

Data warehousing e OLAP

Profº. Elifranio, Msc



FINALIDADE DO DATA WAREHOUSING

- Os bancos de dados tradicionais não são otimizados para acesso de dados apenas, eles equilibram o requisito de acesso a dados com a necessidade de garantir a integridade destes.
- Muitas pessoas só precisam de acesso de leitura aos dados, mas ainda necessitam de acesso rápido a um volume maior de dados.
- A maioria dos dados requeridos para a análise do **data warehouse** vêm de vários bancos de dados. Como muitas das análises realizadas são recorrentes e previsíveis, os vendedores de software e o pessoal de suporte de sistemas projetam sistemas para dar suporte a essas funções.

FINALIDADE DO DATA WAREHOUSING (CONT.)

- Há uma grande necessidade de ferramentas que fornecem informações aos tomadores de decisão para tomarem suas decisões rapidamente e confiavelmente com base em dados históricos.
- Essa funcionalidade é alcançada pelo *Data warehousing*, **processamento analítico on-line (OLAP)**.

INTRODUÇÃO, DEFINIÇÕES E TERMINOLOGIA

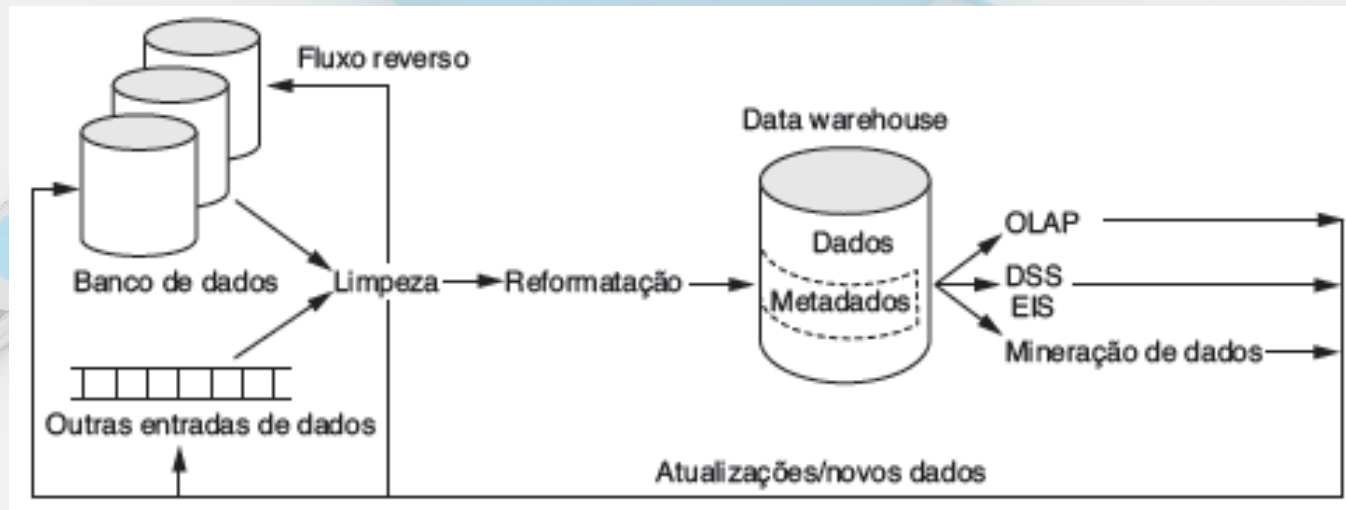
- W. H. Inmon caracterizou um **data warehouse** como:
- *“uma coleção de dados orientada a assunto, integrada, não volátil, variável no tempo para o suporte às decisões da gerência”*

INTRODUÇÃO, DEFINIÇÕES E TERMINOLOGIA

- Os *data warehouses* têm a característica distintiva de servir principalmente para aplicações de apoio à decisão.
 - Bancos de dados tradicionais são transacionais.
- Vários tipos de aplicações são aceitos:
 - **OLAP** (processamento analítico on-line) é um termo usado para descrever a análise de dados complexos do data warehouse.
 - **DSS** (sistemas de apoio à decisão) também conhecidos como EIS (sistemas de informações executivas) ajudam os principais tomadores de decisões de uma organização com dados de nível mais alto em decisões complexas e importantes.
 - **Mineração de dados** é usada para *descoberta do conhecimento*, o processo de procurar novo conhecimento imprevisto nos dados.

ESTRUTURA CONCEITUAL DE UM DATA WAREHOUSE

- O processo de data warehouse envolve
 - Limpeza e reformatação dos dados
 - OLAP
 - Mineração de dados



COMPARAÇÃO COM BANCOS DE DADOS TRADICIONAIS

- Data warehouses são otimizados principalmente para o acesso de dados apropriado.
 - Bancos de dados tradicionais são transacionais e otimizados para mecanismos de acesso e medidas de garantia de integridade.
- Data warehouses costumam apoiar a análise de série temporal e tendência, ambas exigindo mais dados históricos.
- Em comparação com os bancos de dados transacionais, os data warehouses são não voláteis.

COMPARAÇÃO COM BANCOS DE DADOS TRADICIONAIS (CONT.)

- Em sistemas transacionais, as transações são a unidade e o agente de mudança no banco de dados; ao contrário, a informação do data warehouse é muito menos detalhada e atualizada de acordo com uma escolha cuidadosa de política de atualização, normalmente incremental.

CARACTERÍSTICAS DOS DATA WAREHOUSES

- Visão conceitual multidimensional
- Dimensionalidade genérica
- Dimensões e níveis de agregação ilimitados
- Operações irrestritas entre dimensões
- Tratamento dinâmico de matriz esparsa
- Arquitetura cliente-servidor
- Suporte para múltiplos usuários
- Acessibilidade
- Transparência
- Manipulação de dados intuitiva
- Desempenho de relatório consistente
- Recurso de relatório flexível



CLASSIFICAÇÃO DOS DATA WAREHOUSES

- Data warehouses geralmente são uma ordem de magnitude maior que os bancos de dados de origem.
- O imenso volume de dados é um problema, no qual os Data Warehouses podem ser classificados.
 - Data warehouses em nível empresarial
 - São imensos projetos que exigem investimento maciço de tempo e recursos.
 - Data warehouses virtuais
 - Oferecem visões de bancos de dados operacionais que são materializados para acesso eficiente.
 - Data marts
 - Em geral são voltados para um subconjunto da organização, como um departamento, e possuem um foco mais estreito.

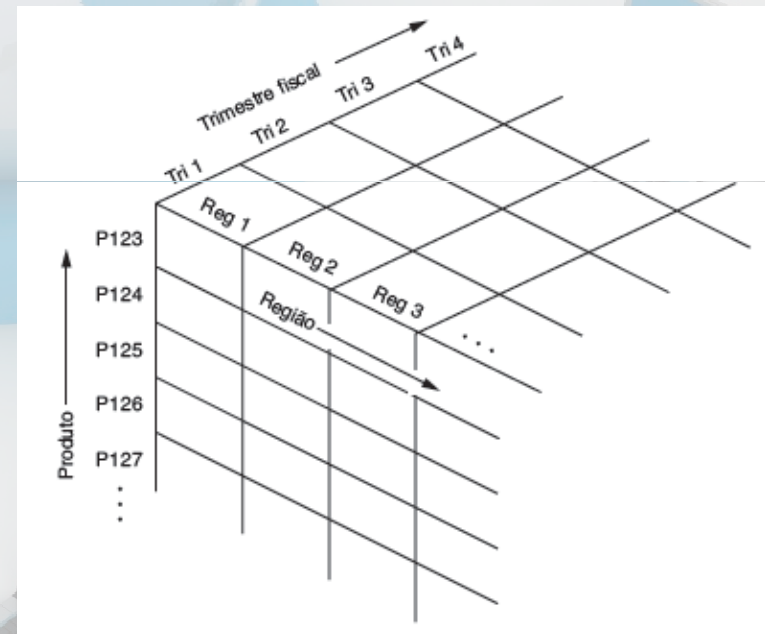
MODELAGEM DE DADOS PARA DATA WAREHOUSES

- Banco de dados tradicionais geralmente lidam com dados bidimensionais (semelhante a uma planilha padrão).
 - Entretanto, a performance de consulta em um modelo de armazenamento de dados multidimensional é muito mais eficiente.
- Data warehouses podem tirar vantagem dessa característica, dado que são:
 - Não voláteis
 - O grau de previsibilidade das análises que serão utilizadas neles é alto.

MODELAGEM DE DADOS PARA DATA WAREHOUSES (CONT.)

- Exemplo bidimensional *versus* multidimensional

	Região		
	Reg 1	Reg 2	Reg 3
Produto	P123		
	P124		
	P125		
	P126		
	⋮		
	⋮		



MODELAGEM DE DADOS PARA DATA WAREHOUSES (CONT.)

- Vantagens de um modelo multidimensional
 - Os modelos multidimensionais atendem prontamente a visões hierárquicas no que é conhecido como exibição roll-up ou exibição drill-down.
 - Os dados podem ser consultados diretamente em qualquer combinação de dimensões evitando consultas de banco de dados complexas.

ESQUEMAS MULTIDIMENSIONAIS

- Esquemas multidimensionais são especificados usando:
 - **Tabela de dimensão**
 - Consiste em tuplas de atributos da dimensão.
 - **Tabela de fatos**
 - Pode ser imaginada como tendo tuplas, uma para cada fato registrado. Esse fato contém alguma(s) variável(is) observada(s) e a(s) identifica com ponteiros para tabelas de dimensão. A tabela de fatos contém os dados, e as dimensões identificam cada tupla nesses dados.

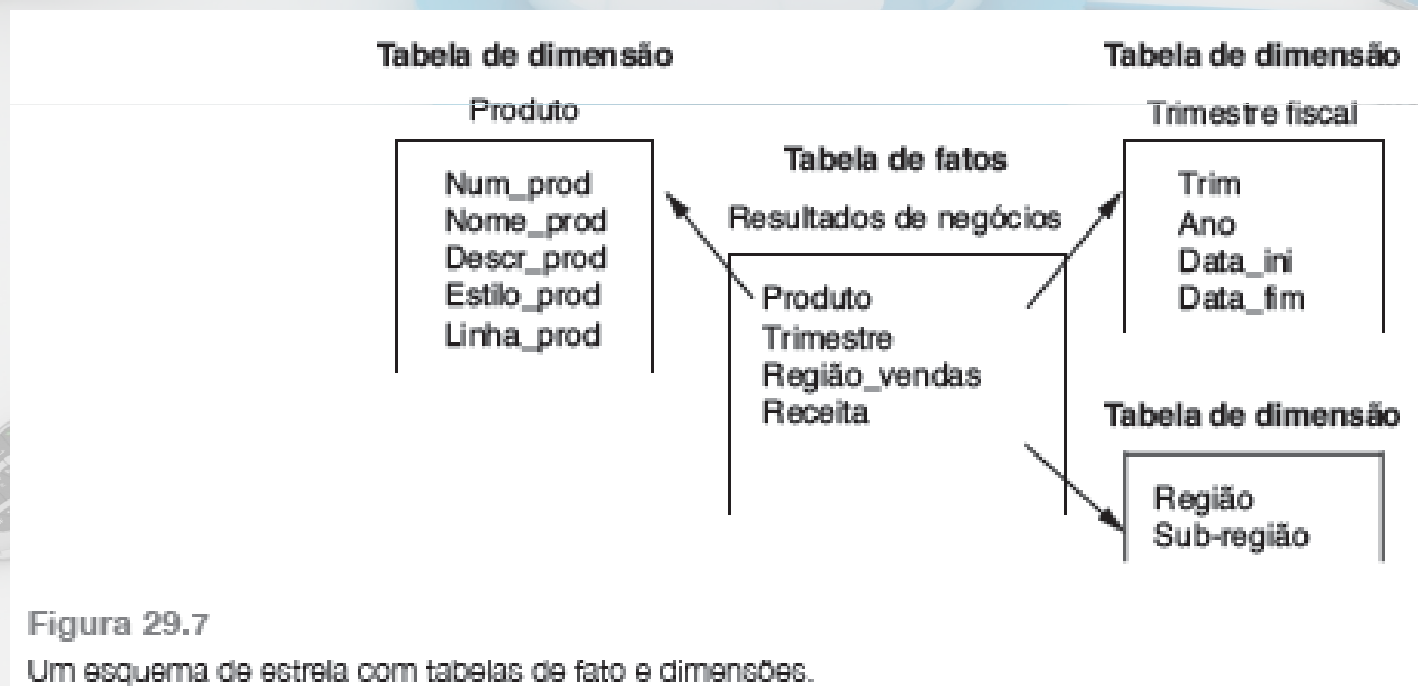
ESQUEMAS MULTIDIMENSIONAIS (CONT.)

- Dois esquemas multidimensionais comuns são
 - **Esquema estrela:**
 - Consiste em uma tabela de fatos com uma única tabela para cada dimensão
 - **Esquema floco de neve:**
 - É uma variação do esquema estrela em que as tabelas dimensões de um esquema estrela são organizadas em uma hierarquia ao normalizá-las.

ESQUEMAS MULTIDIMENSIONAIS (CONT.)

○ Esquema estrela:

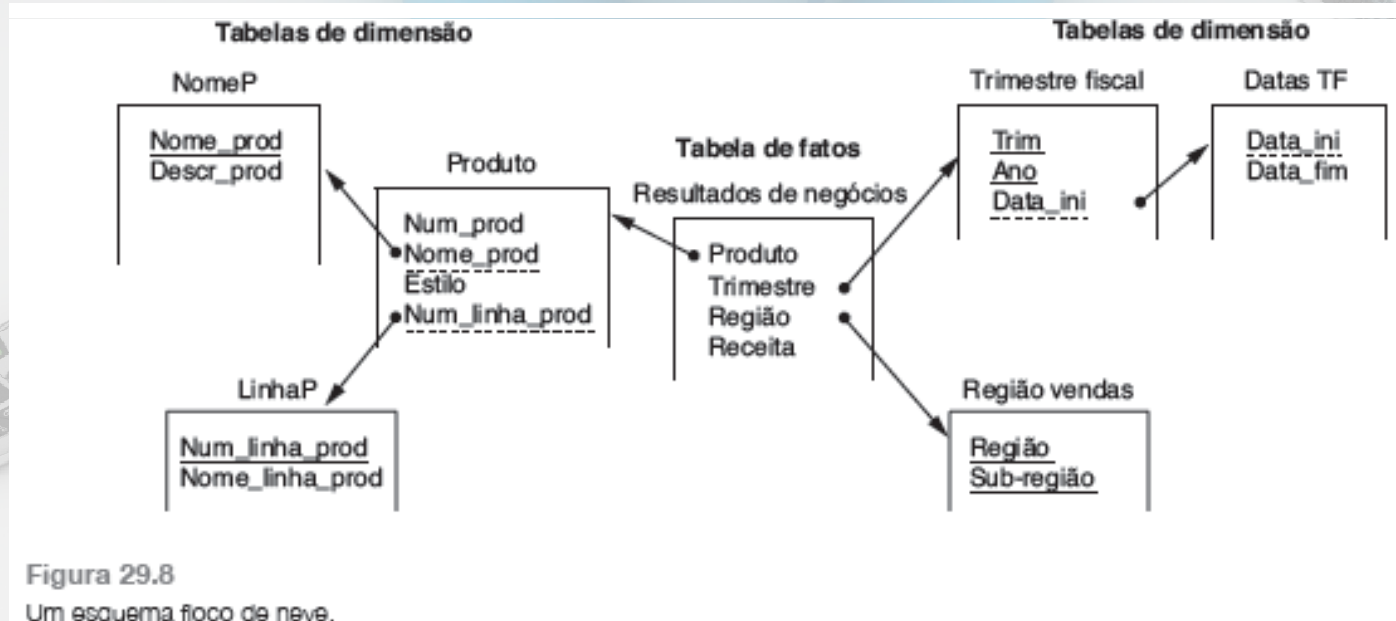
- Consiste em uma tabela de fatos com uma única tabela para cada dimensão.



ESQUEMAS MULTIDIMENSIONAIS (CONT.)

○ Esquema floco de neve:

- É uma variação do esquema estrela em que as tabelas dimensões de um esquema estrela são organizadas em uma hierarquia ao normalizá-las.



ESQUEMAS MULTIDIMENSIONAIS (CONT.)

○ Constelação de fatos

- É um conjunto de tabelas de fatos que compartilham algumas tabelas de dimensão. As constelações de fatos limitam as possíveis consultas para o armazém.

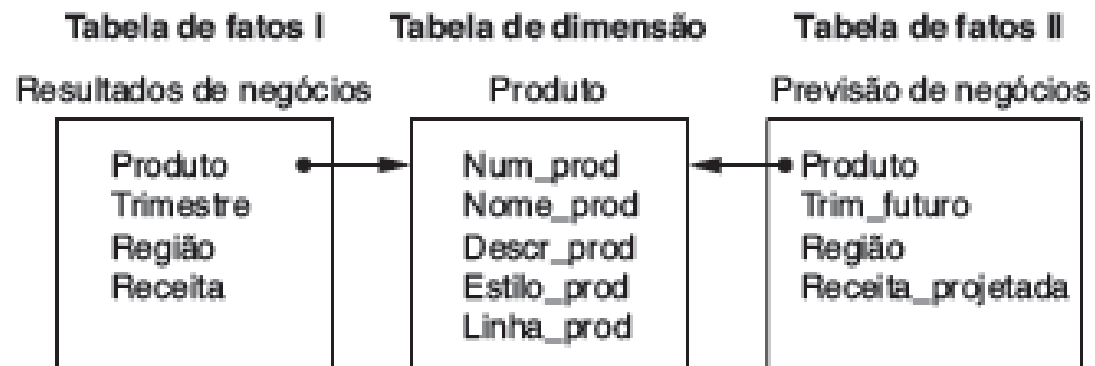


Figura 29.9

Uma constelação de fatos.

ESQUEMAS MULTIDIMENSIONAIS (CONT.)

○ Indexação

- O armazenamento do data warehouse também utiliza técnicas de indexação para dar suporte ao acesso de alto desempenho.
- Uma técnica chamada **indexação de bitmap** constrói um vetor de bits para cada valor em um domínio.
- Ela funciona muito bem para domínios de baixa cardinalidade.

CRIANDO UM DATA WAREHOUSE



- Na construção de um data warehouse, os responsáveis deverão ter uma visão ampla do uso antecipado do armazém.
 - O projeto deve aceitar a **consulta ocasional**.
 - Um esquema apropriado seria escolhido para refletir o uso antecipado.

CRIANDO UM DATA WAREHOUSE (CONT.)

- O projeto de um Data Warehouse envolve os seguintes passos.
 - Aquisição de dados para o armazém.
 - Assegurar que o Armazenamento de Dados encontre os requerimentos de consulta eficientemente.
 - Considerar completamente o ambiente em que o data warehouse reside.

CRIANDO UM DATA WAREHOUSE (CONT.)

- Aquisição de dados para o armazém
 - Os dados precisam ser extraídos de várias fontes heterogêneas.
 - Os dados precisam ser formatados por coerência dentro do armazém.
 - Os dados precisam ser limpos para garantir a validade.
 - Difícil de automatizar o processo de limpeza.
 - Fluxo reverso, atualizando os dados com dados limpos.

CRIANDO UM DATA WAREHOUSE (CONT.)

- Aquisição de dados para o armazém (cont.)
 - Os dados precisam ser ajustados ao modelo de dados do armazém.
 - Os dados precisam ser carregados no armazém.
 - Deve ser considerada um projeto adequado para política de renovação.

CRIANDO UM DATA WAREHOUSE (CONT.)

- Armazenamento dos dados de acordo com o modelo de dados do armazém
- Criação e manutenção das estruturas de dados exigidas
- Criação e manutenção dos caminhos de acesso apropriados
- Fornecimento de dados variáveis no tempo à medida que novos dados são incluídos
- Suporte à atualização dos dados do armazém.
- Atualização dos dados
- Eliminação dos dados

CRIANDO UM DATA WAREHOUSE (CONT.)

- Projeções de uso
- O ajuste do modelo de dados
- Características das fontes disponíveis
- Projeto do componente de metadados
- Projeto de componente modular
- Projeto de facilidade de gerenciamento e mudança
- Considerações de arquitetura distribuída e paralela
 - Armazém distribuído *versus* armazém federado

FUNCIONALIDADE DE UM DATA WAREHOUSE

- Funcionalidade que pode ser esperada:
 - **Roll-up:** Os dados são resumidos com generalização cada vez maior.
 - **Drill-Down:** Níveis cada vez maiores de detalhes são revelados.
 - **Giro:** A tabulação cruzada é realizada.
 - **Slice and dice:** Operações de projeção são realizadas nas dimensões.
 - **Ordenação:** Os dados são ordenados por valor ordinal.
 - **Seleção:** Os dados estão disponíveis por valor ou intervalo.
 - **Atributos derivados:** Atributos são calculados por operações sobre valores armazenados e derivados.

DATA WAREHOUSES *VERSUS* VISÕES

- Visões e data warehouses são semelhantes porque ambos têm extratos apenas de leitura dos bancos de dados.
- Contudo, os data warehouses são diferentes das visões das seguintes maneiras :
 - Os data warehouses existem como armazenamento persistente, em vez de serem materializados por demanda.
 - Os data warehouses normalmente não são relacionais, mas sim multidimensionais.
 - Os data warehouses podem ser indexados para otimizar o desempenho.
 - Data warehouses oferecem suporte específico de funcionalidade.
 - Data warehouses oferecem grande quantidade de dados, mais do que está contido em um banco de dados.

DIFICULDADES DE IMPLEMENTAÇÃO

DATA WAREHOUSES

- O tempo inicial necessário para construir um data warehouse é enorme
 - Potencialmente toma anos para construir e manter eficientemente um data warehouse.
- Tanto qualidade quanto a consistência dos dados são questões importantes.
- Projeções de uso devem ser revisadas de maneira contínua para refletir os requisitos atuais.
 - O armazém também deve ser projetado para acomodar o acréscimo e o atrito das fontes de dados sem um reprojeto importante
- A administração de um data warehouse exigirá habilidades muito mais amplas do que as necessárias para a administração do banco de dados tradicional.

TÓPICOS EM ABERTO DO DATA WAREHOUSING

- Limpeza de dados, indexação, particionamento e visões podem ter nova atenção com a perspectiva do data warehousing.
- Automatização de
 - Aquisição de dados
 - Gerenciamento da qualidade dos dados
 - Seleção e construção de trilhas de acesso e estruturas
 - Automanutenção
 - Otimização de funcionalidade e desempenho
- Incorporação apropriada de regras de domínio e negócios nos processos de criação e manutenção de armazéns.

RESUMO

- Metas do data warehousing
- Introdução, definições e terminologia
- Comparação com bancos de dados tradicionais
- Características dos data warehouses
- Classificação dos data warehouses
- Esquemas multidimensional
- Construção de um data warehouse
- Funcionalidade de um data warehouse
- Warehouse *versus* data views
- Dificuldades de implementação e tópicos em

REFERÊNCIAS

- Nataathe, et al. Sistemas de Banco de Dados. 6^a ed. Editora Pearson. 2010