

UNIVERSIDADE DE BRASÍLIA

Faculdade do Gama

Sistemas de Banco de Dados 2

Tecnologias de Banco de Dados (TI-BD)

Banco de Dados de Colunas

Rômulo Vinícius de Souza –

15/0147601

Brasília, DF

2019

1. Definição da tecnologia utilizada

O banco de dados de coluna surgiu devido a uma maior necessidade de realizar transações e visualizar alguns dados de forma mais rápida. Para isso, utiliza uma orientação vertical, salvando os dados em colunas, ao invés de linhas.

O banco de dados com o qual estamos habituado, o tradicional banco relacional, armazena todo o conjunto de dados relacionado àquela tabela em uma mesma linha e, portanto, se for conhecido o identificador único daquela tupla, é possível encontrar todos os outros dados relacionados. Com o banco de dados em colunas, cada atributo que pertenceria a linha será armazenado em uma única tabela. Dessa forma, é preciso que haja um ID virtual em cada coluna (mantendo a posição do atributo na coluna), a fim de que haja a correlação entre as colunas, indicando que aquelas colunas pertencem a um registro específico. O mesmo pode ser observado a seguir:

PESSOA_ID	PESSOA_NOME	PESSOA_IDADE
1	MACHADO DE ASSIS	44
2	CARLOS DRUMMOND DE ANDRADE	85
3	MONTEIRO LOBATO	66
4	CECÍLIA MEIRELES	63
5	CASTRO ALVES	24

Figura 1: Armazenamento por linhas [1]

PESSOA_ID	PESSOA_IDADE
1	44
2	85
3	66
4	63
5	24

PESSOA_NOME
MACHADO DE ASSIS
CARLOS DRUMMOND DE ANDRADE
MONTEIRO LOBATO
CECÍLIA MEIRELES
CASTRO ALVES

Figura 2: Armazenamento por colunas [1]

2. Objetivos principais da tecnologia utilizada

O principal objetivo da tecnologia utilizada é a melhora do desempenho. Os bancos de colunas são utilizados para otimizar situações em que o banco de dados relacional não é suficiente. Além da melhora do desempenho em consultas, devido ao menor tráfego de dados entre o disco e a memória, o banco de dados de coluna utiliza um menor espaço de armazenamento.

3. Vantagens na tecnologia utilizada

O banco de dados de colunas possui algumas vantagens em relação a bancos de dados relacionais. Dentre elas, podemos citar:

- **Capacidade de compressão dos dados:** Ao analisar a compressão de um banco onde os registros são armazenados em linha, encontraremos em uma mesma linha diferentes tipos (domínios), o que torna o processo de compressão mais complicado. Já no banco orientado a colunas, cada coluna irá conter o mesmo tipo (domínio) de dado. De acordo com algumas pesquisas, o nível de compressão alcançado em bancos orientados a colunas chega a ser de 60% a 70% mais eficiente que nos bancos orientados a linhas.

- **Consultas de agregação:** Devido a sua estrutura, banco de dados de colunas possuem uma boa performance em realizar consultas de agregação (SUM, AVG, COUNT, etc).

- **Escalabilidade:** Bancos de dados de colunas resultam em uma boa escalabilidade. Além disso, são apropriados para realizar processamento paralelo em larga escala, o que é gerado através de dados espalhados dentre várias máquinas em cluster.

- **Rápido carregamento e consulta:** Bancos de dados de colunas carregam extremamente rápido, quando comparado a bancos de dados convencionais. Eles podem carregar milhões de linhas em segundos e rapidamente realizar operações.

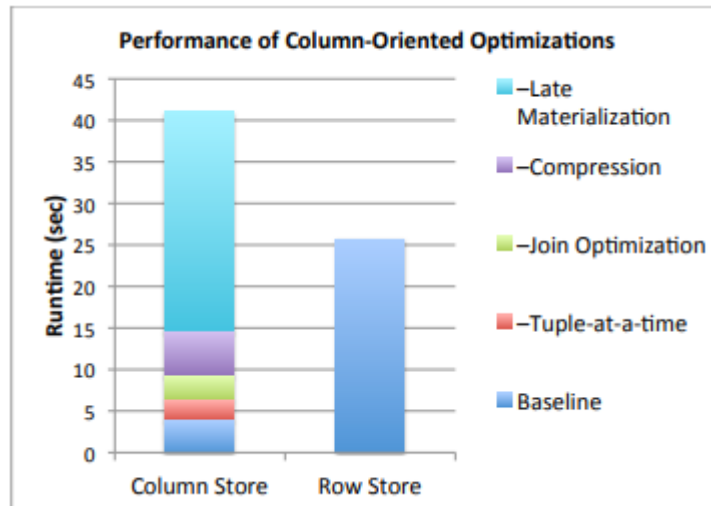


Figura 3: Performance banco de dados de colunas X banco de dados relacional [2]

4. Desvantagens da tecnologia utilizada

Dentre as principais desvantagens do bancos de dados de coluna, podemos citar:

- Pode apresentar dificuldade em realização de transações, ou até mesmo não possibilitá-la.
- Consultas com JOIN em tabelas pode reduzir drasticamente a performance, haja vista que as colunas estão em tabelas distintas.
- Alteração de dados no banco de dados de colunas pode afetar também a performance do sistema, pois caso seja desejada a alteração em diversas colunas, será necessário a descompressão de diversos arquivos diferentes. Bancos de dados de coluna são preferíveis quando se tem dados que não serão alterados com frequência.

5. Exemplos de uso interessantes em empresas, projetos ou instituições dessa tecnologia de Banco de dados pesquisada.

MariaDB

MariaDB é um dos bancos de dados open-source mais populares no mundo, criado pelos próprios desenvolvedores do MySQL. Ele surgiu através de um fork do MySQL, após este ser adquirido pela Oracle. Muitas empresas famosas o utilizam, como Wikipedia, WordPress e Google.

MariaDB transforma dados em informação estruturada para uma grande variedade de aplicações, desde bancos a websites. É muito utilizado devido a sua velocidade, escalabilidade e robustez, com grandes variedades de motores, plug-ins e ferramentas para cada tipo de uso.



Figura 4: Logo MariaDB

CrateDB

O banco de dados CrateDB combina a familiaridade do SQL com a escalabilidade e flexibilidade do NoSQL. Com ele, pode-se utilizar SQL para processar qualquer tipo de dado, seja ele estruturado ou não. Além disso, apresenta uma alta performance na realização de queries, até mesmo incluindo JOINS e agregações.



Figura 5: Logo CrateDB

ClickHouse

O ClickHouse é um Sistema Gerenciador de Banco de Dados orientado a colunas para processamento analítico online. Sua performance se destaca no mercado, conseguindo processar centenas de milhões ou até mais de um bilhão de linhas e dezenas de gigabytes de dados por segundo.



Figura 6: Logo ClickHouse

Apache HBase

O HBase é um banco de dados de colunas não relacional, open-source, distribuído e feito em java, muito utilizado no processamento de grandes tabelas de dados. O HBase permite escalabilidade linear e modular, além de fornecer uma grande variedade de ferramentas para o gerenciamento de tabelas.



Figura 7: Logo Apache HBase

6. Referencias Bibliográficas

- [1] – MERIAT, Vitor. Column Store x Row Store e o In-Memory. Disponível em: <https://www.vitormeriat.com.br/2015/07/17/column-store-x-row-store-e-o-in-memory/>. Acesso em 08 set. 2019.
- [2] – ABADI, D. El al. The Design and Implementation of Modern Column-Oriented Database Systems. Ed. Now. 2012.
- [3] – SGBD relacionais orientados a coluna: uma nova roupagem ao Data Warehousing- Parte 01. Disponível em: <https://www.devmedia.com.br/sqbd-relacionais-orientados-a-coluna-uma-nova-roupagem-ao-data-warehousing-parte-01/11349>. Acesso em 08 set 2019.
- [4] – ISAIAS, Barroso. Banco de dados orientado a colunas. Disponível em: <https://isaiasbarroso.wordpress.com/2012/06/20/banco-de-dados-orientado-a-colunas/>. Acesso em 08 set 2019.
- [5] – What is a Column Store Database? Disponível em: <https://database.guide/what-is-a-column-store-database/>. Acesso em 08 set 2019.
- [6] – COLUMN-ORIENTED DATABASE TECHNOLOGIES. Disponível em: <https://www.dbbest.com/blog/column-oriented-database-technologies/>. Acesso em 08 set 2019.
- [7] – TOP 9 COLUMN-ORIENTED DATABASES. Disponível em: <https://www.predictiveanalyticstoday.com/top-wide-columnar-store-databases/>. Acesso em 08 set 2019.