


cin.ufpe.br



Centro de Informática
U · F · P · E

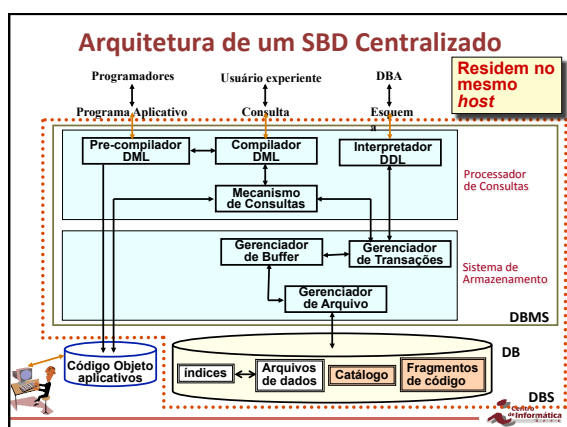
Bancos de Dados Distribuídos e Móveis
Aula 01

Ana Carolina Salgado -acs@cin.ufpe.br
Bernadette Farias Lóscio -bfl@cin.ufpe.br

UNIVERSIDADE FEDERAL DE PERNAMBUCO

Revisão: conceitos básicos de BD

- Um **banco de dados** é uma coleção de dados interrelacionados.
- Um **sistema gerenciador de banco de dados** é um componente de software projetado para garantir o armazenamento e o gerenciamento de bancos de dados.
 - Acesso
 - Controle de Concorrência
 - Recuperação
 - Armazenamento



Motivação para distribuição

- **SBD centralizado**
 - Reside em um **único sistema computacional**
 - Hardware (CPU + Memória + Disco)
 - Sistema operacional
 - Exemplo
 - Mainframes
- **Cenário década de 80**
 - Empresa globalizada (moderna)
 - Departamentos e/ou filiais espalhados pelo mundo
 - Manipulação de grandes volumes de dados
 - Computação departamental (*downsizing*)



Processamento Distribuído

- **Sistema computacional distribuído**
 - Conjunto de dispositivos computacionais
 - **Interconectados** por uma rede de comunicação
 - **Cooperam** na execução de suas tarefas
 - São capazes de executar programas de forma **autônoma**



Processamento Distribuído

- **Critérios de distribuição**
 - **Função**
 - Funções de um sistema computacional podem ser delegadas a diversos módulos de software e de hardware
 - **Controle**
 - O controle de várias tarefas pode ser distribuído em diferentes sistemas computacionais
 - **Dados**
 - Dados utilizados por várias aplicações podem ser distribuídos em diferentes sites de processamento



Distribuição em SBDs

- SBD cliente-servidor (Distribuição de função)
 - Tecnologia cliente-servidor
 - Identificar as funcionalidades a serem suportadas
 - Dividir as funções em duas classes:
 - Funções do servidor (server functions)
 - Funções do cliente (client functions)
 - Tecnologia de banco de dados
 - Distribuir funcionalidades do DBMS entre
 - servidor
 - clientes

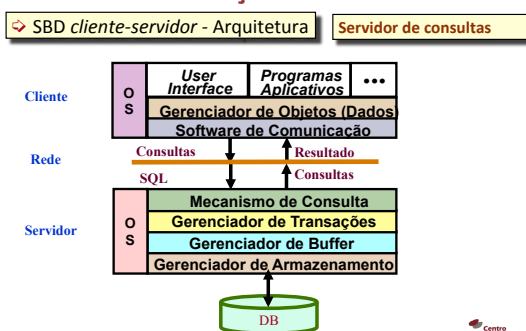


Distribuição em SBDs

- SBD cliente-servidor (Distribuição de função)
 - Funcionalidades do servidor (*back-end*)
 - Processamento e otimização de consulta
 - Gerenciamento de transações
 - Gerenciamento de armazenamento
 - Funcionalidades dos clientes (*front-end*)
 - Execução de programas aplicativos
 - Interface com usuário
 - Módulo do DBMS
 - Gerenciamento dos dados armazenados no cache dos clientes
 - » Gerenciamento de bloqueios (carregados no cache)
 - » Consistência de cache (*cache coherence*)



Distribuição em SBDs



Distribuição em SBDs

- SBD cliente-servidor - Arquitetura
 - Como compatibilizar o acesso de um aplicativo a servidores (SGBD)?
 - Resp: Usando Open Database Connectivity – ODBC
 - Padrão de APIs para garantir a conectividade entre clientes e servidores de BDs
 - Permite que clientes gerem comandos SQL para serem enviados para execução nos servidores
 - Aplicações de BDs fazem chamadas a funções ODBC
 - As funções são específicas para cada SGBD (Oracle, Informix, etc ...)
 - Qualquer cliente que possua uma interface ODBC pode se conectar a qualquer servidor que suporte a interface



Distribuição em SBDs

- SBDs Paralelos (Distribuição de controle)
 - Tecnologia de processamento paralelo
 - Explorar computadores com múltiplos processadores
 - Executar aplicativos utilizando-se os múltiplos processadores cooperativamente
 - Melhorar performance
 - Maior grau de paralelismo
 - Menor tempo de resposta
 - Tecnologia de banco de dados
 - Aumentar a largura de banda de I/O através de paralelismo



Distribuição em SBDs

- SBDs Paralelos
 - Pontos críticos no desenvolvimento de SBDs paralelos
 - Alocação de dados
 - Processamento de consulta paralelo
 - Otimização de consulta paralela
 - Alocação (n discos)
 - Particionamento
 - Fragmentação horizontal
 - Subconjunto de tuplas



Distribuição em SBDs

- **SBDs Distribuídos** (Distribuição de dados)
 - Distribuição de dados
 - Através de departamentos
 - Possibilita o armazenamento de dados onde eles são gerados e » mais utilizados
 - Torna mais eficiente
 - Manutenção
 - Acesso mais eficiente (quem mais utiliza)
 - Manutenção de cópias de dados em diferentes sites
 - Segurança
 - Disponibilidade de dados
 - Sistemas de BDs distribuídos
 - Administrar e manipular dados distribuídos geograficamente e administrativamente através de diversos SBDs

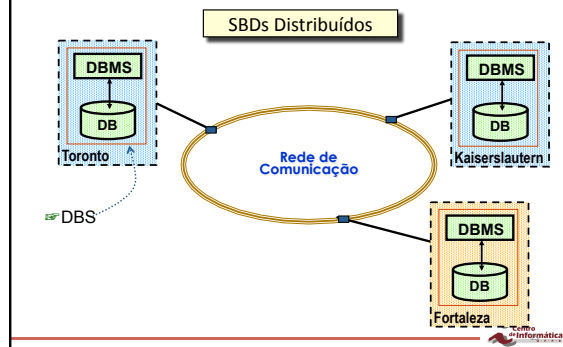


Distribuição em SBDs

- **SBDs Distribuídos**
 - BDs distribuídos
 - Conjunto de DBs (DDBs)
 - Logicamente interrelacionados
 - Distribuídos através de uma rede de computadores
 - DBMS distribuído (DDBMS)
 - Componente de software que permite o gerenciamento do DDBS
 - Torna a distribuição transparente para o usuário
 - Acesso
 - Controle de Concorrência
 - Recuperação
 - Armazenamento
 - Uma cópia do DBMS distribuído em cada site



Distribuição em SBDs



Distribuição em SBDs

- **Exemplo - Oracle8**
 - SBDD para o Banco MyBank
 - Agências em Fortaleza, Kaiserslautern e Toronto
 - Para cada agência
 - Sistema de Banco de Dados
 - Cópia do SGBD Distribuído (Oracle8)
 - Banco de dados
 - » Esquema local
 - Tabela Contas_Agência
 - Software de comunicação (rede)
 - » Compatibilizar a comunicação entre os diversos SBDs distribuídos na rede



Distribuição em SBDs

- **Exemplo - Oracle8 (cont.)**
 - Tabela Agência reside no site Toronto
 - Agência(cod-ag, nome, gerente, ender, cidade)
 - Cada SBD no SBDD
 - Representa um serviço de banco de dados
 - disponível na rede
 - Nome único que identifique o serviço
 - Nome único do serviço de BD Oracle
 - Nome global do banco de dados



Distribuição em SBDs

- **Exemplo - Oracle8 (cont.)**
 - Composição do nome global do banco de dados
 - Primeiro componente
 - Nome local do BD (definido durante sua criação)
 - MyBank
 - » Fortaleza, Toronto, Kaiserslautern
 - Segundo componente
 - Domínio da rede de banco de dados
 - MyBank
 - Nomes globais no domínio MyBank
 - Fortaleza.MyBank
 - Toronto.MyBank
 - Kaiserslautern.MyBank



Distribuição em SBDs

- **Exemplo - Oracle8 (cont.)**
 - No SBD do site Fortaleza
 - Consultar contas corrente da agência Toronto com saldo maior que \$ 5000


```
SELECT * FROM Contas_Agência@Toronto.MyBank
WHERE saldo > 5000
```
 - No SBD Kaiserslautern
 - Listar número de conta corrente e nome da agência de contas em Fortaleza com saldo menor que \$990
 - Lembre-se que a tabela Agência reside em Toronto


```
SELECT f.num-conta, t.nome-ag
FROM Contas_Agência@Fortaleza.MyBank f,
Agência@Toronto.MyBank t
WHERE f.cod-ag=t.cod-ag and f.saldo < 990
```



Distribuição em SBDs

- SBDs Distribuídos
 - **Desafios**
 - Projeto de BDs distribuídos
 - Localização dos dados
 - » Distribuir tabelas
 - » Fragmentar (particionar) tabelas e distribuir fragmentos ou
 - » Replicar
 - Processamento de consulta distribuído
 - Localização de dados
 - Custos de comunicação
 - Benefícios de paralelismo (intra-query)



Distribuição em SBDs

- Integração de SBDs Heterogêneos e Distribuídos
 - **Integrar múltiplos SBDs**
 - Distribuídos
 - Heterogêneos
 - Autônomos
 - **Heterogeneidade**
 - Suporte tecnológico
 - Hardware
 - Sistema operacional
 - Protocolo de rede
 - DBMS
 - Modelo de dados
 - Linguagem de consulta
 - Mecanismo de transações

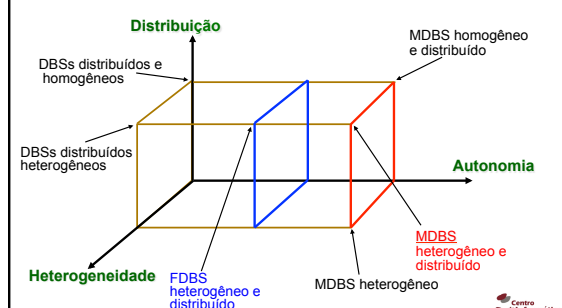


Distribuição em SBDs

- Integração de SBDs Heterogêneos e Distribuídos
 - **Autonomia**
 - Autonomia de projeto
 - Modelo de dados
 - Linguagem de consulta
 - Protocolos de controle de concorrência, recovery/logging, commit
 - Autonomia de comunicação
 - Como e quando responder
 - Autonomia de execução
 - Total controle sobre a execução de transações
 - **Propostas**
 - Sistemas de Bancos de Dados Federados (FDBSs)
 - Sistemas de Banco de Dados Múltiplos (MDBSs)



Distribuição em SBDs



Sistema de BD Distribuído

- **Banco de Dados Distribuído**
 - Coleção de múltiplos bancos de dados logicamente relacionados e distribuídos ao longo de uma rede de computadores
 - Coleção de dados que **pertence logicamente a um sistema**, mas que **fisicamente** está **espalhada** por vários computadores, possivelmente localizados em diferentes locais, conectados por uma rede de comunicação
- **Sistema Gerenciador de Bancos de Dados Distribuídos**
 - Software que gerencia um banco de dados distribuído enquanto torna a distribuição transparente para o usuário



Características dos SGBDD

- A independência de dados
 - Aplicações e visões dos usuários são **imunes a mudanças** nas formas de armazenamento de dados
 - A **forma de armazenamento** é independente da ótica de uso **das aplicações e dos usuários**



Características dos SGBDD

- Transparência de distribuição
 - Aplicações e usuários não precisam conhecer a **localização** dos bancos de dados nas redes de computadores ou lidar com os **protocolos utilizados** na comunicação entre os diferentes sistemas
 - As **diferenças** entre as formas de representação de dados utilizados pelos sistemas componentes devem ser **resolvidas pelo SGBDD**



Características dos SGBDD

- Transparência de fragmentação
 - Aplicações e usuários têm sempre uma **visão única dos dados**, independente da maneira como esses dados estão partidos nos diferentes servidores
- Transparência de réplicas
 - Usuários e aplicações possuem a **visão de que só existe uma única cópia dos dados**
 - A **consistência** entre as cópias deve ser mantida pelo SGBDD



Características dos SGBDD

- SGBDD ao receber **requisição de operação** definida sobre visão externa do esquema global
 - Transforma em uma operação sobre o **esquema global**
 - **Identifica os locais** onde estão os dados correspondentes
 - Deriva as **operações locais** necessárias
 - Emite as requisições de **execução** para os **diferentes nós** (transação distribuída)



Características dos SGBDD

- SGBDD ao receber requisição de operação definida sobre visão externa do esquema global (Cont.)
 - **Coordena** o processo de **execução**
 - **Compõe resultados** parciais obtidos dos diferentes nós
 - **Apresenta de forma integrada** à aplicação do usuário



Razões para Distribuição

- **Natureza distribuída** de algumas aplicações
 - Um Banco / Companhia tipicamente tem várias agências/filiais
 - Dados específicos a cada agência / filial podem ser mantidos na própria agência / filial
 - Usuários de cada agência / filial acessam apenas os dados da própria agência / filial - Usuários locais



Razões para Distribuição

- **Natureza distribuída** de algumas aplicações (Cont.)
 - A direção geral do Banco / Companhia pode precisar acessar os dados armazenados em várias dessas agências / filiais - Usuários globais



Razões para Distribuição

- Disponibilidade e Confiabilidade aumentadas
 - **Confiabilidade** - probabilidade do sistema estar no ar em um momento particular no tempo
 - **Disponibilidade** - probabilidade do sistema estar continuamente disponível durante um dado intervalo de tempo



Razões para Distribuição

- Permite **compartilhamento**, ainda mantendo controle local
 - Controle de dados e software pode ser feito localmente, enquanto outros dados podem ser acessados remotamente



Razões para Distribuição

- **Performance** melhorada - Um BD grande é distribuído em BD locais menores
 - Consultas e transações locais acessam menor volume de dados
 - Cada local (site) terá um número menor de transações executando do que se todas estivessem sendo executadas em um único local



Razões para Distribuição

- **Performance** melhorada - Um BD grande é distribuído em BD locais menores (Cont.)
 - Transações que acessam vários locais podem ter o processamento em diferentes locais prosseguindo em paralelo



Distribuição acrescenta complexidade ao SGBD

- **Acessar locais remotos e transmitir** consultas e dados entre os diferentes nós via uma rede de comunicação
- Guardar detalhes da distribuição dos dados e da replicação deles no **catálogo do SGBD**
- Ter **estratégias de execução de consultas e transações** que acessam dados de mais de um local



Distribuição acrescenta complexidade ao SGBD

- Decidir qual a cópia dos **dados replicados** que deve ser acessada
- Manter a **consistência** de todas as cópias de dados replicados
- **Recuperar falhas** em locais individuais e outros tipos de falha como a de um *link* de comunicação



Formas Básicas de BD Distribuído

- Abordagem *top-down*
- Abordagem *bottom-up*
- Abordagem combinada



Abordagem Top-down

- Projeto de um novo banco de dados
 - Busca da **redução dos custos** e da melhoria do **desempenho** das aplicações
 - **Limites da capacidade** de um banco de dados centralizado atingidos
 - Armazenamento
 - Processamento
 - Tempo de acesso



Abordagem Top-down (Distribuição de Dados)

- Particionar o BD em unidades lógicas (fragmentos)
 - **Fragmentação**
 - **Replicação**



Abordagem Top-down (Distribuição de Dados)

- Fragmentação de Relações
 - É um processo para o **particionamento de BD em banco de dados menores sem replicação**
 - O resultado é um conjunto de fragmentos de relações, os quais devem ser **alocados aos diferentes bancos de dados locais**



Abordagem Top-down (Distribuição de Dados)

- Fragmentação (Cont.)
 - **Alocação dos fragmentos** das relações aos diferentes locais (servidores)
 - Relação
 - Unidades lógicas menores
 - Fragmentação horizontal de relações
 - Fragmentação vertical de relações



Abordagem *Top-down* (Distribuição de Dados)

- Fragmentação **Horizontal**
 - As tuplas de uma relação são divididas em subconjuntos, e cada subconjunto é alocado a um banco de dados diferente
- Fragmentação **Vertical**
 - Produz um conjunto de relações distintas a partir de uma mesma relação
- Fragmentação **mista**
 - As fragmentações horizontal e vertical podem ser combinadas no mesmo projeto



Abordagem *Top-down* (Distribuição de Dados)

- **Replicação** de Dados
- Útil para melhorar a disponibilidade de dados
 - Replicação Total
 - Replicação de todo o BD em cada nó do sistema distribuído
 - Sistema opera completamente se pelo menos um nó estiver operando



Abordagem *Top-down* (Distribuição de Dados)

- **Replicação** de Dados (Cont.)
 - Replicação Parcial
 - Alguns fragmentos do BD podem ser replicados, enquanto outros não
 - O número de cópias de cada fragmento pode ir de uma ao número de nós
 - Fragmentos não precisam ser disjuntos



Abordagem *Bottom-up*

- Processo de **integração de bancos** de dados já existentes e independentes
 - Ambientes onde proliferaram os **bancos de dados** departamentais e **individuais**
 - Integração gera **bancos de dados globais virtuais**
 - Preserva os investimentos já feitos em aplicações, possibilitando ainda uma **visão integrada** dos dados fisicamente dispersos



Abordagem *Bottom-up*

- Manipulação “Simultânea” de Dados Heterogêneos
 - **Multi-Banco de Dados (MBD)**
 - **Bancos de Dados Federados**
 - Outras...
- Processamento de Operações que Acessam Diferentes BD (Operações Globais)



Abordagem *Bottom-up*

- Razões
 - **Sistemas Legados**
 - Organizações possuindo sistemas desenvolvidos há algum tempo (até 1990, aproximadamente)
 - Tecnologias aplicadas bastante diferentes das atuais



Abordagem *Bottom-up*

- Razões
 - Migração entre Plataformas
 - Evolução de hardware e software disponibilizando novas facilidades
 - Right-sizing / Down-sizing / Up-sizing
 - » Busca de eficiência
 - » Facilidade de manutenção dos programas de aplicação
 - » Redução de custos (se possível)



Abordagem *Bottom-up*

- Razões (Necessidade de convivência)
 - Migração entre Plataformas (Cont.)
 - Processo de longa duração
 - Correção de rumos
 - Alteração de objetivos originais



Abordagem *Bottom-up*

- Razões (Acesso integrado)
 - Mudanças Organizacionais
 - Fusão de duas ou mais empresas
 - Visão integrada da nova organização
 - Adoção de modelos de gestão
 - Autonomia aos componentes organizacionais



Abordagem *Bottom-up*

- Razões (Eliminar as ilhas de informação)
 - Evolução da Tecnologia de Redes de Comunicação
 - Integração dos ambientes computacionais anteriormente isolados



Abordagem *Bottom-up*

- Razões (SGBD de diferentes fabricantes)
 - Evolução da Tecnologia de Bancos de Dados Relacionais
 - Disseminação da utilização de SGBD relacionais



Abordagem *Bottom-up*

- Razões
 - Evolução da tecnologia de microprocessadores
 - Processadores de capacidade cada vez mais elevada a custos cada vez mais baixos
 - Aplicações para as mais diferentes áreas de atividade das organizações



Abordagem Combinada

- Geralmente como **consequência do esforço de migração de sistemas de banco de dados** residentes em ambientes de grande porte para outras plataformas.

