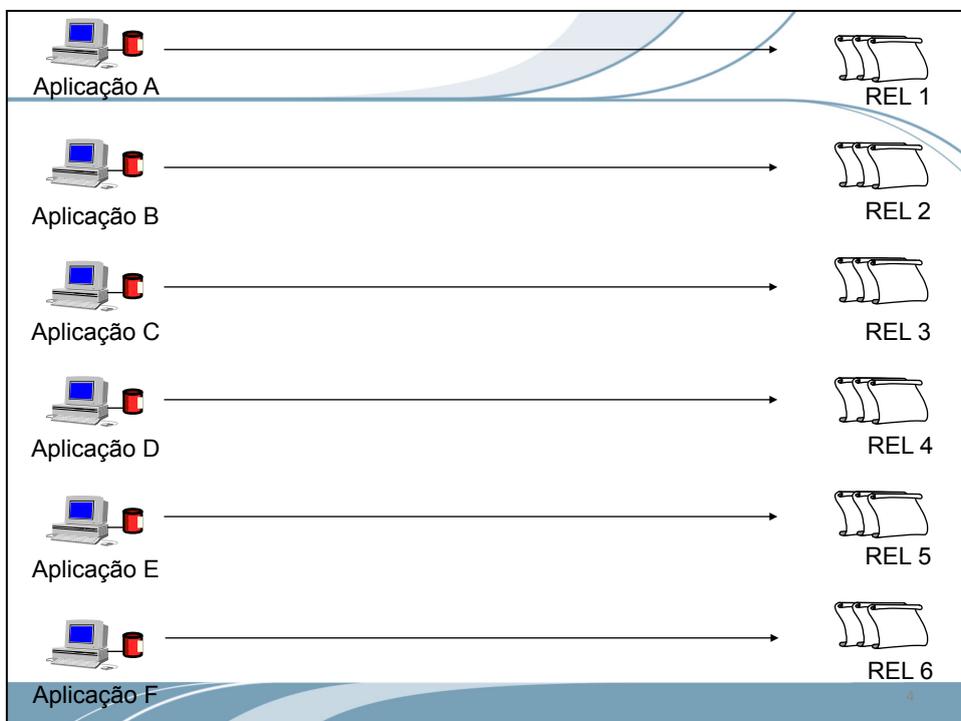
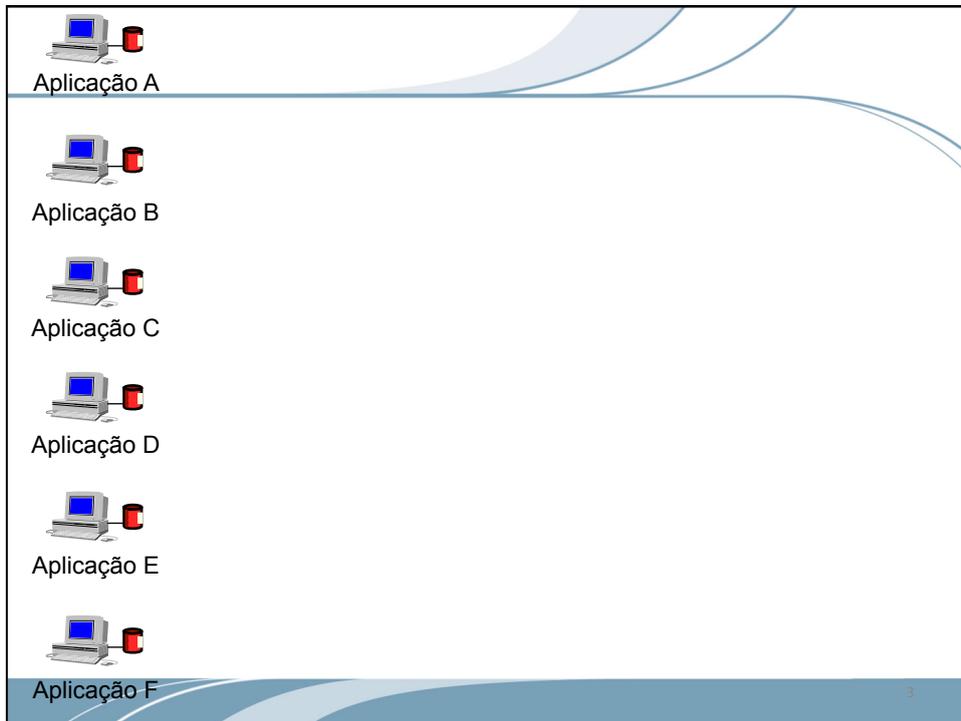


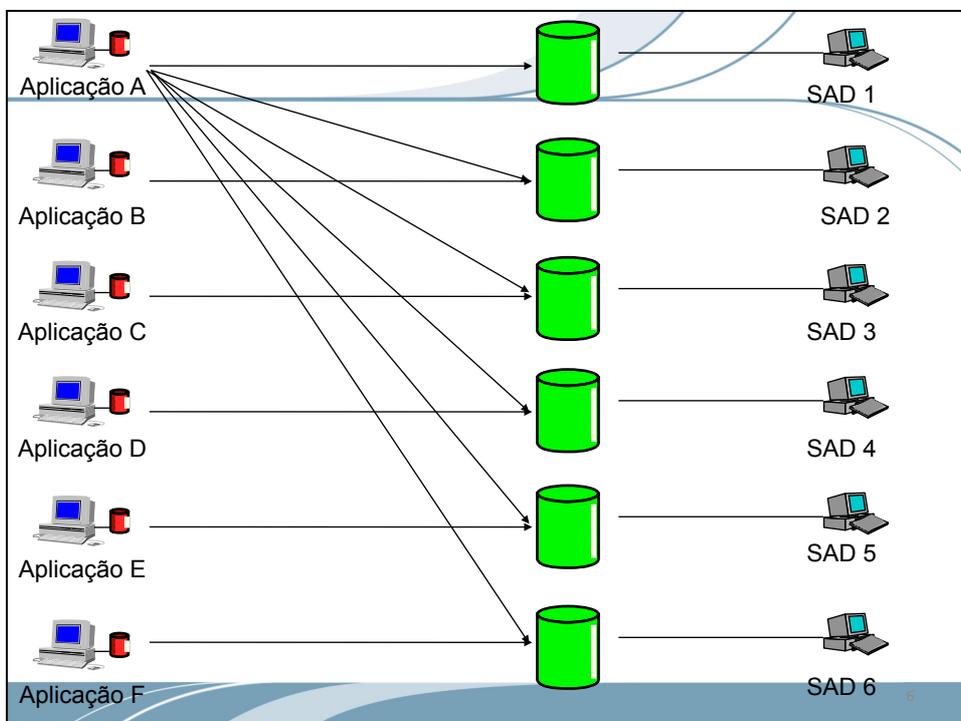
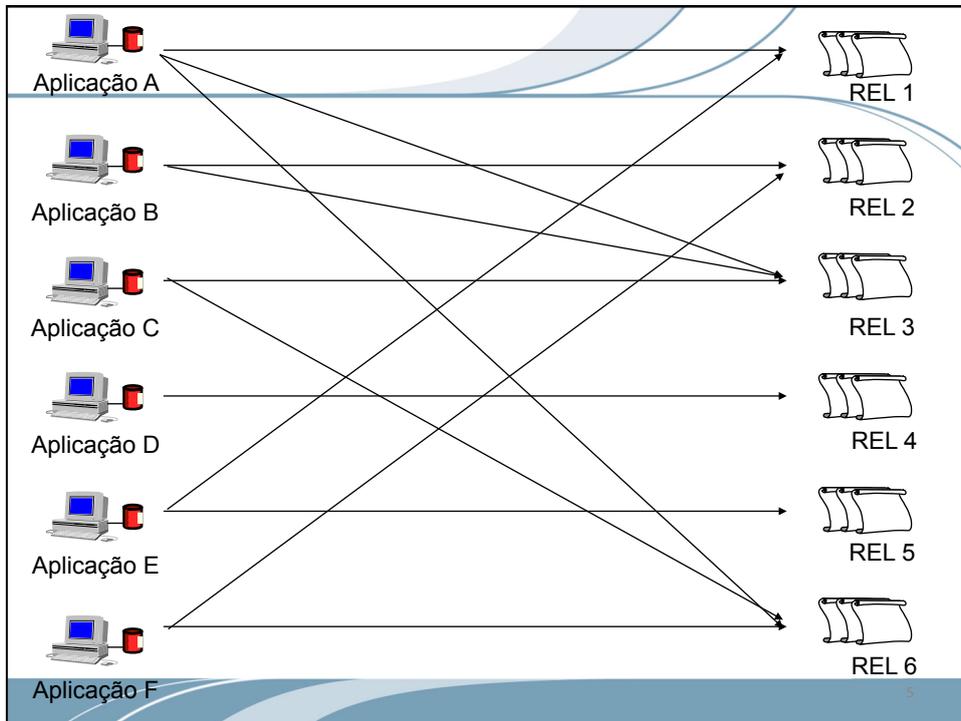
DataSpaces

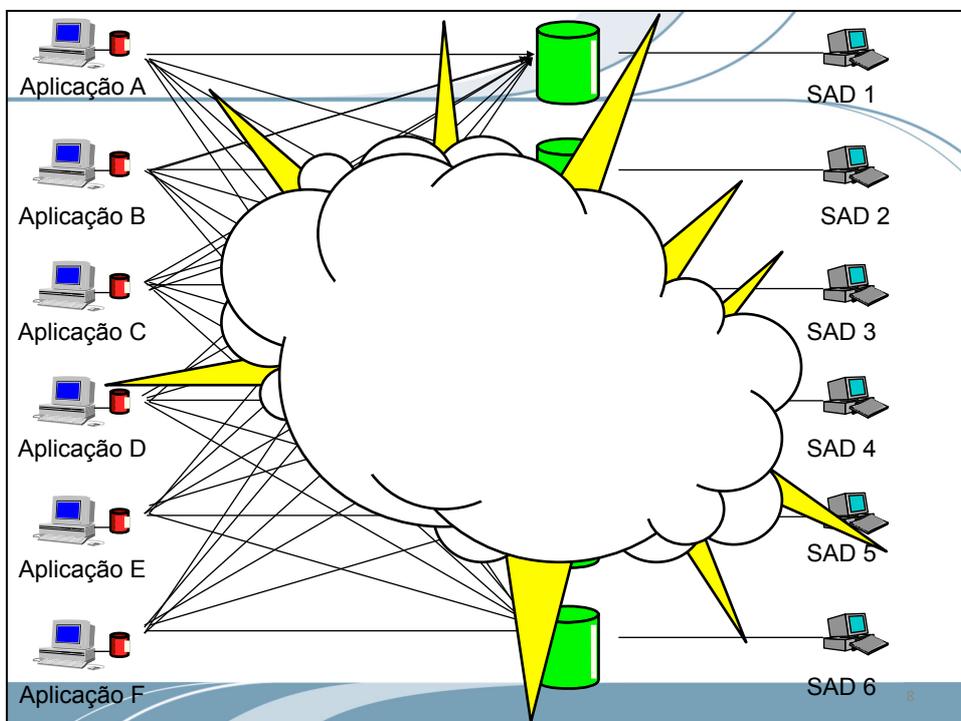
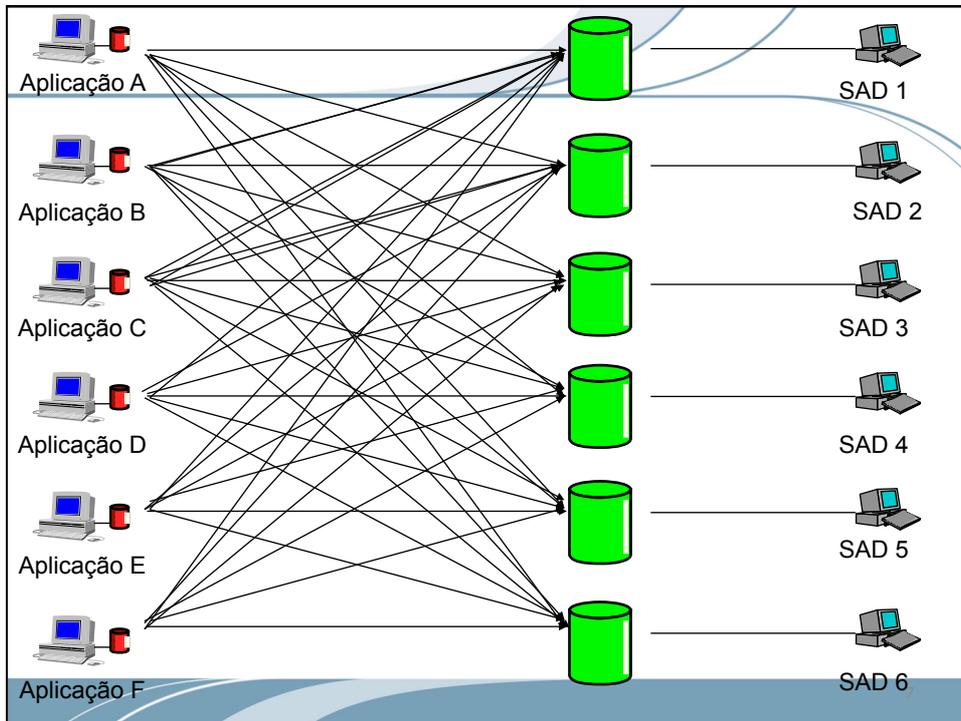
Edilson Ferreira da Silva
Maria Carolina Revoredo Martiniano

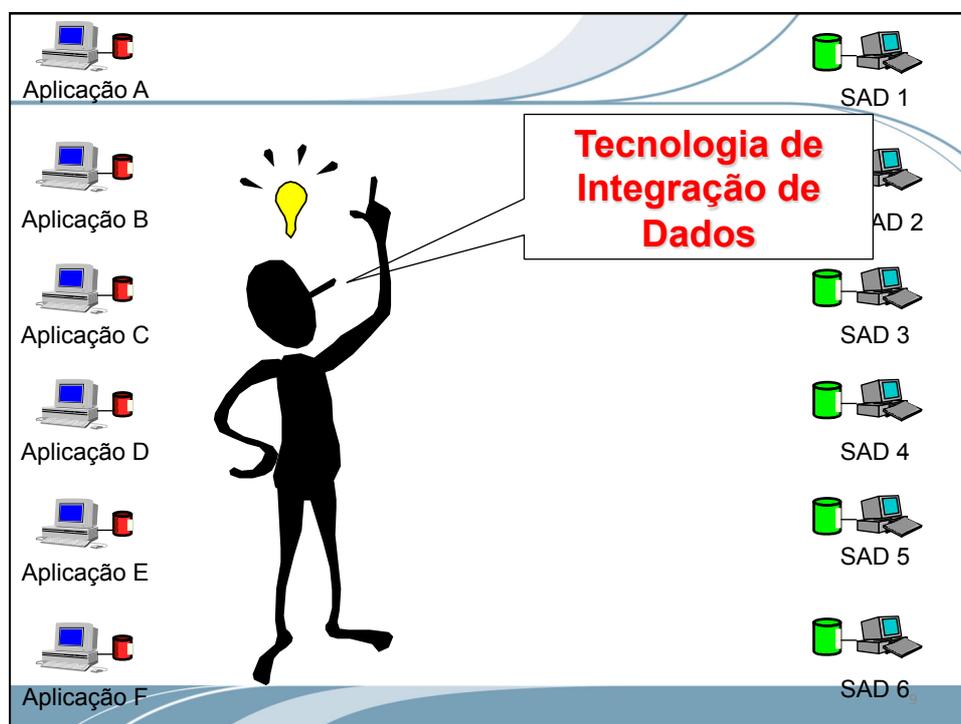
ROTEIRO

- Introdução
- Integração
 - O problema da Integração
 - Abordagens de Integração
- DataSpaces
- Pay-as-You-Go
- Conclusão









Abordagens de Integração

- Virtual
 - Informações consultadas nas fontes apenas quando as consultas são realizadas (Runtime)
 - Dados atualizados
- Materializada
 - Informações são recuperadas, integradas e armazenadas em um repositório
 - Dados disponíveis imediatamente

Processo de extração

Virtual	Materializada
<p>Passo 1: o sistema de integração recebe uma consulta, determina o conjunto de fontes de dados necessárias para respondê-la e gere subconsultas apropriadas para cada fonte de dados.</p>	<p>Passo 1: as informações relevantes de cada fonte dados são extraídas, traduzidas e filtradas, em seguida estas informações integradas e armazenadas em um repositório.</p>
<p>Passo 2: o sistema coleta os resultados das subconsultas de cada fonte de dados, executa as traduções apropriadas, filtra, integra as informações e retorna para usuário.</p>	<p>Passo 2: quando uma consulta é requisitada ao sistema de integração, a consulta é avaliada diretamente no repositório, sem a necessidade de acessar as fontes de dados originais.</p>

Arquiteturas para integração

- **Arquitetura clássica**
 - Abordagem Federada

- **Arquitetura para integração de dados na web**
 - Arquitetura de mediadores
 - Arquitetura de Data Warehouse

Arquitetura Clássica

- Vantagens
 - Único esquema global
 - Oferece uma visão única dos dados
- Desvantagens
 - Não é adequada quando as fontes são dinâmicas
 - Esquema pode tornar-se muito grande e difícil de manter

Arquitetura Clássica

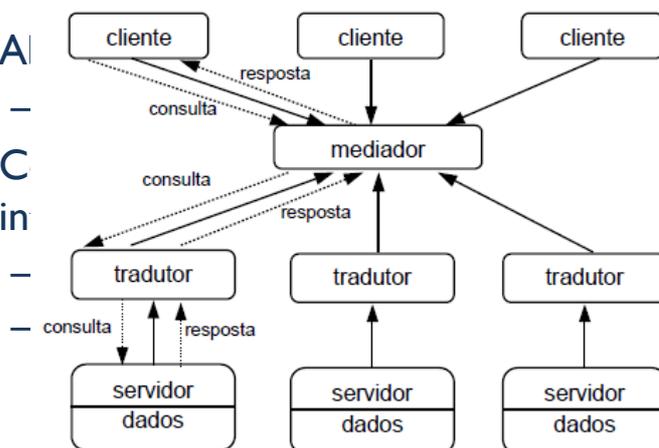
- Banco de dados federados
 - Coleção de sistemas de banco de dados
 - Cooperação de sistemas independentes
 - Múltiplos esquemas globais integrados
 - Estáticos e definidos a priori
 - Também não é adequada para fontes de dados dinâmicas

Integração de dados na web

- Forma alternativa aos tradicionais sistemas
- Abordagem de processamento *multitier*
 - Cliente (interfaces)
 - Fonte de dados
 - *Middleware*

Mediadores

- A
-
- C
- in
-
-



Data Warehousing

- Abordagem materializada
- ETL
 - Extração
 - Transformação
 - Carga
- Manutenção
 - Rematerialização da visão
 - Manutenção Incremental

Desafios na Integração dos Dados

- Modelagem de Dados
- Reformulação de Consultas
- Manutenção do sistema de integração

Modelagem de Dados

- Visão Global
 - Cada componente no esquema da visão integrada tem uma correspondência com o esquema da fonte
 - Reformulação de consultas simples
- Visão Local
 - Cada informação na fonte de dados tem uma correspondência com um esquema na visão integrada
 - Facilita a manutenção da fonte de dados
- Modelagem de dados com XML
 - Facilidade para converter dados
 - Flexibilidade

Reformulação de Consultas

- Tradução de consultas para as fontes de dados
 - Baseada descrições das fontes de dados
 - Correspondência entre esquemas
- Manter a correspondência correta
- Selecionar fontes relevantes

Manutenção do SID

- Fontes
 - Autônomas
 - Dinâmicas
- Aplicações também podem ser dinâmicas
 - Alterações de requisitos em relação ao sistema
- Manutenção dos tradutores
- Atualização quando ocorrem modificações no esquema global

DATASPACE

- O que são Dataspaces?
- DataSpace Support Platform (DSSP)
- Fases do Ciclo de Vida

O QUE SÃO DATASPACES?

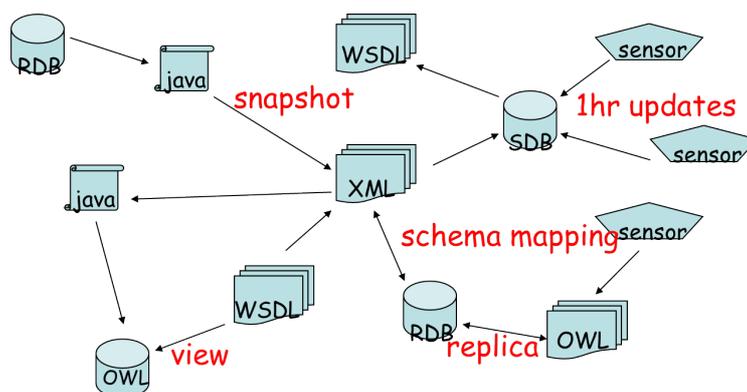
- Motivação
 - Gerenciamento de fontes de dados heterogêneas.
 - Necessidade de localizar dados relevantes e descobrir relacionamentos entre as fontes.

O QUE SÃO DATASPACES?

- Abstração para gerenciamento de dados
- Abordagem de coexistência de dados
- Sem necessidade de integração semântica completa, inicialmente
- Resposta com melhor esforço
- Pay-as-you-go
- DSSP

O QUE SÃO DATASPACES?

- Modelo Lógico



- Conjunto de participantes e relacionamentos

O QUE SÃO DATASPACES?

- Modelo Lógico

- Conjuntos de Participantes

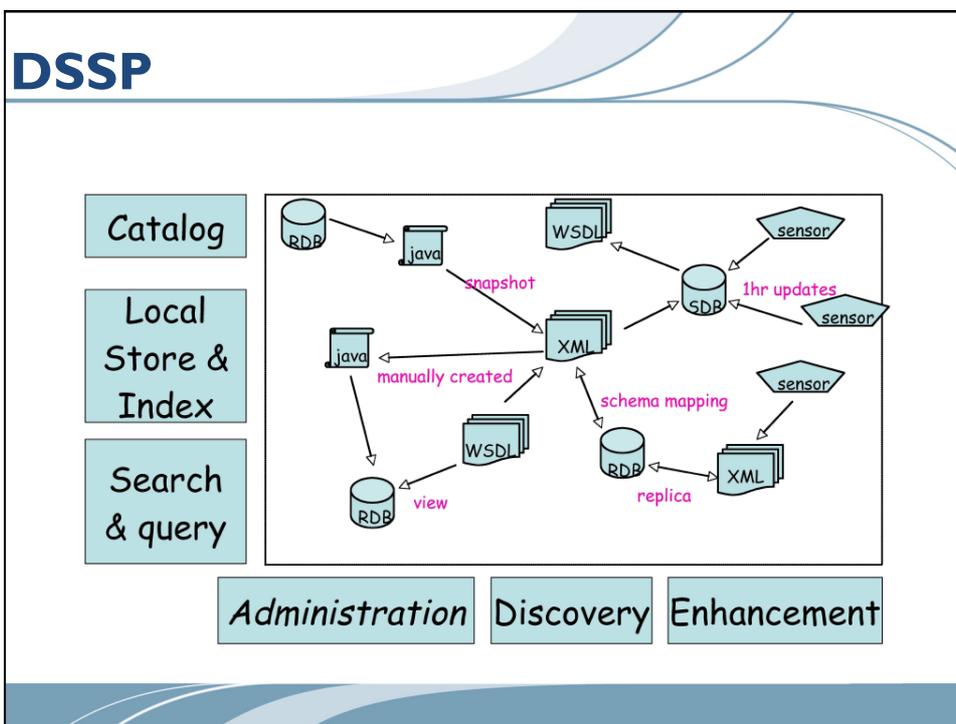
- Fontes de dados individuais
- Divergências no suporte a linguagens de consulta
- Fontes estruturadas, semiestruturadas ou não estruturadas
- Diferentes níveis de suporte a atualização

O QUE SÃO DATASPACE?

- Modelo Lógico
 - Conjuntos de Relacionamentos
 - Devem ser capazes de modelar qualquer tipo de relacionamentos
 - Modelar de forma ampla

DSSP

- Características
 - Suporte aos dados de todas as fontes do dataspaces
 - Proporciona pesquisa, consulta, atualização e administração do dataspaces
 - Consultas com diferentes níveis de detalhe
 - Ferramentas para maior integração



DSSP

- Catalog
 - Descrição de todos os participantes, informações básicas e semântica
 - Repositório de metadados

DSSP

- Local Store & Index
 - Criar associações eficientes de consultas entre objetos de dados em diferentes participantes
 - Melhorar os acesso às fontes de dados que tem padrões de acesso limitados
 - Responder a determinadas perguntas sem acessar a fonte de dados reais
 - Fornecer um grande suporte a disponibilidade e recuperação de dados

DSSP

- Search & Query
 - Consultar tudo
 - Consulta estruturada
 - Consultas aos metadados
 - Monitoramento

DSSP

- **Discovery**
 - Identifica os participantes
 - Cria relacionamentos automaticamente
 - Aperfeiçoamento de relacionamentos

FASES DO CICLO DE VIDA

- **Inicialização**
- **Teste/Avaliação**
- **Implantação**
- **Manutenção**
- **Uso**
- **Melhoria**

FASES DO CICLO DE VIDA

- Inicialização
 - Identificação das fontes de dados que serão acessadas e integradas.
 - As dimensões estão relacionadas com o processo e suas entradas, ou com as saídas do processo.
 - Matching – Correspondências entre elementos e atributos.
 - Mapping – Regras para executar as traduções entre esquemas.

FASES DO CICLO DE VIDA

- Inicialização
 - Identificação de fontes
 - Gerais ou específicas
 - Não estruturadas, semi-estruturadas, estruturadas.
 - Locais ou distribuídas
 - Integração de esquemas
 - Manual, semi-automático ou automático
 - União ou merge de esquemas

FASES DO CICLO DE VIDA

- Inicialização
 - Matching e sua identificação
 - Podem ser correspondências entre fontes locais ou entre fontes locais e o esquema de integração
 - Manual, semi-automático ou automático
 - Grau de incerteza
 - Entradas da identificação : esquema, instancia ou dados de treinamento.

FASES DO CICLO DE VIDA

- Inicialização
 - Mapping e sua identificação
 - Sempre expressos entre esquemas locais e o esquema de integração
 - Manual, semi-automático ou automático
 - Grau de incerteza
 - Entradas da identificação : instancia , matching ou dados de treinamento.

FASES DO CICLO DE VIDA

- Inicialização
 - Dados resultantes
 - Virtual, parcialmente materializada ou materializada
 - Grau de incerteza
 - Duplicatas ou conflitos: pode coexistir ou não

FASES DO CICLO DE VIDA

- Uso
 - Avaliação de consulta pode ser completa ou parcial
 - Combinação de resultados caso múltiplas fontes sejam consultadas
 - Verificação de incerteza

FASES DO CICLO DE VIDA

- **Manutenção e Melhoria**
 - Manutenção lida com o fato das fontes de dados serem autônomas
 - Melhoria visa aperfeiçoar os resultados de integração ao longo do tempo
 - Mesmos passos da fase de inicialização, além de feedback para o usuário e novos alinhamentos e mapeamentos

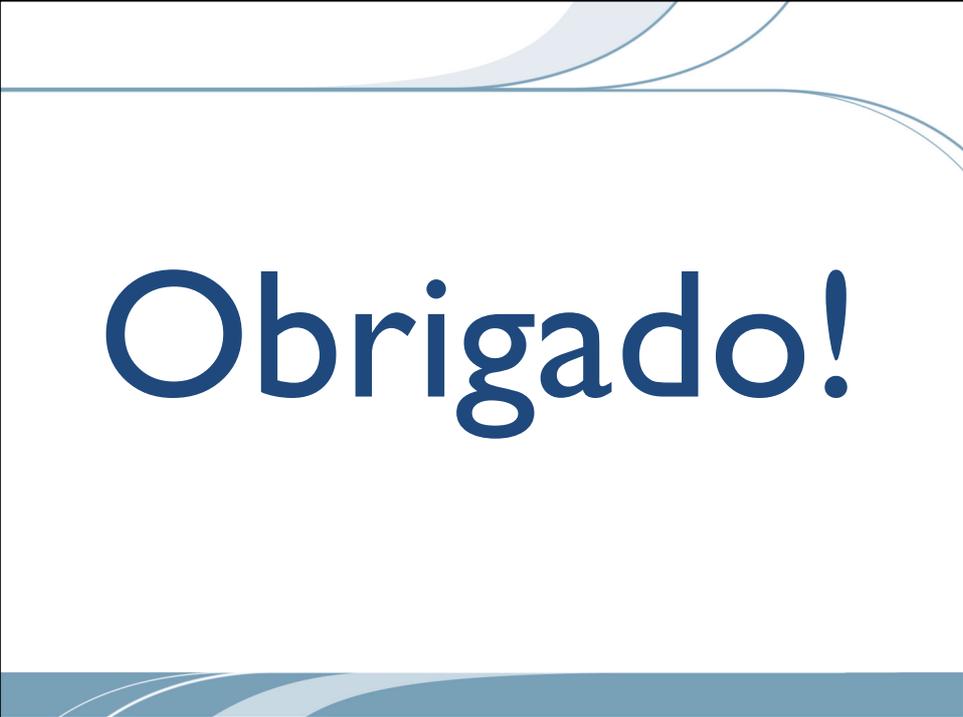
FASES DO CICLO DE VIDA

- **Manutenção e Melhoria**
 - **Manutenção**
 - DSMS precisam ser capazes de lidar com a evolução das fontes de dados
 - DSMS deve exigir pouco ou nenhum esforço manual para responder a essas mudanças
 - **Melhoria**
 - Aquisição de feedback sobre diferentes recursos de dados : alinhamentos, mapeamentos, esquemas e consultas
 - Uso de feedback implícito e explícito

CONCLUSÃO

- Necessidade de gestão de dados entre fontes heterogêneas e o inter-relacionamento entre os dados.
- Dataspace e DSSP podem ser uma possível solução para os problemas existentes na comunidade de BD.

Dúvidas?



Obrigado!