

UNIVERSIDADE DE BRASÍLIA

Faculdade do Gama

Sistemas de Banco de Dados 2

**Tecnologias de Banco de Dados (TI-BD)**

**Bancos de Dados Distribuídos**

**Nome: Vinicius Ferreira Bernardo de Lima**

**Matrícula: 15/0151331**

Brasília, DF

2019

## 1. Definição de Banco de Dados Distribuídos

Um banco de dados distribuído consiste em uma coleção de diversos outros banco de dados que se encontram dispersos por uma rede de computadores interconectados que não compartilham componentes físicos entre si. Apesar das diversas localizações físicas, há um gerenciamento centralizado, no qual qualquer estação cliente conectada à rede possui acesso.

O gerenciamento do banco de dados distribuído é realizado através de um software denominado de Sistema de Gerenciamento de Banco de Dados Distribuído (SGBDD). Sua principal função é realizar o gerenciamento dos múltiplos banco de dados distribuídos pela rede de maneira transparente, ou seja, para um usuário final, vai parecer que o banco de dados se encontra centralizado num único servidor.

Os locais em que se encontram distribuídos os bancos de dados de um mesmo sistema podem estar fisicamente próximos (como num mesmo prédio) ou fisicamente distantes, conectados por meio de uma rede WAN, montada sobre uma estrutura VPN.

Essa distribuição dos dados tem como principais objetivos: maximizar o paralelismo existente em um ambiente distribuído, minimizar o tráfego de dados entre os nós e aumentar a tolerância a falhas do sistema através da replicação e fragmentação de dados.

### 1.1 Replicação e Fragmentação de Dados

O SGBDD mantém várias réplicas dos dados, armazenados em servidores diferentes. Consequentemente teremos maior disponibilidade, maior paralelismo, menos tráfego entre os servidores, atualizações mais lentas, recuperação e controle de concorrência mais complexos. A replicação dos dados pode ser feita de duas maneiras em um sistema distribuído:

- **Total:** o banco de dados distribuído terá cópias completas em todos os nós servidores do sistema;
- **Parcial:** o banco de dados terá partes replicadas pelos nós do sistema.

Na construção de um banco de dados distribuído, existem três técnicas utilizadas na divisão do banco em unidades lógicas, denominados de fragmentos. Durante o processo de fragmentação do banco de dados, é necessário decidir quais unidades lógicas serão fragmentadas e distribuídas. Essas três técnicas são descritas a seguir:

- **Fragmentação horizontal:** agrupar registros em diversos subconjuntos baseado em um ou mais atributos da relação. Por exemplo, em uma empresa que possua as relações Departamento e Funcionário, onde cada funcionário pertença a um único departamento, os registros da relação Funcionário podem ser fragmentados ou agrupados para cada registro de departamento existente;
- **Fragmentação vertical:** a relação é dividida em outros subconjuntos que contém colunas específicas. Cada novo subconjunto deve compartilhar um campo (como chave primária ou candidata) para que seja possível reconstruir as informações originais;
- **Fragmentação híbrida:** uma mistura das duas técnicas anteriores, contendo características de ambas.

Além das técnicas de fragmentação descritas acima, também pode ser definido uma distribuição em que os subconjuntos são formados por tabelas inteiras. Por exemplo, as tabelas Cliente e Produto de uma empresa podem estar armazenados no servidor de vendas conforme ilustra a imagem a seguir:

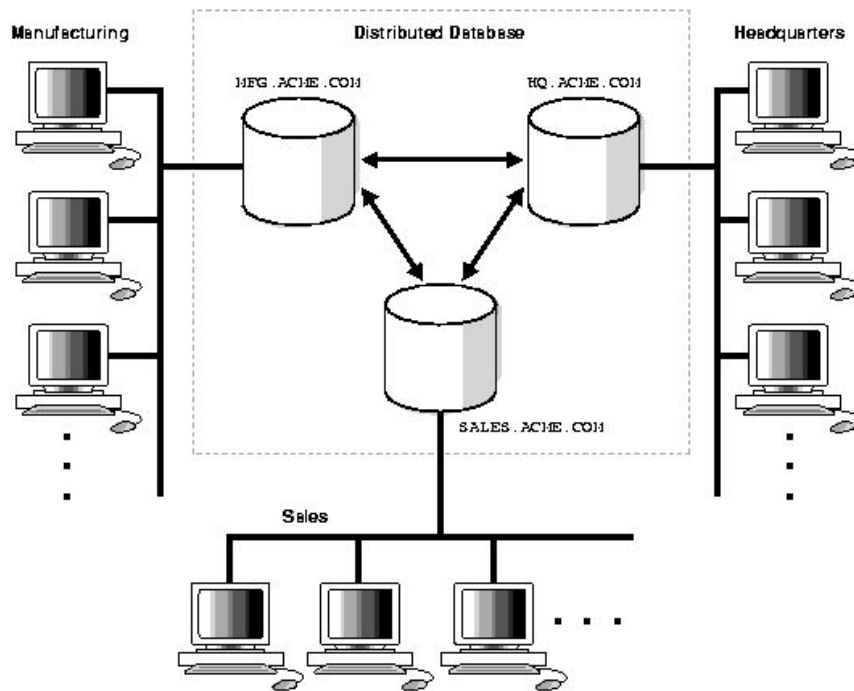


Figura 1: Ilustração de um banco de dados distribuído de uma empresa.

## 1.2 Gerenciamento de Transações

Como solução dos problemas de distribuição e localização dos dados, existe um componente do sistema denominado Gerenciador de Transações que é responsável por analisar requisições de terminais clientes e direcioná-las para o servidor apropriado. Assim que a requisição é respondida pelo(s) servidor(es), o Gerenciador de Transações sintetiza as respostas e devolve o resultado final para ao o usuário. Existem três enfoques possíveis para a organização e funcionamento dos Gerenciadores de Transações em um sistema:

- **Primeiro enfoque:** cada Gerenciador de Transações do sistema contém seu próprio catálogo e uma cópia do catálogo global de servidores do sistema;
- **Segundo enfoque:** cada Gerenciador de Transações mantém seu próprio catálogo de sistema, mas um examina o catálogo do outro quando necessário;
- **Terceiro enfoque:** um único catálogo global do sistema no qual o Gerenciador de Transações realiza consultas.

## 2. Banco de Dados Distribuído Homogêneo

Os bancos de dados distribuídos podem ser divididos em duas categorias: homogêneos e heterogêneos. Em um sistema de banco de dados distribuído homogêneo, todos os servidores físicos que fazem parte da rede interconectada contém o banco de dados implementado de maneira semelhante. O sistema operacional, o sistema gerenciador de banco de dados e as estruturas de dados são iguais ou compatíveis em todos os servidores, tornando assim, o sistema homogêneo mais fácil de projetar e gerenciar.

Em um sistema homogêneo, todos os servidores com seu próprio banco de dados estão cientes um do outro mutuamente e concordam entre si no processamento de solicitações de um usuário. Os seguintes critérios devem ser atendidos para um banco de dados distribuído homogêneo:

- As estruturas de dados usadas em cada serviço com seu próprio banco de dados devem ser iguais ou compatíveis entre si;
- O sistema gerenciador de banco de dados em cada serviço deve ser o mesmo ou compatível.

## 3. Banco de Dados Distribuído Heterogêneo

Em contraste, em um sistema de banco de dados heterogêneo, servidores do mesmo sistema podem utilizar SGBD diferentes que podem levar a problemas no processamento de transações e consultas. Além disso, cada servidor pode implementar modelos de dados diferentes, tornando necessário a realização de traduções na comunicação de servidores do mesmo sistema. Em resumo, a heterogeneidade do modelo de dados pode ocorrer nos seguintes aspectos:

- **Heterogeneidade técnica:** cada servidor do sistema distribuído pode conter diferentes sistema operacionais, diferentes protocolos de comunicação e ainda diferentes sistemas gerenciadores de banco de dados;
- **Heterogeneidade do modelo de dados:** cada servidor pode ter diferentes maneiras de representar e armazenar os mesmos dados. Por exemplo, há

chances de atributos similares serem definidos com nome ou tamanhos diferentes;

## 4. Controle de concorrência

O controle de acesso concorrente de usuários e a recuperação de dados são fatores importantes em Banco de Dados Distribuído. Além disso, é essencial que as transações executadas simultaneamente possam ser realizadas em sequência para que a consistência dos dados após a realização das transações seja garantida. Existem dois tipos de controle de concorrência: controle baseado em cópia distinta e controle baseado em votação.

### 4.1 Controle por Cópia Distinta

Esse método consiste na extensão de técnicas para o controle de concorrência existente em um banco de dados centralizado. A ideia principal desse método é designar cópias distintas de um item de dado e associar a ela os bloqueios. O servidor que contém essa cópia recebe todas as solicitações de bloqueio e desbloqueio. No contexto de banco de dados distribuídos, o controle de cópia distinta ocorre de três maneiras:

- **Nó principal:** um nó do sistema é designado como supervisor das operações de bloqueio e desbloqueio. Como todas as solicitações de bloqueio e desbloqueio são direcionadas para esse nó, podem ocorrer gargalos no sistema;
- **Nó principal com backup:** similar ao anterior com a diferença que outro nó de backup é definido. Pode acarretar na queda de desempenho já que os dois nós devem ser mantidos sincronizados;
- **Cópia primária:** semelhante aos anteriores com a diferença que este procura distribuir a carga de solicitações de bloqueio entre vários nós do sistema.

### 4.2 Controle por Votação

No controle de concorrência em um sistema distribuído baseado em votação, não há uma cópia distinta do dado. O nó que precisa de um bloqueio envia uma

solicitação a todos os outros nós que possuem uma cópia do item. Cada uma delas mantém seu próprio bloqueio e a concessão pode ser autorizada ou negada pelos demais nós.

Se a transação que solicitou o bloqueio receber concessão pela maioria dos nós que possuem a cópia, então ela segura o bloqueio e informa às demais essa ocorrência. Se em determinado espaço de tempo não for recebido concessão pela maioria dos nós, a solicitação é cancelada e esse cancelamento é informado aos outros nós. Esse método de controle de concorrência acarreta num elevado tráfego de dados pela rede.

## 5. Banco de Dados Centralizado x Banco de Dados Distribuídos

A seguir são listadas as vantagens de um Banco de Dados Distribuídos em comparação com um centralizado:

- **Desenvolvimento modular:** se o sistema precisar ser expandido para novos locais, em um sistema de banco de dados centralizado, a ação exigirá um esforço alto e interrupções no serviço já existente. No sistema de banco de dados distribuídos o trabalho envolverá simplesmente a adição de novos servidores conectados no sistema distribuído;
- **Mais confiável:** no caso de falhas no banco de dados centralizado, o sistema inteiro é interrompido. No entanto, em sistemas distribuídos, quando um componente falha, o funcionamento do sistema continua em desempenho reduzido;
- **Menor custo de comunicação:** em sistemas de banco de dados distribuídos, se os dados estiverem localizados próximos onde são mais requisitados, os custos para manipulação dos dados poderão ser minimizados. Isso não é viável em sistemas centralizados já que usuários finais podem estar muito distante fisicamente do servidor centralizado.

A seguir são listadas as desvantagens de um banco de dados distribuídos em comparação com um centralizado:

- **Necessidade de softwares complexos:** os SGBDD além de precisarem possuir as funções comuns a um SGBD centralizado, ele precisa de funcionalidades extras, tais como: capacidade de processar consultas distribuídas e gerenciamento de consultas e transações que acessam dados a partir de vários locais;
- **Sobrecarga de processamento e tráfego:** mesmo operações simples podem exigir um enorme tráfego de dados na rede do sistema e processamento em vários outros nós para que a consistência dos dados seja mantida. Em sistemas centralizados, o único tráfego exigido é o de requisição e resposta de uma transação entre o servidor e cliente;
- **Integridade dos dados:** a necessidade de atualização de dados em vários servidores pode apresentar problemas de integridade dos dados caso ocorra alguma falha na operação de atualização. Em sistemas centralizados, esse problema é inexistente já que não é necessário uma atualização em servidores diferentes.

## 6. Exemplo de uso de Banco de Dados Distribuídos

As tecnologias de Banco de Dados Distribuídos conhecidas são: oracle e postgres-xl. Um caso bem sucedido do uso de sistema distribuído é o da empresa Netflix no qual, com sua arquitetura de microsserviços, consegue uma alta escalabilidade para os vários sistemas fundamentais para o funcionamento da plataforma de streaming de mídia digital, tais como o sistema de recomendação e de autenticação.



## 7. Referências Bibliográficas

Distributed Database Architecture. Oracle Help Center. Disponível em: <[https://docs.oracle.com/cd/B28359\\_01/server.111/b28310/ds\\_concepts001.htm#ADMIN12074](https://docs.oracle.com/cd/B28359_01/server.111/b28310/ds_concepts001.htm#ADMIN12074)>. Acesso em: 09 de set. de 2019.

Scalable Open Source PostgreSQL-based Database Cluster. Postgres-XL. Disponível em: <<https://www.postgres-xl.org/>>. Acesso em: 09 de set. de 2019.

Scalable Open Source PostgreSQL-based Database Cluster. Postgres-XL. Disponível em: <<https://www.postgres-xl.org/>>. Acesso em: 09 de set. de 2019.

Netflix Open Source Software Center. Netflix. Disponível em: <<https://netflix.github.io/>>. Acesso em: 09 de set. de 2019.

Mastering Chaos - A Netflix guide to microservices. Netflix. Disponível em: <<https://www.youtube.com/watch?v=CZ3wluvmHeM>>. Acesso em: 09 de set. de 2019.

ALVES, William Pereira. Banco de dados. São Paulo Erica 2014.