

VISUALIZATION  
TOOLS

# IMDB MOVIES' ETL PIPELINE



Team members:

Isaias Lopez

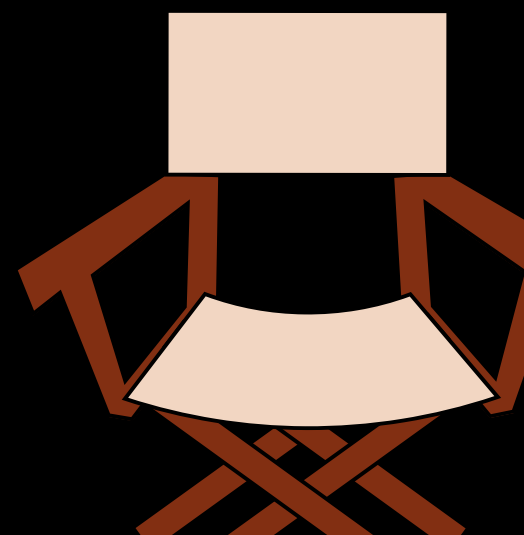
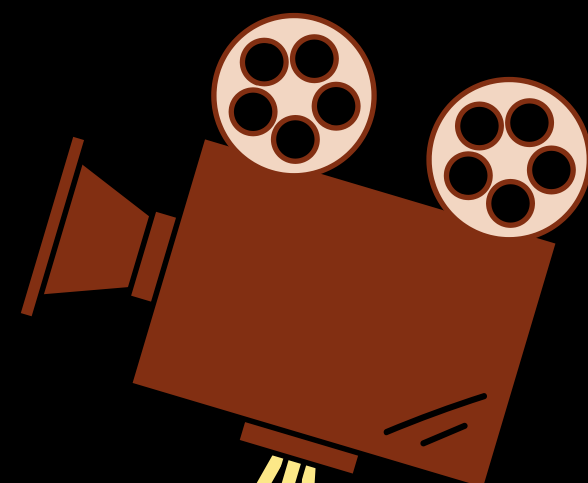
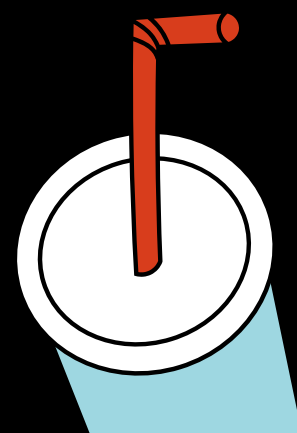
Joaquin Murguia

Valeria Paredes

Damariis Pech

Ana Paula Ramirez

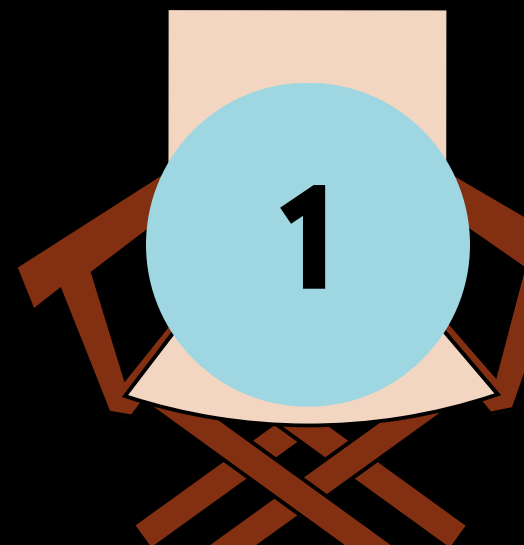
Krishna Sandoval





# IMDB

IMDB is an internet platform that hosts thousands of movies information, reviews and synopsis. When correctly used, the massive amount of information that it generates can shed light into insights that would otherwise be difficult to get. The goal of this project is to build a robust and flexible solution to answer these questions, whose insights can tell us if a movie will be good even before it premieres, or even if a cinema project is worth investing on.



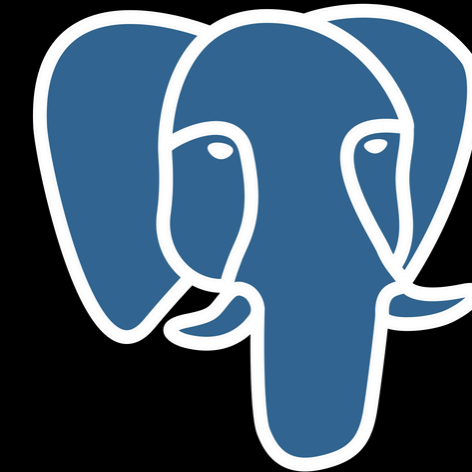
# TOOLS



PYTHON



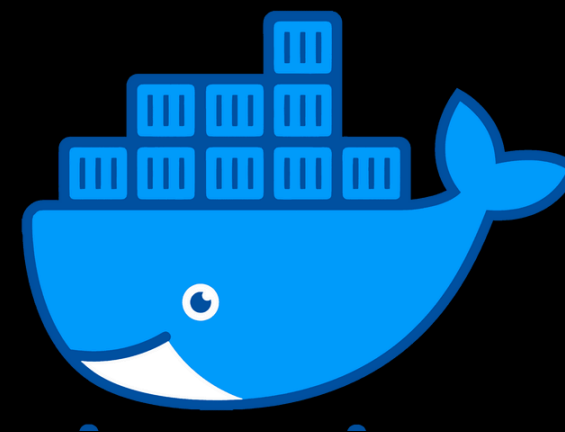
PYTHON  
LIBRARIES



POSTGRESQL



MONGODB



DOCKER



WEB  
DEVELOPMENT

2

# SCHEMA

movie_info		
Key	Field	Data Type
PK	imdb_title_id	str
	title	str
	year	int
	duration	int
	description	str

production_info		
Key	Field	Data Type
PK, FK	imdb_title_id	JSONB
	director	JSONB
	writer	JSONB
	production_company	str
	actors	JSONB
	country	JSONB
	language	JSONB

rating_info		
Key	Field	Data Type
PK, FK	imdb_title_id	str
	avg_vote	float
	votes	int
	reviews_from_users	int
	reviews_from_critics	int

etl_metadata		
Key	Field	Data Type
PK	id	int
	key	str
	value	str
	updated_at	datetime

3

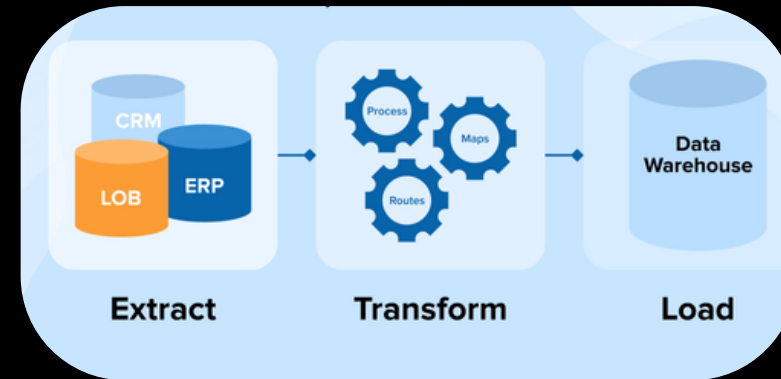
# PROCESS

## 01. Find the dataset

kaggle



## 02. ETL



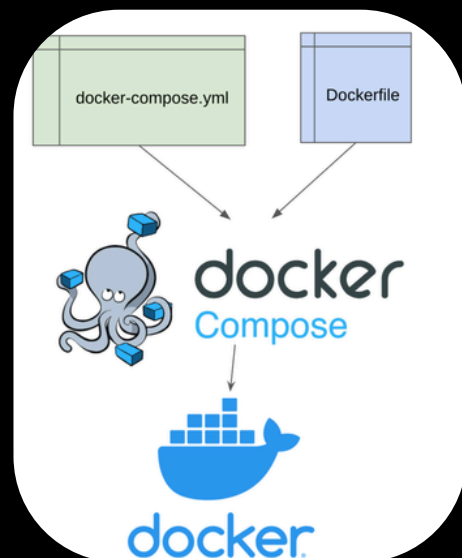
## 03. PostgreSQL



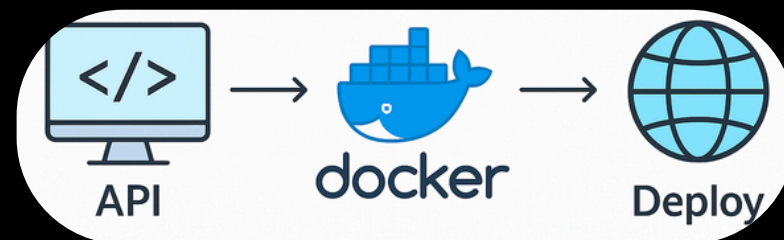
## 04. MongoDB



## 05. Docker



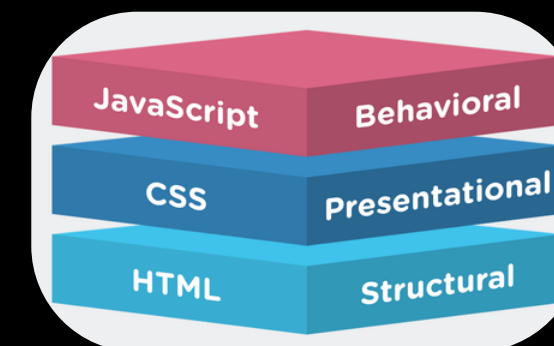
## 06. Fastapi



## 07. Statistics



## 08. Data Visualization



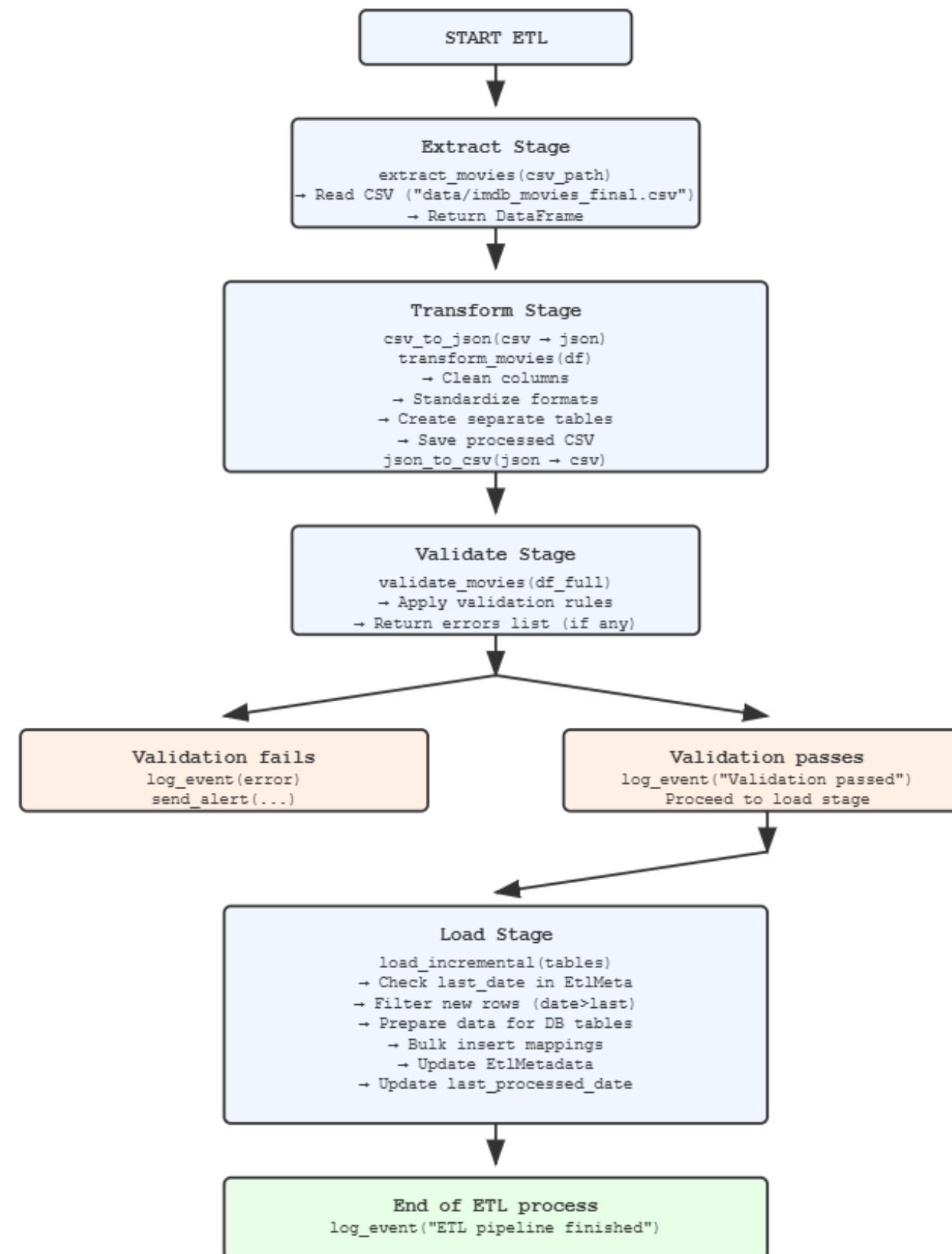
D3.js

## 09. Dashboard





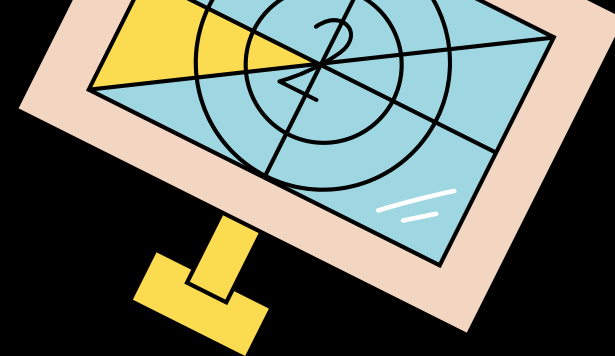
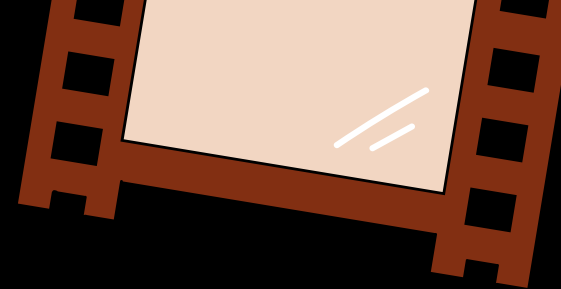
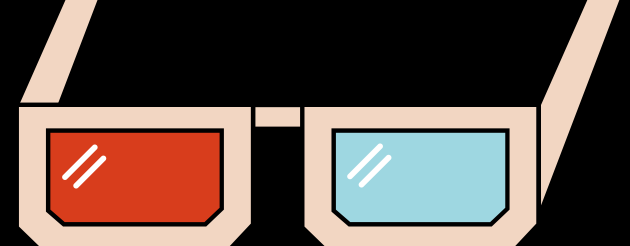
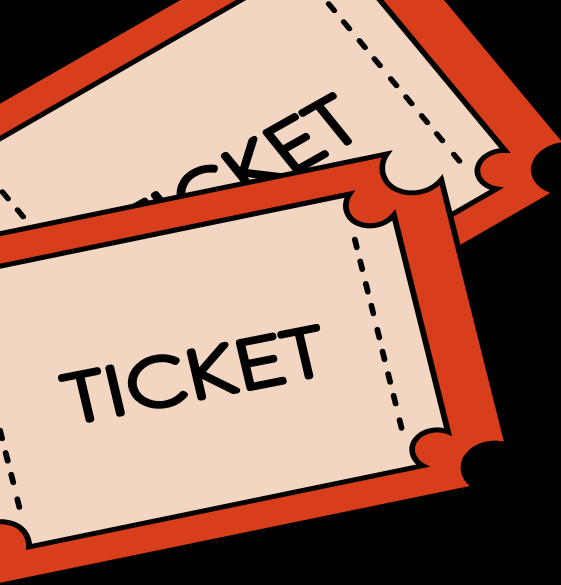
# ETL DIAGRAM



# MONGODB VS POSTGRES

```
status  database  message                               execution_time
-----  -
success PostgreSQL ETL process completed 39.46 seconds
```

```
status      : success
database    : MongoDB
message     : MongoDB ETL process completed
records_loaded : 85736
execution_time : 21.99 seconds
```



THANK YOU FOR YOUR  
ATTENTION

