

**UNIVERSIDADE DE BRASÍLIA**

**Faculdade do Gama**

**Sistemas de Banco de Dados 2**

**Trabalho Final (TF)**

**Extract Transform Load (ETL)**

**Victor Buendia Cruz de Alvim 19/0020601**

**Lucas Ursulino Boaventura 18/0114093**

**Yudi Yamane de Azevedo 16/0149410**

**Brasília, DF**

**2022**

## **a) Contextualização (base de dados + ETL)**

A sigla ETL significa *Extract, Transform and Load* na língua inglesa e pode ser traduzida para o português como Extrair, Transformar e Carregar.

O processo de realização de ETLs tem por principal objetivo realizar a extração de dados de uma ou mais fontes de origem e persistir os dados transformados em um repositório de dados alvo depois de transformá-los fazendo limpezas, padronizações e adequações que garantam a integridade dos dados. Isso é feito para deixar os dados preparados para o uso em análises e relatórios que auxiliem na tomada de decisão da empresa.

Neste trabalho final, escolhemos a base de dados “Popular Movies Datasets - 58098 Movies” para aplicar na prática os conceitos de ETL trabalhando com uma base de dados real. Essa base elenca filmes e atividades relacionadas (avaliações, tagueamentos, categorias...) a eles da plataforma MovieLens entre os anos de 1995 a 2018.

**LINK PARA A BASE DE DADOS:**

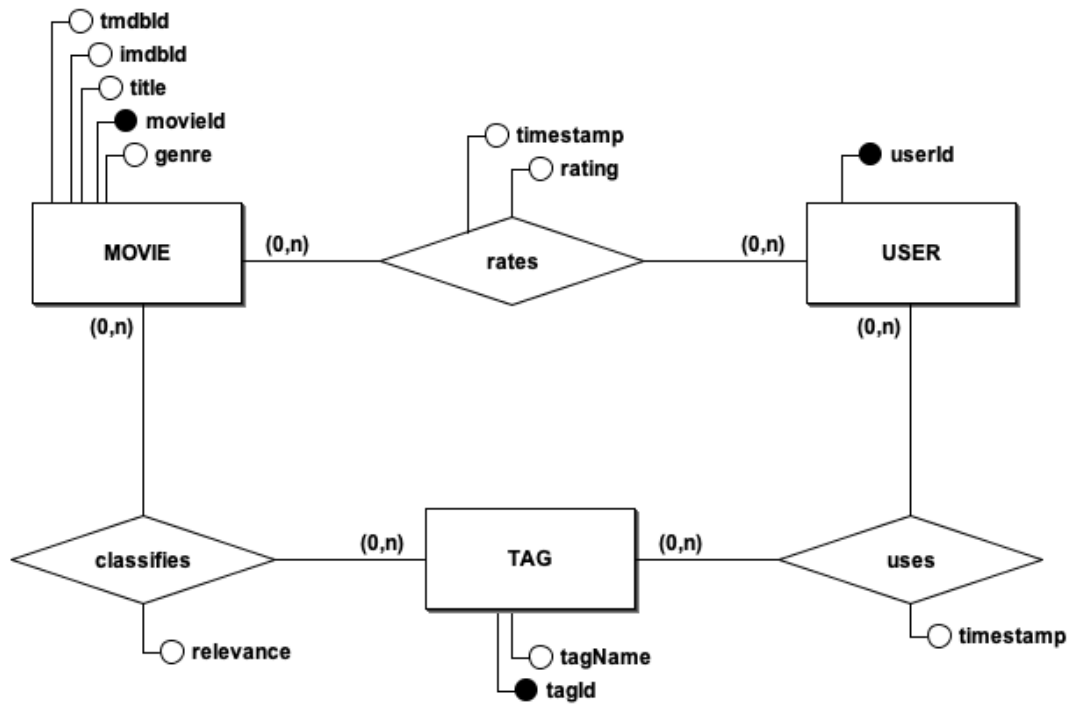
<https://www.kaggle.com/datasets/whenamancodes/popular-movies-datasets-58000-movies?resource=download&select=ratings.csv>

**LINK PARA O REPOSITÓRIO COM O NOTEBOOK EM PYTHON:**

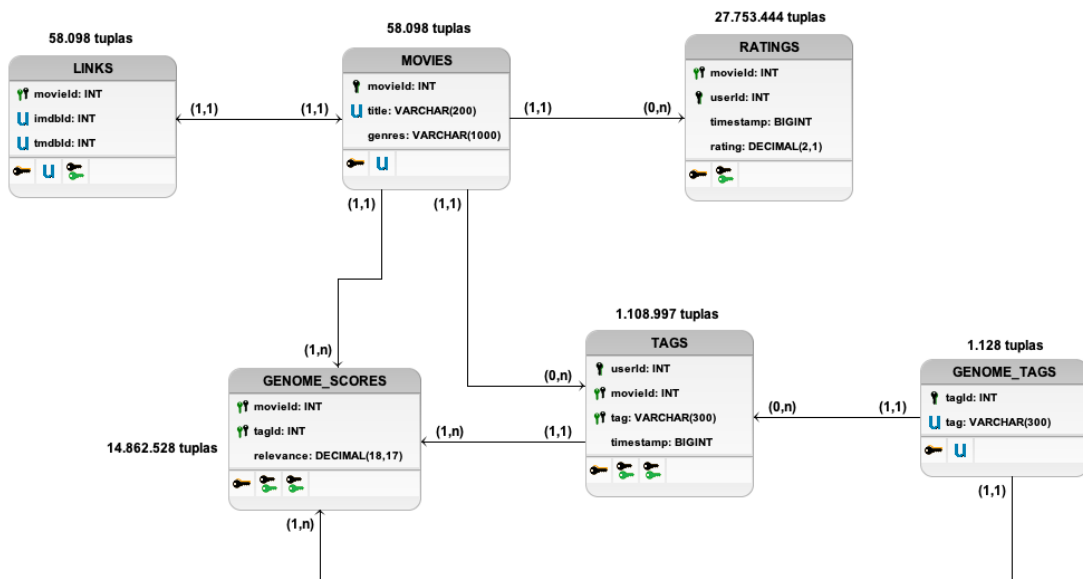
[https://github.com/Victor-Buendia/TF\\_G\\_Pratico](https://github.com/Victor-Buendia/TF_G_Pratico)

**VÍDEO DA APRESENTAÇÃO FINAL:** <https://youtu.be/vectdHPmH4c>

## DER



## DLD



# DICIONÁRIO DE DADOS

ENTIDADE	LINKS			
DESCRIÇÃO	Tabela com correspondência dos IDs dos filmes nas plataformas IMDB e TMDB.			
TUPLAS	58.098			
ATRIBUTO	PROPRIEDADES DO ATRIBUTO	TIPO DE DADO	TAMANHO	DESCRIÇÃO DO ATRIBUTO
movielfd	Chave primária; chave estrangeira; obrigatório.	INT	-	Identificador único do filme.
imdbld	Valor único.	INT	-	Identificador único do filme na plataforma IMDB.
tmdbld	Valor único.	INT	-	Identificador único do filme na plataforma TMDB.

ENTIDADE	MOVIES			
DESCRIÇÃO	Tabela com as informações dos filmes na plataforma.			
TUPLAS	58.098			
ATRIBUTO	PROPRIEDADES DO ATRIBUTO	TIPO DE DADO	TAMANHO	DESCRIÇÃO DO ATRIBUTO
movielfd	Chave primária; obrigatório.	INT	-	Identificador único do filme.
title	Obrigatório; valor único.	VARCHAR	200	Nome do filme.
genres	Obrigatório.	VARCHAR	1000	Todos os gêneros do filme.

ENTIDADE	RATINGS			
DESCRIÇÃO	Tabela com o banco de todas as avaliações feitas aos filmes.			
TUPLAS	27.753.444			
ATRIBUTO	PROPRIEDADES DO ATRIBUTO	TIPO DE DADO	TAMANHO	DESCRIÇÃO DO ATRIBUTO
movielfd	Chave primária; chave estrangeira; obrigatório.	INT	-	Identificador único do filme.

userId	Chave primária; obrigatório.	INT	-	Identificador único do usuário que avaliou o filme.
timestamp	Obrigatório.	BIGINT	-	Número UNIX que indica o momento em que o filme foi avaliado.
rating	Obrigatório.	DECIMAL(2,1)	-	Avaliação do usuário.

ENTIDADE	GENOME_SCORES			
DESCRIÇÃO	Tabela com a relevância de cada filme dentro das tags que é classificado.			
TUPLAS	14.862.528			
ATRIBUTO	PROPRIEDADES DO ATRIBUTO	TIPO DE DADO	TAMANHO	DESCRIÇÃO DO ATRIBUTO
moviedId	Chave primária; chave estrangeira; obrigatório.	INT	-	Identificador único do filme.
tagId	Chave primária; chave estrangeira; obrigatório.	INT	-	Identificador único da tag de classificação do filme.
relevance	Obrigatório.	DECIMAL(18,17)	-	Relevância do filme na tag.

ENTIDADE	TAGS			
DESCRIÇÃO	Tabela com os tagueamentos dos filmes pelos usuários.			
TUPLAS	1.108.997			
ATRIBUTO	PROPRIEDADES DO ATRIBUTO	TIPO DE DADO	TAMANHO	DESCRIÇÃO DO ATRIBUTO
moviedId	Chave primária; chave estrangeira; obrigatório.	INT	-	Identificador único do filme.
userId	Chave primária; obrigatório.	INT	-	Identificador único do usuário que avaliou o filme.
tag	Chave primária; chave estrangeira;	VARCHAR	300	Tag usada para classificar o filme.

	obrigatório.			
timestamp	Obrigatório.	BIGINT	-	Número UNIX que indica o momento em que o filme foi classificado com a tag.

<b>ENTIDADE</b>	GENOME_TAGS			
<b>DESCRIÇÃO</b>	Tabela com a identificação e nome das tags disponíveis.			
<b>TUPLAS</b>	1.128			
<b>ATRIBUTO</b>	<b>PROPRIEDADES DO ATRIBUTO</b>	<b>TIPO DE DADO</b>	<b>TAMANHO</b>	<b>DESCRIÇÃO DO ATRIBUTO</b>
tag	Valor único; obrigatório.	VARCHAR	300	Tag usada para classificar o filme.
tagId	Chave primária; obrigatório.	INT	-	Identificador único da tag de classificação do filme.