

UNIVERSIDADE DE BRASÍLIA
Faculdade do Gama

Sistemas de Banco de Dados 2

Tecnologias de Banco de Dados (TI-BD)

Bancos de Dados Distribuídos

Luís Guilherme Gaboardi Lins - 180022962

Brasília, DF
2023

a) Definição de Bancos de Dados Distribuídos;

Um banco de dados distribuído (BDD) é uma coleção de múltiplos bancos de dados lógicos que são distribuídos por diferentes computadores em uma rede. Os bancos de dados podem ser geograficamente distribuídos, conectados por uma rede de comunicação, mas apresentados ao usuário final como um banco de dados único e centralizado.

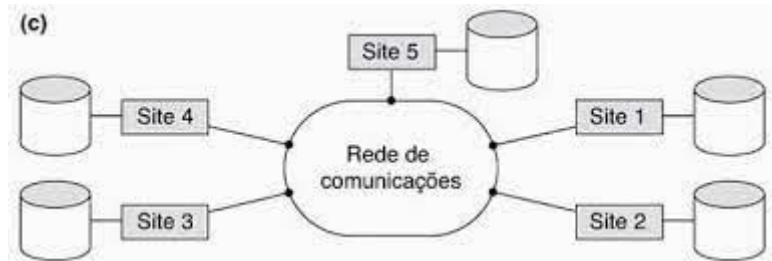
Um BDD consiste de dois ou mais bancos de dados que são conectados por uma rede de comunicação. Cada banco de dados individual contém um subconjunto dos dados totais. A arquitetura BDD pode ser centralizada, onde um servidor central controla todos os nós da rede, ou pode ser descentralizada, onde não há um servidor central, mas cada nó é igualmente importante.

Um dos principais componentes de uma arquitetura BDD é o software de gerenciamento de banco de dados distribuído (SGBD), que é responsável por gerenciar a distribuição de dados e garantir a consistência dos dados em todos os nós. O SGBD permite que os usuários acessem e manipulem dados em qualquer nó do BDD como se estivessem trabalhando com um único banco de dados centralizado.

A distribuição de dados em um BDD pode ser feita de diferentes maneiras, por exemplo, por meio de fragmentação horizontal, onde tuplas são divididas em diferentes nós, por fragmentação vertical, onde colunas são divididas em diferentes nós, ou por replicação, onde os dados são armazenados em múltiplos nós para garantir maior disponibilidade e tolerância a falhas.

Esta tecnologia pode ser encontrada em muitas indústrias, tais como finanças, saúde e telecomunicações, e está em constante evolução para atender às demandas cada vez maiores de armazenamento e gerenciamento de dados distribuídos.

Figura 1 – Representação gráfica de um BDD



Fonte: Material de aula de Tiago De Melo, 2012.

b) Características e conceitos relevantes;

Transações são operações que executam um conjunto de ações em um banco de dados, geralmente incluindo inserções, atualizações e exclusões de dados. No contexto de um banco de dados distribuído, as transações são executadas em diferentes nós do sistema, o que pode levar a desafios adicionais para garantir a integridade dos dados e a consistência das transações.

Para garantir a consistência e integridade dos dados em um banco de dados distribuído, é comum aplicar as características ACID, que são:

- **Atomicidade:** uma transação é uma operação atômica, ou seja, é executada como uma única unidade de trabalho. Se uma parte da transação falhar, todas as partes da transação são revertidas, deixando o banco de dados em seu estado original.
- **Consistência:** uma transação deve manter a consistência dos dados, ou seja, os dados no banco de dados devem estar em um estado válido antes e depois da transação. Qualquer violação de integridade dos dados deve ser evitada.
- **Isolamento:** uma transação deve ser isolada de outras transações, o que significa que cada transação deve ser executada como se fosse a única

transação no sistema. Isso evita que uma transação interfira em outra transação.

- **Durabilidade:** uma transação concluída deve ser durável, o que significa que o resultado da transação deve ser armazenado permanentemente no banco de dados, mesmo que ocorra uma falha no sistema.

No contexto de um banco de dados distribuído, as características ACID podem ser mais desafiadoras de implementar devido à natureza distribuída do sistema. Por exemplo, o isolamento de transações pode ser difícil de manter quando várias transações são executadas simultaneamente em diferentes nós. Para lidar com isso, existem vários tipos de transação que são usados em um banco de dados distribuído:

- **Transações locais:** transações que afetam apenas um único nó em um banco de dados distribuído. Essas transações são mais fáceis de garantir a consistência e a atomicidade, mas não são escaláveis em sistemas distribuídos de grande escala.
- **Transações distribuídas:** transações que afetam múltiplos nós em um banco de dados distribuído. Essas transações são mais complexas, pois é necessário garantir a consistência e a atomicidade em todos os nós afetados, mas são mais escaláveis em sistemas distribuídos de grande escala.
- **Transações replicadas:** transações que são replicadas em vários nós para garantir a disponibilidade e a confiabilidade dos dados. Essas transações podem ser escaláveis e confiáveis, mas podem levar a problemas de consistência se não forem implementadas corretamente.

Além dos tipos de transação, existem várias funções adicionais que são usadas em bancos de dados distribuídos para garantir a consistência e a atomicidade das transações:

- Coordenador de transações: um componente do sistema que é responsável por coordenar transações distribuídas e garantir que todas as partes da transação sejam concluídas com sucesso.
- Registro de transações: um log que registra todas as transações executadas no banco de dados. Isso é usado para garantir a durabilidade das transações, permitindo que o sistema se recupere de falhas.
- Controle de concorrência: a técnica utilizada para garantir que várias transações em um banco de dados possam ser executadas simultaneamente sem causar conflitos ou inconsistências nos dados.

c) Objetivos principais de Banco de Dados Distribuídos;

De acordo com Korth et al. (2014), o objetivo principal dos bancos de dados distribuídos é permitir que os dados sejam distribuídos em diferentes locais físicos, melhorando o desempenho, a escalabilidade e a disponibilidade do sistema de banco de dados. A distribuição de dados permite que as consultas e transações sejam processadas localmente em cada nó, o que pode melhorar significativamente o desempenho do sistema, já que não é necessário transportar todos os dados para um local centralizado para processamento.

Além disso, DeWitt et al. (1990) afirmam que a distribuição de dados em diferentes locais pode tornar o sistema mais escalável e disponível, permitindo que mais usuários acessem o sistema simultaneamente e reduzindo a probabilidade de interrupções do sistema. A distribuição de dados também pode melhorar a segurança do sistema, já que é possível replicar os dados em diferentes locais, tornando-os mais resistentes a falhas e a ataques maliciosos.

Outro objetivo importante dos bancos de dados distribuídos, conforme destacado por Ozsu e Valduriez (2011), é fornecer maior confiabilidade do sistema. Como os dados são distribuídos em diferentes nós, a falha de um nó não afetará todo o sistema, mas apenas a parte que está localizada naquele nó. Isso significa que, em caso de falha em um nó, os dados ainda estarão

disponíveis em outros locais, mantendo a continuidade das operações do sistema.

Ademais, a distribuição dos dados pode permitir que os usuários acessem os dados de forma mais rápida e eficiente, independentemente da sua localização geográfica. Por exemplo, uma empresa pode manter um banco de dados distribuído para seus funcionários em diferentes países, permitindo que eles acessem as informações relevantes com maior rapidez e sem a necessidade de transferir grandes quantidades de dados pela rede.

d) Vantagens de Banco de Dados Distribuídos;

Os bancos de dados distribuídos oferecem várias vantagens em relação aos bancos de dados relacionais tradicionais. Entre essas vantagens, podemos destacar a escalabilidade, a disponibilidade e a tolerância a falhas.

Em relação à escalabilidade, um estudo realizado por Ozsu e Valduriez (2011) afirma que os bancos de dados distribuídos permitem que o sistema seja dimensionado horizontalmente, ou seja, é possível adicionar novos servidores e nós à medida que a demanda por dados aumenta, sem a necessidade de grandes investimentos em hardware de alto desempenho. Já os bancos de dados relacionais têm mais dificuldade em lidar com grande volume de dados e usuários simultâneos, exigindo investimentos maiores em infraestrutura.

Outra vantagem dos bancos de dados distribuídos é a disponibilidade, já que os dados estão distribuídos em vários nós, se um deles falhar, os dados ainda estão disponíveis em outros nós. De acordo com Tanenbaum e Van Steen (2017), isso torna os bancos de dados distribuídos mais tolerantes a falhas e mais adequados para aplicações críticas, que não podem ficar indisponíveis por longos períodos de tempo.

Por fim, a tolerância a falhas é uma vantagem importante dos bancos de dados distribuídos. Como mencionado anteriormente, se um nó falhar, os

dados ainda estão disponíveis em outros nós, o que torna o sistema mais confiável. Segundo Bernstein et al. (2017), a redundância de dados nos bancos de dados distribuídos aumenta a confiabilidade do sistema e minimiza a perda de dados em caso de falhas.

e) Desvantagens de Banco de Dados Distribuídos;

Comparando as desvantagens de um BDD em relação a um banco de dados relacional, podemos citar algumas diferenças significativas. Em primeiro lugar, um BDD é mais complexo de ser gerenciado em comparação a um banco de dados relacional, devido à necessidade de lidar com múltiplos nós e a necessidade de garantir a consistência dos dados distribuídos. Isso pode requerer uma equipe de gerenciamento mais especializada e também pode aumentar os custos de infraestrutura.

Em segundo lugar, a complexidade do gerenciamento de um BDD pode aumentar a probabilidade de falhas no sistema, uma vez que um nó defeituoso pode afetar o funcionamento do sistema como um todo. Isso pode requerer mecanismos adicionais de tolerância a falhas, o que pode aumentar ainda mais a complexidade do sistema.

Por fim, outra desvantagem é que, em algumas situações, pode ser mais difícil garantir a integridade dos dados em um BDD. Isso ocorre porque, em um sistema distribuído, diferentes partes dos dados podem estar armazenadas em diferentes nós, o que pode dificultar a manutenção da integridade dos dados em caso de atualizações simultâneas.

f) Exemplos de usos de Banco de Dados Distribuídos em empresas, organizações, projetos ou instituições;

Um exemplo interessante de uso de banco de dados distribuídos é o sistema de gerenciamento de estoque e vendas da Amazon, que utiliza o Amazon DynamoDB, um banco de dados NoSQL distribuído. O DynamoDB é

projetado para alta disponibilidade e escalabilidade automática, permitindo que a Amazon gerencie facilmente um grande número de transações em seu site.

Outro exemplo de banco de dados distribuído é o Apache Cassandra, um banco de dados NoSQL distribuído e de código aberto. Ele é usado por empresas como Apple, Netflix, eBay e Spotify para armazenar grandes quantidades de dados em clusters distribuídos de servidores. O Cassandra é projetado para ser altamente escalável e disponível, com tolerância a falhas e replicação de dados automática.

Um caso de insucesso na utilização de banco de dados distribuídos é o da empresa de tecnologia Rackspace, que em 2014 teve uma falha em seu banco de dados NoSQL distribuído Cassandra, que resultou na perda de dados de seus clientes. A empresa admitiu que a falha ocorreu devido a um erro humano na configuração do cluster de bancos de dados, destacando a importância da configuração adequada e gerenciamento de bancos de dados distribuídos.

g) Bibliografias Pesquisadas

BERNSTEIN, P. A. et al. Principles of Transaction Processing. Morgan Kaufmann, 2017.

CERI, Stefano; PELAGATTI, Giuseppe. Distributed Databases: Principles and Systems. New York: McGraw-Hill, 1984.

CHAO, K. M., & SINGH, D. K. (1996). Distributed database systems: where are we now? Computer, 29(12), 49-58.

DESHAPANDE, Ankita P.; KULKARNI, Nitin A. A Survey on Replication Techniques for Distributed Databases. International Journal of Computer Science and Network Security, v. 10, n. 9, p. 204-211, 2010.

DEWITT, D. J. et al. The Gamma Database Machine Project. IEEE Transactions on Knowledge and Data Engineering, v. 2, n. 1, p. 44-62, 1990.

ELMASRI, R., & NAVATHE, S. B. (2016). Fundamentals of database systems. Pearson.

KORTH, H.; SILBERSCHATZ, A.; SUDARSHAN, S. Sistema de Banco de Dados. São Paulo: Cengage Learning, 2014.

KRZYZANOWSKI, Paul. Distributed Transactions and Two-Phase Commit. Disponível em: <http://www.cs.rutgers.edu/~pxk/417/notes/transactions.html>. Acesso em: 17 abr. 2023.

KULKARNI, D. D.; SANE, S. S. Concurrency Control in Distributed Databases. International Journal of Computer Applications, v. 7, n. 6, p. 19-26, 2010.

OZSU, M. T.; VALDURIEZ, P. Princípios de Sistemas de Banco de Dados Distribuídos. São Paulo: Campus, 2011.

TANENBAUM, A. S.; VAN STEEN, M. Distributed Systems: Principles and Paradigms. Pearson, 2017.