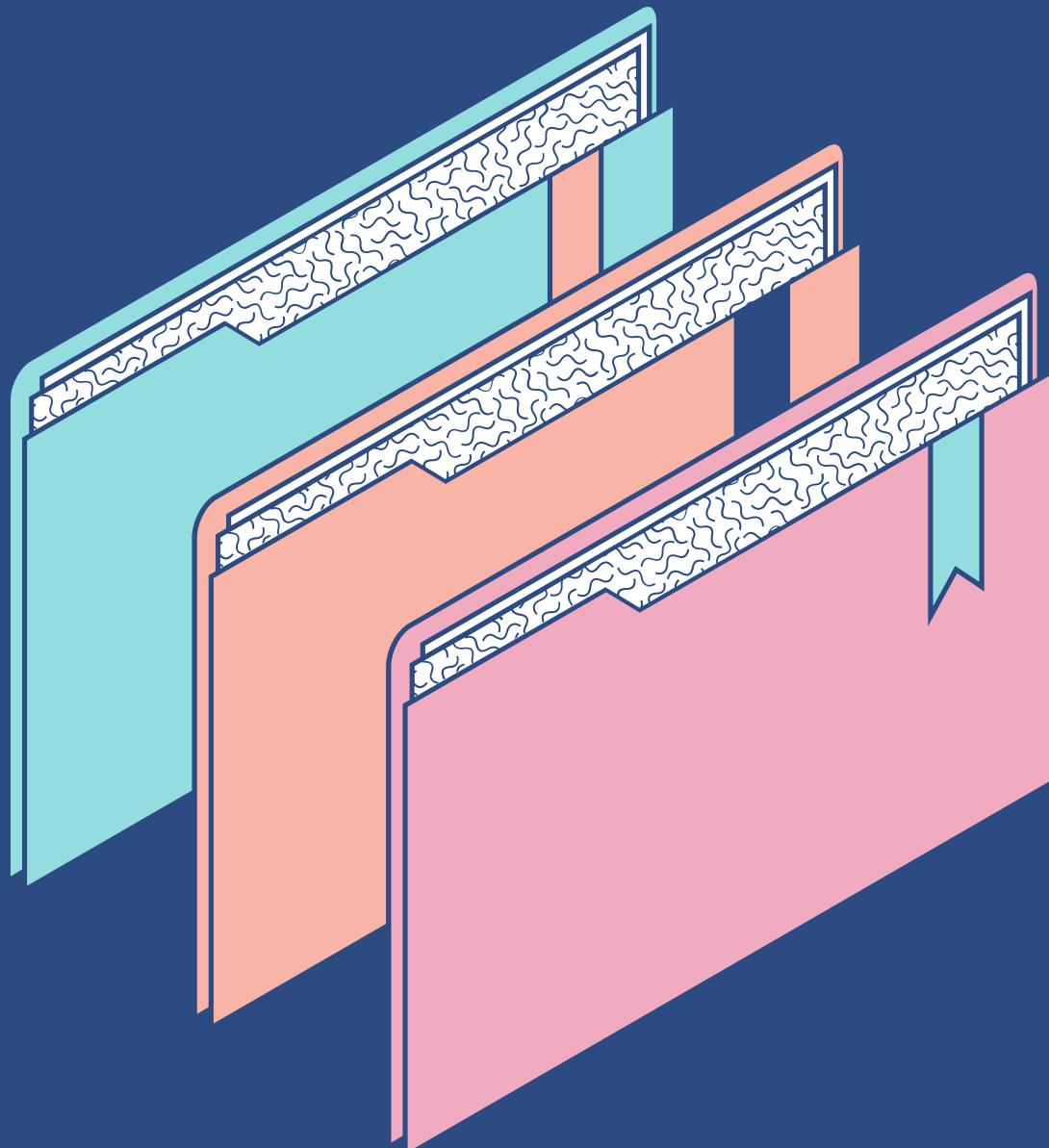


person_id	id	name
23441	tm74391	Paul Satterfield
23442	tm74391	Hamilton Luske
23443	tm74391	Jim Handley
6164	tm74391	Ford Beebe
23444	tm74391	T. Hee
23446	tm74391	Wilfred Jackson
23440	tm74391	William Roberts
13776	tm74391	Ben Sharpsteen
23438	tm74391	James Algar
23437	tm74391	Samuel Armstrong

person_id	id	name	character
23433	tm74391	Deems Taylor	Narrator - Narrative Introductions
5910	tm74391	Walt Disney	Mickey Mouse (segment 'The Sorcerer's Apprentice') (voice)
23436	tm74391	Julietta Novis	Soloist (segment 'Ave Maria') (singing voice)
23434	tm74391	Leopold Stokowski	Himself - Conductor of The Philadelphia Orchestra
6134	tm67803	Adriana Caselotti	Snow White (voice) (uncredited)
6141	tm67803	Lucille La Verne	Queen / Witch (voice) (uncredited)
6142	tm67803	Harry Stockwell	Prince (voice) (uncredited)
6135	tm67803	Roy Atwell	Doc (voice) (uncredited)
6139	tm67803	Pinto Colvig	Sleepy / Grumpy / Dopey (hiccup) (voice) (uncredited)
6140	tm67803	Otis Harlan	Happy (voice) (uncredited)
6136	tm67803	Scotty Mattraw	Bashful (voice) (uncredited)
6137	tm67803	Billy Gilbert	Sneezy (voice) (uncredited)

PRE-PROCESSING

TITLES.CSV



1. Presenza di valori null in corrispondenza delle voci inerenti ai film per l'attributo *seasons*.
2. Scissione del file in: *show.csv* e *movie.csv*
3. Rimozione attributo ridondante *type* dai file creati.
4. In *movie.csv*: presenza superflua dell'attributo *seasons*.

titles.csv --> show.csv & movie.csv



id	title	type	description	release_year	age_certification	runtime	genres	production_countries	seasons	imdb_id	imdb_score	imdb_votes	tmdb_popularity	tmdb_score
tm74391	Fantasia	MOVIE	Walt Disney's ti	1940	G	120	['animation',	['US']		tt0032455	7.7	94681.0	57.751	7.4
tm67803	Snow White a	MOVIE	A beautiful girl,	1937	G	83	['fantasy', 'fa	['US']		tt0029583	7.6	195321.0	107.137	7.1
tm82546	Pinocchio	MOVIE	Lonely toymake	1940	G	88	['animation',	['US']		tt0032910	7.5	141937.0	71.16	7.1
tm79357	Bambi	MOVIE	Bambi's tale un	1942	G	70	['animation',	['US']		tt0034492	7.3	140406.0	68.136	7.0

id	title	description	release_year	age_certification	runtime	genres	production_countries	seasons	imdb_id	imdb_score	imdb_votes	tmdb_popularity	tmdb_score
ts22470	The Muppet	Go behind the cur	1976	TV-G	25	['comedy',	['GB']	5.0	tt0074028	8.4	24278.0	12.640999999999998	8.1
ts30718	Schoolhouse	Schoolhouse Rock!	1973		3	['animati	['US']	7.0	tt0069627	8.2	3905.0	2.987	7.8
ts28051	The Mickey	A variety show fea	1955	TV-PG	31	['family', '	['US']	3.0	tt0047757	7.6	779.0	7.194	5.8
ts24939	Spider-Woma	When Jessica Dre	1979	TV-PG	21	['animati	['US']	1.0	tt0197911	5.8	790.0	19.081	8.0

id	title	description	release_year	age_certification	runtime	genres	production_countries	imdb_id	imdb_score	imdb_votes	tmdb_popularity	tmdb_score
tm74391	Fantasi	Walt Disney's	1940	G	120	['animation	['US']	tt0032455	7.7	94681.0	57.751000000000005	7.4
tm67803	Snow V	A beautiful gir	1937	G	83	['fantasy', '	['US']	tt0029583	7.6	195321.0	107.137	7.1
tm82546	Pinocchio	Lonely toymal	1940	G	88	['animation	['US']	tt0032910	7.5	141937.0	71.16	7.1
tm79357	Bambi	Bambi's tale u	1942	G	70	['animation	['US']	tt0034492	7.3	140406.0	68.13600000000001	7.0

Fasi di sviluppo del progetto



OPERAZIONI CRUD

C-CREATE, U-UPDATE, D-DELETE

C - Inserimento di un attore e di un suo relativo film.

C - Inserimento di un nuovo regista e una nuova serie tv da lui diretta.

U - Aggiornamento di score e numero di stagioni di una determinata serie.

U - Aggiornamento della certificazione d'età per tutti i film prodotti tra il 1900 e il 1950.

D - Eliminazione di un film dato e di tutte le informazioni circa regista e attori ad esso associati.

OPERAZIONI CRUD

R - RETRIEVE WITH USER INPUT

- R - Restituzione delle informazioni di un film con titolo inserito da utente.
- R - Restituzione del numero di stagioni relativo ad una data serie tv.
- R - Restituzione di una serie tv prodotta in un determinato arco di tempo.
- R - Restituzione di tutte le serie tv aventi un numero specifico di stagioni.
- R - Restituzione delle serie tv con specifica certificazione d'età in cui un dato attore ha partecipato.
- R - Restituzione tutti i film diretti da un determinato regista.
- R - Restituzione di tutte le serie tv in cui ha partecipato un determinato attore.
- R - Restituzione tutti i film che un regista ha diretto in un intervallo di tempo dato, di dato genere e imdb_score maggiore di un dato valore.
- R - Restituzione per un regista di tutti i film e le serie tv che ha diretto e i relativi attori.
- R - Restituzione delle serie tv o dei film in cui è presente un determinato attore dove, se si tratta di film, devono durare almeno 30 minuti altrimenti, se si tratta di serie tv, gli episodi devono almeno 30 minuti, sono stati rilasciati dal 2000 in poi negli US e sono di un particolare genere.
- R - Restituzione delle descrizioni di film prodotti in Canada in cui un dato attore ha dato voce a dei personaggi.

OPERAZIONI CRUD

R - RETRIEVE WITHOUT USER INPUT

- R - Restituzione del film che ha tmdb_score 8 e tmbd_popularity 100.
- R - Restituzione di un film di genere animation di durata maggiore di 100 minuti.
- R - Restituzione di tutti gli attori segnati "uncredited".
- R - Restituzione tutte le serie tv e i film prodotti negli US che hanno un imdb_score 7.0 e tmdb_score 8.

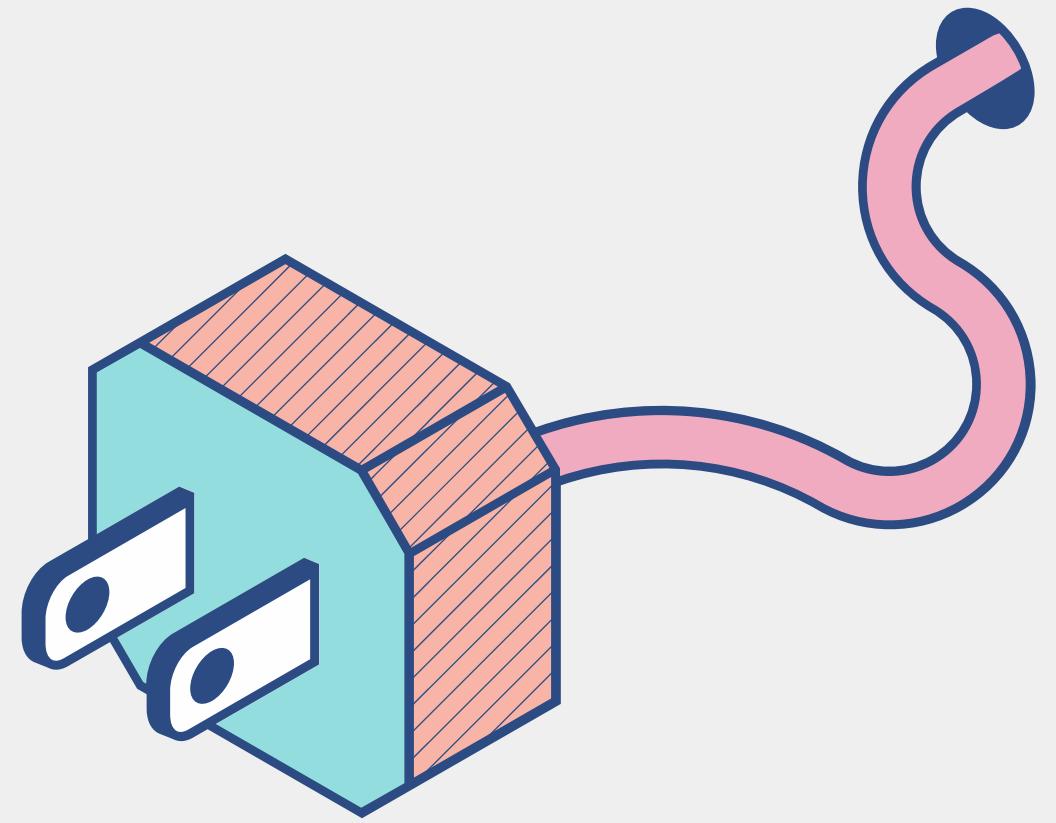
Fasi di sviluppo del progetto



Ora vediamo l'esecuzione del software implementato...

Il DBMS NoSQL utilizzato è MongoDB.

Il software è stato implementato in linguaggio python 3.9 con il supporto della libreria Pymongo; si presenta con una semplice CLI.



Grazie per l'attenzione!