



SISTEMA GERENCIADOR DE BANCO DE DADOS

AULA 6

Profª Vívian Ariane Barausse de Moura



CONVERSA INICIAL

O objetivo da aula é introduzir os principais conceitos e temas sobre os bancos de dados distribuídos. Para isso, é preciso conhecer e distinguir os conceitos básicos relacionados ao BDD (banco de dados distribuídos), os requisitos considerados fundamentais para o funcionamento e a utilização de um BDD. Serão abordadas as vantagens e desvantagens de funcionamento e utilização do BDD e também aspectos referentes à arquitetura de um banco de dados distribuído.

TEMA 1 – BANCO DE DADOS DISTRIBUÍDOS

1.1 Conceitos sobre banco de dados distribuídos

Um banco de dados distribuído pode ser definido, de acordo com Alves (2014, p. 133), “como uma coleção de diversos outros bancos de dados que se encontram espalhados por uma rede de computadores e são logicamente inter-relacionados”. O autor destaca que, mesmo com as diversas localizações físicas, um gerenciamento centralizado permite um acesso a partir de qualquer estação cliente que esteja conectada à rede.

Nesse modelo existe também um *software* responsável por todo o gerenciamento do banco de dados, denominado de *Sistema de Gerenciamento de Banco de Dados Distribuído* (SGBDD) e, conforme Alves (2014), uma de suas principais funções é permitir que os diversos bancos distribuídos pela rede sejam manipulados de forma transparente, ou seja, não cabe ao usuário conhecer os detalhes da distribuição dos dados nem suas localizações. Assim, se ele executa uma consulta, vai parecer que o banco de dados se encontra centralizado num único servidor. Outra função importante é o processamento uniforme de consultas e transações.

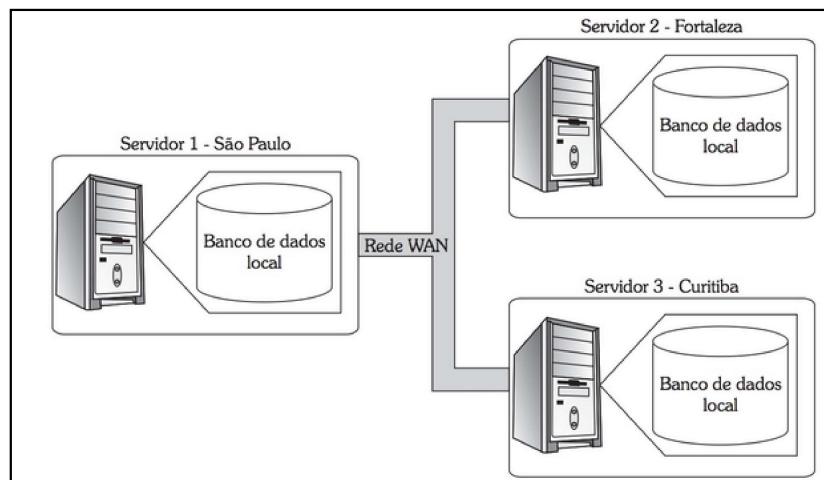
Elsmari e Navathe (2011, p. 590) definem um banco de dados distribuído – BDD “como uma coleção de múltiplos bancos de dados logicamente inter-relacionados, distribuídos por uma rede de computadores, e um sistema de gerenciamento de banco de dados distribuído – SGBDD”. O SGBDD é um sistema de *software* que gerencia um banco de dados distribuído e torna a distribuição transparente ao usuário.



Ao relacionar o BDD com a rede de computadores, é pertinente destacar que o banco de dados distribuído é diferente dos arquivos web da internet, pois, conforme Elsmari e Navathe (2011, p. 590) defendem, “as páginas da web são uma coleção muito grande de arquivos armazenados em diferentes nós em uma rede, a internet, com inter-relacionamentos entre os arquivos, representados por hyperlinks”, porém os autores destacam que as funções comuns do gerenciador de banco de dados, como o processamento de consulta uniforme e o processamento de transação ainda não se aplicam a esse cenário, no entanto, com as mudanças tecnológicas, os autores acreditam que muito em breve os BDD da internet serão realidade.

A Figura 1 apresenta um exemplo de banco de dados distribuído com acesso por meio de uma rede WAN.

Figura 1 – Exemplo de BDD



Fonte: Alves, 2014, p. 138.

Para Ramakrishnan e Gehrke (2008, p. 605), em um sistema de banco de dados distribuídos,

os dados são fisicamente armazenados em vários *sites* e cada *site* normalmente é gerenciado por um SGBD capaz de executar independente dos outros *sites*. A localização dos itens de dados e o grau de autonomia dos sites individuais têm um impacto significativo sobre todos os aspectos do sistema, incluindo a otimização e o processamento de consultas, o controle de concorrência e a recuperação.

Ao comparar com os bancos de dados paralelos, segundo Ramakrishnan e Gehrke (2008), a distribuição dos dados é governada por fatores como a posse local e a maior disponibilidade, como também questões de desempenho.



Enquanto o paralelismo é motivado pelo desempenho, vários problemas distintos motivam a distribuição de dados, conforme apresentadas no Quadro 1:

Quadro 1 – Fatores para distribuição dos dados

Maior disponibilidade	Se um <i>site</i> contendo uma relação para de funcionar, a relação continuará a estar disponível se uma cópia for mantida em outro <i>site</i> .
Acesso distribuído aos dados	Uma empresa pode ter escritórios em várias cidades. Embora os analistas talvez precisem acessar dados correspondentes a diferentes sites, normalmente encontramos uma característica de localidade nos padrões de acesso (por exemplo, um gerente de banco provavelmente vai pesquisar as contas de clientes na agência local) e essa característica pode ser explorada pela distribuição apropriada dos seus dados.
Análise de dados distribuídos	As empresas querem examinar todos os dados disponíveis para elas, mesmo que estejam armazenados em vários <i>sites</i> e em vários sistemas de banco de dados. O suporte para esse acesso integrado envolve muitos problemas; até mesmo permitir o acesso a dados amplamente distribuídos pode ser um desafio.

Fonte: Ramakrishnan; Gehrke, 2008, p. 605.

Ramakrishnan e Gehrke (2008) destacam que, em um sistema de banco de dados distribuído, os dados são armazenados em vários *sites* e cada *site* é gerenciado por um SGBD executado independentemente dos outros *sites*. De acordo com esses autores, a visão clássica de um sistema de banco de dados distribuído é que o sistema deve tornar o impacto da distribuição dos dados transparente (Ramakrishnan; Gehrke, 2008), conforme as propriedades apresentadas no Quadro 2, que são consideradas desejáveis:

Quadro 2 – Propriedade esperada de um SGBDD

Independência dos dados distribuídos	Os usuários devem ser capazes de fazer consultas sem especificar onde as relações referenciadas/ou cópias ou fragmentos das relações estão localizadas. Esse princípio é uma extensão natural da independência física e lógica dos dados. Além disso, as consultas que abrangem vários <i>sites</i> devem ser sistematicamente otimizadas, com base no custo, levando em conta os custos de comunicação e as diferenças nos custos da computação local.
Atomicidade da transação distribuída	Os usuários devem ser capazes de escrever transações que acessam e atualizam dados em vários sites, exatamente como se escrevessem transações em dados puramente locais. Em particular, os efeitos de uma transação entre os sites devem continuar a ser atômicos; isto é, todas as alterações persistem se a transação é efetivada e nenhuma persiste se ela é cancelada.

Fonte: Ramakrishnan; Gehrke, 2008, p. 613.



Ramakrishnan e Gehrke (2008) defendem que, embora essas propriedades sejam desejáveis em geral, “em certas situações, como quando os sites são conectados por meio de uma rede de longa distância lenta, essas propriedades não são eficientemente atingíveis”. Segundo os autores, tem-se argumentado que, quando os sites são globalmente distribuídos, essas propriedades nem mesmo são desejáveis, e não existe nenhum consenso real sobre quais devem ser os objetivos de projeto dos bancos de dados distribuídos, e o setor está evoluindo em resposta às necessidades dos usuários (Ramakrishnan; Gehrke, 2008).

ElsMari e Navathe (2011) complementam que o termo *sistema de gerenciamento de banco de dados distribuído* pode descrever diversos sistemas que diferem um do outro em muitos aspectos. O que eles têm em comum é o fato de os dados e o software serem distribuídos por vários sites conectados por alguma forma de rede de comunicação. Existe uma série de tipos de SGBDDs e critérios e fatores que tornam alguns destes sistemas diferentes.

Segundo os autores, Elsmari e Navathe (2011), o primeiro fator a considerar é o grau de homogeneidade do software do SGBDD, que será homogêneo se todos os servidores, SGBD locais individuais, usarem softwares idênticos e todos os usuários (clientes) utilizarem softwares idênticos; caso contrário, ele é chamado de heterogêneo. Outro fator relacionado à homogeneidade é o grau de autonomia local: se não houver provisão para o site local funcionar como um SGBD independente, então o sistema não tem autonomia local, no entanto se seu acesso direto por transações locais a um servidor for permitido, o sistema tem algum grau de autonomia local.

Com esses conceitos expostos, Elsmari e Navathe (2011) definem as classificações de bancos de dados distribuídos como:

- Sistemas de bancos de dados centralizados tradicionais.
- Sistemas de bancos de dados distribuídos puros.
- Sistemas de bancos de dados federados.
- Sistemas de multibancos de dados ou bancos de dados peer-to-peer.

TEMA 2 – REQUISITOS DE UM BDD

ElsMari e Navathe (2011, p. 590) destacam a distinção entre os bancos de dados distribuídos dos sistemas multiprocessadores que usam



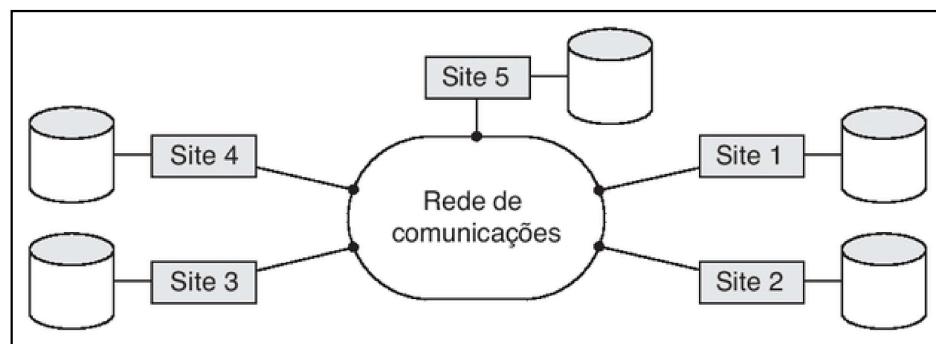
armazenamento compartilhado (memória primária ou disco) para um banco de dados ser chamado de *distribuído* às condições mínimas apresentadas no Quadro 3, que devem ser satisfeitas:

Quadro 3 – Condições mínimas para um BDD

Conexões de nós de banco de dados por uma rede de computadores	Existem vários computadores, chamado <i>site</i> ou <i>nós</i> . Esses <i>sites</i> devem ser conectados por uma rede de computadores básicas para transmitir dados e comandos entre <i>sites</i> . Conforme apresentada na Figura 2. A figura 2 será abordada com mais detalhes no Tema 5, que abordará as questões pertinentes à arquitetura de um BDD.
Inter-relação lógica dos bancos de dados conectados	É essencial que as informações nos bancos de dados sejam relacionadas logicamente.
Ausência de restrição de homogeneidade entre os nós conectados	Não é necessário que todos os nós sejam idênticos em relação aos dados, <i>hardware</i> e <i>software</i> .

Fonte: Elsmari; Navathe, 2011, p. 590.

Figura 2 – Rede de comunicações



Fonte: Elsmari e Navathe, 2011, p. 597

Elsmari e Navathe (2011) afirmam que os *sites* podem estar localizados nas proximidades físicas, dentro do mesmo prédio ou um grupo de prédios e conectados por uma rede local, ou podem estar distribuídos geograficamente por grandes distâncias e conectados por uma rede de longa distância ou rede remota. As redes locais costumam utilizar sem fios ou cabos e as de longa distância utilizam linhas telefônicas ou satélites. Existe a possibilidade de usar uma combinação de redes. O que vai diferenciar é a topologia que define os caminhos de comunicação diretos entre os *sites*. É importante destacar que o tipo e a topologia da rede vão impactar no desempenho e consequentemente nas estratégias para o processamento de consulta distribuído e o projeto de banco de dados distribuídos. Os autores destacam, no entanto, que para as



questões arquitetônicas de alto nível, não vai importar o tipo de rede utilizada e sim que cada *sítio* seja capaz de se comunicar direta ou indiretamente com todos os outros *sítios*. (ElsMari; Navathe, 2011)

Alguns conceitos são fundamentais para entender a relação do banco de dados distribuídos. Um deles é o de *transparência*, que, de acordo com Elsmari e Navathe (2011, p. 591), “estende a ideia geral de ocultar detalhes da implementação dos usuários finais, um sistema altamente transparente oferece muita flexibilidade ao usuário final/desenvolvedor, pois requer pouco ou nenhum conhecimento dos detalhes básicos de sua parte”. No BDD, os dados e o *software* são distribuídos por vários *sítios* conectados por uma rede de computadores, de modo que tipos adicionais e transparências são introduzidos.

Outra questão fundamental é a autonomia, pois, segundo Elsmari e Navathe (2011, p. 592), “determina a extensão da qual os nós individuais ou BDs em um BDD conectado pode operar independentemente, um alto grau de autonomia desejável para maior flexibilidade e manutenção personalizada de um individual”. A autonomia pode ser aplicada ao projeto de comunicação e execução. A *autonomia de projeto* está relacionada à independência do uso do modelo de dados e a técnicas de gerenciamento de transação entre os nós. A *autonomia de comunicação* determina a extensão a qual cada nó pode decidir sobre o compartilhamento de informações com outros nós, e a *autonomia de execução* refere-se à independência dos usuários para atuarem conforme desejarem.

ElsMari e Navathe (2011, p. 596) exemplificam que a autonomia de comunicação se refere à capacidade de decidir se irá se comunicar com outro SBD componente. Autonomia da execução refere-se à capacidade de um SBD componente executar operações locais sem interferência das operações externas por outros SBDs componentes e sua capacidade de decidir a ordem em que serão executadas. A autonomia de associação de SBD de componente implica que ele tem a capacidade de decidir se e quando compartilhar de sua funcionalidade (operações que ele suporta) e recursos (dados que ele gerencia) com outros SBDs componentes.

As relações de confiabilidade e disponibilidade são citadas como potenciais vantagens ao se utilizar banco de dados distribuídos. Elsmari e Navathe (2011, p. 592) definem “a confiabilidade como a probabilidade de um sistema estar funcionando (não parado) em certo ponto no tempo, enquanto a



disponibilidade é a probabilidade de que o sistema esteja continuamente disponível durante um intervalo de tempo”.

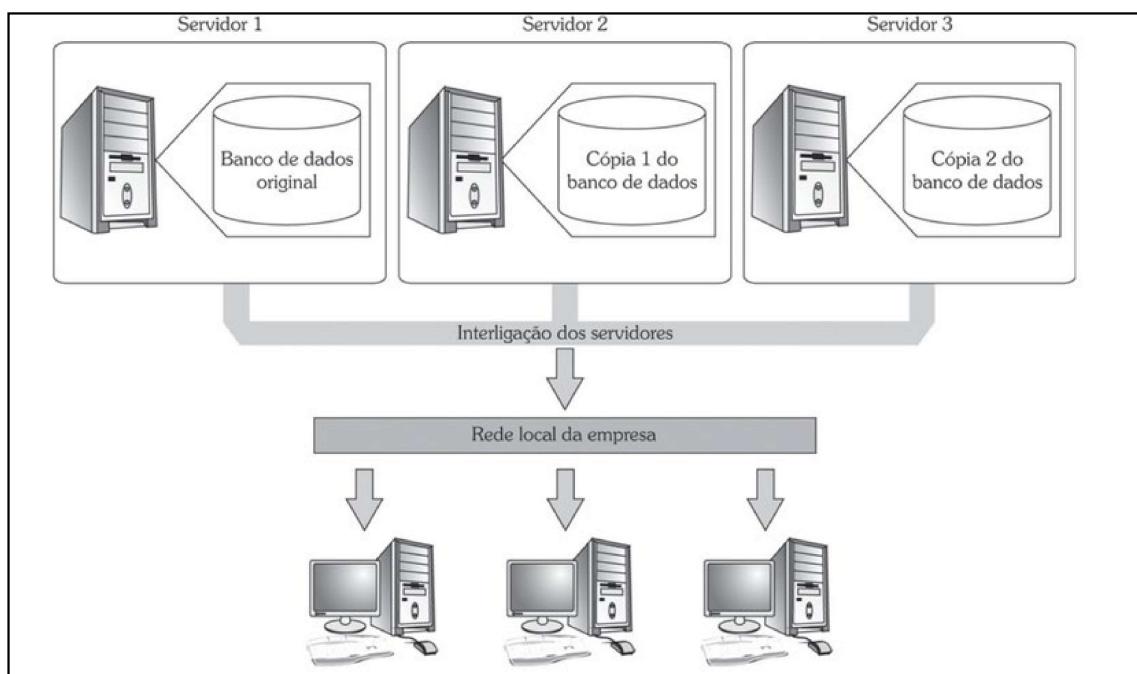
Os autores relacionam a confiabilidade e disponibilidade do banco de dados aos defeitos e erros e falhas associadas a ele, e definem “uma falha pode ser descrita como um desvio de um comportamento do sistema daquele especificado a fim de garantir a execução correta das operações, os erros constituem o subconjunto de estados do sistema as que causam a falha, que por sua vez a falha é a causa de um erro”. (Elsmari e Navathe, 2011, p. 592).

De acordo com Elsmari e Navathe (2011), para que a construção de um sistema seja confiável, podem ser adotadas várias técnicas. Uma técnica citada pelos autores como comum é a tolerância a correções, que reconhece que as falhas ocorreram e projeta mecanismos que podem detectar e remover falhas antes que elas possam resultar em uma falha do sistema. Outra técnica mais rigorosa tenta garantir que o sistema final não tenha falha alguma, sendo realizado por meio de um processo de projeto abrangente, seguido por controle de qualidade de testes abrangente.

Elsmari e Navathe (2011, p. 592) defendem que um SGBDD confiável “tolera falhas dos componentes básicos e processos solicitações do usuário desde que a concorrência do banco de dados não seja violada”. Para isso, é necessário que o gerenciador de recuperação do SGBDD lide com as falhas que surgem das transações *hardware* e redes de comunicação e definem que as “falhas de *hardware* podem ser aquelas que resultam em perda de conteúdo da memória principal ou perda de conteúdo do armazenamento secundário, as falhas de comunicação ocorrem devido a erros associados a mensagens e falhas na linha e os erros de mensagem podem incluir sua perda adulteração ou chegada fora de ordem no destino”.(Elsmari; Navathe, 2011, p. 592)

Alves (2014) defende que outra característica que pode ser encontrada num ambiente de banco de dados distribuído é a replicação de dados (Figura 3), que, conforme o autor, “consiste na duplicação total ou parcial do banco de dados em mais de uma localização ou ponto da rede e é útil para melhorar a disponibilidade dos dados, uma vez que agiliza as consultas globais na recuperação de dados” Alves (2014, p. 138). O autor aponta como efeito colateral o fato de que pode tornar o sistema todo mais moroso nas operações de atualização, as quais devem ser replicadas nas cópias.

Figura 3 – Exemplo de banco de dados replicado



Fonte: Alves, 2014, p. 138.

Outra questão fundamental de um BDD é que, conforme abordado por Alves (2014), o controle de acesso concorrente de usuários e a recuperação de dados são fatores muito importantes num SGDB distribuído e não devem ser relegados a um segundo plano, pois por meio deles é possível tratar alguns problemas que não são encontrados em ambiente centralizado, como apresentado no Quadro 4:

Quadro 4 – Problemas abordados no controle de usuário e recuperação de dados

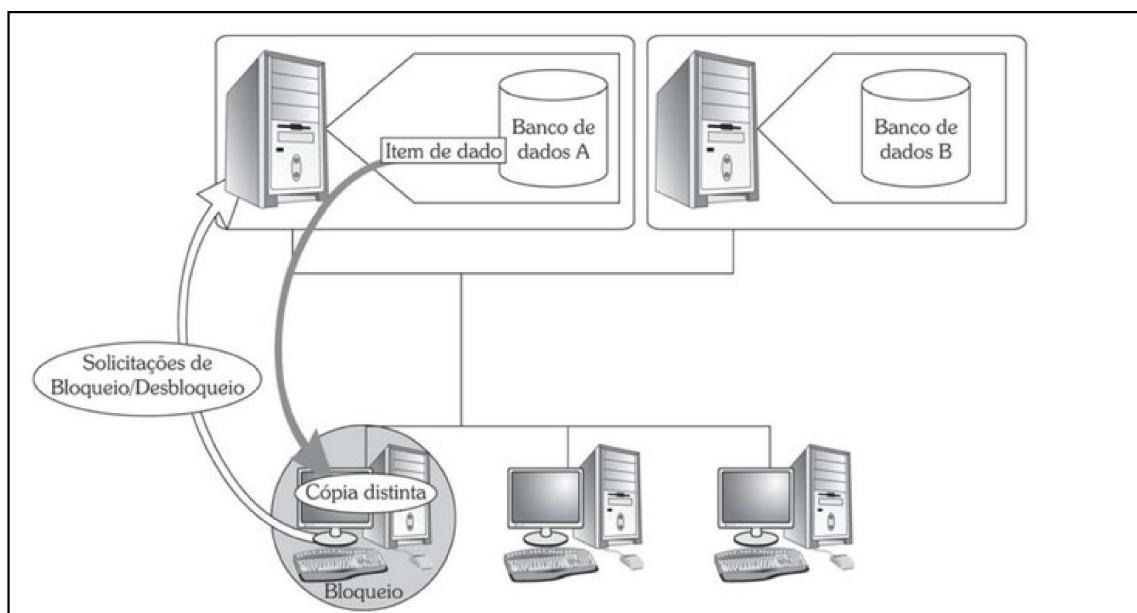
Manipulação de várias cópias de itens de dados:	O controle de concorrência permite manter a consistência dos dados entre as cópias e a recuperação torna uma cópia consistente com as outras se o ponto da rede no qual se encontra a cópia apresentar alguma falha.
Falhas em pontos da rede	Mesmo que um ou mais pontos da rede apresentem falhas e parem de funcionar, o SGBDD deve ter a capacidade de continuar trabalhando normalmente.
Falhas na comunicação	Se ocorrerem falhas na comunicação de dados pela rede, não devem interferir no funcionamento geral do sistema, uma vez que ele deve ter capacidade de lidar com elas.
Falha em commit	O sistema deve ser capaz de lidar com falha de commit de uma transação. O <i>commit</i> encerra uma transação completa.
Deadlock	O sistema deve lidar com problemas relacionados a <i>deadlock</i> , uma situação em que mais de usuário está acessando o mesmo registro simultaneamente.

Fonte: Alves, 2014, p. 140.



Alves (2014, p. 140) define que, para o controle de concorrência, existem dois tipos: “controle baseado em cópia distinta de item de dado e controle baseado em votação”. Segundo o autor, no primeiro tipo se encontram técnicas que são extensões já existentes em bancos de dados centralizados (Alves, 2014). Uma dessas extensões se aplica ao contexto de bloqueio centralizado. A ideia principal é designar cópias distintas de um item de dado e associar a elas os bloqueios. O servidor que contém essa cópia recebe todas as solicitações de bloqueio e desbloqueio, conforme o esquema da Figura 4.

Figura 4 – Controle de concorrência



Fonte: Alves, 2014, p. 141.

Elsmari e Navathe (2011) indicam que algumas funções aumentam a complexidade de um SGBDD em relação a um SGBD centralizado, pois a distribuição leva uma maior complexidade no projeto e implementação do sistema, os autores destacam que o SGBD de precisa ser capaz de oferecer as funções listadas no Quadro 5:



Quadro 5 – Funções adicionais dos bancos de dados distribuídos

Acompanhar a distribuição de dados	A capacidade de acompanhar a distribuição de dados, fragmentação e replicação ao expandir o catálogo do SGBDD.
Processamento de consulta distribuído	A capacidade de acessar sites remotos e transmitir consultas e dados entre os vários sites por meio de uma rede de comunicação.
Gerenciamento de transação distribuído	A capacidade de criar estratégias de execução para consultas e transações que acessem dados de mais de um site e de sincronizar o acesso aos dados distribuídos, mantendo a integridade do banco de dados geral.
Gerenciamento de dados replicado	A capacidade de decidir qual cópia de um item de dados replicado acessar e de manter a consistência das cópias de um item de dados replicado.
Recuperação de banco de dados distribuído	A capacidade de recuperar-se de falhas no site individual e de novos tipos de falhas, como a dos <i>links</i> de comunicação.
Segurança	As transações distribuídas precisam ser executadas com o gerenciamento apropriado da segurança dos dados e dos privilégios de autorização/acesso dos usuários.
Gerenciamento de diretório (catálogo) distribuído	Um diretório contém informações (metadados) sobre os dados no banco de dados. O diretório pode ser global, para o BDD inteiro, o local para cada site. O posicionamento e a distribuição do diretório são reações de projeto e Diretriz.

Fonte: Elsmari; Navathe, 2011, p. 593.

TEMA 3 – VANTAGENS

Elsmari e Navathe (2011) apontam que as organizações lançam mão do gerenciamento de banco de dados distribuídos por diversos motivos e listam as vantagens de acordo com o Quadro 6.



Quadro 6 – Vantagens dos bancos de dados distribuídos

Maior facilidade e flexibilidade de desenvolvimento da aplicação	O desenvolvimento e a manutenção de aplicações em sites geograficamente distribuídos de uma organização são facilitados devido à transparência da distribuição e controle dos dados.
Maior confiabilidade e disponibilidade	Isso é obtido pelo isolamento de falhas ao seu site de origem, sem afetar os outros bancos de dados conectados à rede. Quando os dados e o software de SGBDD são distribuídos por vários sites, um destes pode apresentar falha enquanto outros continuam a operar. Apenas os dados e software que existem no site defeituoso não poderão ser acessados. Isso melhora tanto a confiabilidade quanto à disponibilidade. Uma melhoria ainda maior é obtida pela devida à replicação dos dados e software em mais de um site. Em um sistema centralizado, uma falha em um único site torna o sistema inteiro indisponível a todos os usuários. Em um banco de dados distribuído, alguns dos dados podem ficar inalcançáveis, mas os usuários ainda podem ser capazes de acessar outras partes do banco de dados. Se os dados no site que apresentou falhas tiverem sido duplicados em outro site antes da falha, então o usuário não será afetado de forma alguma.
Maior desempenho	Um SGBD distribuído fragmenta o banco de dados ao manter os dados mais próximos de onde eles são mais necessários. A localização de dados reduz a disputa pela CPU e serviços de E/S e, ao mesmo tempo, reduz os atrasos no acesso envolvido nas redes remotas. Quando um banco de dados grande é distribuído por vários sites, existem bancos de dados menores em cada site. Como resultado, consultas e transações locais que acessam dados em um único site possuem melhor desempenho por causa dos bancos de dados locais menores. Além disso, cada site tem um número menor de transações executando do que se todas as transações fossem submetidas a um banco de dados centralizado. Ainda, o paralelismo entre consultas e dentro da consulta pode ser alcançado ao executar várias consultas em diferentes sites, ou ao desmembrar uma consulta em uma série de subir consultas executadas em paralelo. Isso contribui para melhorar o desempenho.
Expansão mais fácil	Em um ambiente distribuído, a expansão do sistema em matéria de inclusão de mais dados, aumento do tamanho do banco de dados ou inclusão de mais processadores é muito mais fácil.

Fonte: Elsmari; Navathe, 2011, p. 592.

Elsmari e Navathe (2011) destacam que a transparência leva um comprometimento entre a facilidade de uso e o custo de *overhead* da provisão da transparência, devido ao fato de que a transparência total oferece ao usuário global uma visão de SBDD de inteiro como se fosse um único sistema centralizado. A transparência é fornecida como um complemento à autonomia, que dá aos usuários um controle mais preciso sobre os bancos de dados locais. Os recursos de transparência podem ser implementados como uma parte da



linguagem do usuário, que pode traduzir os serviços requisitados em operações apropriadas. Além disso, a transparência afeta os recursos que devem ser fornecidos pelo sistema operacional e pelo SGBD.

Alves (2014, p. 134) complementa, defendendo que “para tirar vantagens reais de um banco de dados distribuído, é necessário que estejam presentes, além das funções comuns a um SGBD centralizado, as seguintes funções extras”:

Quadro 7 – Funções necessárias de um BDD

- Controle de distribuição, fragmentação e replicação dos dados por meio de extensão no catálogo do SGBDD.
- Capacidade de processar consultas distribuídas, acessando, por meio de uma rede de comunicação, localizações remotas para extrair os dados.
- Gerenciamento de consultas e transações que acessam dados a partir de vários locais, ao mesmo tempo em que mantém o acesso sincronizado e a integridade do próprio banco de dados.
- Capacidade de decidir qual cópia de um item de dado que se encontra replicado deve ser considerada.
- Recuperação do banco de dados distribuído quando houver alguma falha ou colapso do sistema (os comumente denominados *crashes*).
- Segurança na execução de transações distribuídas e gerenciamento de privilégios de autorização/acesso por parte dos usuários.
- Gerenciamento do catálogo do banco de dados distribuído, de forma que contenha informação referente às localizações dos bancos. Esse catálogo pode ser geral a todo o banco ou local para cada nó da rede envolvida do processo.

Fonte: Alves, 2014, p. 134.

TEMA 4 – DESVANTAGENS

De acordo com Elsmari e Navathe (2011), as desvantagens do banco de dados distribuídos estão relacionadas ao tipo de heterogeneidade presente nos sistemas de gerenciamento de banco de dados federados e podem surgir de várias fontes, mas geram problemas que estão listados no Quadro 4:



Quadro 8 – Desvantagens causadas por problemas com sistemas de gerenciamentos de bancos de dados federados

Diferenças nos modelos de dados	Os bancos de dados em uma organização vêm de uma série de modelo de dados, incluindo os chamados modelos legados, o modelo de dados relacional, o modelo de dados de objeto e até mesmo arquivos. As capacidades de modelagem variam e lidar com os modelos uniformemente por meio de um único esquema global ou processá-los em uma única linguagem é algo desafiador. Mesmo que dois bancos de dados sejam ambos do ambiente do SGBDR, a mesma informação pode ser representada como um nome de atributo, como o nome de relação ou como um valor de banco de dados diferentes. Isso exige um mecanismo inteligente de processamento de consultas, que possa relacionar informações com base nos metadados.
Diferenças nas restrições	As facilidades de restrição para especificação e a implementação variam de um sistema para outro. Existem recursos comparáveis que devem ser reconciliados na construção de um esquema global. Por exemplo, os relacionamentos dos modelos ER são representados como restrições de integridade referencial ao modelo relacional. Triggers podem precisar ser usados para implementar certas restrições no modelo relacional. O esquema global também deve lidar com conflitos em potencial entre as restrições.
Diferenças nas linguagens de consulta	Até com o mesmo modelo de dados, as linguagens e suas versões variam. Por exemplo: a SQL tem diversas versões, como SQL-89, SQL-92, SQL-99 e SQL:2008, e cada sistema tem o próprio conjunto de tipos de dados, operadores de comparação, recursos de manipulação de strings, etc.
Heterogeneidade semântica	Ocorre quando existem diferenças no significado, interpretação e o uso intencionado dos mesmos dados ou dados relacionados. A heterogeneidade semântica entre os sistemas de bancos de dados componentes criam maior obstáculo do projeto de esquemas globais de bancos de dados heterogêneos.
Autonomia de projeto do sistema de bancos de dados	Refere-se à liberdade de escolher os parâmetros de projeto, como: representação e a nomeação; universo de discurso do qual os dados são retirados; o conhecimento significado interpretação subjetiva dos dados; as restrições de transação e de diretriz; e as derivações de resumos. Parâmetros que irão afetar a complexidade do sistema de banco de dados federados.

Fonte: Elsmari; Navathe, 2011, p. 595.

Ainda analisando problemas que trazem desvantagens, Elsmari e Navathe (2011), ao realizarem uma visão geral sobre o controle de concorrência e recuperação em banco de dados distribuídos, elencaram diversos problemas que não são encontrados em um ambiente SGBD centralizado, conforme apresentado no Quadro 9.



Quadro 9 – Desvantagens causadas por problemas encontrados no controle de concorrência e recuperação de um BDD

Lidar com múltiplas cópias dos itens de dados	O método de controle de concorrência é responsável por manter a consistência entre essas cópias. O método de recuperação é responsável por tornar uma cópia coerente com outras cópias se o site em que a cópia é armazenada falhar e recuperar mais tarde.
Falha de sites individuais	O SGBDD deve continuar a operar com seus sites em execução, se possível, quando um ou mais sites individuais falharem. Quando um site se recupera, seu banco de dados local precisa ser atualizado com o restante dos sites antes que se junte novamente ao sistema.
Falha dos links de comunicação	O sistema precisa ser capaz de lidar com a falha de um ou mais dos <i>links</i> de comunicação que conectam os sites. Um caso extremo desse problema é que pode haver particionamento da rede. Isso divide os sites em duas ou mais partições, onde os sites dentro de cada partição só podem se comunicar entre si e não com sites de outras partições.
Confirmação distribuída	Pode haver problemas com a confirmação de uma transação que está acessando bancos de dados armazenados em vários sites se alguns destes falharem durante o processo de confirmação.
Deadlock distribuído	Pode haver <i>deadlock</i> entre vários sites, de modo que as técnicas para lidar com os <i>deadlocks</i> precisam ser estendidas para levar isso em consideração.

Fonte: Elsmari; Navathe, 2011, p. 613.

Elsmari e Navathe (2011) destacam que, apesar de serem fatores problemáticos, as técnicas de controle de concorrência e recuperação precisam lidar com esses e outros problemas.

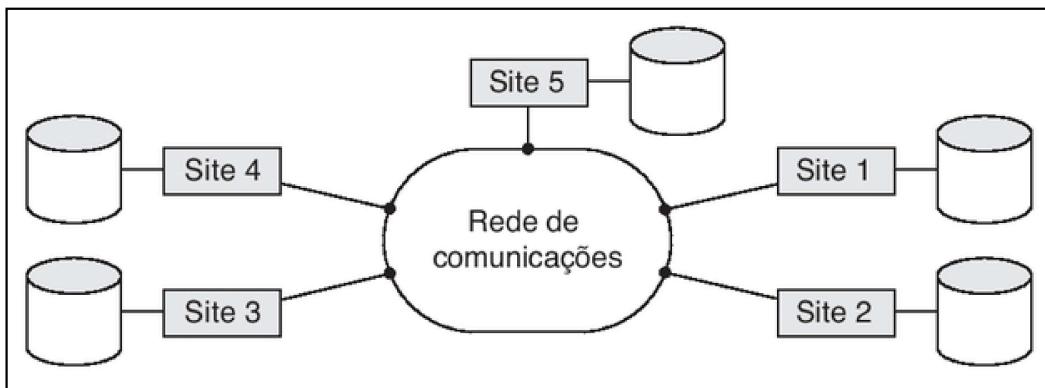
TEMA 5 – ARQUITETURA DE BDD

De acordo com Alves (2014), o modelo de banco de dados distribuído é fortemente baseado na arquitetura cliente/servidor e utiliza muito dos conceitos empregados nela. É comum falar em sistemas multiprocessados quando se trata de bancos de dados distribuídos.

A figura 5 representa a arquitetura de um banco de dados verdadeiramente distribuído.



Figura 5 – Arquitetura de BDD



Fonte: Elsmari; Navathe, 2011, p. 597.

Para compreender a distinção entre *arquiteturas de bancos de dados paralela* e *arquitetura de banco de dados distribuída*, Elsmari e Navathe (2011) utilizam exemplos que abordam essas relações. Os autores destacam que, embora ambas as arquiteturas estejam presentes na indústria atualmente, existem manifestações das arquiteturas distribuídas que estão continuamente evoluindo entre as grandes empresas. Eles reiteram que arquitetura paralela é a mais comum na computação de alto desempenho, onde há uma necessidade de arquiteturas multiprocessadas para enfrentar o volume de dados passando por aplicações de processamento de transação e *warehousing*.

Segundo Elsmari e Navathe (2011), são comumente utilizados dois tipos principais de arquiteturas de sistemas multiprocessador, representados no Quadro 10.

Quadro 10 – Dois tipos de arquiteturas de sistema multiprocessador

Arquitetura de memória compartilhada (altamente acoplada)	Arquitetura de disco compartilhado (livremente acoplada)
Múltiplos processadores compartilham armazenamento secundário disco e também memória principal.	Múltiplos processadores compartilham armazenamento secundário (disco), mas cada um tem a própria memória principal.

Fonte: Elsmari; Navathe, 2011, p. 596.

De acordo com Elsmari e Navathe (2011), essas arquiteturas permitem que os processadores se comuniquem sem *overhead* de trocar mensagens por uma rede. Os sistemas de gerenciamento de bancos de dados desenvolvidos que utilizam esses tipos de arquitetura são chamados de *sistemas de*

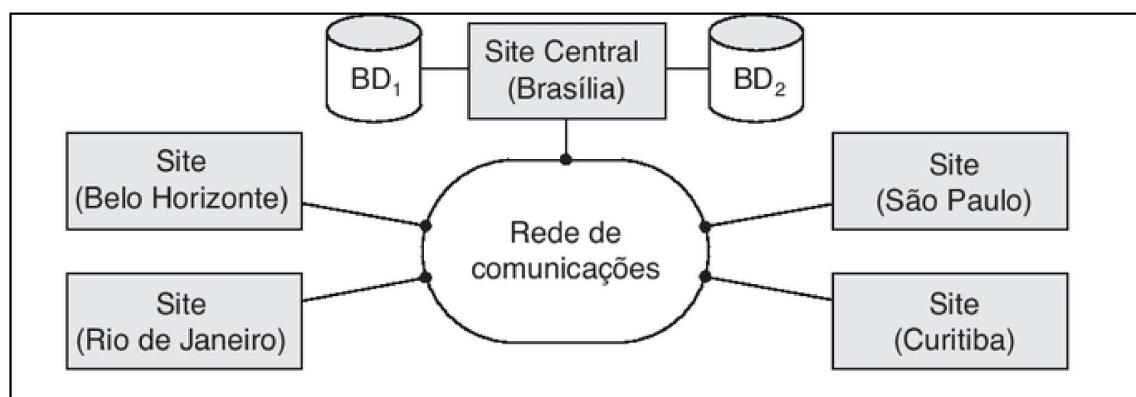


gerenciamento de banco de dados paralelo, em vez de SGBDDs, pois utilizam a tecnologia de processadores paralelos. Para Alves (2014, p. 134), os bancos desenvolvidos com base em uma dessas arquiteturas, representadas na tabela 4, são normalmente denominados “Sistemas de Bancos de Dados Paralelos, uma vez que é utilizado processamento paralelo”.

De acordo com Alves (2014), a outra arquitetura que pode ser encontrada quando se fala em bancos de dados distribuídos é a denominada de *processamento distribuído*, em que cada processador possui sua própria memória principal e seu meio de armazenamento secundário. Eles se comunicam por meio de uma rede de alta velocidade para executar suas tarefas de processamento. A diferença principal entre essa arquitetura e um modelo de banco de dados distribuído é o fato de todos os componentes serem homogêneos (de um mesmo fabricante ou fornecedor), o que não ocorre com bancos distribuídos, em que é comum a existência de plataformas de *hardware* e de *software* totalmente diferentes entre si.

A Figura 6 representa um banco de dados centralizado com acesso distribuído.

Figura 6 – Arquitetura em rede com banco de dados centralizado em um dos sites com acesso distribuído



Fonte: Elsmari; Navathe, 2011, p. 597.

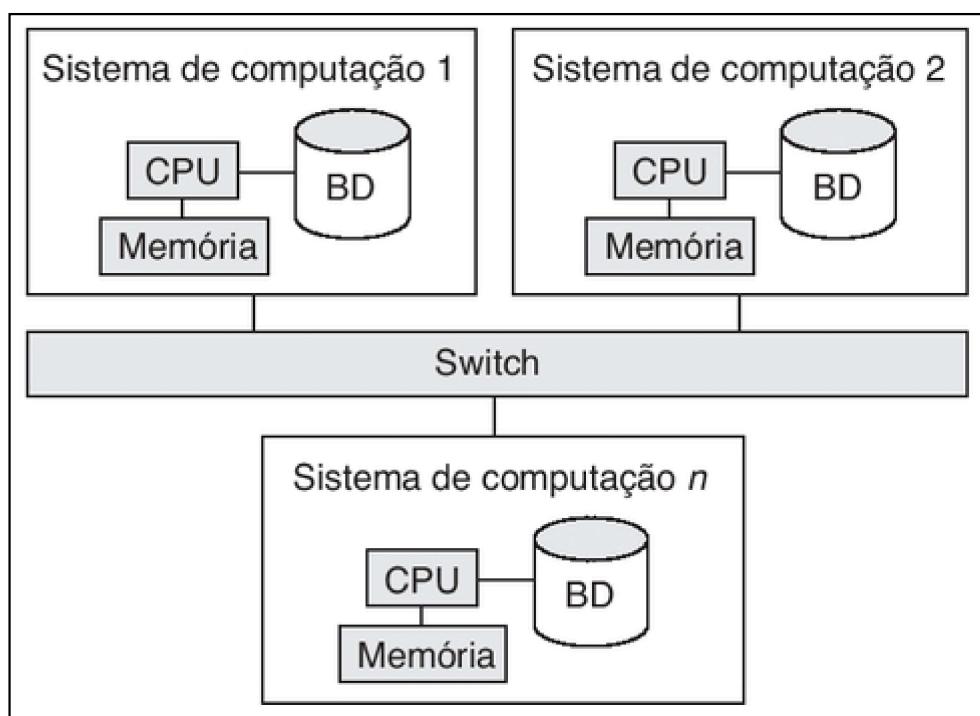
Outro tipo de arquitetura de processador, de acordo com Elsmari e Navathe (2011, p. 596), “é chamado de arquitetura nada compartilhado, cada processador tem a própria memória principal e secundária (disco), não existe memória comum e os processadores se comunicam por uma rede de interconexão de alta velocidade (barramento ou switch) ”.



Embora na *arquitetura nada compartilhado* seja semelhante ao ambiente de computação de banco de dados distribuídos, existem diferenças importantes no modo de operação. Nos sistemas de multiprocessador nada compartilhado, assimetria e homogeneidade de nós, isso não acontece no ambiente de banco de dados distribuído, no qual a heterogeneidade do *hardware* e do sistema operacional em cada nó é muito comum. A *arquitetura nada compartilhado* também é considerada um ambiente para banco de dados paralelos.

A Figura 7 ilustra um banco de dados paralelo (nada compartilhado).

Figura 7 – Arquitetura nada compartilhado



Fonte: Elsmari; Navathe, 2011, p. 597.

Alves (2014) salienta que os locais em que se encontram distribuídos os bancos podem estar fisicamente próximos (como num mesmo edifício) ou geograficamente distantes (bairro, cidade, estado ou mesmo país), conectados por meio de uma rede WAN, montada sobre uma estrutura VPN (Internet em banda larga ASDL, ISDN, a cabo ou a rádio) ou com a tecnologia *Frame Relay* (que tem um custo muito mais elevado que o da VPN). O autor defende que “a escolha do tipo e da topologia de rede é de altíssima importância, uma vez que esses itens podem ter grande influência no desempenho dos processamentos de consultas num sistema distribuído” (Alves, 2014, p. 135).



FINALIZANDO

Nesta aula, abordamos as questões pertinentes aos bancos de dados distribuídos. Esse é um assunto muito amplo e aqui discutimos algumas das técnicas básicas usadas com os bancos de dados distribuídos, abordamos alguns requisitos e foram definidos conceitos considerados essenciais para o BDD. Apresentamos algumas vantagens em potencial e também algumas relações que apresentam desvantagem quanto à utilização dos bancos de dados distribuídos, pois alguns problemas adicionais devem ser tratados em ambiente distribuído, que não aparecem no ambiente centralizado. Em relação à arquitetura, apresentamos a distinção da arquitetura de sistemas paralela e distribuída e também uma arquitetura genérica dos bancos de dados distribuídos e alguns aspectos relacionados ao tema.



REFERÊNCIAS

- ALVES, W. P. **Banco de dados**. São Paulo: Érica, 2014.
- ELSMARI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. 6. ed. São Paulo: Pearson Education do Brasil, 2011.
- RAMAKRISHNAN, R.; GEHRKE, J. **Sistemas de gerenciamento de bancos de dados**. 3. ed. Porto Alegre: McGraw Hill, 2008.