




Integração de Dados e Warehousing

Integração de Esquemas: MULTIDB

Fernando Fonseca
Ana Carolina

Centro de Informática
UNIVERSIDADE FEDERAL DE PERNAMBUCO
70 ANOS
CENTRO DE INFORMÁTICA


Cln.ufpe.br



Introdução

- Cenário motivador do surgimento de Multibanco de Dados (MULTIDB)
 - ◆ Avanços na tecnologia de SGBD
 - Sistemas economicamente viáveis
 - ◆ Uso em grande escala
 - ◆ Dados armazenados não em função de uma aplicação (ou mais), porém de modo a representar de forma a mais completa possível um determinado universo
 - ◆ Diversos BD armazenando os dados de segmentos distintos e, aparentemente, totalmente independentes, da mesma organização

Cln.ufpe.br




Introdução

- Cenário motivador ... (Cont.)
 - ◆ Evolução das necessidades por informação mais completa
 - Aplicações consultam diversos BD distintos
 - ◆ É exigido do usuário o conhecimento dos diversos SGBD utilizados
 - ◆ Manutenção da consistência volta a ser feita pelos aplicativos
 - ◆ Internet

→

Interoperabilidade de sistemas de banco de dados


Cln.ufpe.br



Abordagens para Interoperabilidade

- Abordagens para desenvolver aplicações requerendo acesso a fontes heterogêneas de dados
 - ◆ Migração de dados de um sistema para outro
 - ◆ Gateways para pares específicos de sistemas de gerenciamento de dados
 - ◆ MultiBanco de Dados Federados


Cln.ufpe.br



Abordagens para Interoperabilidade

- Migração de dados de um sistema para outro
 - ◆ Principalmente para ter dados de um sistema lidos por outro
 - ◆ Substituição de um SGBD por outro
 - ◆ Exemplo
 - IMS Extract → IBM: SQL/DS e DB2/MVS (Hierárquico → Relacional)


Cln.ufpe.br



Abordagens para Interoperabilidade


- Migração... (Cont.)
 - ◆ Problemas
 - Migração apenas para tornar dados disponíveis para outro sistema
 - Pode ser necessária a migração para o sistema de origem, de dados que foram modificados no sistema de destino

Cln.ufpe.br

 **Abordagens para Interoperabilidade**


- Migração...
 - ◆ Problemas (Cont.)
 - Se o sistema do qual todos os dados migraram for desativado
 - ◆ Aplicações podem precisar de ser convertidas para rodar no novo sistema (manual/automaticamente)
 - ◆ Uma emulação do sistema original pode ser necessária no sistema para o qual os dados migraram

Cln.ufpe.br 7

 **Abordagens para Interoperabilidade**


- Gateways para pares específicos de sistemas de gerenciamento de dados
 - ◆ Traduz uma consulta na linguagem de um dos elementos do par numa consulta na linguagem do outro elemento e a submete a esse elemento
 - INGRES → DEC RMS; INGRES → dBase
 - ORACLE → IMS; ORACLE → RMS

Cln.ufpe.br 8


 **Abordagens para Interoperabilidade**

- Gateways... (Cont.)
 - ◆ Problemas
 - Não dão suporte ao gerenciamento de transações - são chaveadores e tradutores de consultas
 - Nenhum dos sistemas coordena controle de concorrência nem recuperação de transações que envolvam atualizações em ambos os sistemas


Cln.ufpe.br 9

 **Abordagens para Interoperabilidade**

- Gateways...
 - ◆ Problemas (Cont.)
 - Não trata dos problemas de homogeneização das diferenças estruturais e representacionais entre diferentes esquemas


 **Manipulação "Simultânea" de Dados Heterogêneos**

Cln.ufpe.br 10

 **Abordagens para Interoperabilidade**

- MultiBanco de Dados Federados
 - ◆ Solução mais viável e geral
 - ◆ Um Sistema de Multi-Banco de Dados (MBD) é um sistema de banco de dados que reside sobre sistemas de arquivos e SGBD existentes chamados de Sistemas de Banco de Dados Locais e dá a ilusão a seus usuários de que é um único SGBD

Cln.ufpe.br 11

 **Abordagens para Interoperabilidade**

- MultiBanco de Dados Federados (Cont.)
 - ◆ Mantém um esquema global sobre o qual os usuários emitem consultas e atualizações
 - ◆ Todos os dados residem nos sistemas locais, no MBD reside apenas o esquema global
 - ◆ O esquema global é construído pela consolidação (integração) dos esquemas dos sistemas locais


Cln.ufpe.br 12



Abordagens para Interoperabilidade

- **MultiBanco de Dados Federados (Cont.)**
 - ◆ Esse processo em geral requer homogeneização das diferenças de esquemas (conflitos)
 - ◆ O MBD traduz a consulta global (ou atualização) em consultas (ou atualizações) para serem enviadas aos sistemas locais apropriados para o processamento de fato, junta os resultados obtidos e gera o resultado final

Cln.ufpe.br 13



Abordagens para Interoperabilidade


- **MultiBanco de Dados Federados (Cont.)**
 - ◆ O MBD coordena o commit ou abort de transações dos sistemas locais que as processaram para manter a consistência dos dados nos sistemas locais
 - ◆ MBD controlam gateways múltiplos, um para cada sistema local

Cln.ufpe.br 14



MultiBanco de Dados


Cln.ufpe.br 15



MultiBanco de Dados

- **Objetivos**
 - ◆ Deve simplificar a necessidade de uma conversão em bloco e migração de dados de uma fonte para outra
 - ◆ Não deve requerer mudanças no software dos sistemas locais (LBD) - Autonomia de Projeto. Para cada LBD, o MBD deve aparecer como outra aplicação ou usuário


Cln.ufpe.br 16



MultiBanco de Dados

- **Objetivos (Cont.)**
 - ◆ Não pode impedir nenhum dos sistemas locais de ser utilizado na forma nativa
 - ◆ LBD continua a trabalhar para transações que requerem dados de um único BD local
 - ◆ MBD é utilizado para transações que requerem acessos a mais de uma fonte
 - ◆ Deve tornar possível a usuários e aplicações interagir com o MBD numa linguagem de BD

Cln.ufpe.br 17



MultiBanco de Dados

- **Objetivos (Cont.)**
 - ◆ MBD não deve trabalhar com as diferentes interfaces de linguagens dos LBD
 - ◆ Deve proteger usuários e aplicações da heterogeneidade dos ambientes operacionais dos LBD
 - Hardware, sistemas operacionais, protocolos de rede
 - ◆ Deve dar suporte a transações distribuídas envolvendo leituras e atualizações sobre diferentes BD

Cln.ufpe.br 18



MultiBanco de Dados

- **Objetivos (Cont.)**
 - ◆ Deve tornar disponíveis todas as facilidades providas por sistemas padrão de banco de dados
 - Definição de esquemas
 - Consultas não-procedurais
 - Otimização automática de consultas
 - Controle de concorrência
 - Recuperação
 - Controle de integridade
 - Autorização de acesso

Cln.ufpe.br
19



MultiBanco de Dados

- **Objetivos (Cont.)**
 - ◆ Não deve introduzir virtualmente mudanças nas operações e administração de qualquer um dos LBD
 - ◆ Deve prover desempenho em tempo de execução que se aproxime do desempenho de um BD distribuído homogêneo

MBD é um SEBD completo

Cln.ufpe.br
20




MultiBanco de Dados

- **Arquitetura Básica**
 - ◆ Um só nó MBD dirigindo múltiplos LBD

Um Sistema MBD é uma arquitetura Cliente-Servidor: múltiplos clientes do MBD Interagem com um único servidor MBD

Cln.ufpe.br
21



MultiBanco de Dados


- **Arquitetura Básica (Cont.)**
 - ◆ Ferramentas de acesso a BD e de desenvolvimento de aplicações são fornecidas com o MBD

Usuários podem desenvolver aplicações que acessem múltiplos LBD

- ◆ Gerenciamento de toda informação de controle

Desde o esquema global à submissão e ao gerenciamento de transações que invoquem um ou mais sistemas de LBD

Cln.ufpe.br
22



MultiBanco de Dados


- **Arquitetura Básica (Cont.)**
 - ◆ Há um driver para cada LBD

Cada driver reside no mesmo nó que o próprio LBD. MBD e drivers comunicam-se por meio de um subsistema de comunicação

- ◆ MBD é completamente não-vinculado aos LBD

Nenhuma modificação é necessária aos LBD. Usuários dos LBD e aplicações podem acessar diretamente seus dados por meio desses LBD

Cln.ufpe.br
23



MultiBanco de Dados

- **Arquitetura Básica (Cont.)**
 - ◆ Banco de dados global

Uma coleção de visões definidas sobre os LBD, nos quais dados são de fato guardados

- ◆ Processamento de consultas

Consultas e atualizações no MBD são traduzidas em consultas e atualizações para processamento nos LBD que armazenam os dados de interesse


Cln.ufpe.br
24



MultiBanco de Dados

- Arquitetura Básica (Cont.)
 - Processamento de Consultas
 - Devendo providenciar
 - Transformação de formatos
 - Casamento de informação
 - Pós-processamento com junção inter-BD


Cln.ufpe.br
25



MultiBanco de Dados

- Arquitetura Básica (Cont.)
 - Gerenciamento de transações
 - Distribuído sobre os LBD de modo que todas as atualizações em uma transação no MBD, mesmo operando em múltiplos LBD, são **simultaneamente** efetivadas (commit) ou eliminadas (abort)
 - O gerenciamento de transações deve preservar as tradicionais propriedades de **atomicidade e serializabilidade**


Cln.ufpe.br
26



Heterogeneidade Esquemática em MBD

- Proliferação de BD baseados em diferentes modelos impede os usuários de terem acesso uniforme aos dados a partir de fontes heterogêneas
- MBD provêm um **ambiente uniforme** no qual o usuário pode acessar dados de componentes heterogêneos de BD
- O MBD é uma federação de BD componentes desenvolvidos **de forma independente**


Cln.ufpe.br
27



Heterogeneidade Esquemática em MBD

- MBD provê uma camada **homogeneizante** para cada BD componente, de modo a dar a ilusão de um sistema de BD homogêneo
- Um elemento de homogeneização importante é o esquema do MBD (**global**) que integra os esquemas dos BD componentes


Cln.ufpe.br
28



Processo de Integração de Esquemas

- Derivação de **um só** esquema a partir de um **conjunto** de esquemas por meio de uma sequência de funções, cada uma delas resolvendo um **conflito esquemático**
 $esq_integ_proc: esquema_1 \times esquema_2 \times \dots \times esquema_n \rightarrow esquema_integrado$
- Não existe** uma metodologia geral para enumeração completa e classificação sistemática de técnicas de resolução para conflitos esquemáticos

Cln.ufpe.br
29



Processo de Integração de Esquemas

- Possibilidades Básicas de Integração
 - Sistemas relacionais
 - Sistema relacional e sistema orientado a objetos
 - Sistemas orientados a objetos
 - Sistema relacional e sistema legado
 - Sistema orientado a objetos e sistema legado

Cln.ufpe.br
30

Processo de Integração de Esquemas

- MBD - Modelo de Dados
 - ◆ Abrigar construções existentes nos modelos locais
 - ◆ Deve ser utilizado por um SGBD que possua a funcionalidade requerida para atuar como um SGBD global
 - ◆ Objetivo principal é fornecer uma visão integrada dos dados

Cln.ufpe.br 31

Processo de Integração de Esquemas

- Conflitos Esquemáticos em um MBD
 - ◆ Um dado conceito pode ser expresso de várias maneiras em diferentes LBD
Nomes, estruturas, tipos de dados, ...

↓

- ◆ Impossível manipular dados em um MBD com linguagens tradicionais

↓

- ◆ Extensões às tradicionais DML e DDL para definir esquemas de MBD → Resolução de Conflitos

EX.: BDR → BDOO

Cln.ufpe.br 32

Processo de Integração de Esquemas

- Conflitos Esquemáticos em um MBD (Cont.)
 - ◆ Problemas
 - Esquemas dos LBD usam diferentes estruturas para representar a mesma informação
 - Diferentes esquemas usam uma estrutura similar, mas especificações diferentes para representar a mesma informação
 - Generalização e agregação, bem como métodos
 - Inclusão de entidades e atributos (Generalização)
 - Tipos de dados
 - Composição (Agregação)

Cln.ufpe.br 33

Processo de Integração de Esquemas

- Conflitos
 - ◆ Entidade X Entidade
 - Um - a - um
 - ◆ Nome da Entidade
 - Nomes diferentes para entidades equivalentes
 - Mesmo nome para entidades diferentes
 - ◆ Estrutura da Entidade
 - Atributos faltando
 - Atributos faltando, mas implícitos

Cln.ufpe.br 34

Processo de Integração de Esquemas


- Conflitos
 - ◆ Entidade X Entidade
 - Um - a - um (Cont.)
 - ◆ Restrições
 - ◆ Inclusão de Entidades
 - Muitos - para - muitos

Cln.ufpe.br 35

Processo de Integração de Esquemas

- Conflitos (Cont.)
 - ◆ Atributo X atributo
 - Um - a - um
 - ◆ Nome do atributo
 - ◆ Restrições de atributo
 - ◆ Restrições de integridade
 - ◆ Tipos de dados
 - ◆ Composição
 - ◆ Valores default
 - ◆ Inclusão de atributo
 - ◆ Métodos
 - Muitos - para - muitos


Cln.ufpe.br 36



Processo de Integração de Esquemas

- Conflitos (Cont.)
 - ◆ Entidade X Atributo
 - Representação diferente para dados equivalentes
 - ◆ Expressão diferente denotando a mesma informação
 - ◆ Unidades diferentes
 - Diferentes níveis de precisão

Cln.ufpe.br 37



Processo de Integração de Esquemas

- Conflitos (Cont.)
 - ◆ Generalização
 - Inclusão de Entidades

Uma entidade de um LDB está logicamente incluída em outra entidade de outro LDB

 - ◆ Ex.: Estudante em um LDB e Est_Posgrad em outro
 - ◆ Pior caso: Hierarquia de herança de um BDOO a ser integrada com a hierarquia de herança de outro BDOO

Cln.ufpe.br 38




Processo de Integração de Esquemas

- Conflitos
 - ◆ Generalização (Cont.)
 - Inclusão de Atributo

Relacionamentos de inclusão existem entre dois ou mais atributos

 - ◆ Ex.: Em um LDB nome_do_filho pode ser visto como se fosse incluído em nome_criança em outro LDB
 - ◆ Pode ser usado para induzir uma hierarquia de herança natural entre as entidades correspondentes

Cln.ufpe.br 39




Processo de Integração de Esquemas

- Conflitos de Generalização/Agregação
 - ◆ Tipos de dados

Os domínios ou tipos são diferentes para atributos semanticamente equivalentes em BD distintos

 - Ex.: cpf pode ser char(11) em um LDB e integer em outro
 - Mais geral ainda pode ocorrer com BDOO
 - Ex.: depto numa classe Professor de um LDB pode ser char(20), enquanto num outro LDB pode ser do tipo de dados Departamento, uma classe definida pelo usuário

Cln.ufpe.br 40




Processo de Integração de Esquemas

- Conflitos de Generalização/Agregação (Cont.)
 - ◆ Composição

Conceitos similares são representados em um modelo de dados como uma agregação, enquanto não são no outro

 - Ex.: Atributo disciplina de uma classe Cursa em um LDB com o domínio Disciplina que tem um atributo prereq, cujo domínio é set_of(Disciplina). Mais complicado ainda se o outro LDB fosse relacional

Cln.ufpe.br 41



Processo de Integração de Esquemas

- Conflitos (Cont.)
 - ◆ Métodos
 - Só ocorre em integrações envolvendo BDOO ou sistemas objeto-relacionais
 - Como é parte da definição de uma classe (ou tipo de objeto), pode ser tratado como no caso de um atributo
 - ◆ Falta método (faltando atributo)
 - ◆ Polimorfismo (conflito de nome)
 - ◆ Argumentos de tipos diferentes (conflito de tipo)
 - ◆ Especialização de métodos em subclasses (inclusão de atributo)

Cln.ufpe.br 42