

UNIVERSIDADE DE BRASÍLIA

Faculdade do Gama

Sistemas de Banco de Dados 2

Tecnologias de Banco de Dados (TI-BD)

Bancos de Dados Distribuídos

Thiago França Vale Oliveira - 17/0114929

Brasília, DF

2023

O que são Bancos de Dados Distribuídos?

Existem várias arquiteturas nas quais um banco de dados pode ser construído, dentre elas existe a arquitetura distribuída. Os bancos de dados distribuídos diferentemente de um banco de dados centralizado, onde todos os dados são gerenciados e acessados a partir de um único servidor central, são bancos de dados que armazenam dados em vários computadores ou servidores interconectados por uma rede. Essa rede pode ser tanto local quanto global, permitindo que os dados possam estar fisicamente localizados em diferentes lugares geográficos.

Em um banco de dados distribuído, os dados são divididos em fragmentos e cada fragmento é armazenado em um servidor diferente, esses servidores são chamados de nós. Cada nó é responsável por gerenciar seu próprio fragmento de dados, mas todos os servidores devem cooperar e se comunicar para manter a consistência dos dados em todo o sistema. Isso significa que, quando um usuário solicita dados de um banco de dados distribuído, a solicitação pode ser atendida por vários nós diferentes, que trabalham juntos para fornecer os dados solicitados.

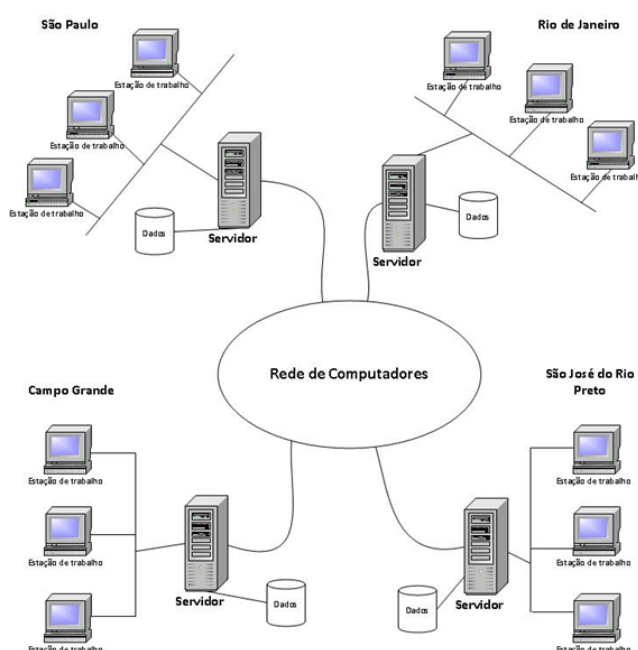


Figura 1. Arquitetura de um banco de dados distribuído

Um aspecto importante dos bancos de dados distribuídos é a distribuição dos dados, ou seja, a maneira de como os dados são divididos e armazenados pelos nós. Existem várias maneiras de distribuir os dados em um banco de dados distribuído:

- Fragmentação horizontal: Os dados são divididos em linhas e cada linha é armazenada em um nó diferente. Por exemplo, em um banco de dados com uma tabela Clientes, pode-se dividir os clientes por região geográfica e armazenar os clientes de Brasília em um nó, os clientes de São Paulo em outro e assim por diante.
- Fragmentação vertical: Os dados são divididos em colunas e cada coluna é armazenada em um nó diferente. Por exemplo, em um banco de dados com a tabela Clientes, a partir dos seus atributos, pode-se dividir as colunas em "idCliente", "nome", "email" e armazenar cada uma em um nó diferente.
- Replicação: Os dados são copiados em vários nós para aumentar a disponibilidade e a redundância. Se um servidor falhar, os dados ainda estarão disponíveis em outro. A replicação também pode melhorar o desempenho, permitindo que várias cópias de um conjunto de dados sejam armazenadas em vários nós para minimizar o tempo de acesso aos dados.
- Combinação de fragmentação e replicação: Os dados são divididos em várias partições e cada partição é replicada em vários nós. Isso permite que as consultas sejam executadas em várias partições em paralelo, enquanto a replicação garante a disponibilidade dos dados, mesmo que um ou mais nós apresentem falhas.

Vale ressaltar que cada método de distribuição apresenta suas vantagens e desvantagens, e a escolha de cada um deles depende do desempenho desejado, a quantidade de dados armazenados, a segurança e integridade dos dados, entre outros fatores.

Para que servem os Bancos de Dados Distribuídos?

Os objetivos e finalidades dos bancos de dados distribuídos são basicamente os mesmos que os dos bancos de dados centralizados, que são o armazenamento e a gestão eficiente de grandes volumes de dados. No entanto, os bancos de dados distribuídos têm algumas metas específicas que são mais relevantes para sua arquitetura e funcionalidade distribuída. Algumas dessas metas incluem o desempenho, escalabilidade, disponibilidade, flexibilidade e segurança.

Além de apresentar essas características, todas as transações em um banco de dados com essa arquitetura precisam respeitar as propriedades ACID (Atomicidade, Consistência, Isolamento, Durabilidade).

Vantagens dos Bancos de Dados Distribuídos

Os bancos de dados distribuídos apresentam várias vantagens em relação a outras arquiteturas como a centralizada. Algumas das principais vantagens são:

- **Desempenho:** Os bancos de dados distribuídos podem fornecer melhor desempenho ao distribuir dados em vários servidores e permitir que as operações de leitura e gravação sejam executadas em paralelo.
- **Escalabilidade:** Os bancos de dados distribuídos são altamente escaláveis, o que significa que a capacidade de armazenamento e processamento pode ser facilmente expandida adicionando mais servidores caso necessário.
- **Disponibilidade:** A distribuição dos dados em vários servidores torna o sistema mais tolerante a falhas, pois se um servidor falhar, os dados ainda estarão disponíveis em outros servidores. Isso aumenta a disponibilidade do sistema e reduz o tempo de inatividade.
- **Flexibilidade:** Os bancos de dados distribuídos podem ser gerenciados em diferentes locais, permitindo que os usuários acessem os dados de qualquer lugar do mundo. Isso se torna muito útil quando uma empresa atua em vários países e precisa compartilhar dados por toda a parte.
- **Custo:** Os bancos de dados distribuídos podem ser mais econômicos do

que os bancos de dados centralizados, uma vez que é mais importante investir na quantidade do que na qualidade do hardware, permitindo o uso de equipamentos mais baratos porém em maior escala no lugar de uma máquina mais potente porém bem mais cara.

Desvantagens dos Bancos de Dados Distribuídos

Apesar de apresentar várias vantagens para aqueles que escolherem a arquitetura distribuída, também existem desvantagens que requerem atenção na escolha desse modelo para o armazenamento dos dados. Algumas dessas desvantagens são:

- **Complexidade:** O gerenciamento de um banco de dados distribuído é mais complexo do que um banco de dados centralizado por exemplo, uma vez que envolve a coordenação, replicação e a consistência de dados em vários servidores ao mesmo tempo.
- **Conflitos de dados:** Os bancos de dados distribuídos podem apresentar conflitos de dados, que podem ocorrer quando os mesmos dados são alterados em diferentes locais, resultando em inconsistência de dados.
- **Segurança:** A arquitetura distribuída pode aumentar a exposição a ataques, já que os dados são distribuídos em vários servidores e podem ser mais difíceis de proteger igualmente quando comparado a uma arquitetura centralizada, por exemplo.
- **Latência de rede:** Os bancos de dados distribuídos podem sofrer de latência de rede, já que os dados precisam ser transmitidos através de uma rede para serem acessados em diferentes locais. Isso pode afetar o desempenho e a velocidade de resposta do sistema.

Uso de Bancos de Dados Distribuídos pelo mundo

O uso da arquitetura distribuída em um banco de dados, normalmente

vem por parte de empresas de grande porte que possuem filiais em vários locais diferentes, dentre elas é comum ver instituições financeiras e empresas de comércio eletrônico, pois tiram muito proveito das vantagens apresentadas.

Uma empresa muito expressiva que faz o uso dessa arquitetura é a Amazon. A Amazon utiliza uma ampla variedade de tecnologias de banco de dados distribuídos para gerenciar seus vários serviços, incluindo o Amazon S3 (Simple Storage Service), o Amazon RDS (Relational Database Service) e o Amazon DynamoDB (um banco de dados NoSQL altamente escalável). Esses bancos de dados distribuídos são usados para armazenar e gerenciar uma enorme quantidade de informações, como informações de pedidos de clientes, informações de inventário e dados de transações. A capacidade de escalabilidade dos bancos de dados distribuídos é fundamental para eles, pois a empresa cresce mais a cada dia e precisa lidar com grandes volumes de dados em tempo real.

Além das tecnologias usadas pela própria Amazon, existem vários outros bancos de dados que fazem uso dessa arquitetura. Alguns dos mais populares sendo:

- MongoDB: Um banco de dados NoSQL orientado a documentos, projetado para escalabilidade horizontal em ambientes de nuvem.
- Google Bigtable: Um banco de dados distribuído em colunas, desenvolvido pelo Google para gerenciar grandes volumes de dados em ambientes de alta carga.
- Microsoft Azure Cosmos DB: Um banco de dados distribuído, multimodelo e escalável projetado para atender a diversas necessidades de aplicativos, como web, móvel e IoT.

Referência Bibliográfica

Özsu, M.T. Valduriez, P. Principles of Distributed Database Systems, 3ª Edição, Prentice Hall, 2010.

FILETO, Renato. *Banco de Dados Distribuídos*. 2009. 30f. Artigo Científico – Universidade Federal de Santa Catarina, Santa Catarina, 2009.

ROSARIO, Luiz G. O que é Banco de Dados Distribuído? IMasters. Disponível em: <https://imasters.com.br/banco-de-dados/o-que-e-banco-de-dados-distribuido/?trace=1519021197&source=single> Acesso em: 16 de abr. 2023.

ALMEIDA, Roniere. O que é um Banco de Dados Distribuído? Devmedia. Disponível em: <https://www.devmedia.com.br/o-que-e-um-banco-de-dados-distribuido/24762> Acesso em 16 de abr. 2023.

AMAZON WEB SERVICES, INC. Amazon's Database Strategy. Disponível em: <https://aws.amazon.com/pt/relational-database/>. Acesso em: 17 abr. 2023.

MICROSOFT. Azure Cosmos DB, 2023. Disponível em: <https://azure.microsoft.com/en-us/services/cosmos-db/>. Acesso em: 17 abr. 2023.

GOOGLE CLOUD. What is Google Cloud Bigtable?, 2023. Disponível em: <https://cloud.google.com/bigtable/>. Acesso em: 17 abr. 2023.