

Disciplina: Sistemas de Banco de Dados 2	Turma: A
Professora: Vandor Rissoli	Período: 2º/2019
Alunos: Welison Lucas Almeida Regis Lieverton Santos Silva	Matrícula: 17/0024121 17/0039251

TRABALHO FINAL - DATASET DE FILMES

1. BASE DE DADOS

A base de dados escolhida se trata de um dataset com filmes que tiveram lançamento anterior ao ano de 2017. O dataset é composto por filmes com descrições sobre título, gênero, visão geral, elenco, equipe, data de lançamento, duração, popularidade etc. A base de dados possui 26 milhões de registros de classificações de 270.000 usuários para os 45.000 filmes cadastrados. As classificações estão em uma escala de 1 a 5.

Base de dados disponível em: <https://www.kaggle.com/rounakbanik/the-movies-dataset>

2. JUSTIFICATIVA

Escolheu-se a **base da equipe 1**. Tem-se como principais motivos para escolha da base: possui os arquivos “.csv” separados individualmente por “entidade”, o que a torna mais maleável; tem muitos atributos interessantes em cada um de seus arquivos, alguns com tipos de dados diferentes aos comuns, como os jsons; possui uma quantidade adequada de dados, o que impulsiona a finalidade das atividades desenvolvidas no tema do trabalho final; o dataset tem uma boa consistência de dados em geral, pois há raros casos de tuplas com falta de informações relevantes.

1. DIAGRAMA ENTIDADE-RELACIONAMENTO (DE-R)

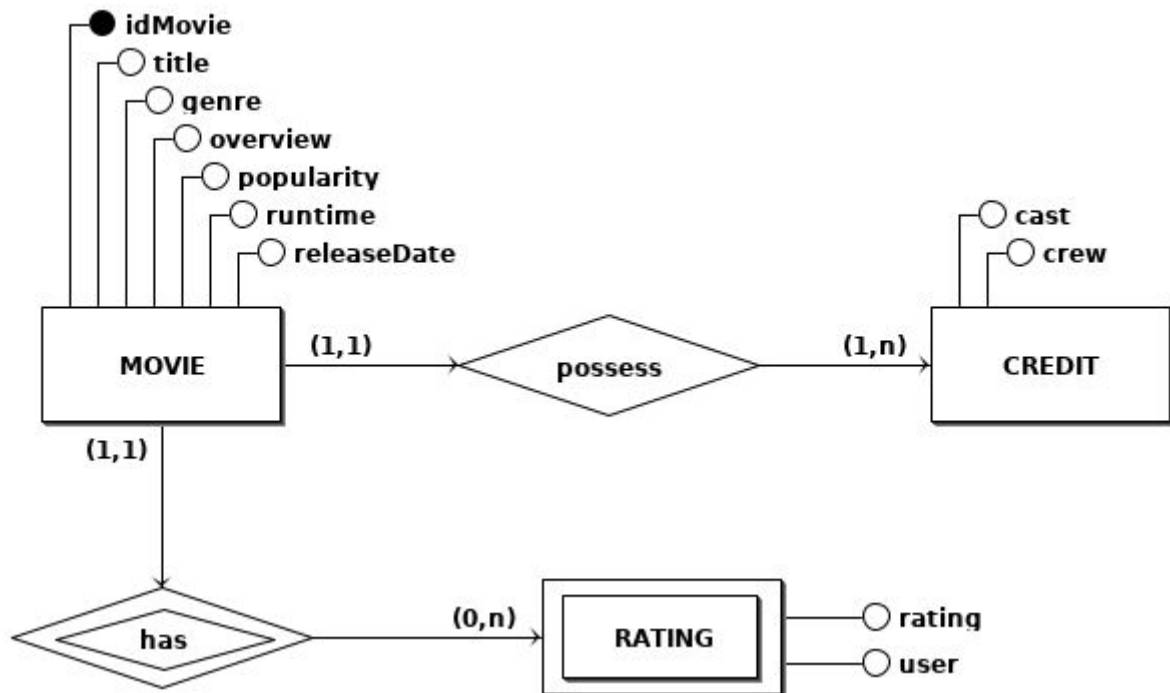


Figura 1 - Diagrama entidade-relacionamento do projeto. Fonte: do autor, 2019.

2. DIAGRAMA LÓGICO (DL)

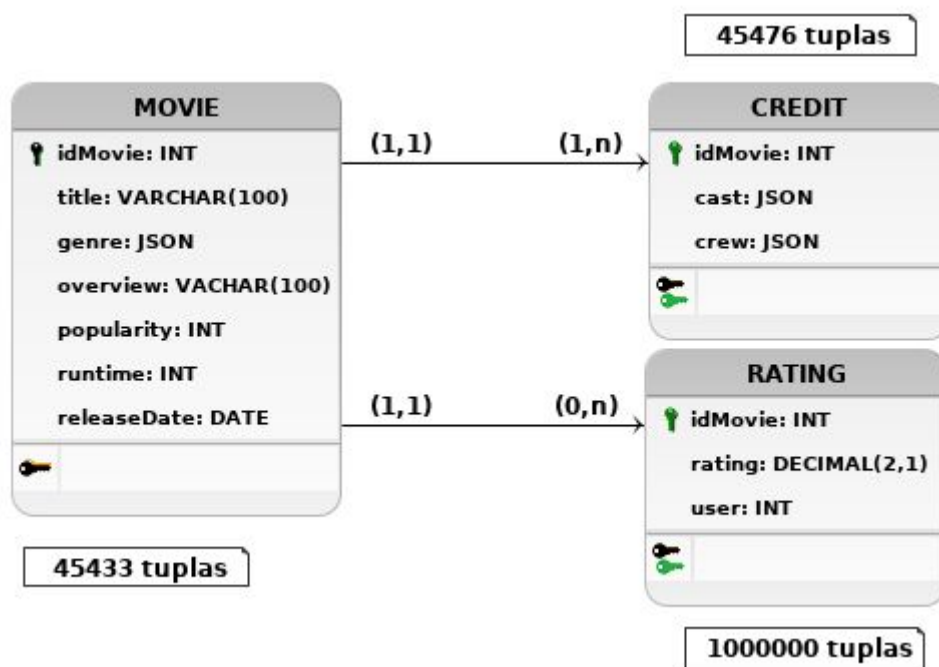


Figura 2 - Diagrama lógico do projeto. Fonte: do autor, 2019.

3. DICIONÁRIO DE DADOS

Entidade: MOVIE				
Descrição: responsável por armazenar informações descritivas sobre os filmes.				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
idMovie	chave primária; obrigatório.	INT	-	Identifica unicamente um filme na base
title	obrigatório.	VARCHAR	110	Título do filme em inglês
genre	obrigatório.	JSON	-	Gêneros ao qual o filme pertence
overview	obrigatório.	VARCHAR	1000	Sinopse do filme
popularity	obrigatório.	DECIMAL	11, 7	Popularidade de bilheteria do filme
runtime	obrigatório.	INT	-	Duração do filme
releaseDate	obrigatório.	DATE	-	Data de lançamento do filme.

Entidade: CREDIT				
Descrição: responsável por armazenar o elenco e a equipe para os filmes da base.				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
idMovie	chave estrangeira; obrigatório.	INT	-	Identifica unicamente um filme na base
cast	obrigatório.	JSON	-	Elenco de dado filme
crew	obrigatório.	JSON	-	Equipe técnica de dado filme

Entidade: RATING				
Descrição: responsável por armazenar as avaliações de diferentes usuários para os filmes disponíveis.				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
idMovie	chave estrangeira; obrigatório.	INT	-	Identifica unicamente um filme na base
rating	obrigatório.	DECIMAL	2, 1	Avaliação do usuário, que varia de 1 a 5
user	obrigatório.	INT	-	Identifica unicamente o usuário que realizou a avaliação.