

UNIVERSIDADE DE BRASÍLIA  
Faculdade do Gama

Sistemas de Banco de Dados 2

**Trabalho Final**  
Bancos de Dados em Colunas

Sara Campos - 17/0045269

Brasília, DF  
2023

## **Definição**

Bancos de dados colunares, também conhecidos como bancos de dados orientados a colunas, são sistemas de gerenciamento de banco de dados (SGBDs) que organizam os dados em colunas, em oposição aos bancos de dados relacionais tradicionais, que organizam os dados em linhas.

Ao contrário dos bancos de dados relacionais, onde as linhas contêm informações relacionadas, os bancos de dados colunares armazenam valores semelhantes juntos em uma coluna, independentemente das linhas em que eles aparecem. Isso resulta em uma compactação melhor dos dados e pode levar a uma redução significativa no espaço de armazenamento necessário.

No entanto, é importante observar que os bancos de dados colunares podem não ser adequados para todos os cenários. Eles tendem a ser mais eficientes em consultas de leitura do que em operações de gravação ou atualização de dados. Além disso, a modelagem e o design de um banco de dados colunar são diferentes dos bancos de dados relacionais tradicionais, exigindo considerações específicas ao desenvolver e otimizar a estrutura de dados.

## **Objetivos**

Os bancos de dados colunares visam melhorar o desempenho, eficiência de armazenamento e flexibilidade das consultas analíticas, permitindo uma análise eficaz de grandes volumes de dados, atendendo às necessidades de aplicativos nos quais esses fatores são fundamentais.

Os principais objetivos dos bancos de dados colunares são:

1. Melhor desempenho em consultas analíticas: Os bancos de dados colunares são otimizados para consultas analíticas, especialmente aquelas que envolvem operações de agregação e análise de grandes volumes de dados. Ao armazenar os valores de cada coluna de forma contígua, é possível acessar apenas as colunas necessárias para uma consulta específica, reduzindo a quantidade de dados lidos e melhorando o desempenho das consultas.

2. **Eficiência de armazenamento:** Os bancos de dados colunares geralmente têm uma melhor taxa de compressão em comparação com os bancos de dados relacionais, devido à organização dos dados por coluna. Valores semelhantes são armazenados juntos, permitindo uma melhor compressão de dados e, conseqüentemente, a redução do espaço de armazenamento necessário.
3. **Escalabilidade e paralelismo:** Os bancos de dados colunares são escaláveis horizontalmente, o que significa que podem lidar com grandes volumes de dados distribuídos em várias máquinas. Além disso, eles se beneficiam do paralelismo, pois consultas podem ser executadas simultaneamente em diferentes colunas, resultando em um melhor desempenho em ambientes de processamento distribuído.
4. **Integração com ferramentas de análise:** Os bancos de dados colunares são frequentemente usados em conjunto com ferramentas de análise e business intelligence. Eles podem ser integrados facilmente com sistemas de visualização de dados, aplicativos analíticos e outras ferramentas que exigem acesso rápido e eficiente aos dados.

## **Vantagens e desvantagens**

Os bancos de dados colunares têm se destacado como uma alternativa eficiente e poderosa para o armazenamento e análise de dados em cenários analíticos, proporcionando benefícios significativos para consultas complexas e operações de análise. No entanto, assim como qualquer tecnologia, os bancos de dados colunares apresentam vantagens e desvantagens a serem consideradas.

### **Vantagens:**

1. **Desempenho aprimorado para consultas analíticas:** Os bancos de dados colunares são otimizados para consultas analíticas, permitindo um desempenho superior em operações de agregação, filtragem e análise de dados. Ao armazenar os valores de cada coluna de forma contígua, o acesso seletivo às colunas necessárias é mais rápido, resultando em consultas mais eficientes.

2. Compressão de dados: A organização por colunas nos bancos de dados colunares permite uma melhor taxa de compressão de dados em comparação com os bancos de dados relacionais. Valores semelhantes são armazenados juntos, o que facilita a compressão e reduz o espaço de armazenamento necessário.
3. Escalabilidade: Os bancos de dados colunares são altamente escaláveis horizontalmente, o que significa que podem lidar com grandes volumes de dados distribuídos em várias máquinas. Isso permite que os bancos de dados colunares cresçam facilmente à medida que a quantidade de dados aumenta.
4. Integração com ferramentas de análise: Os bancos de dados colunares podem ser facilmente integrados a ferramentas de análise e business intelligence, como sistemas de visualização de dados e aplicativos analíticos. Essa integração permite a análise rápida e eficiente dos dados armazenados nos bancos de dados colunares.

#### Desvantagens:

1. Desempenho inferior para operações de gravação: Enquanto os bancos de dados colunares são otimizados para consultas analíticas, eles podem ter um desempenho inferior em operações de gravação e atualização de dados. A estrutura de armazenamento por coluna pode tornar as operações de gravação mais lentas em comparação com os bancos de dados relacionais.
2. Complexidade do design: O design e a modelagem de bancos de dados colunares podem ser mais complexos do que os bancos de dados relacionais tradicionais. A estrutura de armazenamento por coluna pode exigir considerações específicas ao projetar o esquema do banco de dados, o que pode aumentar a complexidade de desenvolvimento e manutenção.
3. Requerimento de conhecimentos especializados: Para aproveitar ao máximo os benefícios dos bancos de dados colunares, pode ser necessário ter conhecimentos especializados em sua modelagem, otimização e uso eficiente. Isso pode exigir recursos adicionais de treinamento e habilidades técnicas específicas.

4. Limitações em cenários transacionais: Os bancos de dados colunares geralmente são mais adequados para cargas de trabalho analíticas e consultas ad hoc complexas. Em cenários transacionais com altos volumes de inserções, atualizações e exclusões de dados, os bancos de dados relacionais tradicionais podem ser mais eficientes. Os bancos de dados colunares podem não ter o mesmo desempenho em transações devido à sua estrutura de armazenamento e otimização focada em consultas analíticas.

## **Histórias**

Uma história interessante de uso de bancos de dados colunares é o caso do Pinterest, uma plataforma de compartilhamento de imagens e descoberta visual. O Pinterest enfrentou desafios em lidar com grandes quantidades de dados gerados pelos usuários, incluindo imagens, descrições, comentários e interações.

Para otimizar a análise de dados e melhorar o desempenho das consultas, o Pinterest adotou um banco de dados colunar chamado Apache Druid. O Apache Druid foi projetado especificamente para consultas analíticas em tempo real e análise exploratória de dados.

Ao utilizar o banco de dados colunar, o Pinterest conseguiu realizar consultas complexas e interativas em grandes volumes de dados de maneira mais eficiente. Isso permitiu a visualização em tempo real de dados de engajamento do usuário, como métricas de compartilhamento, cliques e interações com os Pins.

Com a implantação do banco de dados colunar, o Pinterest conseguiu acelerar a análise de dados, fornecer recomendações de conteúdo mais relevantes e aprimorar a experiência geral do usuário. Além disso, a capacidade de escalabilidade horizontal do banco de dados colunar permitiu que o Pinterest lidasse com o crescimento exponencial dos dados gerados pelos usuários.

Esse caso demonstra como o uso de bancos de dados colunares pode trazer benefícios reais para empresas que lidam com grandes volumes de dados e precisam realizar análises rápidas e interativas. O Pinterest aproveitou as vantagens do banco de dados colunar para obter insights valiosos, aprimorar sua plataforma e proporcionar uma experiência atraente aos usuários.

Outro exemplo real do uso de bancos de dados colunares é o caso da empresa de streaming de música Spotify. O Spotify enfrentava o desafio de processar e

analisar rapidamente grandes volumes de dados de streaming de música de milhões de usuários em todo o mundo.

Para lidar com essa demanda, o Spotify optou por implementar um banco de dados colunar chamado Apache Cassandra. O Apache Cassandra é projetado para lidar com grandes volumes de dados distribuídos em várias máquinas, fornecendo escalabilidade horizontal e alta disponibilidade.

Com o banco de dados colunar, o Spotify foi capaz de realizar análises complexas em tempo real sobre os hábitos de audição dos usuários. Eles puderam identificar tendências musicais, criar playlists personalizadas com base nos gostos dos usuários e aprimorar a recomendação de músicas.

Além disso, o Spotify aproveitou a capacidade de escalabilidade do banco de dados colunar para lidar com picos de demanda durante eventos ao vivo, lançamentos de álbuns populares e playlists virais. Isso garantiu uma experiência de streaming contínua e confiável para os usuários, mesmo em momentos de alto tráfego.

A utilização do banco de dados colunar ajudou o Spotify a otimizar suas operações de análise de dados, fornecer recomendações mais precisas e aprimorar a personalização do conteúdo musical. Isso contribuiu para a satisfação dos usuários e impulsionou o crescimento contínuo do serviço de streaming de música.

Essa história destaca como os bancos de dados colunares, como o Apache Cassandra, podem ser aplicados com sucesso em um contexto de streaming de música para melhorar a análise de dados, personalização e oferecer uma experiência excepcional aos usuários.

Outro exemplo real do uso de bancos de dados colunares é o caso da empresa de streaming de música Spotify. O Spotify enfrentava o desafio de processar e analisar rapidamente grandes volumes de dados de streaming de música de milhões de usuários em todo o mundo.

Para lidar com essa demanda, o Spotify optou por implementar um banco de dados colunar chamado Apache Cassandra. O Apache Cassandra é projetado para lidar com grandes volumes de dados distribuídos em várias máquinas, fornecendo escalabilidade horizontal e alta disponibilidade.

Com o banco de dados colunar, o Spotify foi capaz de realizar análises complexas em tempo real sobre os hábitos de audição dos usuários. Eles puderam

identificar tendências musicais, criar playlists personalizadas com base nos gostos dos usuários e aprimorar a recomendação de músicas.

Além disso, o Spotify aproveitou a capacidade de escalabilidade do banco de dados colunar para lidar com picos de demanda durante eventos ao vivo, lançamentos de álbuns populares e playlists virais. Isso garantiu uma experiência de streaming contínua e confiável para os usuários, mesmo em momentos de alto tráfego.

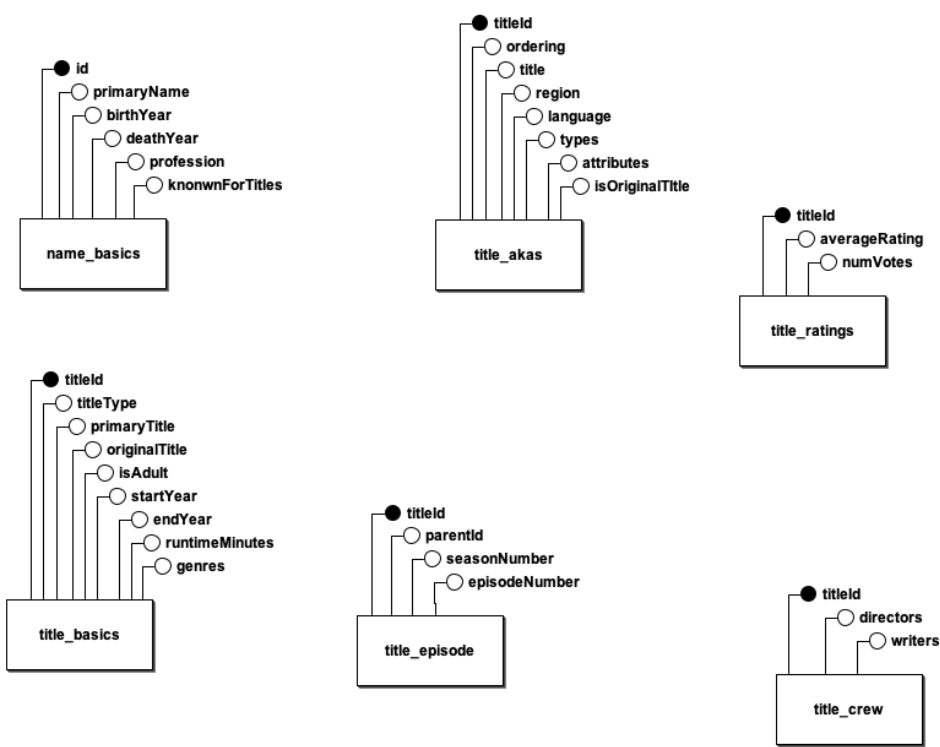
A utilização do banco de dados colunar ajudou o Spotify a otimizar suas operações de análise de dados, fornecer recomendações mais precisas e aprimorar a personalização do conteúdo musical. Isso contribuiu para a satisfação dos usuários e impulsionou o crescimento contínuo do serviço de streaming de música.

Essa história destaca como os bancos de dados colunares, como o Apache Cassandra, podem ser aplicados com sucesso em um contexto de streaming de música para melhorar a análise de dados, personalização e oferecer uma experiência excepcional aos usuários.

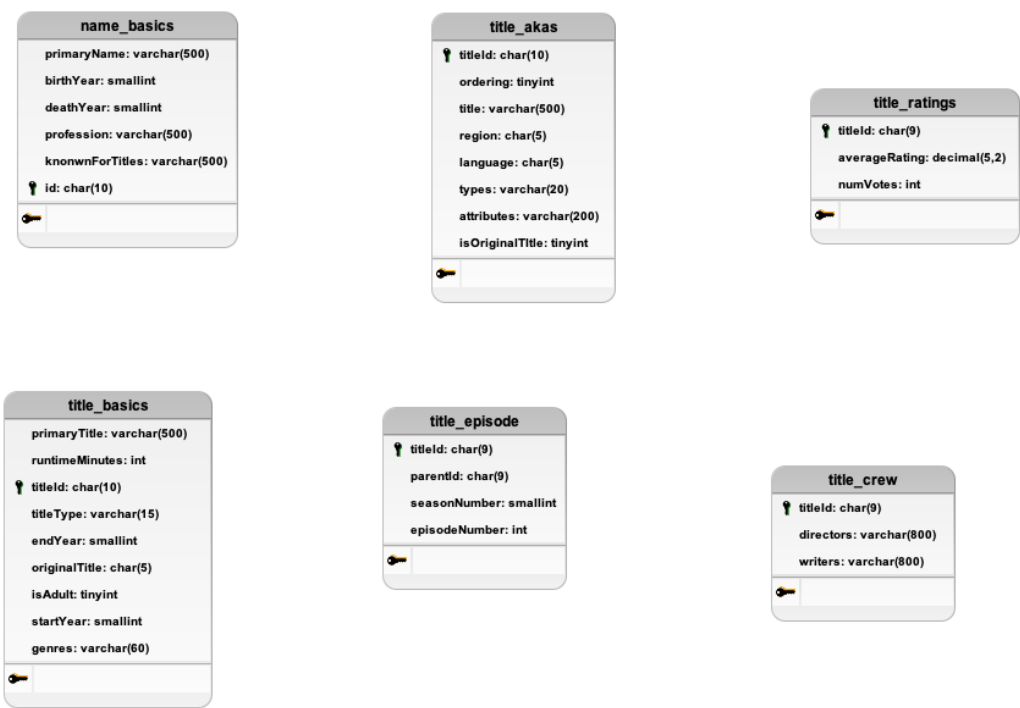
## **Bases de dados colunar**

Através do Github foi encontrado um exemplo de banco de dados colunar usando MariaDB Column Store. No exemplo escolhido, são carregados os dados sobre filmes da plataforma IMDB. A base pode ser acessada pelo link: <https://github.com/geekgogie/columnstore-imdb-data-load/tree/master>

DER



DLD





## Referências bibliográficas

1. KLEPPMANN, Martin. Designing Data-Intensive Applications. 1. ed. Sebastopol, CA: O'Reilly Media, 2017.
2. GitHub - Columnstore IMDB Data Load. Disponível em: <https://github.com/geekgogie/columnstore-imdb-data-load/tree/master>. Acesso em: 12 jun. 2023.
3. MariaDB - ColumnStore. Disponível em: <https://mariadb.com/kb/en/mariadb-columnstore/>. Acesso em: 12 jun. 2023.
4. Amazon Web Services - Banco de dados colunar. Disponível em: <https://aws.amazon.com/pt/nosql/columnar/#:~:text=Banco%20de%20dados%20colunar%20vs.&text=Enquanto%20um%20banco%20de%20dados,da dos%2C%20normalmente%20em%20aplicativos%20anal%C3%ADticos..> Acesso em: 12 jun. 2023.
5. Kondado. O que é um banco de dados colunar? Disponível em: <https://kondado.com.br/blog/blog/2019/10/31/o-que-e-um-banco-de-dados-colunar/>. Acesso em: 12 jun. 2023.