

Análise dos Dados do Dataset IMDb

Introdução a Banco de Dados

Prof. Dr. Sávio Salvarino Teles de Oliveira

Iury Tavares do Monte – 2025200234

João Pedro José Santos da Silva Guedes – 2025200237

Manuel Lucala Zengo – 2025200249

Owen Alves Lima – 2025200255

Pedro Felipe de Moraes Carrijo – 2025200257

Willgnner Ferreira Santos – 2025200271

INF



UFG
UNIVERSIDADE
FEDERAL DE GOIÁS



PODER JUDICIÁRIO
Tribunal de Justiça do Estado de Goiás



Programa de
Residência em TI

Introdução

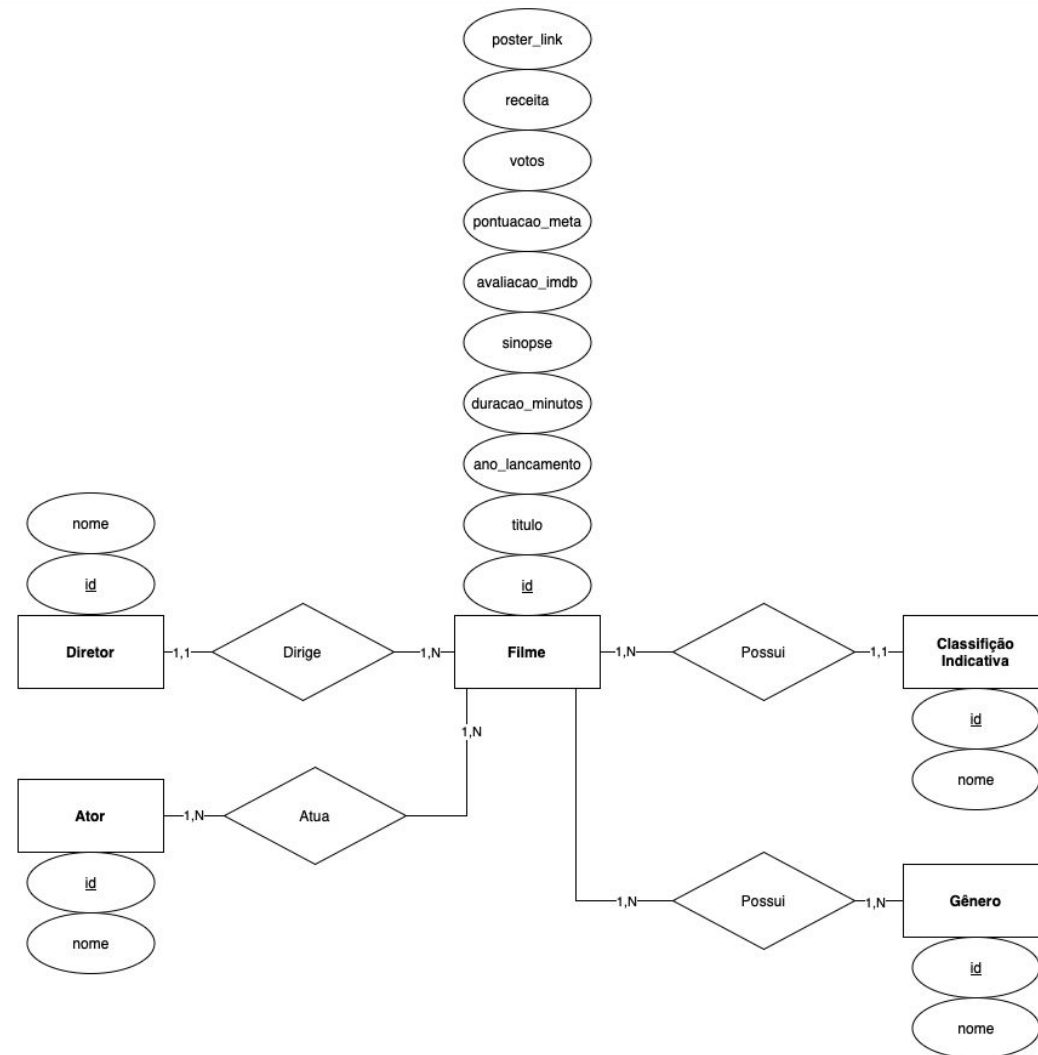
- **Objetivo:** Criar um banco de dados relacional normalizado e um Data Warehouse para análise de dados do IMDb.
- **Dataset:** IMDb Top 1000 filmes
- **Fonte:** Kaggle
(<https://www.kaggle.com/datasets/sonalisingh1411/imdb-top-1000-movies>)

Caso de Uso

- **Domínio:** Plataforma de streaming ou análise cinematográfica
- **Propósito:** Avaliar filmes com base em dados agregados (nota, receita, gênero, diretor, etc.)
- **Motivador:** Tomada de decisão sobre tendências e desempenho de filmes

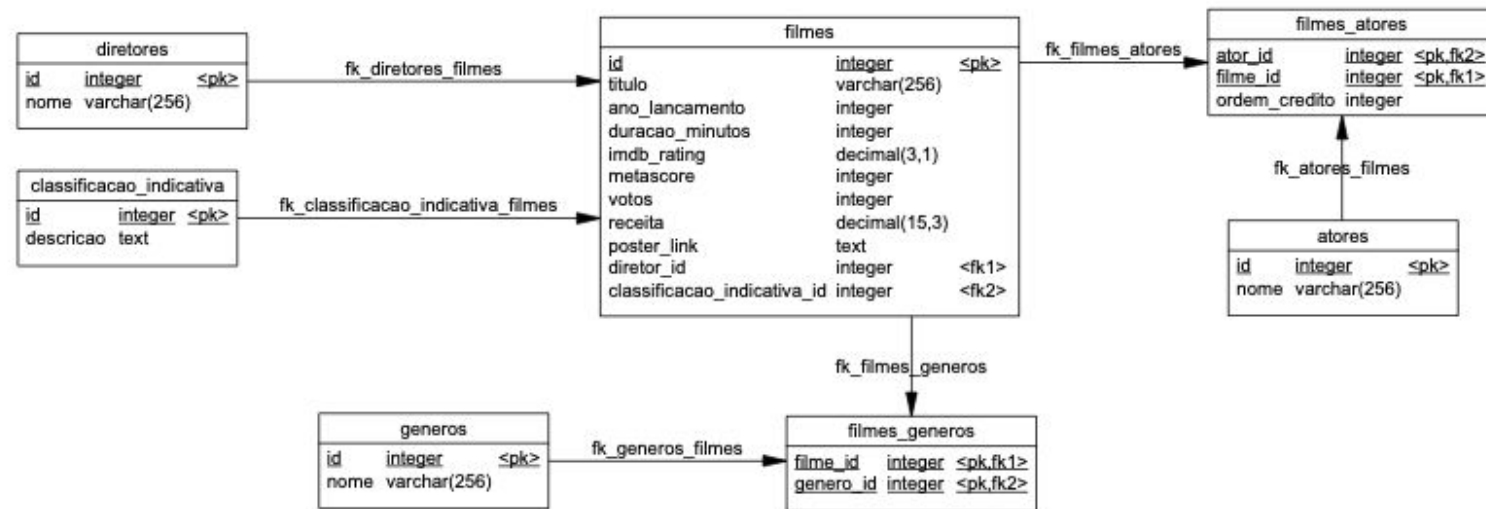
Modelo Entidade-Relacionamento

- Criado a partir do dataset;
- Definição das entidades: Filme, Ator, Diretor, Classificação indicativa e Gênero.
- Modelado com foco na normalização;



Esquema Relacional

- Criação de tabelas associativas;
- Normalização até a 3FN para reduzir redundância e garantir a integridade dos dados.



Consultas SQL

1. Contar quantos filmes existem por cada classificação indicativa

```
SELECT c.descricao AS classificacao,  
COUNT(f.id) AS total_filmes  
FROM filmes f  
JOIN classificacao_indicativa c ON  
f.classificacao_indicativa_id = c.id  
GROUP BY c.descricao  
ORDER BY total_filmes DESC;
```

classificacao	total_filmes
U	234
A	197
UA	175
R	146
PG-13	43
PG	37
Passed	34
G	12
Approved	11
TV-PG	3
GP	2
16	1
Unrated	1
TV-14	1
U/A	1
TV-MA	1
(16 rows)	

Consultas SQL

2. Listar os atores de um filme específico
(Ex.: "The Dark Knight")

```
SELECT f.titulo, a.nome AS ator,  
fa.ordem_credito  
FROM filmes f
```

```
JOIN filmes_atores fa ON f.id = fa.filme_id
```

```
JOIN atores a ON fa.ator_id = a.id
```

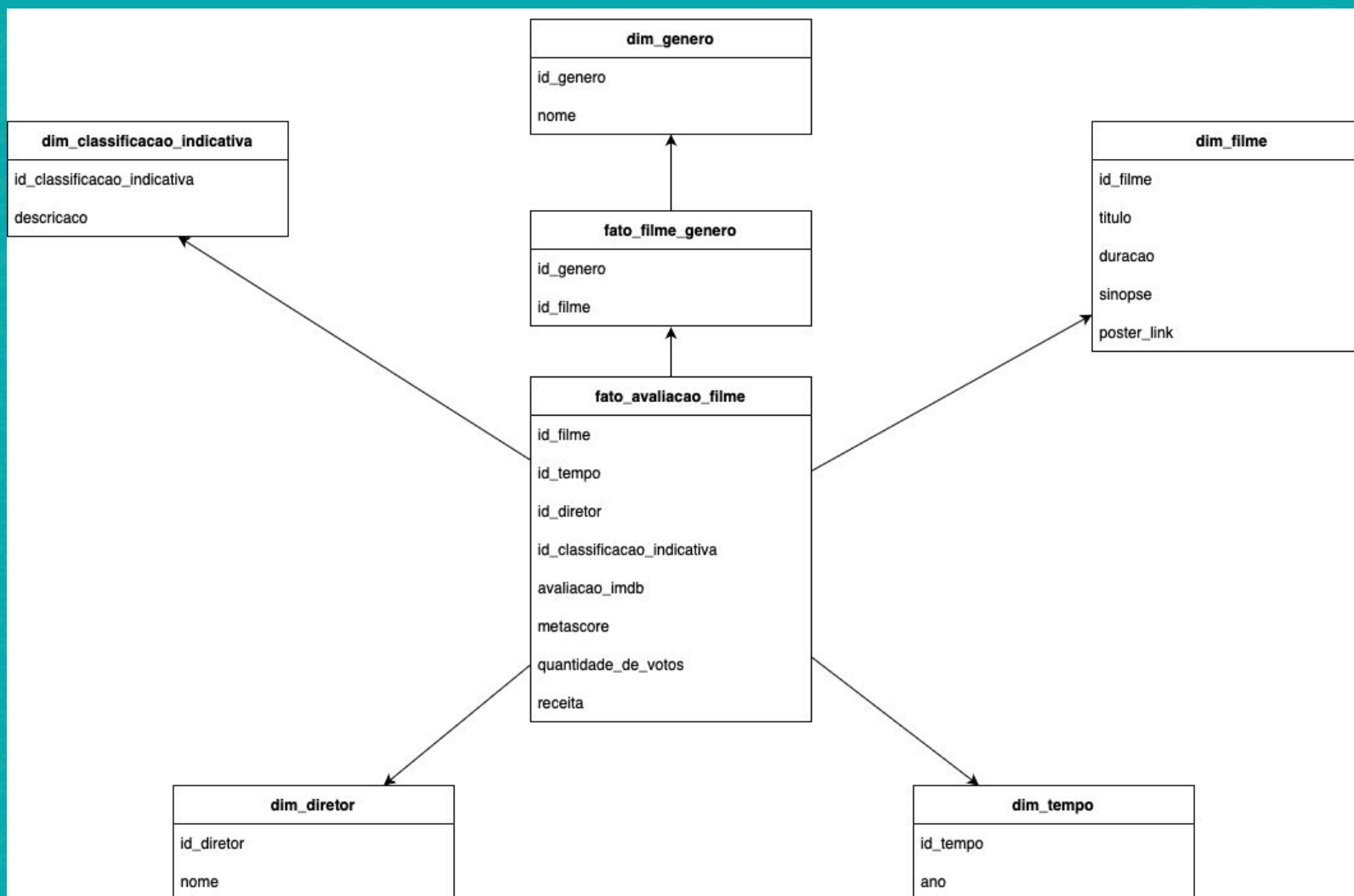
```
WHERE f.titulo = 'The Dark Knight'
```

```
ORDER BY fa.ordem_credito;
```

titulo	ator	ordem_credito
The Dark Knight	Christian Bale	1
The Dark Knight	Heath Ledger	2
The Dark Knight	Aaron Eckhart	3
The Dark Knight	Michael Caine	4

(4 rows)

Esquema Estrela



Consultas OLAP (DW)

1. Média IMDB por gênero

```
SELECT g. nome AS genero,  
ROUND (AVG (f.avaliacao_imdb), 2) AS media_imdb  
FROM fato_avaliacao_filme f  
JOIN fato_filme_genero fg ON f.id_filme = fg.id_filme  
JOIN dim_genero g ON fg.id_genero = g.id_genero  
GROUP BY g. nome  
ORDER BY media_imdb DESC;
```

Result 1 (21)		
<input type="checkbox"/>	genero	media_imdb
<input type="checkbox"/>	Family	8.7
<input type="checkbox"/>	Fantasy	8.6
<input type="checkbox"/>	Adventure	8.6
<input type="checkbox"/>	Musical	8.6
<input type="checkbox"/>	Animation	8.6
<input type="checkbox"/>	Horror	8.6
<input type="checkbox"/>	Mystery	8.6
<input type="checkbox"/>	Film-Noir	8.6
<input type="checkbox"/>	History	8.6
<input type="checkbox"/>	Comedy	8.5
<input type="checkbox"/>	Biography	8.5

Decisões de Projeto

- Escolha do domínio de negócio que todo o grupo fosse familiar;
- Geração automatizada de scripts;
- Utilização do Amazon Redshift.

Conclusão e Aprendizados

- Prática de modelagem conceitual e lógica;
- Experiência com SQL e manipulação de dados reais;
- Aprendizado sobre modelagem dimensional e análise OLAP;
- Aprendizado sobre RedShift e suas tipagens;
- Apoio da IA (ChatGPT) para geração de scripts e estruturação.

Obrigado pela Atenção!

INF

INSTITUTO DE
INFORMÁTICA



UFG
UNIVERSIDADE
FEDERAL DE GOIÁS



PODER JUDICIÁRIO

Tribunal de Justiça do Estado de Goiás



Programa de
Residência em TI