

UNIVERSIDADE DE BRASÍLIA
Faculdade do Gama

Sistemas de Banco de Dados 2

Tecnologias de Banco de Dados (TI-BD)

Bancos de Dados Distribuídos

Nome: Denniel William Roriz Lima

Matrícula: 170161871

Brasília, DF

2023

a) Definição de Sistemas de Bancos de Dados Distribuídos;

De acordo com Ray (2018) em "Distributed Database Systems", um banco de dados distribuído é um conjunto de bancos de dados interconectados logicamente, controlados por um software de gerenciamento distribuído e executados em servidores geograficamente distribuídos.

Em "Database Systems: A Practical Approach to Design, Implementation, and Management", Connolly e Begg (2015) definem banco de dados distribuído como um conjunto de bancos de dados independentes, conectados por uma rede de comunicação, que colaboram entre si para atender às necessidades dos usuários.

Já Tanenbaum e Van Steen (2017), em seu livro "Distributed Systems: Principles and Paradigms", trazem a definição de banco de dados distribuído como um banco de dados logicamente integrado composto por um conjunto de bancos de dados interconectados, localizados em diferentes locais, mas que aparecem para os usuários como um único banco de dados coerente.

Entre os autores, é destacado a ideia de que um banco de dados distribuído é composto por múltiplos bancos de dados interconectados que são gerenciados de forma centralizada e que trabalham em conjunto para atender às necessidades dos usuários. Além disso, elas enfatizam a importância da coerência e da integração dos dados, de modo que os usuários possam acessá-los de forma transparente e consistente, independentemente da localização física dos dados. A figura 1 demonstra uma visão geral de como seria um sistema de banco de dados distribuído.

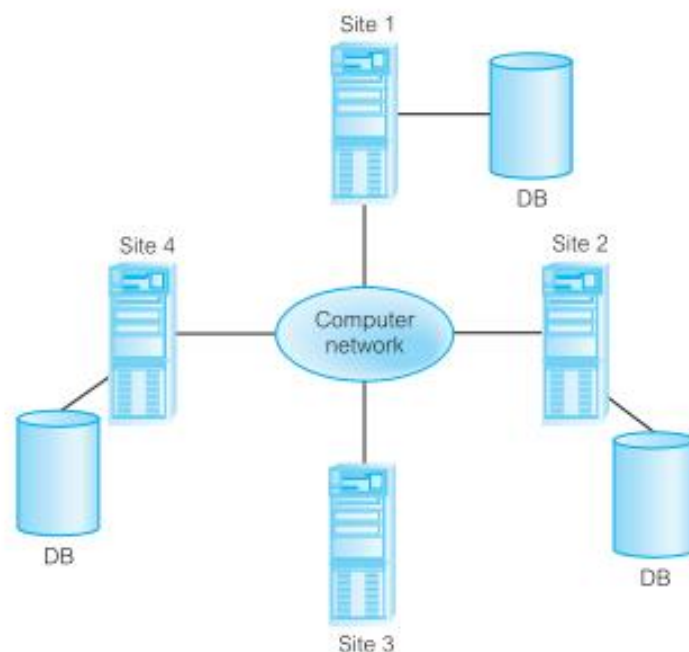


Figura 1 - Visão geral de um banco de dados distribuído, adaptada de Connolly e Begg (2015, p. 788).

b) Objetivos principais do sistema de banco de dados distribuído;

De acordo com Özsu e Valduriez (1991), um sistema de banco de dados distribuídos (SBD) é um conjunto de múltiplos bancos de dados independentes, localizados em computadores separados e interconectados por uma rede de comunicação. O principal objetivo dos SBDs é fornecer uma visão única e uniforme dos dados armazenados em diferentes locais geográficos e de propriedade distintas, permitindo que os usuários acessem e manipulem esses dados como se estivessem em um único sistema centralizado.

Já Ray (2018) aponta que um dos principais objetivos dos SBDs é fornecer escalabilidade e compartilhamento de recursos, permitindo que diferentes usuários e aplicações acessem e atualizem os dados de forma distribuída. Além disso, esses sistemas também visam garantir a disponibilidade e a confiabilidade dos dados, minimizando o impacto de falhas e interrupções no sistema como um todo.

Connolly e Begg (2015) destacam que os SBDs têm como objetivo

principal a integração e a coordenação dos dados armazenados em diferentes locais, permitindo que os usuários acessem e manipulem esses dados de maneira transparente e eficiente. Eles também salientam a importância dos SBDs em fornecer segurança e controle de acesso aos dados distribuídos, protegendo informações sensíveis e garantindo a privacidade dos usuários.

Por fim, Tanenbaum e Van Steen (2017) enfatizam que um dos principais objetivos dos SBDs é fornecer uma plataforma para a execução de aplicações distribuídas e cooperativas. Eles ressaltam que esses sistemas devem suportar transações distribuídas e coordenação de concorrência, permitindo que diferentes usuários e processos acessem e atualizem os dados de forma consistente e confiável. Além disso, os SBDs devem ser capazes de se adaptar a mudanças na topologia da rede e no ambiente de execução, garantindo a escalabilidade e a flexibilidade do sistema.

c) Vantagens do sistema de banco de dados distribuído;

As vantagens de sistemas de banco de dados distribuídos são apresentadas de forma distinta pelos autores consultados.

Segundo Özsü e Valdúriez (1991), algumas vantagens dos sistemas de banco de dados distribuídos são: "múltiplos pontos de acesso, compartilhamento de recursos, alta disponibilidade, desempenho escalável e tolerância a falhas". Eles enfatizam que, por terem dados distribuídos em diferentes sites, esses sistemas podem ser mais resilientes a falhas do que sistemas centralizados.

Para Ray (2018), as vantagens dos sistemas de banco de dados distribuídos incluem: "escalabilidade, disponibilidade, desempenho, compartilhamento de recursos e melhor uso do hardware". Segundo ela, "com um sistema de banco de dados distribuído, é possível adicionar mais recursos em qualquer um dos nós, o que permite aumentar a capacidade do sistema sem comprometer sua disponibilidade ou desempenho".

Connolly e Begg (2015) também destacam algumas vantagens dos sistemas de banco de dados distribuídos, como "a capacidade de lidar com grandes quantidades de dados, a possibilidade de atender a usuários em diferentes locais e o compartilhamento de recursos". Eles ainda ressaltam que

esses sistemas são capazes de "garantir a integridade dos dados distribuídos, a segurança e a privacidade".

Já Tanenbaum e Van Steen (2017) apontam como vantagens dos sistemas de banco de dados distribuídos: "a escalabilidade, a disponibilidade, a eficiência no acesso aos dados e a capacidade de lidar com falhas de hardware e software". Segundo eles, "o uso de vários nós permite aumentar a capacidade de armazenamento e processamento dos dados, bem como a disponibilidade do sistema".

Desta forma, com base nas informações apresentadas por cada autor e em uma comparação com o conceito de banco de dados relacional simples estudado na disciplina de Sistemas de Banco de Dados 1 (SBD1), pode-se afirmar que o sistema de banco de dados distribuído oferece maior disponibilidade e alta tolerância a falhas. Isso se deve ao fato de que, em caso de falha de um nó de dados, outros eles podem continuar funcionando normalmente. Além disso, o sistema distribuído é mais escalável, uma vez que permite a distribuição de processamento e armazenamento em várias máquinas, possibilitando o aumento de capacidade sem comprometer a disponibilidade ou o desempenho.

d) Desvantagens do sistema de banco de dados distribuído;

De acordo com Özsu e Valduriez (1991), as principais desvantagens dos sistemas de banco de dados distribuídos incluem a complexidade de gerenciamento e coordenação de múltiplos bancos de dados distribuídos, a necessidade de comunicação constante entre os diferentes sites e o aumento da vulnerabilidade à falhas e erros.

Já para Ray (2018), as desvantagens dos sistemas de banco de dados distribuídos incluem a dificuldade em manter a consistência dos dados, a necessidade de garantir a integridade dos dados e a segurança das transações distribuídas, além da possibilidade de desempenho inferior em comparação com sistemas centralizados.

Segundo Connolly e Begg (2015), as principais desvantagens dos sistemas de banco de dados distribuídos incluem a complexidade da

implementação e manutenção, a possibilidade de problemas de sincronização de dados, a necessidade de garantir a disponibilidade e confiabilidade dos dados e a possibilidade de aumento de custos devido à necessidade de infraestrutura de rede e hardware.

Por fim, Tanenbaum e Van Steen (2017) mencionam como desvantagens dos sistemas de banco de dados distribuídos a necessidade de gerenciamento complexo de vários bancos de dados, a complexidade do controle de concorrência e a garantia da integridade dos dados, além da possibilidade de aumento da complexidade do projeto devido à necessidade de considerar a distribuição dos dados.

Devido à replicação e aumento da redundância em um banco de dados distribuído, em comparação com um banco de dados relacional simples, a implementação e manutenção podem se tornar mais complexas, uma vez que é necessário lidar com múltiplos bancos de dados e garantir a sincronização entre eles. Além disso, os custos de infraestrutura de hardware e rede são mais elevados que um banco de dados relacional, devido à necessidade de garantir a conectividade e processamento entre os nós distribuídos. Por fim, temos a eventual consistência dos dados, ou seja, a consistência e integridade dos dados são atingidos de forma eventual, isso significa que pode haver momentos em que os dados não estão totalmente sincronizados, diferente do banco de dados relacional que sua sincronização é direta.

e) Exemplo de uso interessante em empresas, organizações, projetos ou instituições.

Os bancos de dados distribuídos têm sido amplamente utilizados por empresas, organizações, projetos e instituições para diversas finalidades, proporcionando vantagens significativas no gerenciamento de grandes volumes de dados em diferentes locais geográficos e garantindo alta disponibilidade, escalabilidade e processamento paralelo.

Um exemplo interessante do uso de bancos de dados distribuídos é o caso do Twitter, que migrou do MySQL para o Apache Cassandra em 2010. Essa migração permitiu ao Twitter enfrentar o desafio de lidar com um aumento

exponencial no número de tweets diários, alcançando uma escalabilidade de 20 vezes em comparação com o ano anterior. Além disso, o tempo de inatividade (downtime) do Twitter foi reduzido de 84 horas em 2008 para apenas 23 horas e 45 minutos em 2010, mantendo uma disponibilidade de aproximadamente 99,72% durante 12 meses. O Apache Cassandra foi fundamental para possibilitar a expansão do Twitter e melhorar sua disponibilidade, garantindo um serviço confiável e de alta performance aos seus usuários.

Outro exemplo interessante é o uso do Amazon Aurora, um serviço de banco de dados relacional distribuído oferecido pela Amazon Web Services (AWS). A Samsung, uma grande empresa de tecnologia, utilizou o AWS Database Migration Service para migrar seus bancos de dados Oracle para o Amazon Aurora. O Amazon Aurora é conhecido por sua capacidade de escalabilidade, desempenho e alta disponibilidade, tornando-se uma escolha estratégica para empresas que precisam gerenciar grandes volumes de dados de forma distribuída. A Samsung foi capaz de aproveitar os recursos do Amazon Aurora para garantir um gerenciamento eficiente de seus dados e otimizar sua infraestrutura de banco de dados.

Em resumo, os bancos de dados distribuídos têm sido amplamente utilizados em empresas, organizações, projetos e instituições para enfrentar desafios relacionados ao gerenciamento de grandes volumes de dados, garantir alta disponibilidade, escalabilidade e processamento paralelo. Exemplos como o do Twitter e da Samsung demonstram como o uso de sistemas de bancos de dados distribuídos pode trazer vantagens competitivas e melhorar a eficiência operacional em diversas áreas de negócio.

REFERÊNCIAS BIBLIOGRÁFICAS

ÖZSU, M. T.; VALDURIEZ, P. Principles of Distributed Database Systems. Prentice Hall, 1991.

RAY, Chhanda. Distributed Database Systems. Boca Raton: CRC Press, 2018.

CONNOLLY, Thomas; BEGG, Carolyn. Database Systems: A Practical Approach to Design, Implementation, and Management. 6th global ed. Harlow: Pearson, 2015.

TANENBAUM, Andrew S.; VAN STEEN, Maarten. Distributed Systems: Principles and Paradigms. 3rd ed. Upper Saddle River: Pearson, 2017.

IT FORUM. Crescimento faz Twitter trocar o MySQL pelo Cassandra. Disponível em: <https://itforum.com.br/noticias/crescimento-faz-twitter-trocar-o-mysql-pelo-cassandra/>. Acesso em: 17 abr. 2023.

AMAZON WEB SERVICES (AWS). Samsung migrates off Oracle to Amazon Aurora. Disponível em: <https://aws.amazon.com/pt/solutions/case-studies/samsung-migrates-off-oracle-to-amazon-aurora/>. Acesso em: 17 abr. 2023.