

# SIMULAÇÃO BANCO DE DADOS NETFLIX

04

Sergio Soares RA: 160105

Thiago Vieira RA: 171196

CP904TIN2

Prof. Marco Antonio Montebello Júnior



## INTRODUÇÃO

O objetivo deste trabalho é simular o banco de dados da Netflix utilizando dados reais fornecidos pela plataforma. Analisamos dados de 2010 a 2021, com o intuito de entender melhor as dinâmicas de armazenamento e recuperação de informações em um ambiente de grande escala. Utilizamos o banco de dados Cassandra e Docker para a implementação da simulação e realização de consultas.

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, film...
1	s2	TV Show	Blood & Water	NaN	Arma Qamata, Kizad Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabil...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, tiratations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

Figura 1. Dataset oficial Netflix.

## JUSTIFICATIVA

Com o crescimento exponencial do volume de dados, compreender as técnicas de armazenamento e recuperação eficientes é crucial. A Netflix, sendo uma das maiores empresas de streaming, oferece um caso de estudo ideal para práticas de banco de dados em larga escala. Este projeto nos permite aplicar e expandir nossos conhecimentos em tecnologias de banco de dados NoSQL, especificamente Cassandra, além de entender a infraestrutura necessária para suportar um sistema de streaming.

## OBJETIVOS

- Importar e estruturar um dataset real da Netflix no Cassandra.
- Configurar um ambiente de banco de dados utilizando Docker.
- Realizar consultas no banco de dados simulado para validar a integridade e eficiência do sistema.
- Comparar o desempenho e as funcionalidades do banco de dados simulado com a estrutura de banco de dados real da Netflix.

## RESULTADOS E VALIDAÇÃO

**Estrutura de Dados:** Criamos uma estrutura de tabela que refletiu as principais necessidades de armazenamento e recuperação da Netflix.

**Consultas:** Realizamos diversas consultas para verificar a consistência, disponibilidade e particionamento dos dados.

**Desempenho:** Os testes de desempenho mostraram que Cassandra é eficaz para o armazenamento e recuperação de grandes volumes de dados, mantendo a alta disponibilidade e escalabilidade.

**Validação:** Os resultados das consultas foram validados comparando com os dados esperados, confirmando a integridade da simulação.

## ORÇAMENTO

Serviços de Nuvem	R\$ 100,00
Recursos de Hardware	R\$ 50,00
Licenças e Softwares	R\$ 50
Total	R\$ 350

Tabela 1. Orçamento simulado.

## CONCLUSÃO

A simulação do banco de dados da Netflix utilizando Cassandra demonstrou a viabilidade e eficiência de bancos de dados NoSQL para grandes volumes de dados. O uso do Docker facilitou a configuração e gerenciamento do ambiente, tornando a simulação mais acessível e reproduzível. Este projeto reforçou a importância de tecnologias modernas de banco de dados e abriu caminho para futuras explorações em escalabilidade e performance.

## AGRADECIMENTOS

Agradecemos especialmente ao prof. Marco Montebello por todo auxílio e orientação.