

Bancos de Dados Distribuídos



1

Conteúdo

- Introdução aos BDs Distribuídos
- Processamento de Consultas Distribuídas
- Transações Distribuídas
- Controle de Concorrência

2

Introdução aos Bancos de Dados Distribuídos

- Bancos de Dados Distribuídos
- Arquitetura de BDs Distribuídos
- Armazenamento dos Dados

3

Bancos de Dados

- Panorama Atual
 - Crescimento do número de usuários
 - Crescimento da quantidade de consultas
 - Automatização de todos os processos dentro de uma empresa ou instituição
 - Maior dependência dos Bancos de Dados
 - Novos tipos de dados, como som e imagem, exigem maior poder de processamento e armazenamento
- Soluções
 - Usar processamento paralelo e distribuído para processar as consultas em bancos de dados

4

Processamento Paralelo

- O que é?
 - Consiste em executar simultaneamente várias partes de um mesmo processo ou aplicação
 - Tornou-se possível a partir do desenvolvimento de sistemas operacionais multi-tarefa, multi-thread e paralelos
- Processos são executados paralelamente:
 - Em um mesmo processador
 - Em uma máquina multiprocessada
 - Em um cluster (máquinas interligadas por uma rede local que se comportam como uma só máquina)

5

Processamento Distribuído

- O que é?
 - Consiste em executar processos / aplicações cooperantes em máquinas diferentes
 - Tornou-se possível a partir da popularização das redes de computadores
- Processos são executados em máquinas diferentes interligadas por uma rede
 - Redes locais
 - Internet
 - Outras redes públicas ou privadas

6

Diferenças

- Acoplamento
 - Sistemas paralelos são fortemente acoplados: compartilham hardware ou se comunicam através de um barramento de alta velocidade
 - Sistemas distribuídos são fracamente acoplados
- Previsibilidade
 - O comportamento de sistemas paralelos é mais previsível; já os sistemas distribuídos são mais imprevisíveis devido ao uso da rede e a falhas

7

Diferenças

- Influência do Tempo
 - Sistemas distribuídos são bastante influenciados pelo tempo de comunicação pela rede; em geral não há uma referência de tempo global
 - Em sistemas paralelos o tempo de troca de mensagens pode ser desconsiderado
- Controle
 - Em geral em sistemas paralelos se tem o controle de todos os recursos computacionais; já os sistemas distribuídos tendem a empregar também recursos de terceiros

8

Vantagens

- Usam melhor o poder de processamento
- Apresentam um melhor desempenho
- Permitem compartilhar dados e recursos
- Podem apresentar maior confiabilidade
- Permitem reutilizar serviços já disponíveis
- Atendem um maior número de usuários
- ...

9

Dificuldades

- Desenvolver, gerenciar e manter o sistema
- Controlar o acesso concorrente a dados e a recursos compartilhados
- Evitar que falhas de máquinas ou da rede comprometam o funcionamento do sistema
- Garantir a segurança do sistema e o sigilo dos dados trocados entre máquinas
- Lidar com a heterogeneidade do ambiente
- ...

10

Classificação dos SGBDs

- SGBDs Locais
 - Acesso ao banco apenas na máquina na qual está instalado o sistema
- SGBDs Cliente-Servidor
 - O servidor pode ser acessado remotamente (através da rede)
- SGBDs Paralelos
 - O servidor utiliza uma máquina paralela (com vários processadores)
- SGBDs Distribuídos
 - Dados armazenados em vários servidores interligados por uma rede de comunicação

11

Bancos de Dados Paralelos

- Máquinas paralelas vem sendo usadas para suportar uma carga maior de trabalho dos SBDs
 - Vários processadores executam as operações em paralelo → uso de controle de concorrência
 - Memória e disco podem ser compartilhados ou não
 - Custo de máquinas multi-processadas está caindo
- Sistemas paralelos são usados para aumentar:
 - Escala: processar mais transações
 - Desempenho: tornar o processamento mais rápido

12

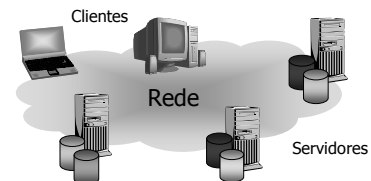
Bancos de Dados Paralelos

- Problema: mais hardware → mais falhas
 - Solução: replicar os dados → controlar consistência
- Paralelismo pode ser usado:
 - Na entrada e saída de dados (I/O)
 - No processamento de consultas
 - No processamento de operações individuais

13

Bancos de Dados Distribuídos

- BDs Distribuídos
 - Os dados estão distribuídos por várias localidades
 - BDs distribuídos se comunicam via rede para acessar os dados e processar as consultas



14

Bancos de Dados Distribuídos

- BDs distribuídos x centralizados
 - BDs centralizados possuem um ponto único de falha
 - BDs distribuídos podem aumentar a robustez do sistema e a disponibilidade dos dados
 - BDs centralizados são limitados pela capacidade de processamento e armazenamento de uma máquina
 - BDs distribuídos podem crescer em escala adicionando novos servidores ao sistema
 - BDs distribuídos são mais sujeitos a apresentar falhas parciais de funcionamento e de segurança
 - BDs distribuídos são mais difíceis de administrar pois os servidores estão em locais diferentes

15

Bancos de Dados Distribuídos

- BDs distribuídos x cliente-servidor
 - BDs distribuídos também podem ser acessados remotamente por clientes
- BDs distribuídos x paralelos
 - Ambos podem processar consultas em paralelo usando os vários processadores disponíveis
 - Em BDs paralelos os processadores podem trocar dados usando discos ou memória compartilhada
 - O uso da rede em BDs distribuídos pode prejudicar o desempenho do sistema ao processar consultas
 - BDs paralelos são mais vulneráveis a algumas falhas, e não são tão escaláveis quanto BDs distribuídos

16

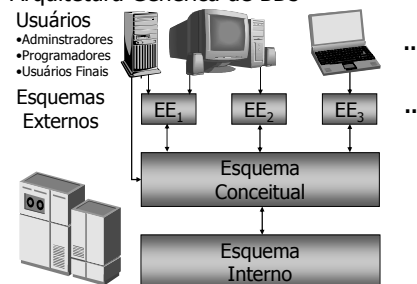
Bancos de Dados Distribuídos

- BDs distribuídos podem ser classificados como:
 - Homogêneos: todos os *sites* usam o mesmo *software*
 - Heterogêneos
 - Usam *software* diferente
 - Podem usar esquemas de dados diferentes
 - Linguagem de consulta pode ser diferente
- Dificuldades a serem superadas
 - Controlar a consistência dos dados armazenados
 - Processar consultas e transações de modo distribuído
 - Controlar a concorrência e resolver conflitos
 - Conciliar as diferenças em sistemas heterogêneos

17

Arquitetura de BDs Distribuídos

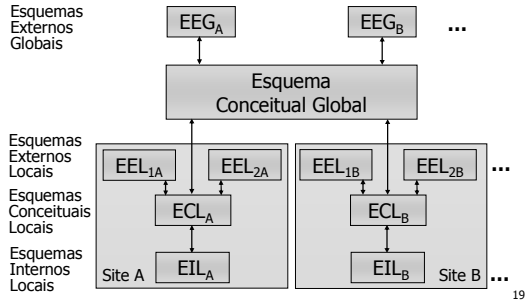
- Arquitetura Genérica de BDs



18

Arquitetura de BDs Distribuídos

Arquitetura Genérica de BDs Distribuídos



19

Arquitetura de BDs Distribuídos

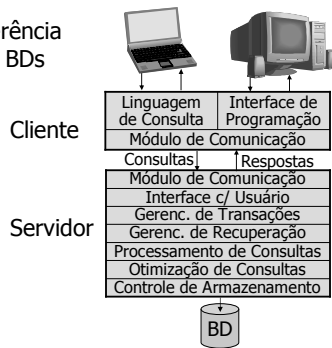
Usuários de BDs Distribuídos

- Tipos de Usuários
 - Administradores locais ou globais
 - Programadores locais ou globais
 - Usuários finais locais ou globais
- Usuários Globais
 - Possuem visão global do sistema
 - Visualizam um esquema externo global
- Usuários Locais
 - Acessam diretamente o servidor local
 - Visualizam um esquema externo local

20

Arquitetura de BDs Distribuídos

Gerência de BDs



21

Arquitetura de BDs Distribuídos

Problemas em BDs distribuídos

- Otimização e processamento de consultas/transações distribuídas requer algoritmos/protocolos adequados
- Mecanismos de controle e gerenciamento devem trabalhar de maneira integrada

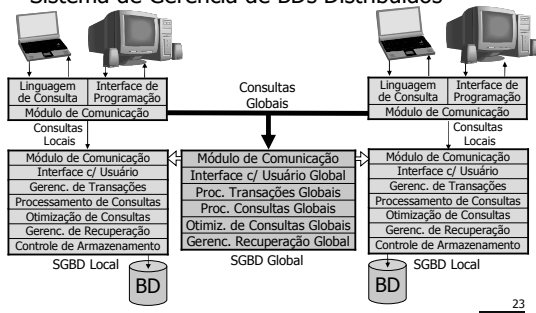
Problemas em BDDs heterogêneos

- Precisamos conciliar as diferenças entre:
 - Modelos lógicos
 - Linguagens de definição e manipulação de dados
 - Formatos de dados: língua, ordenação dos bits, tamanho dos tipos de dados e representação na memória, tabelas de caracteres, etc.

22

Arquitetura de BDs Distribuídos

Sistema de Gerência de BDs Distribuídos



23

Armazenamento dos Dados

Em BDs distribuídos os dados podem ser:

- Replicados
- Fragmentados
- Replicados e Fragmentados

Replicação de Dados

- Uma mesma tabela pode ser armazenada em mais de um servidor
- Vantagens: aumenta a disponibilidade e o paralelismo
- Desvantagem: atualizações devem ser feitas em todos os servidores para manter consistência entre réplicas
- Apresenta bom desempenho nas operações de leitura, mas causa *overhead* nas operações de escrita

24

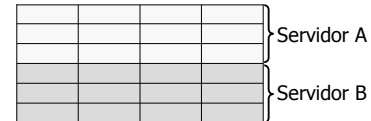
Armazenamento dos Dados

- Fragmentação de Dados
 - Os dados mantidos em uma tabela podem ser divididos em dois ou mais fragmentos
 - Cada fragmento é armazenado em servidor do banco de dados distribuído
 - Métodos usados para fragmentação:
 - Fragmentação horizontal
 - Fragmentação vertical
 - Fragmentação mista

25

Armazenamento dos Dados

- Fragmentação Horizontal
 - Cada fragmento possui um conjunto de tuplas da relação
 - Cada tupla de uma relação precisa ser armazenada em pelo menos um servidor
 - A relação completa pode ser obtida fazendo a união dos fragmentos



26

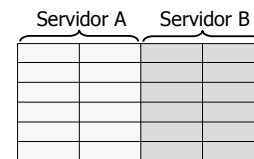
Armazenamento dos Dados

- Técnicas para fragmentação horizontal:
 - *Round-Robin*
 - Tuplas distribuídas uma a uma entre os servidores
 - Divide os dados igualmente
 - *Hash*
 - Usa uma função *hash* para determinar o servidor
 - Boa para processar seleções de igualdade
 - Por faixa
 - Cada servidor armazena as tuplas com um certo atributo dentro de uma faixa de valores
 - Boa para processar seleções de igualdade e para procurar por faixas de valores

27

Armazenamento dos Dados

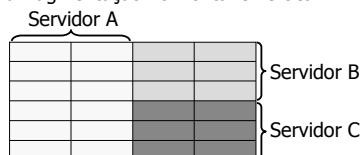
- Fragmentação Vertical
 - Relações são decompostas em conjuntos de atributos mantidos em servidores diferentes
 - Cada fragmento é uma projeção da relação completa
 - A relação completa pode ser obtida fazendo a junção de todos os fragmentos



28

Armazenamento dos Dados

- Fragmentação Mista
 - Combina fragmentação horizontal e vertical



- Fragmentação e Replicação de Dados
 - Dados são fragmentados horizontal ou verticalmente
 - Cada fragmento é mantido em mais de um servidor

29