



Data Warehouse

- Definição
- Ambiente
 - Ferramentas ETL
 - Data Marts
- Arquiteturas
 - Global
 - Data Marts independentes
 - Data Marts integrados
- Implementação
 - Top Down
 - Botton up
 - Combinada
- Metadados

Introdução

- A informação é o melhor recurso do qual empresas podem dispor para tomar decisões
- Obtida analisando históricos sobre vendas, clientes, produtos, etc
- Dados conflitantes de fontes diversas podem gerar informações desencontradas

Introdução

- A quantidade de dados a serem considerados cresce com a expansão do negócio e com o passar do tempo...



- Data Warehouses auxiliam a resolver esses problemas ao prover montantes gigantescos de dados temporais integrados para posterior análise!

Introdução

- Criado pela IBM na década de 60 com o nome Information WareHouse
- Relançado diversas vezes sem sucesso
- O nome Data WareHouse foi dado por William Inmon, considerado o pai desta tecnologia
 - Tornou-se viável com o surgimento de novas tecnologias para armazenar e processar uma grande quantidade de dados.

Definição

Conjunto de dados **agrupados por assunto, integrados**, variável em relação ao **tempo e não volátil**, que serve de suporte para o processo de **tomada de decisões**.

Definição

- **Orientado a Assunto**

- Um Data Warehouse está sempre orientado ao redor do principal assunto da organização
 - Ao contrário de aplicações clássicas, orientadas por processos/funções

- **Integrado**

- Os dados criados dentro de um ambiente de Data Warehouse são integrados
 - A integração beneficia com a convenção consistente de nomes, estrutura consistente de códigos etc

Definição

- **Não volátil**
 - Os dados nunca são excluídos nem alterados de um Data Warehouse
- **Variante no tempo**
 - Data Warehouse apresenta os dados com seu posicionamento em relação ao tempo

Comparativo com BD operacional

Aspecto	BD Operacional	Data Warehouse
Usuários	Funcionários	Alta administração
Utilização	Tarefas Cotidianas	Decisões estratégicas
Padrão de Uso	Previsíveis	Difícil de prever
Princípio de Func.	Com base em transações	Com base em análise de dados
Valores de dados	Valores atuais e voláteis	Valores históricos e imutáveis
Detalhamento	Alto	Sumarizado
Organização dos dados	Orientado a aplicações	Orientado a assunto

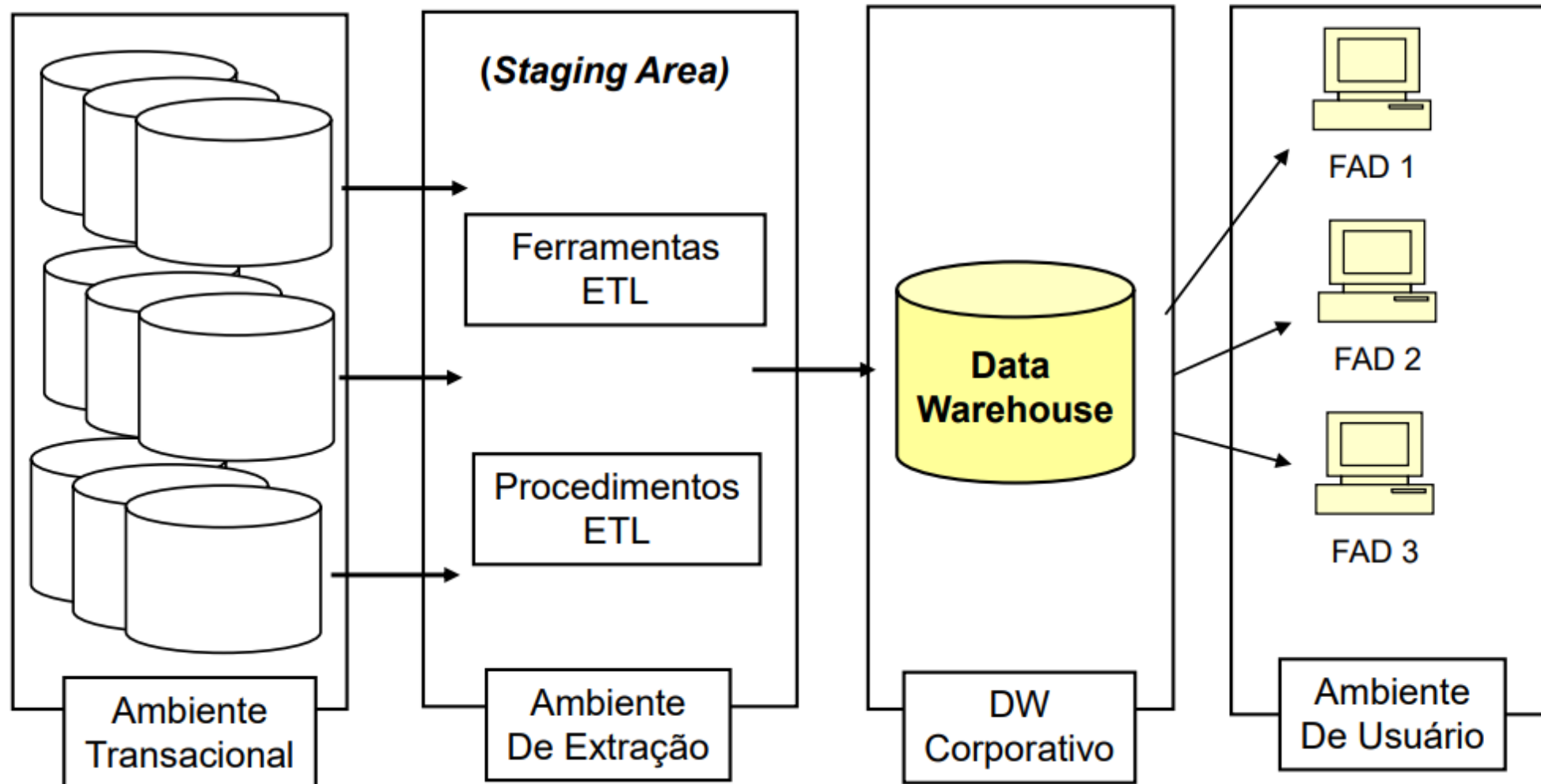
Comparativo com BD operacional

Algumas diferenças adicionais do Data Warehouse para um BD operacional

- Permitem a redundância de dados
- Buscas complexas e ad hoc (personalizadas pelo usuário)
- Modelagem de dados multidimensional

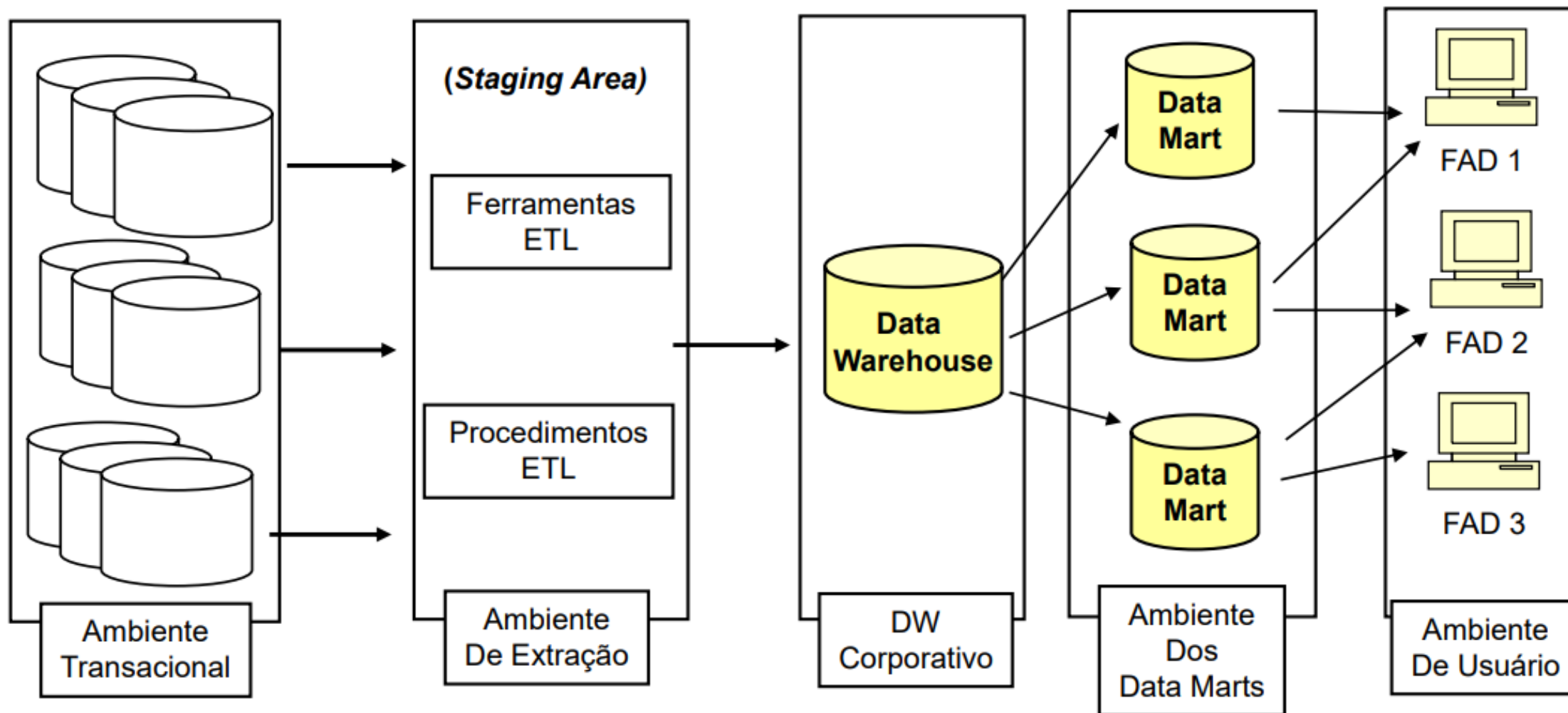
Ambiente de Data Warehouse

- Centralizado



Ambiente de Data Warehouse

- Com Data Marts



Ambiente de Data Warehouse

Extraction, Transformation and Load

- Consiste da integração e limpeza dos dados
- Integração: consolidação dos dados de diversas origens
- Limpeza: rejeição de valores inválidos

Ambiente de Data Warehouse

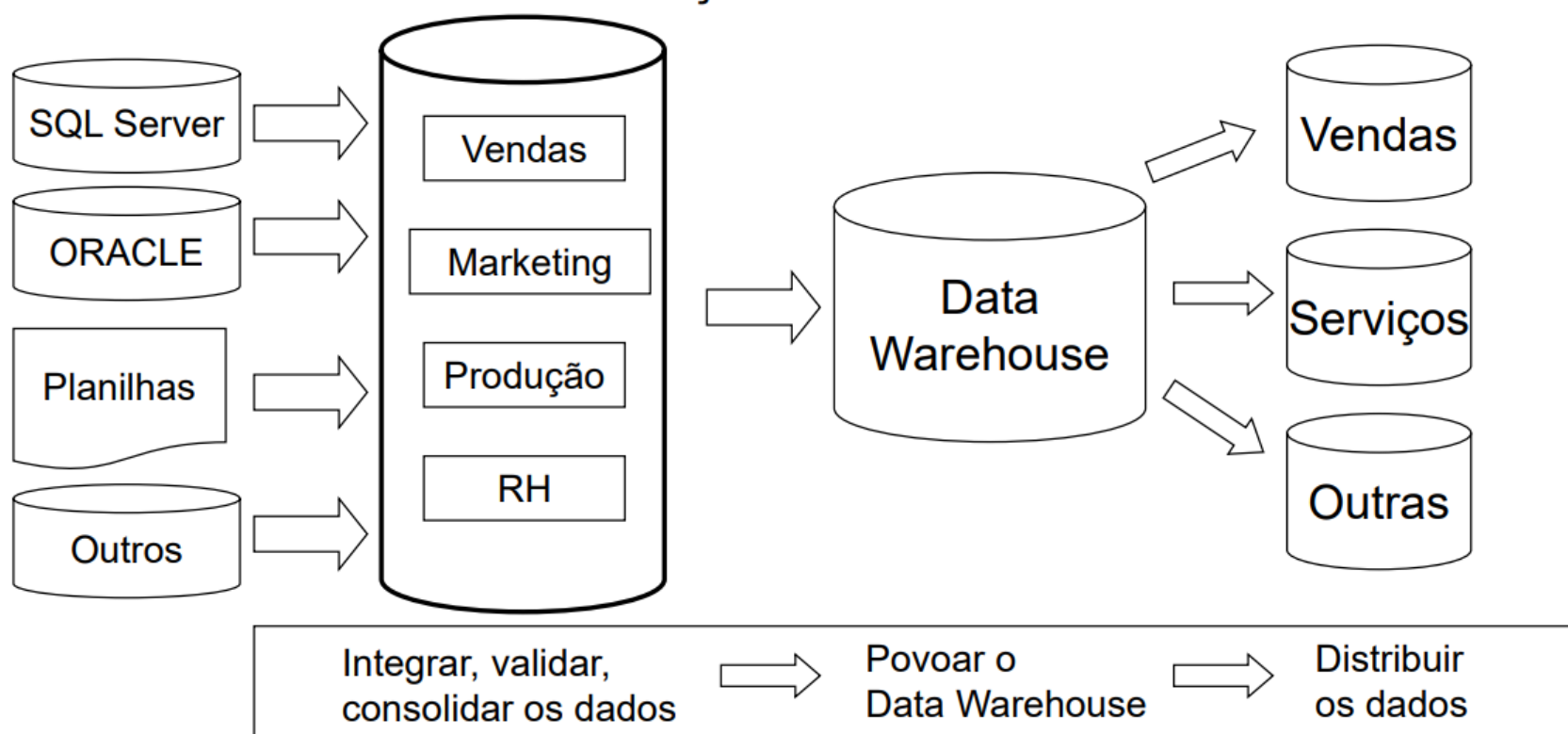
Extraction, Transformation and Load

- Os processos ETL consomem 70% do tempo de desenvolvimento em um projeto de DW
- Estes processos são específicos para cada organização
- Opcionalmente, pode-se ter uma segunda área intermediária, chamada Operational Data Store (ODS)

Ambiente de Data Warehouse

Extraction, Transformation and Load

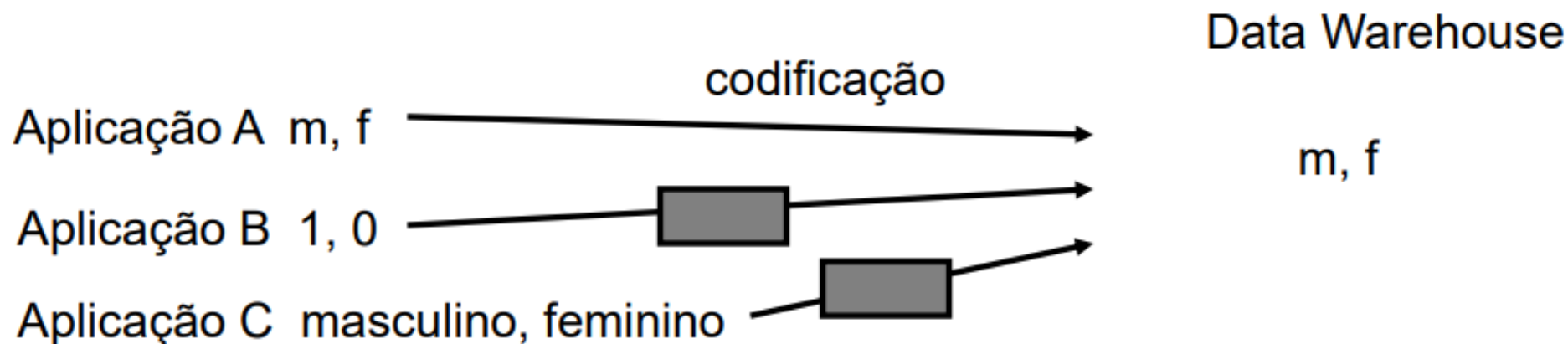
- Carga – receber os dados de diversos Sistemas de Processamento de Transações



Ambiente de Data Warehouse

Extraction, Transformation and Load

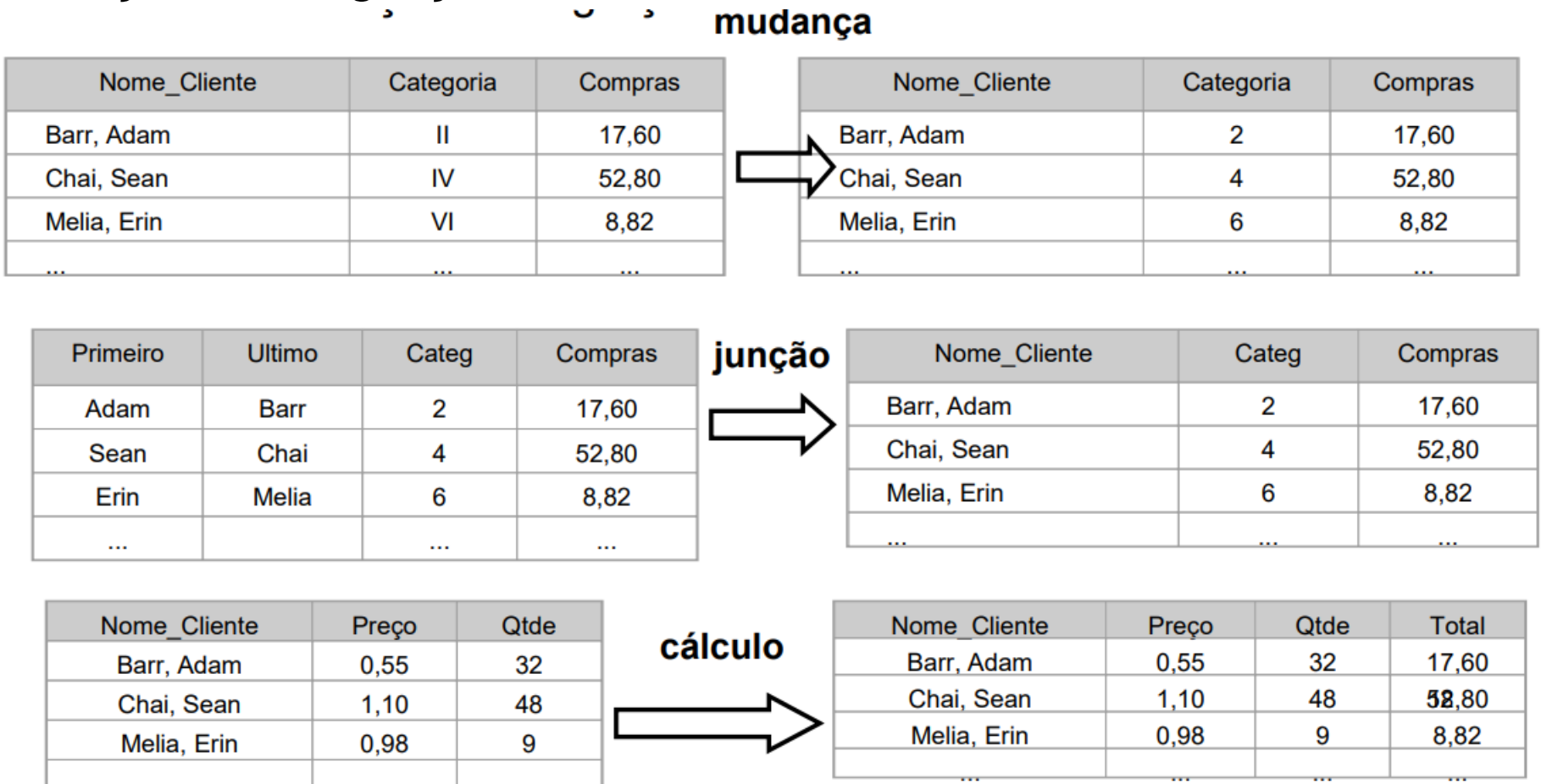
- Transformação e integração - processo de formatação e modificação de dados extraídos de várias origens para transformá-los em informações úteis ao Data Warehouse
- Os dados de origem são consistentes mas apresentados de diferentes formas



Ambiente de Data Warehouse

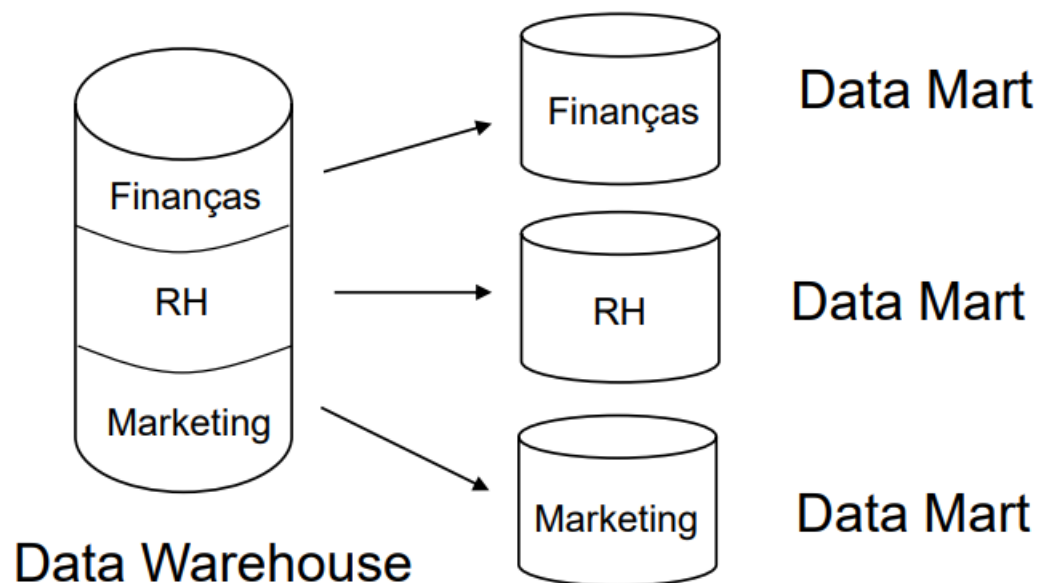
Extraction, Transformation and Load

- Transformação e integração



Ambiente de Data Warehouse – Data Mart

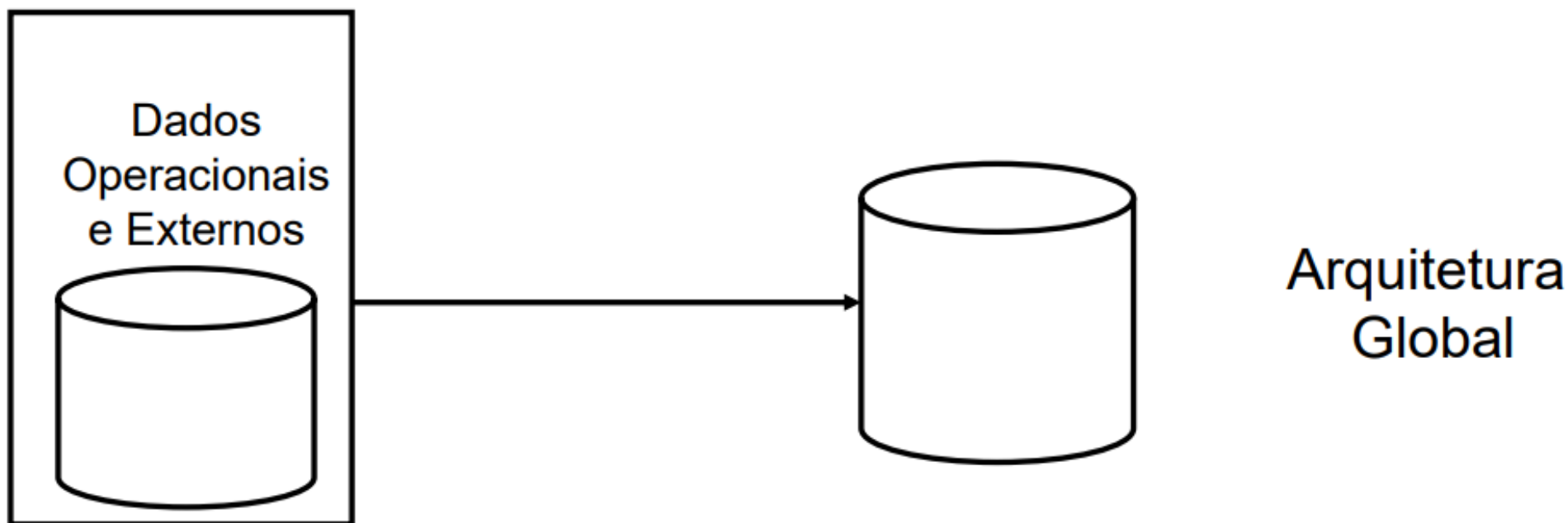
- Data Mart - Subconjunto lógico de um Data Warehouse, um Data Warehouse setorial
 - Geralmente descritos como um subconjunto dos dados contidos em um Data Warehouse extraído para um ambiente separado



Arquitetura de Data Warehouses

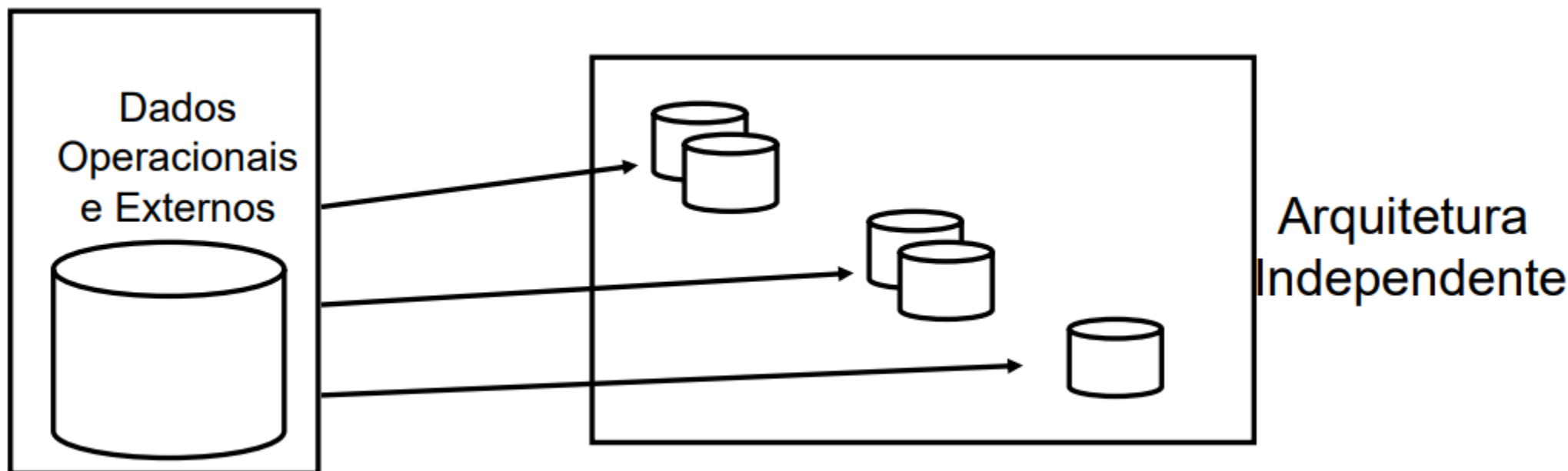
- **Arquitetura global**

- Utiliza um repositório comum de dados, integrado, utilizado por toda a organização



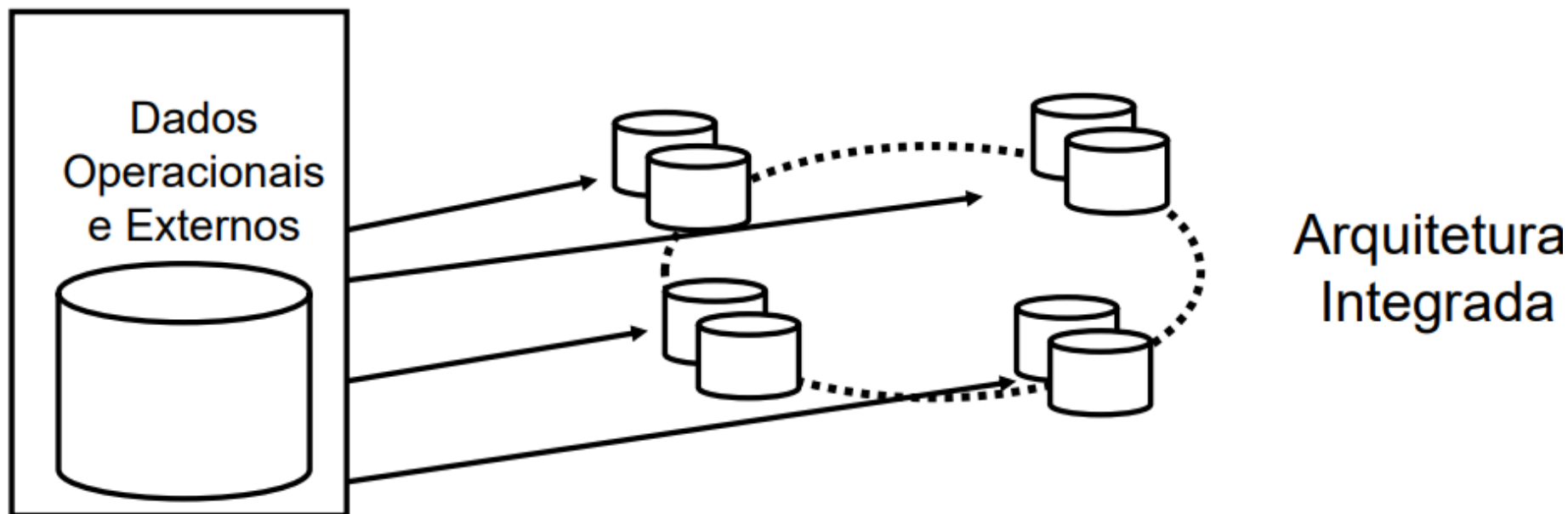
Arquitetura de Data Warehouses

- **Arquitetura de data marts independentes**
 - Possui um data mart para atender a cada departamento em específico
 - Não se tem acesso aos data marts de outros departamentos



Arquitetura de Data Warehouses

- **Arquitetura de data marts integrados**
 - Possui um data mart para atender a cada departamento em específico
 - Os dados são compartilhados entre os data marts de diferentes departamentos



Implementação de Data Warehouses

- **Abordagem Top-down**

- O modo como os dados serão armazenados e consultados nasce do DW e posteriormente são distribuídos entre os Data Marts
- Tem objetivo de atender às necessidades da organização como um todo e não departamentos isolados
- Modelo mais comum de implementação
- Demorada implementação e resultado apenas a longo prazo

Implementação de Data Warehouses

- **Abordagem Bottom-up**
 - Parte dos Data Marts até compor o DW por completo
 - Maior dificuldade na padronização dos dados
 - Implementação mais rápida e manutenção mais fácil devido ao menor tamanho das partes

Implementação de Data Warehouses

- **Abordagem Combinada**

- Combina características de ambas abordagens
- Planejamento geral da estruturação do DW para toda a organização (conforme Top-down)
- Desenvolvimento dos data marts de forma graduada, apresentando funcionalidades parciais
- A criação de cada data mart é padronizada para facilitar a integração dos dados

Metadados

- **“Dados sobre dados”**
- Possuem papel de grande importância nos DW
 - Especialmente na fase de desenvolvimento, onde especificam os dados de variadas fontes..