

# Sistemas de Banco de Dados Projeto, implementação e gerenciamento

## *Capítulo 12*

Sistemas de gerenciamento de banco de  
dados distribuídos

# Objetivos

- **Neste capítulo, você aprenderá:**
  - O que é um sistema de gerenciamento de banco de dados distribuídos (SGBDD) e quais são seus componentes
  - Como a implementação do banco de dados é afetada pelos diferentes níveis de distribuição de dados e processo
  - Como as transações são gerenciadas em um ambiente de banco de dados distribuído
  - Como o projeto é afetado pelo ambiente de banco de dados distribuído

# A Evolução dos Sistemas de Gerenciamento de Banco de Dados Distribuídos

- **Sistema de gerenciamento de banco de dados distribuídos (SGBDD)**
  - Controla o armazenamento e processamento de dados relacionados
  - Sistemas computacionais interconectados
  - Tanto os dados como as funções de processamento são distribuídos entre os diversos locais
- Sistemas de gerenciamento de bancos de dados centralizados para atender a suas necessidades de informações estruturadas

**FIGURA  
12.1**

**Sistema de gerenciamento de banco de dados centralizados**



# Vantagens e Desvantagens do SGBDD

- Vantagens:
  - Os dados ficam localizados próximos do local de maior demanda
  - Maior rapidez de acesso aos dados
  - Maior rapidez de processamento de dados
  - Facilidade de ampliação
  - Aprimoramento das comunicações
  - Custos operacionais reduzidos
  - Interface amigável
  - Menor risco de falha em ponto único
  - Independência do processador

# Vantagens e Desvantagens do SGBDD (cont.)

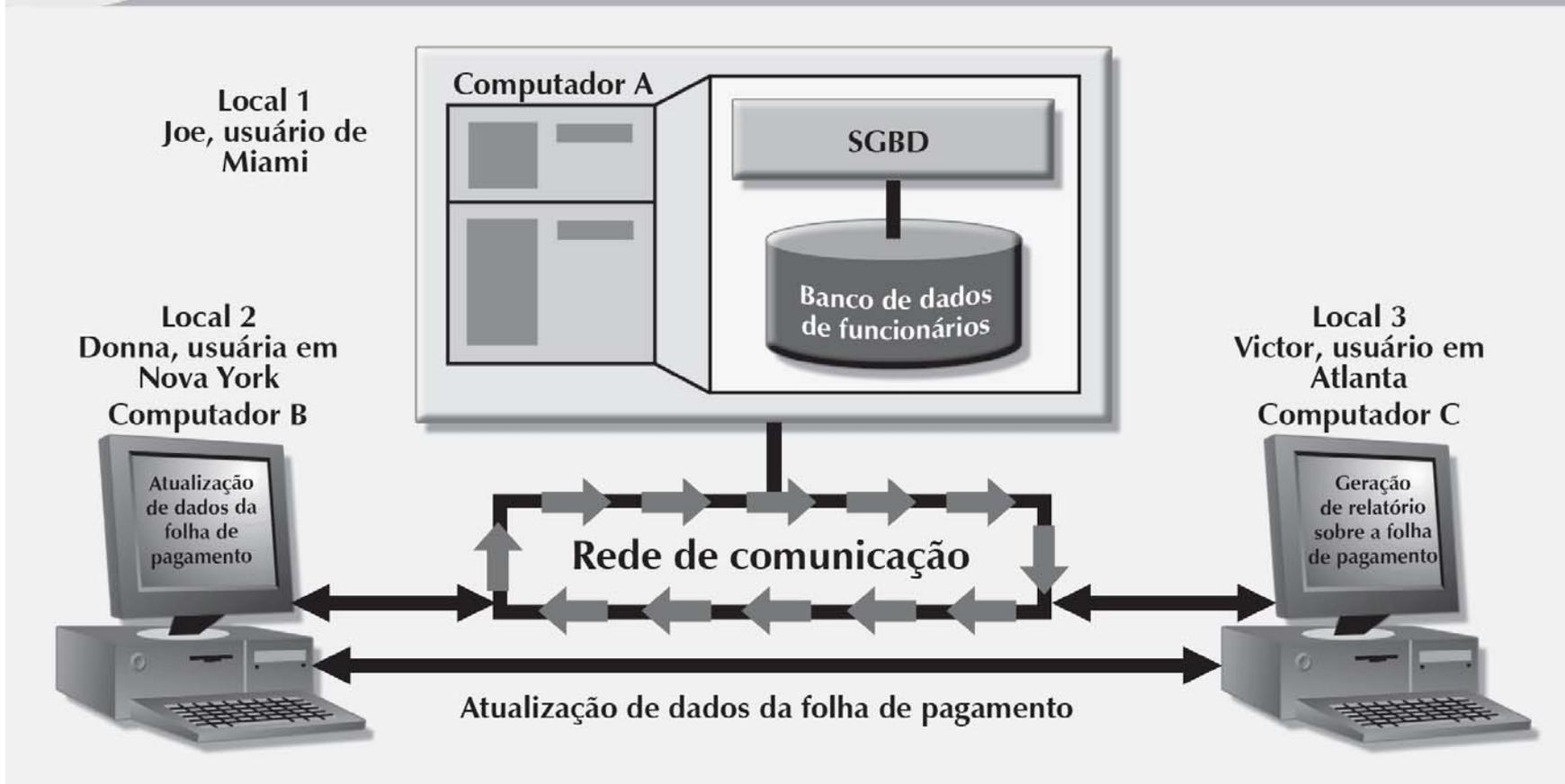
- Desvantagens:
  - Complexidade de gerenciamento e controle
  - Dificuldade tecnológica
  - Segurança
  - Falta de padrões
  - Ampliação das necessidades de armazenamento e infraestrutura
  - Aumento de custos com treinamento
  - Custos

# Processamento Distribuído e Banco de Dados Distribuídos

- **Processamento distribuído**
  - O processamento lógico do banco de dados é compartilhado entre dois ou mais locais fisicamente independentes
    - Conectados por uma rede
- **Banco de dados distribuído**
  - Armazena o banco relacionado logicamente por dois ou mais locais físicos independentes
  - O banco é composto de várias partes conhecidas como **fragmentos de banco de dados**

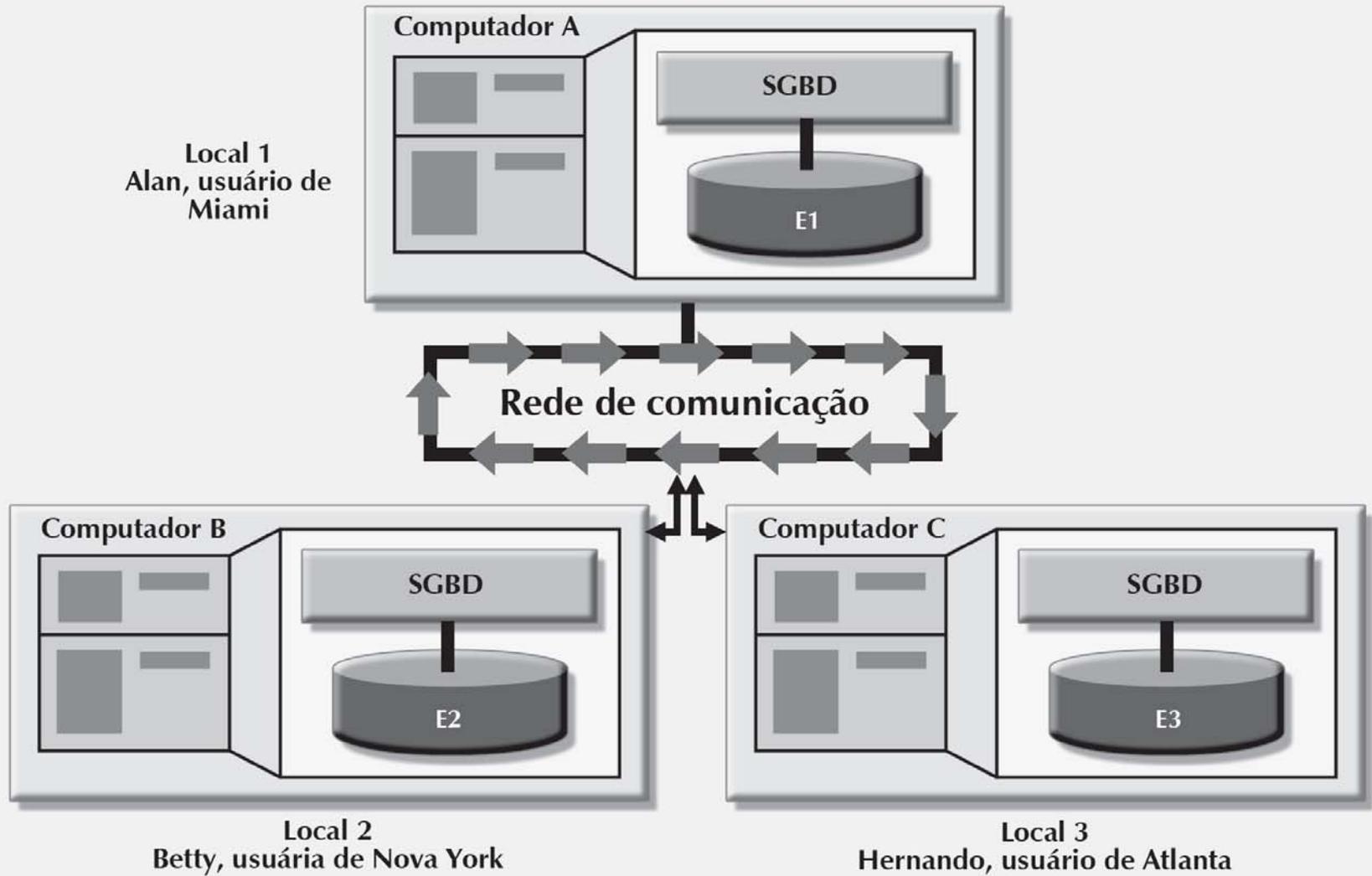
**FIGURA 12.2**

**Ambiente de processamento de dados distribuído**



**FIGURA  
12.3**

**Ambiente de banco de dados distribuído**



# Características dos Sistemas de Gerenciamento de Bancos de Dados Distribuídos

- Interface de aplicação
- Validação
- Transformação
- Otimização de consulta
- Mapeamento
- Interface de E/S

# Características dos Sistemas de Gerenciamento de Bancos de Dados Distribuídos (cont.)

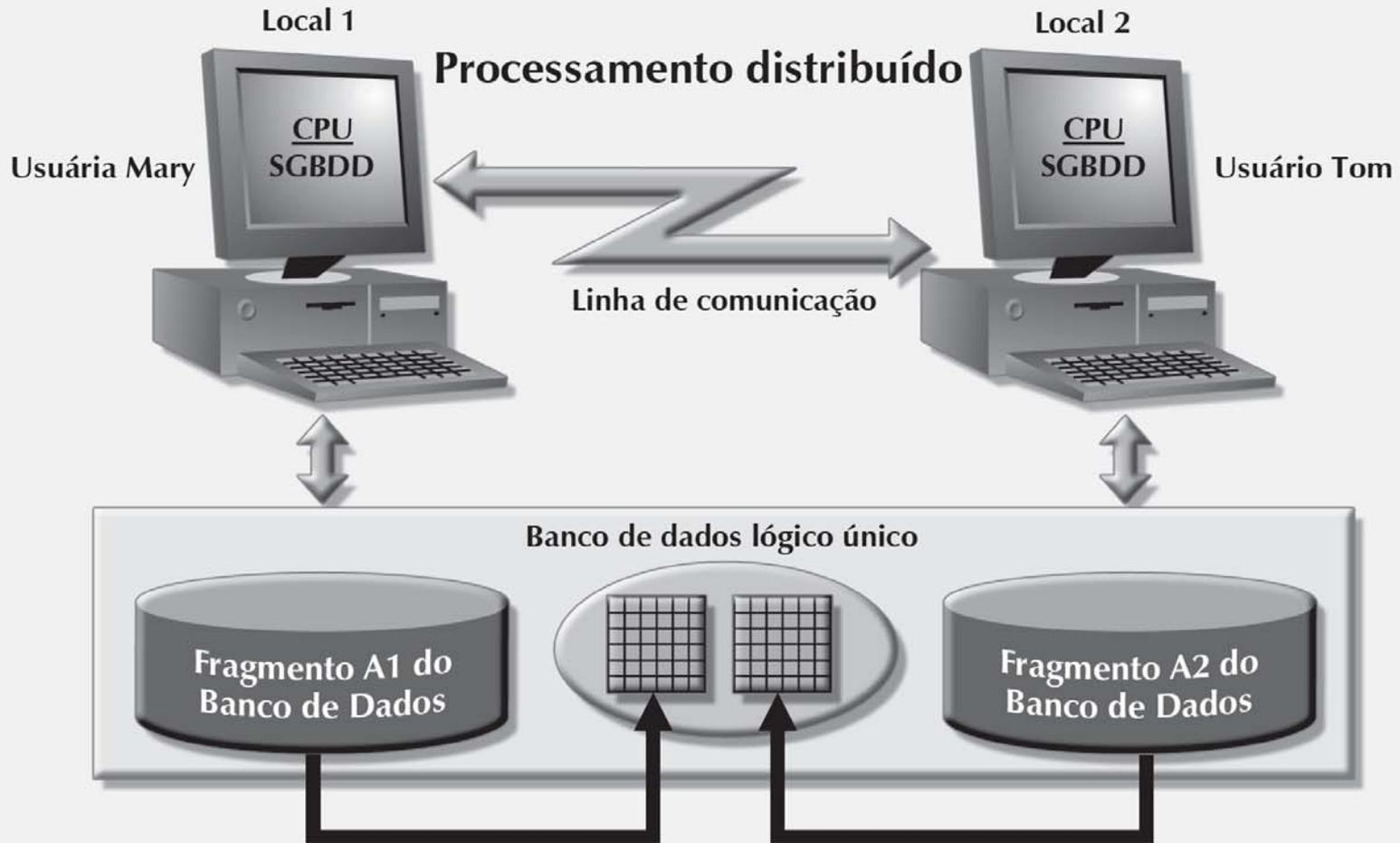
- Formatação
- Segurança
- Backup e recuperação
- Recursos de administração de BD
- Controle de concorrência
- Gerenciamento de transações

# Características dos Sistemas de Gerenciamento de Bancos de Dados Distribuídos (cont.)

- Um sistema de gerenciamento de banco de dados totalmente distribuído deve executar todas as funções de um SGBD centralizado
- Deve lidar com todas as funções necessárias impostas pela distribuição de dados e do processamento
  - Executar essas funções de modo *transparente* para o usuário final

**FIGURA 12.4**

**Sistema de gerenciamento de banco de dados totalmente distribuído**



# Componentes dos SGBDD

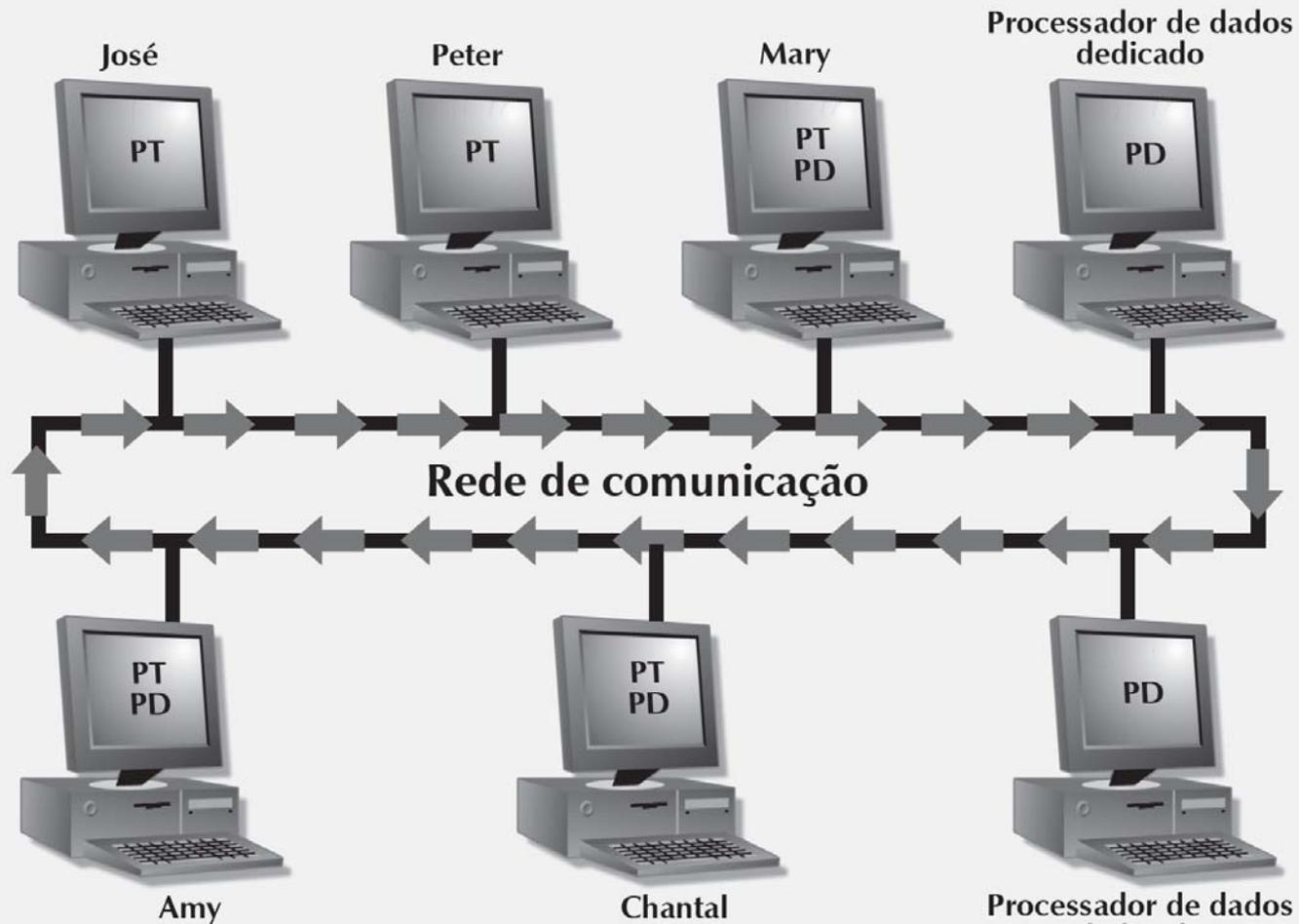
- O SGBDD deve incluir, no mínimo, os seguintes componentes:
  - Estações de trabalho
  - Componentes de hardware e software de rede
  - Meios de comunicação
  - **Processador de transações (PT)**, também conhecido como **processador de aplicações (PA)** e **gerente de transações (GT)**
    - Componente de software encontrado em cada computador que solicita dados

# Componentes dos SGBDD (cont.)

- O SGBDD deve incluir, no mínimo, os seguintes componentes (cont.):
  - **Processador de dados (PD) ou gerente de dados (GD)**
    - Componente de software residente em cada computador, armazenando e recuperando dados existentes no local
    - Pode ser até mesmo um SGBD centralizado

**FIGURA  
12.5**

**Componentes de gerenciamento do sistema de banco de dados distribuído**



**Observação:** Cada PT pode acessar dados em qualquer PD e cada PD trata de todas as solicitações de dados locais de qualquer PT.

# Níveis de Dados e Distribuição de Processos

- Os sistemas atuais de banco de dados podem ser classificados com base em como a distribuição de processos e a distribuição de dados são suportadas

TABELA 12.2 Sistemas de banco de dados: níveis de distribuição de dados e de processos

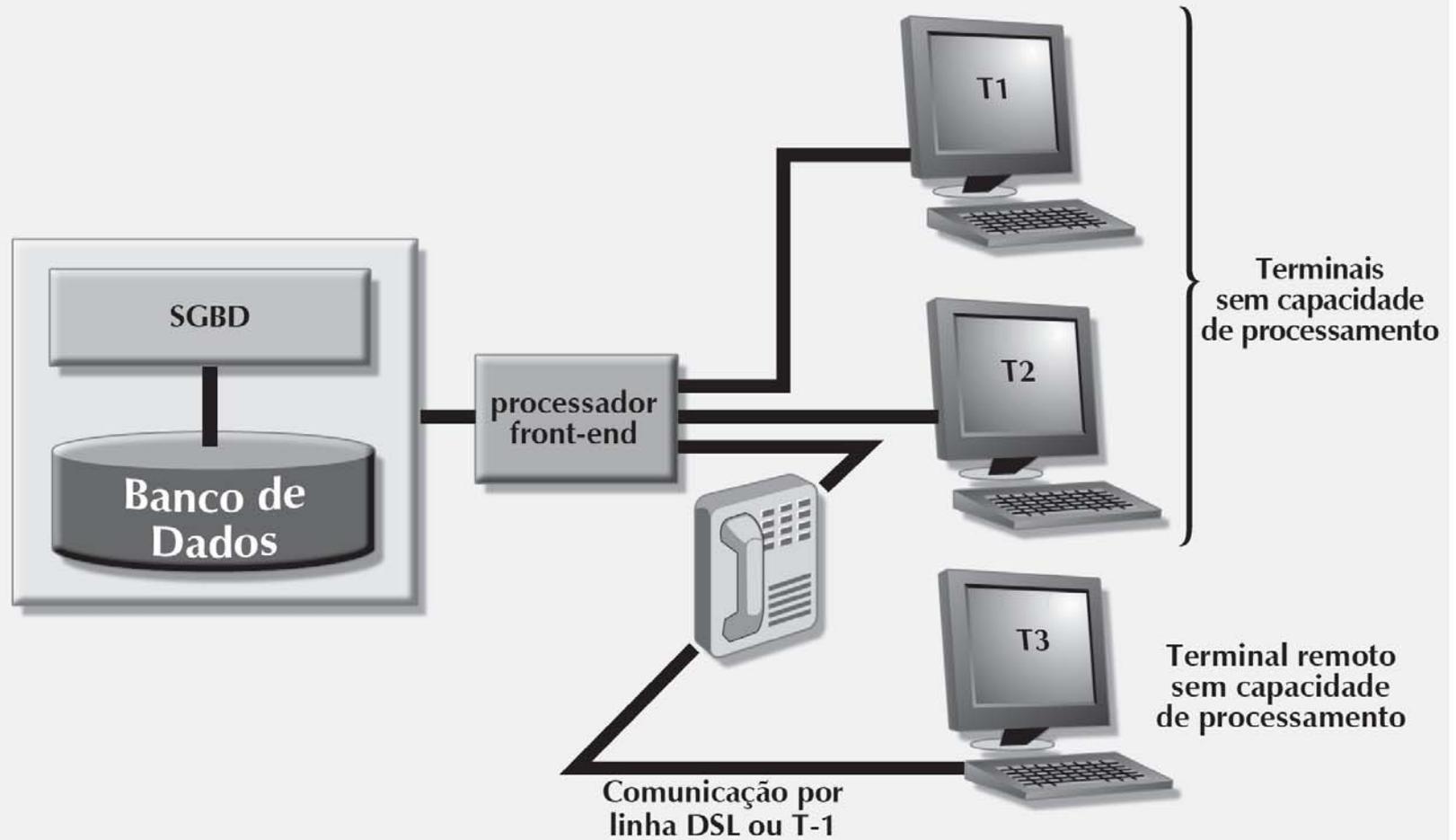
	DADOS EM UM ÚNICO LOCAL	DADOS EM VÁRIOS LOCAIS
Processamento em um único local	SGBD hospedeiro	Não aplicável (Exige processamento em vários locais)
<b>Processamento em vários locais</b>	Servidor de arquivos SGBD cliente/servidor (SGBD em LAN)	Totalmente distribuído SGBDD cliente/servidor

# Processamento em um Único Local, Dados em um Único Local (SPSD)

- Todo o processamento é feito em um computador hospedeiro (servidor com um único processador, servidor com vários processadores, sistema de mainframe)
- Todos os dados são armazenados no sistema do disco local desse computador
- O processamento não pode ser feito pelo usuário final do sistema
- Esse cenário é típico da maioria dos SGBDs de mainframe ou de computadores servidores de médio porte
- O SGBD localiza-se no computador hospedeiro que é acessado por terminais sem capacidade de processamento a ele conectados

**FIGURA  
12.6**

**Processamento em um único local, dados em um único local (centralizado)**

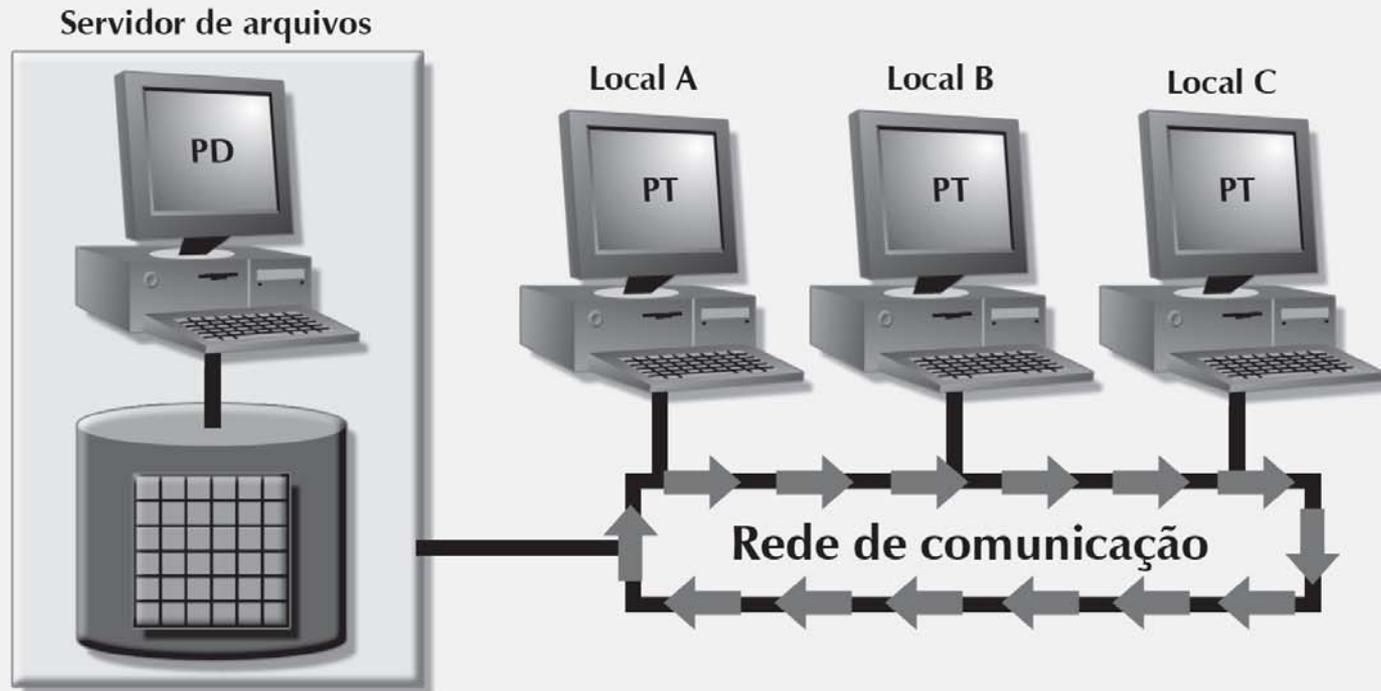


# Processamento em Vários Locais, Dados em um Único Local (MPSD)

- Vários processos são executados em diferentes computadores que compartilham um único depósito de dados
- O cenário MPSD exige um servidor de arquivos que execute aplicações convencionais a serem acessadas por rede
- Muitas aplicações de contabilidade de multiusuário executadas em uma rede de computadores pessoais

**FIGURA  
12.7**

**Processamento em vários locais, dados em um único local**



# Processamento em Vários Locais, Dados em Vários Locais (MPMD)

- SGBD totalmente distribuído com suporte para vários processadores de dados e de transações em diversos locais
- Pode ser classificado como homogêneo ou heterogêneo
- **SGBDDs homogêneos** integram apenas um tipo de SGBD centralizado na rede
- **SGBDDs heterogêneos** integram tipos diferentes de SGBDs centralizados na rede

# Processamento em Vários Locais, Dados em Vários Locais (MPMD) (cont.)

- **SGBDD totalmente heterogêneo**
  - Dá suporte a SGBDs distintos
  - Pode aceitar diferentes modelos (relacional, hierárquico ou em rede)
  - Executados em diversos sistemas de computadores, como mainframes e PCs

**FIGURA 12.8**

**Cenário de banco de dados distribuído heterogêneo**

	Platform	SGBD	Sistema Operacional	Protocolo de Comunicação de Rede
	IBM 3090	DB2	MVS	APPC LU 6.2
	DEC/VAX	VAX rdb	OpenVMS	DECnet
	IBM AS/400	SQL/400	OS/400	3270
	Computador RISC	Informix	UNIX	TCP/IP
	CPU Pentium	Oracle	Windows Server 2003	TCP/IP

# Recursos de Transparência de Banco de Dados Distribuídos

- Esses recursos apresentam a propriedade de permitir que o usuário final sintá-se como o monousuário do banco
- Os recursos de transparência dos SGBDDs são:
  - **Transparência de distribuição**
  - **Transparência de transação**
  - **Transparência de falhas**
  - **Transparência de desempenho**
  - **Transparência de heterogeneidade**

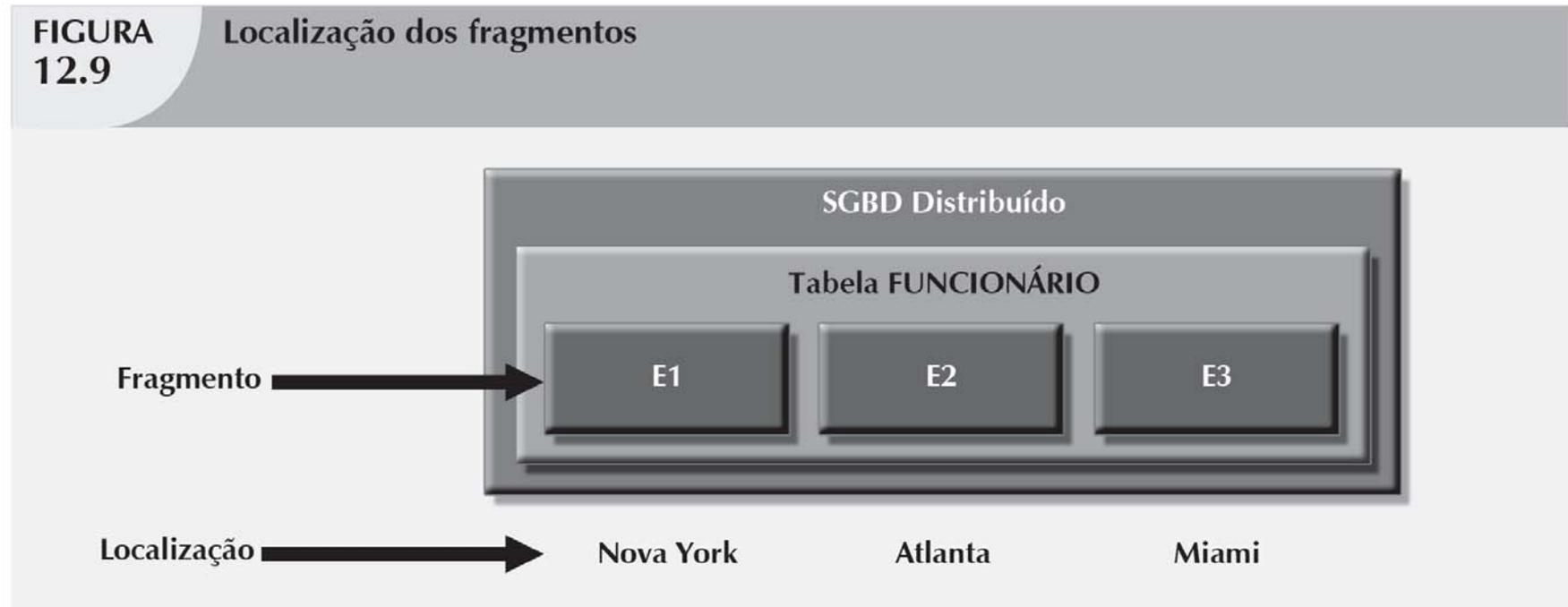
# Transparência de Distribuição

- Permite que um banco de dados fisicamente disperso seja gerenciado como se estivesse centralizado
- Três níveis de transparência de distribuição:
  - **Transparência de fragmentação**
  - **Transparência de localização**
  - **Transparência de mapeamento local**

**TABELA 12.3** Resumo dos recursos de transparência

SE O COMANDO DE SQL EXIGE:			
NOME DO FRAGMENTO?	NOME DA LOCALIZAÇÃO?	ENTÃO, O SGBD DÁ SUPORTE A	NÍVEL DE TRANSPARÊNCIA DE DISTRIBUIÇÃO
Sim	Sim	Mapeamento local	Baixo
Sim	Não	Transparência de localização	Médio
Não	Não	Transparência de fragmentação	Alto

**FIGURA 12.9** Localização dos fragmentos



# Transparência de Transação

- Assegura que as transações de bancos de dados manterão a integridade e a consistência
- A transparência de transação garante que ela só será concluída quando todos os locais envolvidos do banco de dados concluírem suas partes da transação
- Os sistemas de bancos de dados distribuídos exigem mecanismos complexos para gerenciar transações e garantir a consistência e a integridade do banco de dados

# Solicitações Distribuídas e Transações Distribuídas

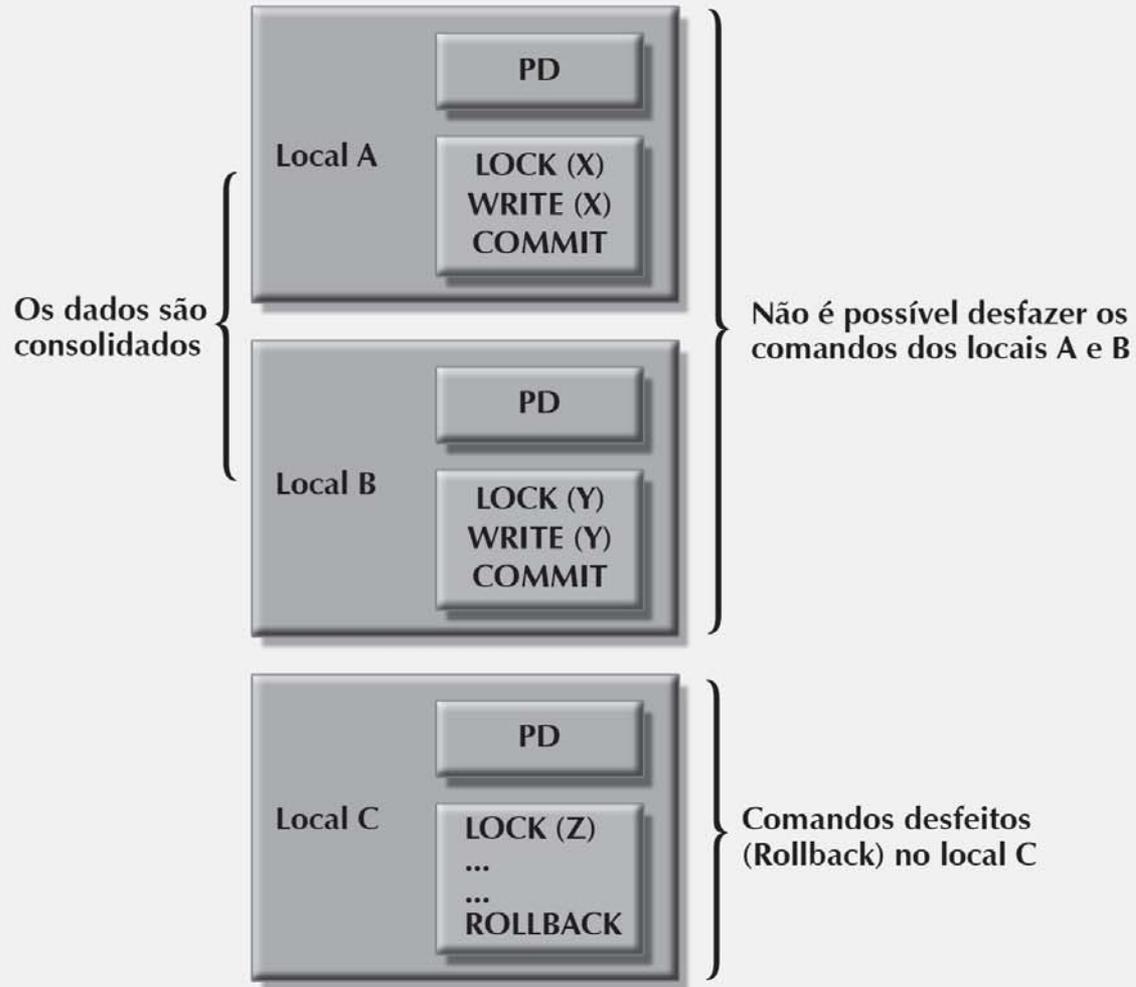
- **Solicitação remota:** permite que um único comando de SQL acesse os dados a serem processados por um único processador remoto do banco de dados
- **Transação remota:** composta de várias solicitações, acessa os dados em um único local remoto
- **Transação distribuída:** permite que uma transação referencie vários locais de PD, remotos ou não
- **Solicitação distribuída:** permite que um único comando de SQL referencie dados localizados em vários locais de PD, remotos ou não

# Controle de Concorrência Distribuída

- O controle de concorrência torna-se especialmente importante no ambiente de banco de dados distribuído
  - É mais provável que operações de vários locais e processos criem inconsistências de dados e transações em deadlock

**FIGURA  
12.15**

**Efeito de um comando COMMIT prematuro**



# Protocolo de Consolidação de Duas Fases

- Os bancos distribuídos possibilitam que uma transação acesse dados em vários locais
- O comando COMMIT final não deve ser emitido até que todos os locais tenham consolidado suas partes da transação
- Exige que o log de entrada da transação em cada PD seja gravado antes do fragmento do banco de dados ser efetivamente atualizado
- O **protocolo FAZER-DESFIZER-REFAZER** é utilizado pelo PD para desfazer ou refazer transações com auxílio das entradas do log de transações do sistema
- Define as operações entre dois tipos de nós: o **coordenador** e um ou mais **subordinados**

# Transparência de Desempenho e Otimização de Consultas

- O objetivo de uma rotina de otimização de consulta é minimizar o custo total associado à execução de consultas
- Esses custos de solicitação são em função de:
  - Custo do tempo (E/S)
  - Custo de comunicação
  - Custo de tempo da CPU
- Deve fornecer transparência de distribuição e transparência de réplica

# Transparência de Desempenho e Otimização de Consultas (cont.)

- **Transparência de réplica**
  - Refere-se à capacidade do SGBDD de ocultar do usuário a existência de várias cópias de dados
- **Otimização de consulta:**
  - Manual ou automática
  - Estática ou dinâmica
  - Baseada em estatísticas ou algoritmo com base em regras

# Projeto de Banco de Dados Distribuídos

- Fragmentação de dados
  - Como separar as bases de dados em fragmentos
- Replicação de dados
  - Quais fragmentos replicar
- Alocação de dados
  - Onde armazenar os fragmentos e réplicas

# Fragmentação de Dados

- Separa um objeto simples em dois ou mais segmentos ou fragmentos
- Cada fragmento pode ser armazenado em qualquer local de uma rede de computadores
- As informações sobre a fragmentação de dados são armazenadas no catálogo de dados distribuídos (CDD)
  - A partir do qual elas são acessadas pelo PT para as solicitações do usuário do processo

# Fragmentação de Dados (cont.)

- Estratégias
  - **Fragmentação horizontal**
    - Divisão de uma relação em subconjuntos (fragmentos) de Tuplas (linhas)
  - **Fragmentação vertical**
    - Divisão de uma relação em subconjuntos de atributos (colunas)
  - **Fragmentação mista**
    - Combinação das estratégias horizontal e vertical

# Replicação de Dados

- Refere-se ao armazenamento de cópias de dados em vários locais servidos por uma rede de computador
- As cópias dos fragmentos podem ser armazenadas nesses diversos locais para atender a necessidades específicas de informação
  - Aprimoramento da disponibilidade de dados e do tempo de resposta
  - Reduz os custos totais de comunicação e consulta
- **Regra da consistência mútua:** todas as cópias de fragmentos de dados sejam idênticas

# Replicação de Dados (cont.)

- **Banco de dados totalmente replicado**
  - Armazena diversas cópias de *cada* um de seus fragmentos em vários locais
  - Todos os fragmentos do banco são replicados
- **Banco de dados parcialmente replicado**
  - Armazena diversas cópias de *alguns* de seus fragmentos em vários locais
- **Banco de dados não replicado**
  - Armazena cada um de seus fragmentos em um único local
  - Não há fragmentos duplicados no banco de dados

# Alocação de Dados

- Descreve o processo de decidir onde posicionar os dados
  - **Alocação centralizada de dados**
    - Todo o banco de dados é armazenado em um único local
  - **Alocação particionada de dados**
    - O banco de dados é dividido em duas ou mais partes separadas (fragmentos) e armazenado em dois ou mais locais
  - **Alocação replicada de dados**
    - Cópias de um ou mais fragmentos do banco de dados são armazenadas em vários locais

# Cliente/servidor *versus* SGBDD

- Modo como os computadores interagem para formar um sistema
- Apresenta um *usuário* de recursos, ou seja, um cliente, e um *fornecedor* de recursos, um servidor
- Pode ser utilizada na implementação de um SGBD em que o cliente seja o PT, e o servidor, o PD

# Cliente/servidor *versus* SGBDD (cont.)

- Vantagens
  - Soluções menos dispendiosas do que as de minicomputadores ou mainframes alternativos
  - Suas soluções permitem que o usuário final utilize a GUI do microcomputador, aprimorando, assim, a funcionalidade e a simplicidade
  - Mais pessoas no mercado de trabalho possuem habilidades com PC do que com mainframe
  - O PC é bem estabelecido no local de trabalho
  - Existem várias ferramentas de análise e consulta de dados para facilitar a interação
  - Há uma considerável vantagem de custos para o desenvolvimento de aplicações offloading do mainframe para PCs poderosos

# Cliente/servidor *versus* SGBDD (cont.)

- Desvantagens
  - Ambiente mais complexo
  - Aumento do número de usuários e de locais de processamento costuma abrir espaço para problemas de segurança
  - O ambiente cliente/servidor torna possível a disseminação do acesso aos dados a um círculo de usuários muito mais amplo
    - Amplia a demanda por pessoal com profundo conhecimento de computadores e aplicativos
    - Os encargos de treinamento elevam o custo de manutenção do ambiente

# Os Doze Mandamentos de C. J Date para Bancos de Dados Distribuídos

- Independência de local
- Independência do local central
- Independência de falhas
- Transparência de localização
- Transparência de fragmentação
- Transparência de replicação

# Os Doze Mandamentos de C. J Date para Bancos de Dados Distribuídos (cont.)

- Processamento de consultas distribuídas
- Processamento de transações distribuídas
- Independência de hardware
- Independência de sistema operacional
- Independência de rede
- Independência de banco de dados

# Resumo

- O banco de dados distribuído armazena dados relacionados logicamente em dois ou mais locais fisicamente independentes
  - Conectados por uma rede de computadores
- Processamento distribuído: a divisão do processamento do banco de dados lógico entre dois ou mais nós de rede
  - Exige processamento distribuído
- Os principais componentes do SGBDD são o processador de transações e o processador de dados

## Resumo (cont.)

- Sistemas de banco de dados atuais:
  - Homogêneo integra apenas um tipo particular de SGBD por meio de uma rede
  - Heterogêneo integra vários tipos diferentes de SGBDs por meio de uma rede
- As características dos SGBDDs são mais bem descritas como um conjunto de transparências:
  - De distribuição, de transação de falha, de heterogeneidade e de desempenho

# Resumo (cont.)

- A transação é composta de uma ou mais solicitações de banco de dados
- O controle de concorrência distribuída é necessário em uma rede de bancos de dados distribuídos
- O SGBD distribuído avalia todas as solicitações de dados para encontrar o melhor caminho de acesso

# Resumo (cont.)

- O projeto de um banco de dados distribuído deve considerar a fragmentação e a replicação de dados
- Um banco de dados pode ser replicado por vários locais diferentes de uma rede
- Os fornecedores costumam identificar o software como produtos de banco de dados cliente/servidor