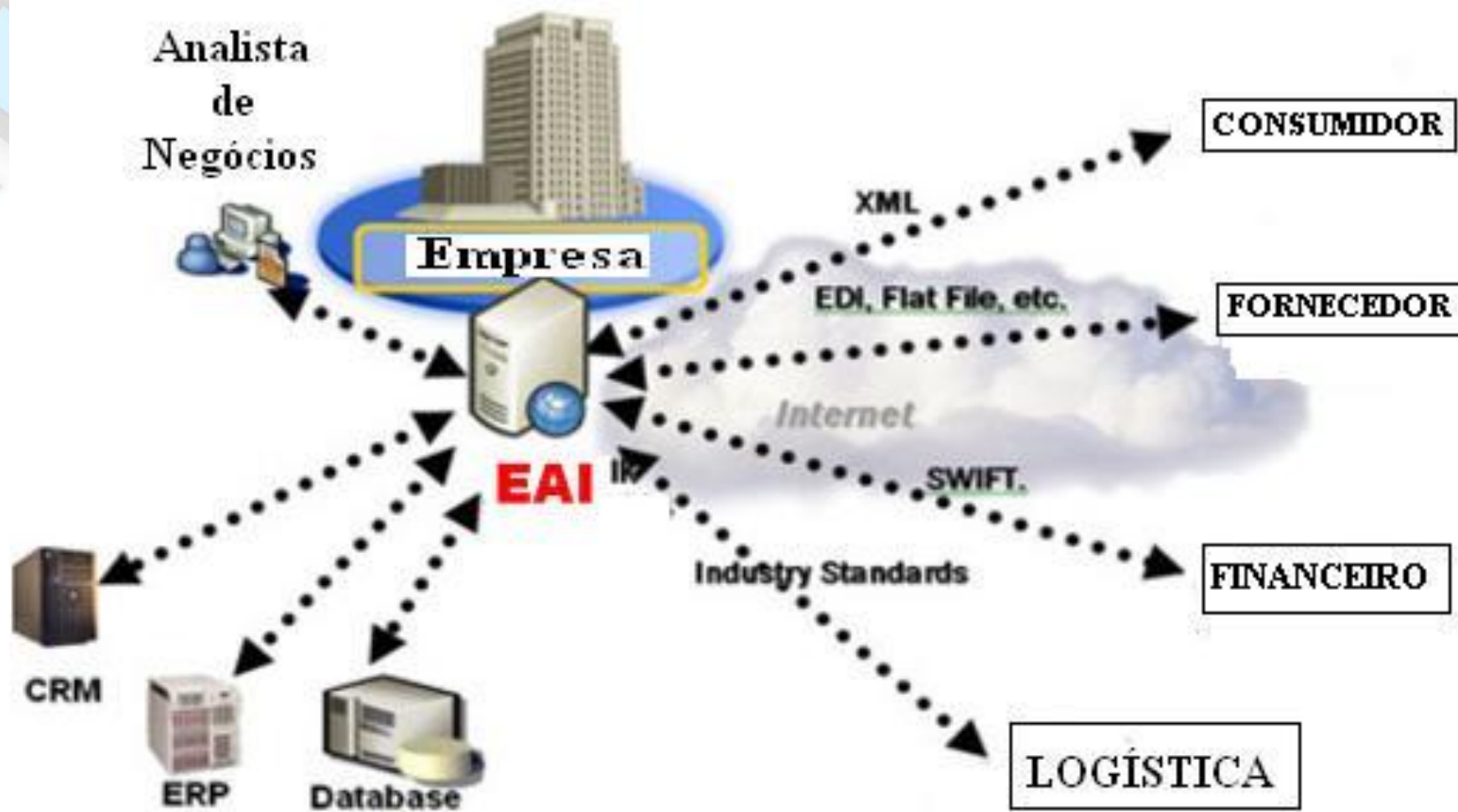


Integração de Dados, Heterogeneidades em Banco de Dados

Profº Elifranio

Cenário



Cenário



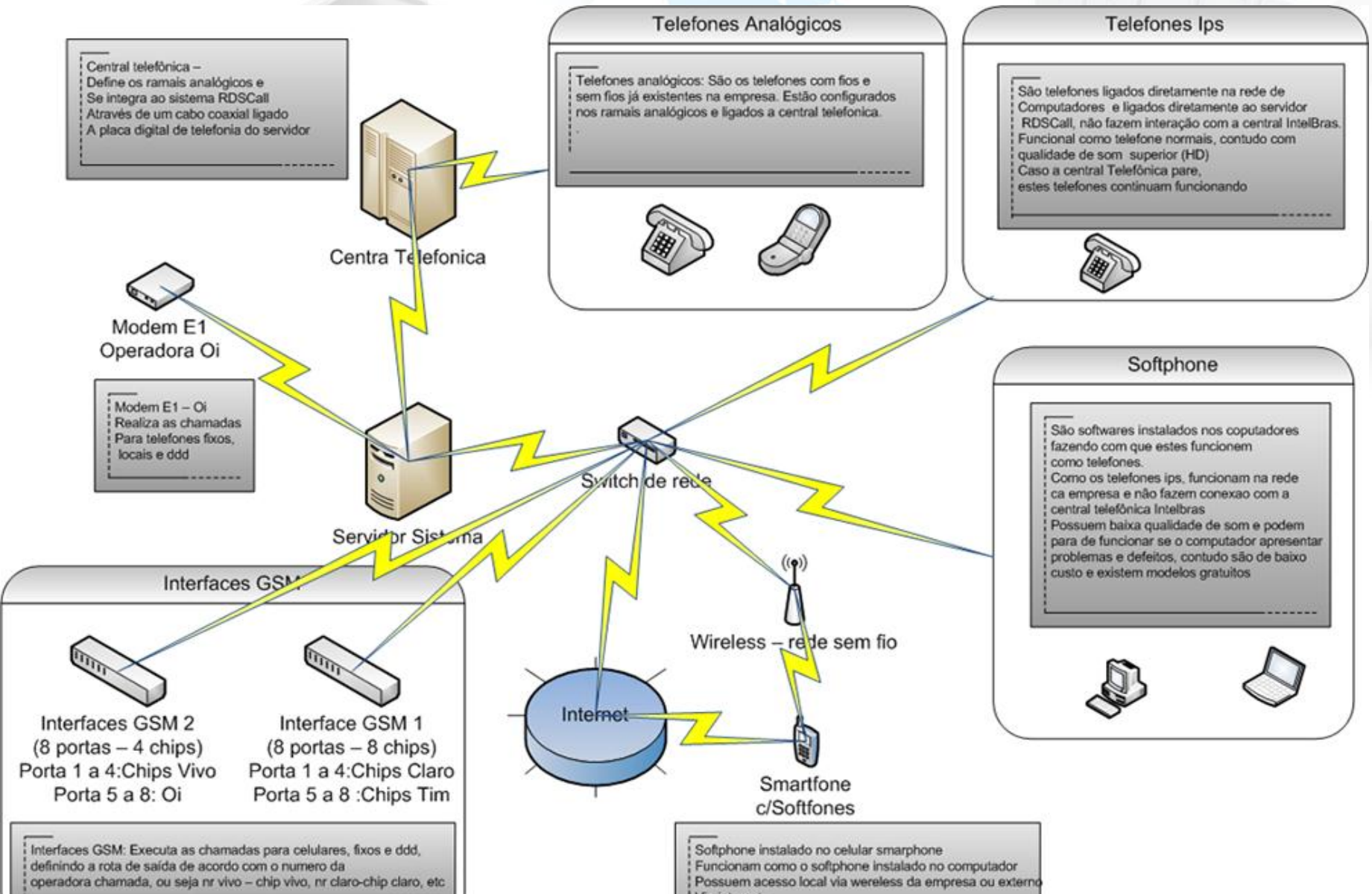
PORTAIS RFID

COLETORES
RFID

COLETORES DE DADOS WIFI, BLUETOOTH E GPRS

LEITORES DE CÓDIGO DE BARRAS ESTACIONÁRIOS

cenário



Cenário



Diferenças entre sistemas multiprocessadores e BDD

Precisamos distinguir 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, as seguintes condições mínimas devem ser satisfeitas:

- **Conexões de nós de banco de dados por uma rede de computadores.**
- **Inter-relação lógica dos bancos de dados conectados.**
- **Ausência de restrição de homogeneidade entre os nós conectados.**

Transparência



O conceito de transparência estende a ideia geral de ocultar detalhes da implementação dos usuários finais. Um sistema altamente transparente oferece muita flexibilidade ao usuário final/desenvolvedor de aplicação, pois requer pouco ou nenhum conhecimento dos detalhes básicos de sua parte.

Transparência (cont.)

Considere o banco de dados de empresa da Figura 3.5, que usamos como exemplo no decorrer do livro. As tabelas FUNCIONARIO, PROJETO e TRABALHA_EM podem ser fragmentadas horizontalmente (ou seja, em conjuntos de linhas, conforme discutiremos na Seção 25.4) e armazenadas com possível replicação, como mostra a Figura 25.1. Os seguintes tipos de transparências são possíveis:

Transparência (cont.)

- **Transparência da organização dos dados (também conhecida como *transparência de distribuição* ou *de rede*).**
- **Transparência de replicação.**
- **„Transparência de fragmentação.**
- **Outras transparências incluem *transparência de projeto* e *transparência de execução***

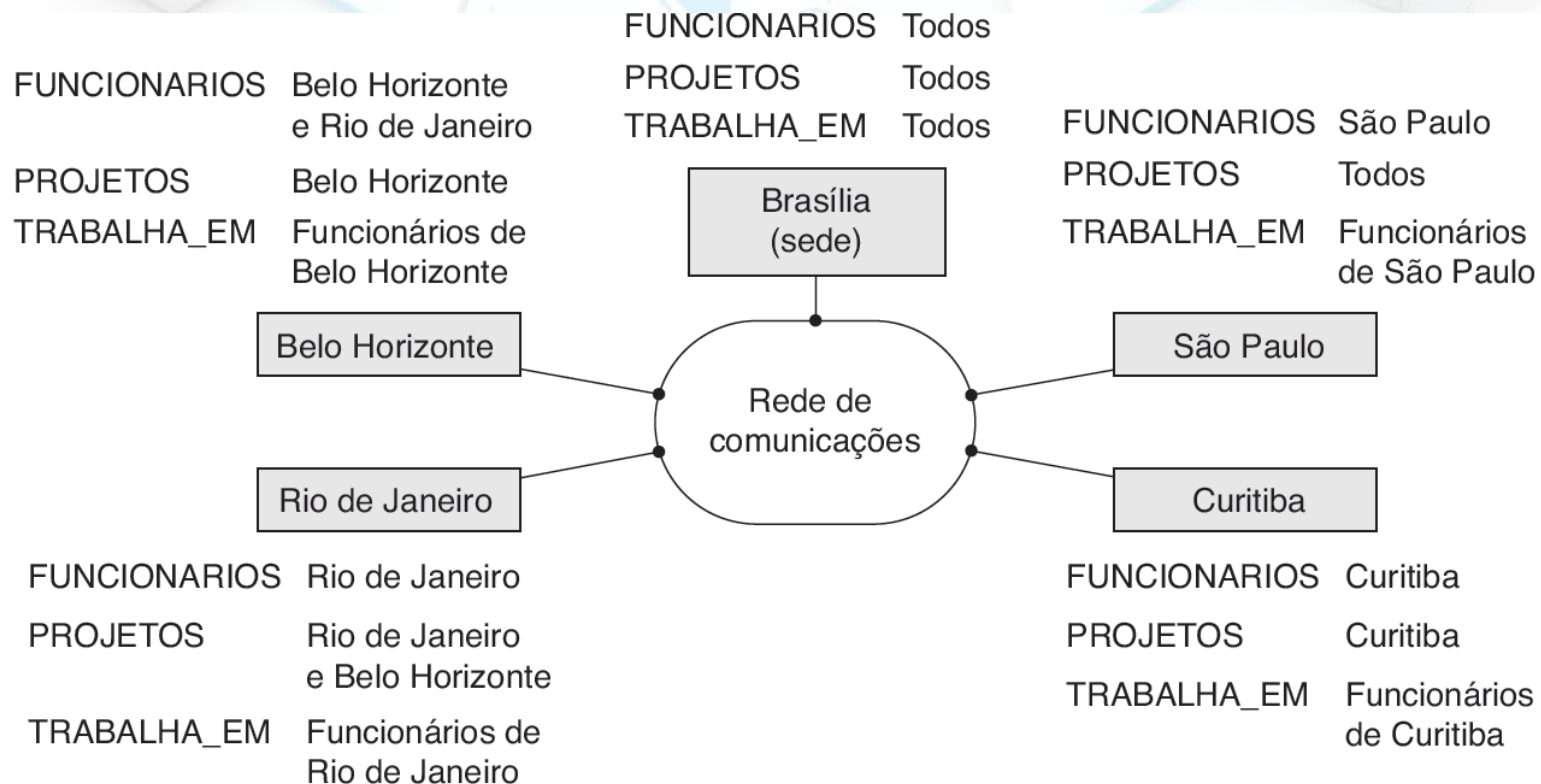


Figura 25.1

Distribuição e replicação de dados entre bancos de dados distribuídos.

Autonomia



A **autonomia** determina a extensão à qual os nós individuais ou BDs em um BDD conectado podem operar independentemente. Um alto grau de autonomia é desejável para maior flexibilidade e manutenção personalizada de um nó individual. A autonomia pode ser aplicada ao projeto, comunicação e execução.

Confiabilidade e disponibilidade



Confiabilidade e disponibilidade são duas das vantagens em potencial mais comuns citadas para bancos de dados distribuídos. A **confiabilidade é definida** em termos gerais 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.

Vantagens dos bancos de dados distribuídos

As organizações lançam mão do gerenciamento de banco de dados distribuído por diversos motivos. Algumas vantagens importantes são listadas a seguir.

- 1. Maior facilidade e flexibilidade de desenvolvimento da aplicação.**
- 2. Maior confiabilidade e disponibilidade.**
- 3. Maior desempenho.**
- 4. Expansão mais fácil.**

Funções adicionais dos bancos de dados distribuídos

A distribuição leva a uma maior complexidade no projeto e implementação do sistema. Para conseguir as vantagens em potencial já listadas, o software de SGBDD precisa ser capaz de oferecer as seguintes funções:

- **Acompanhar a distribuição de dados.**
- **Processamento de consulta distribuído.**
- **Gerenciamento de transação distribuído.**
- **Gerenciamento de dados replicado.**
- **Recuperação de banco de dados distribuído.**
- **Segurança.**
- **Gerenciamento de diretório (catálogo) distribuído.**

Tipos de sistemas de bancos de dados distribuídos

O termo *sistema de gerenciamento de banco de dados distribuído* pode descrever diversos sistemas que diferem um do outro em muitos aspectos. O item principal que todos os sistemas têm em comum é o fato de os dados e software serem distribuídos por vários sites conectados por alguma forma de rede de comunicação.

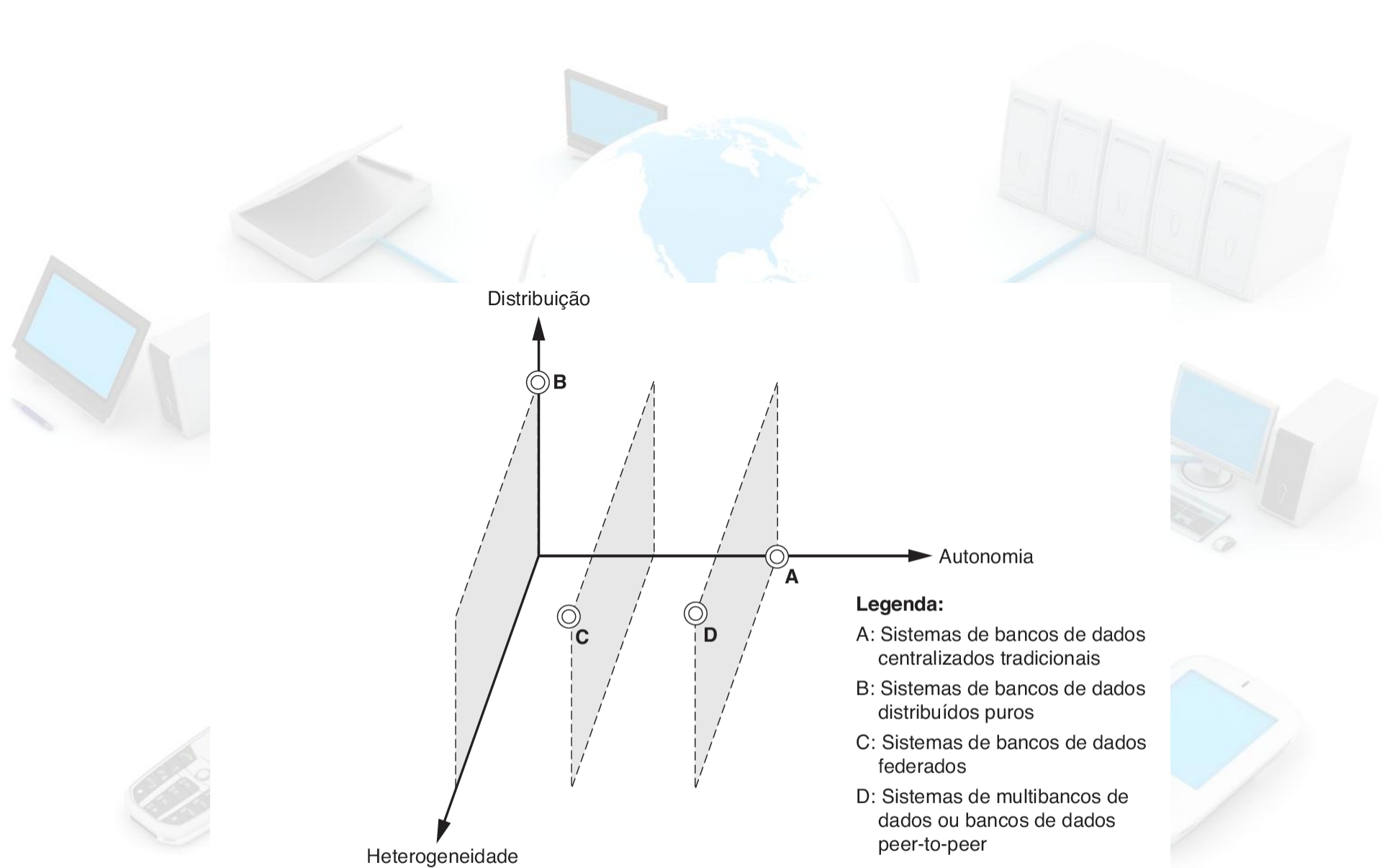


Figura 25.2

Classificação de bancos de dados distribuídos.

Problemas com sistemas de gerenciamento de banco de dados federados

O tipo de heterogeneidade presente nos SBDFs pode surgir de várias fontes. Discutimos primeiro essas fontes e, depois, indicamos como os diferentes tipos de autonomias contribuem para uma heterogeneidade semântica que deve ser resolvida em um SBDF heterogêneo.

- **Diferenças nos modelos de dados.**
- **Diferenças nas restrições.**
- **Diferenças nas linguagens de consulta.**

Problemas com sistemas de gerenciamento de banco de dados federados (cont.)

Heterogeneidade semântica.

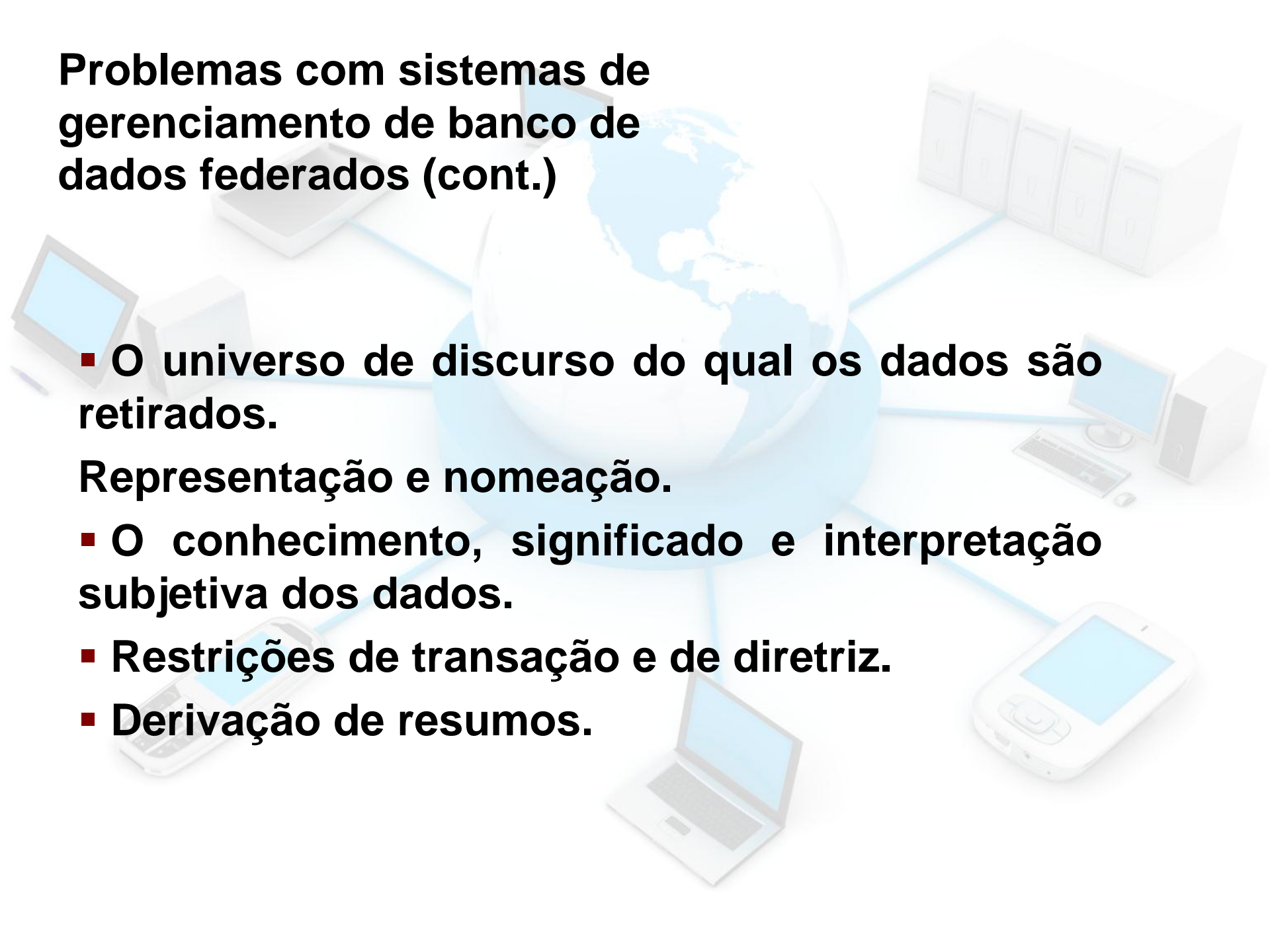
A heterogeneidade semântica ocorre quando existem diferenças no significado, interpretação e uso intencionado dos mesmos dados ou dados relacionados. A heterogeneidade semântica entre os sistemas de banco de dados (SBDs) componentes cria o maior obstáculo no projeto de esquemas globais de bancos de dados heterogêneos.

Problemas com sistemas de gerenciamento de banco de dados federados (cont.)

- **O universo de discurso do qual os dados são retirados.**

Representação e nomeação.

- **O conhecimento, significado e interpretação subjetiva dos dados.**
- **Restrições de transação e de diretriz.**
- **Derivação de resumos.**



Arquiteturas de banco de dados distribuídas

The background features a light blue globe centered on the Atlantic Ocean. A network of thin blue lines radiates from the globe, connecting to various digital devices: a laptop at the top left, a tablet at the top, a server rack at the top right, a desktop computer with a monitor and tower at the middle right, a smartphone at the bottom right, a laptop at the bottom, and another smartphone at the bottom left. The overall theme is global connectivity and distributed systems.

Arquiteturas paralelas *versus* distribuídas

Existem dois tipos principais de arquiteturas de sistema multiprocessador que são comumente utilizados:

- **Arquitetura de memória compartilhada (altamente acoplada).**
- **Arquitetura de disco compartilhado (livremente acoplada).**

Arquitetura geral de bancos de dados distribuídos puros

Na Figura 25.4, que descreve a arquitetura de esquema genérico de um BDD, a empresa é apresentada com uma visão consistente, unificada, que mostra a estrutura lógica dos dados básicos por todos os nós. Essa visão é representada pelo esquema conceitual global (ECG), que oferece transparência de rede.

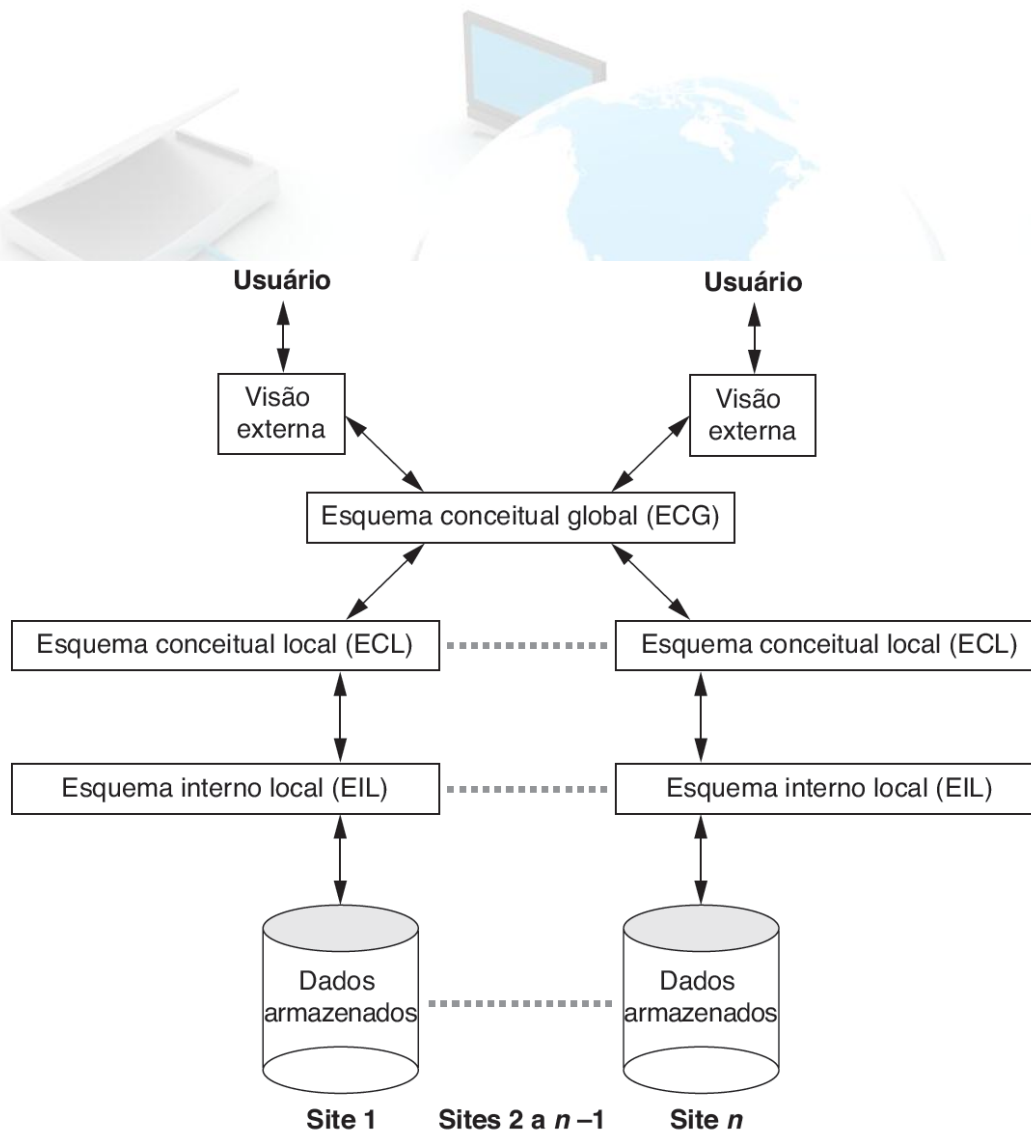


Figura 25.4

Arquitetura do esquema para bancos de dados distribuídos.

Arquitetura do esquema de banco de dados federado

A arquitetura típica do esquema de cinco níveis para dar suporte a aplicações globais no ambiente de SBDF aparece na Figura 25.6. Nessa arquitetura, o **esquema local é o esquema conceitual (definição de banco de dados completa)** de um banco de dados componente, e o **esquema de componente é derivado** ao se traduzir o esquema local para um modelo de dados canônico ou um modelo de dados comum (MDC) para o SBDF.

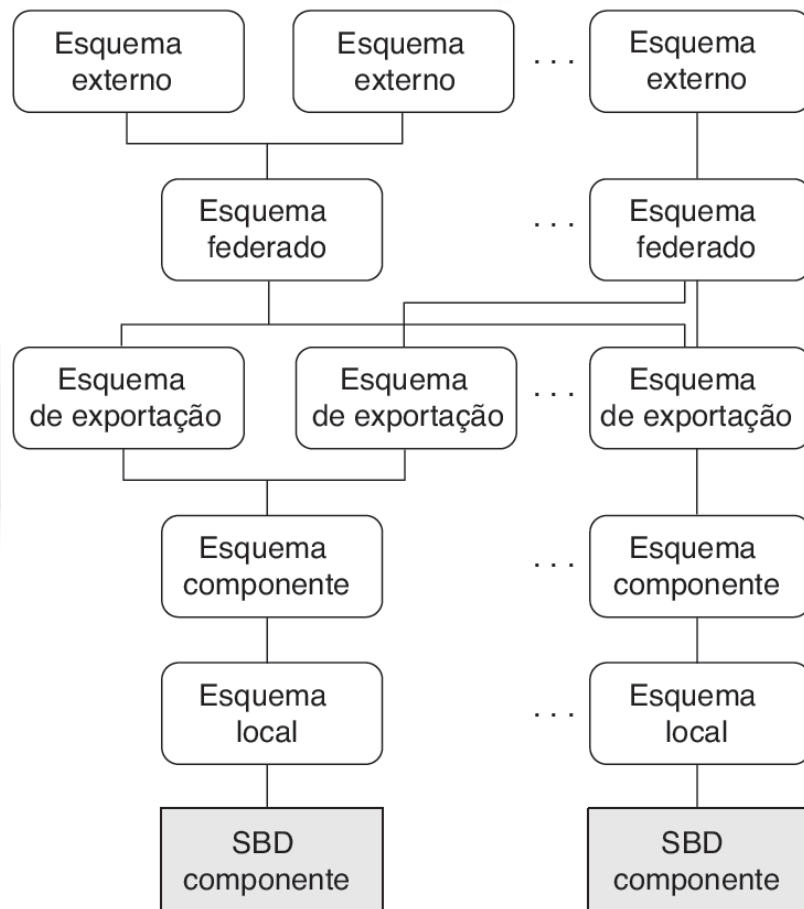


Figura 25.6

A arquitetura de esquema em cinco níveis em um sistema de banco de dados federado (SBDF).

Fonte: Adaptado de Sheth e Larson, 'Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases', *ACM Computing Surveys*, v. 22, n. 3, set. 1990.

Visão geral da arquitetura cliente-servidor de três camadas

Conforme indicado na introdução do capítulo, os SGBDDs em escala completa não foram desenvolvidos para dar suporte a todos os tipos de funcionalidades discutidos até aqui. Em vez disso, aplicações de banco de dados distribuído estão sendo desenvolvidas no contexto das arquiteturas cliente-servidor. Essa arquitetura é ilustrada na Figura 25.7.

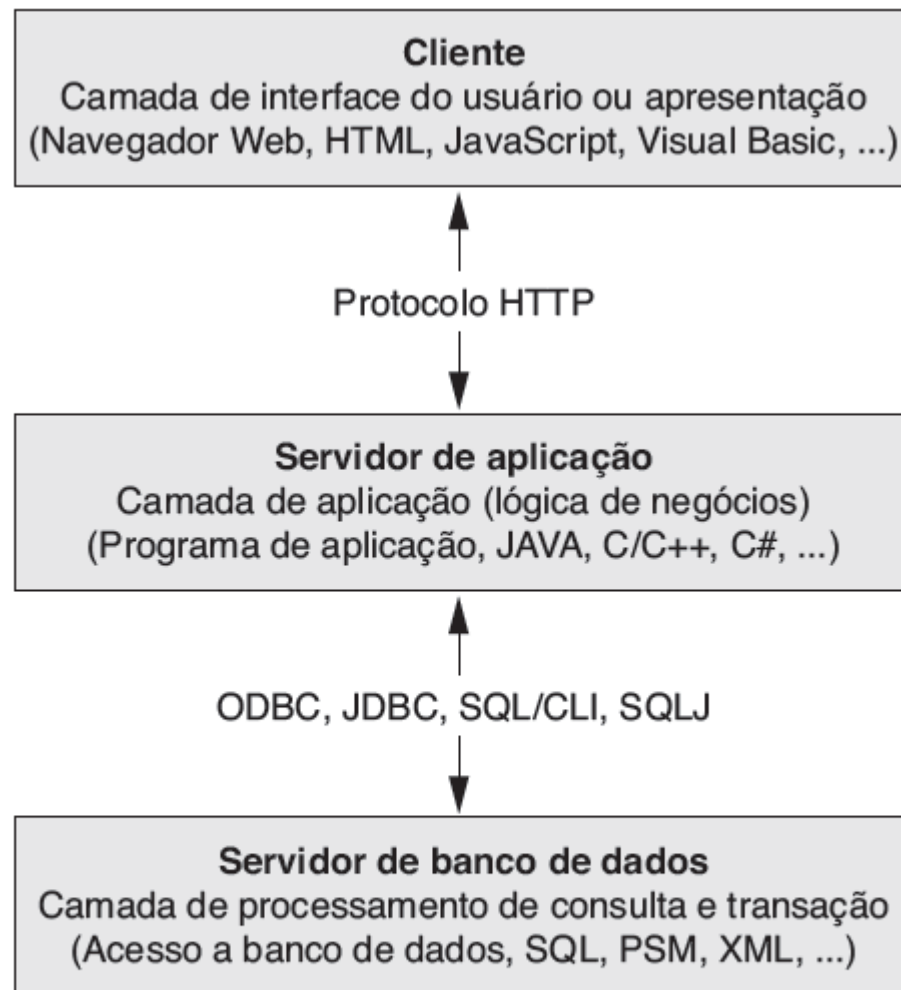


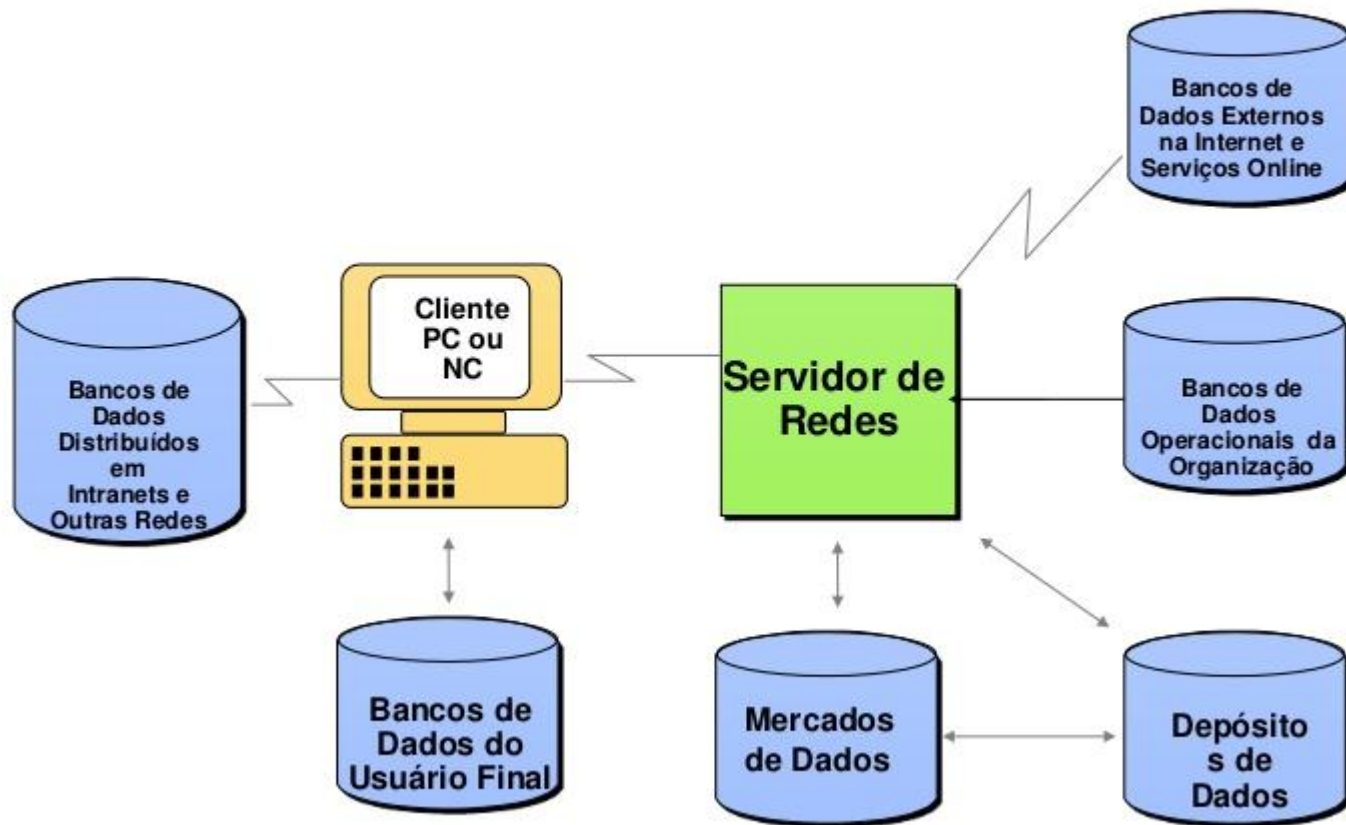
Figura 25.7

Arquitetura cliente-servidor de três camadas.



Heterogeneidade de BD

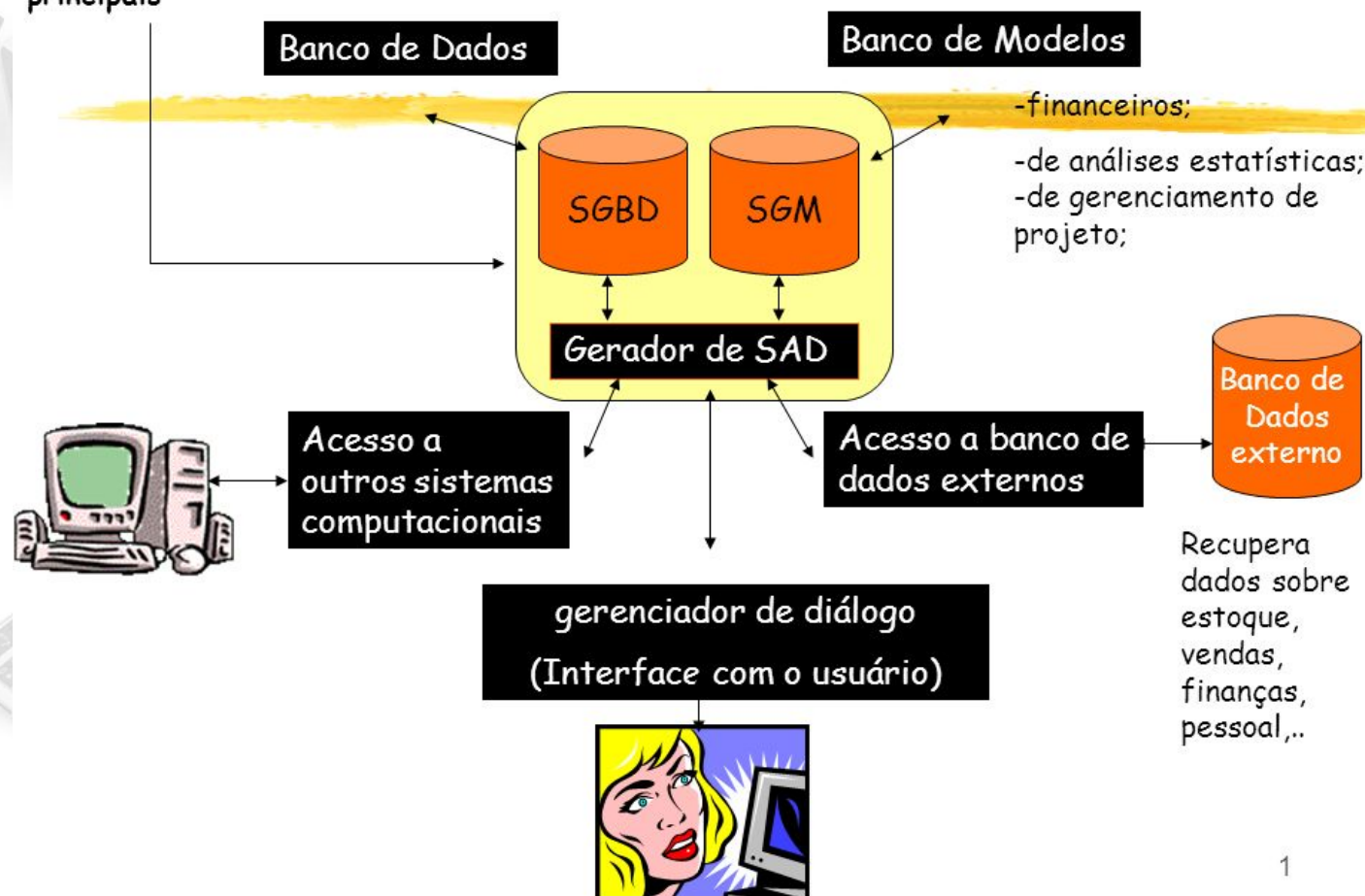
Principais Tipos de Banco de Dados



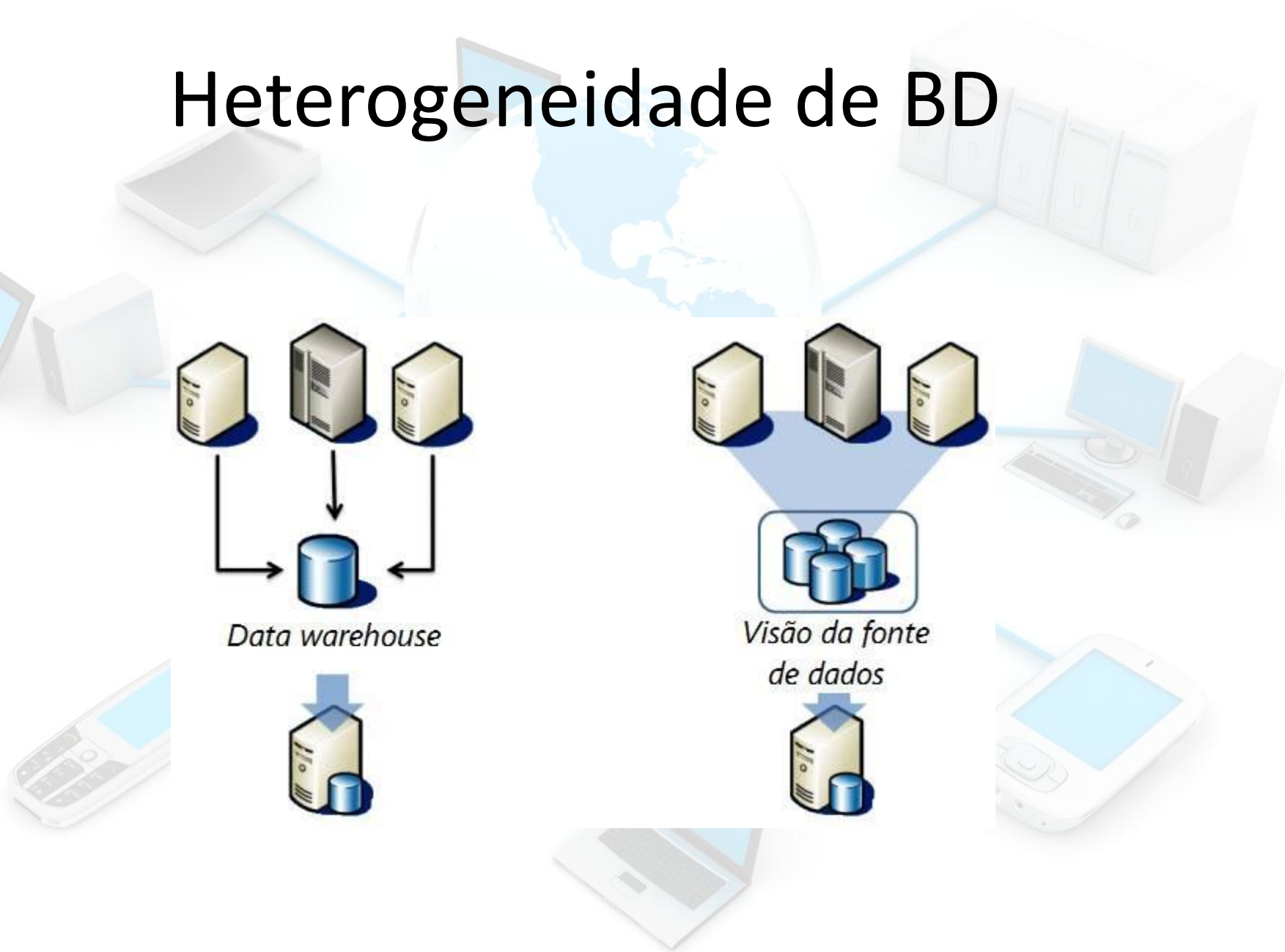
Heterogeneidade de BD

Arquitetura Funcional de SADs

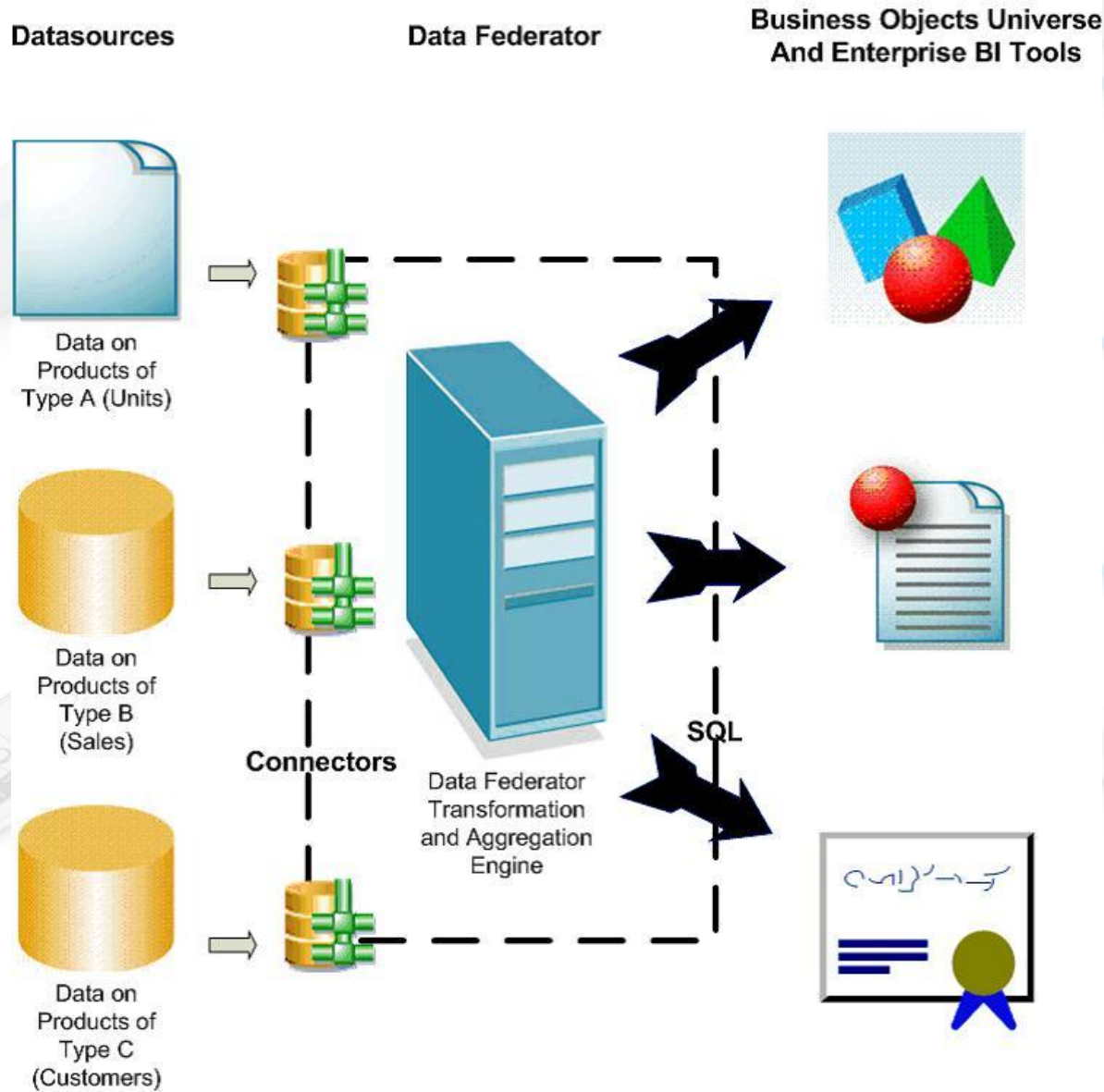
componentes principais



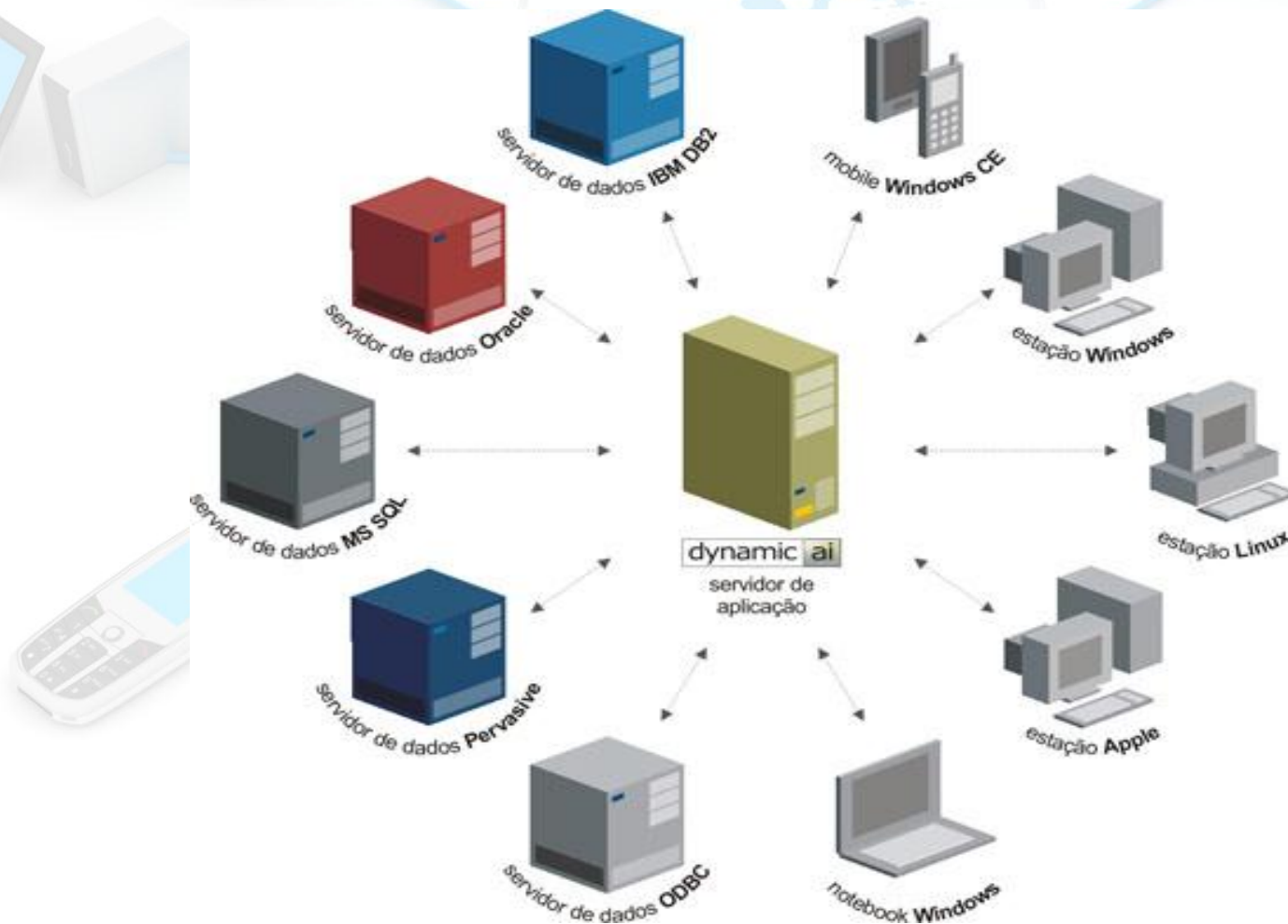
Heterogeneidade de BD



Heterogeneidade de BD

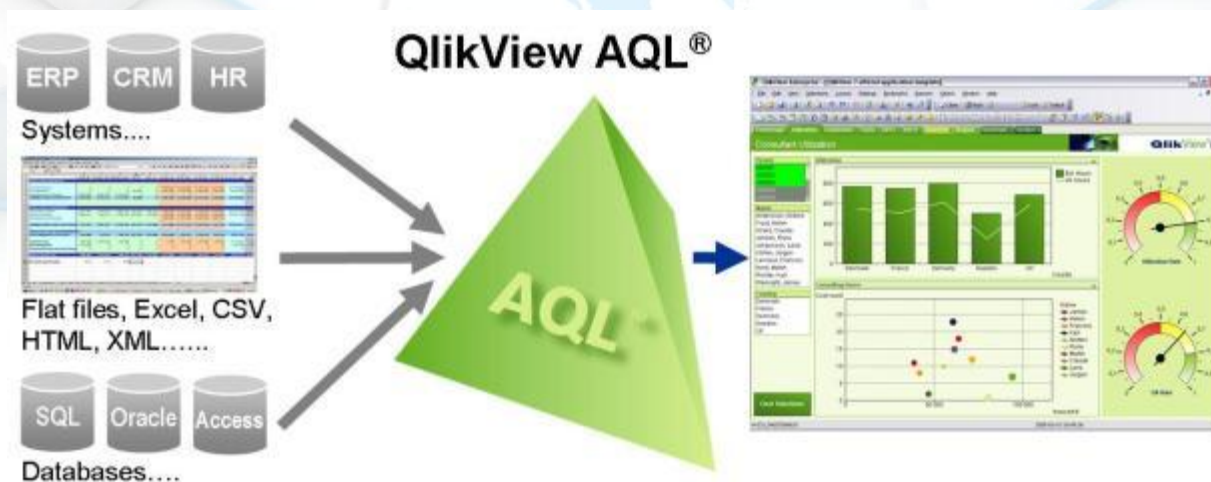


Heterogeneidade de BD



Heterogeneidade de BD

AQL™ (Associative Query Logic)



Com poder de uma inteligência artificial que compõem o AQL, são criados links de relacionamento entre palavras-chave, que permitem o desenvolvimento de análises em diversas dimensões, sem limites para o cruzamento de dados.

Heterogeneidade de BD

Características das tecnologias OLAP

- Exigem o investimento em Datawarehouse;
- Apresentam número limitado de dimensões;
- São inflexíveis e complexos;
- Permitem consultas somente pré definidas;
- Não possuem interface amigável com o usuário;
- Demandam grandes volumes de espaço em disco;
- Alto custo em licenças, implementação e estrutura de máquinas;
- Tempo mínimo de implantação de projetos entre 6 à 12 meses.

Diferenciais do AQL

- Ferramenta de ETL integrada;
- Não há obrigatoriedade de um Datawarehouse;
- Interface extremamente amigável com o usuário;
- Dados processados em memória (RAM) do computador e do servidor;
- Não há processamento em disco;
- Análises on line e off line;
- Permite integração via browser com celulares, palm top e iPhone;
- Flexibilidade para usuários realizarem modificações;
- Capacidade de análise em grandes volumes de dados (64 bits);
- Fácil aprendizagem e uso intuitivo;
- Tempo de implantação de projetos entre uma a 4 semanas;
- Reconhecimento e poder em 64bits;
- Alteração de dimensões em tempo real de análise.

