



Sistemas de informação

enterprise analytics and data
warehousing

Profº Fabiano J. Cury Marques

enterprise analytics and data warehousing

Introdução a Data WareHouse



Agenda

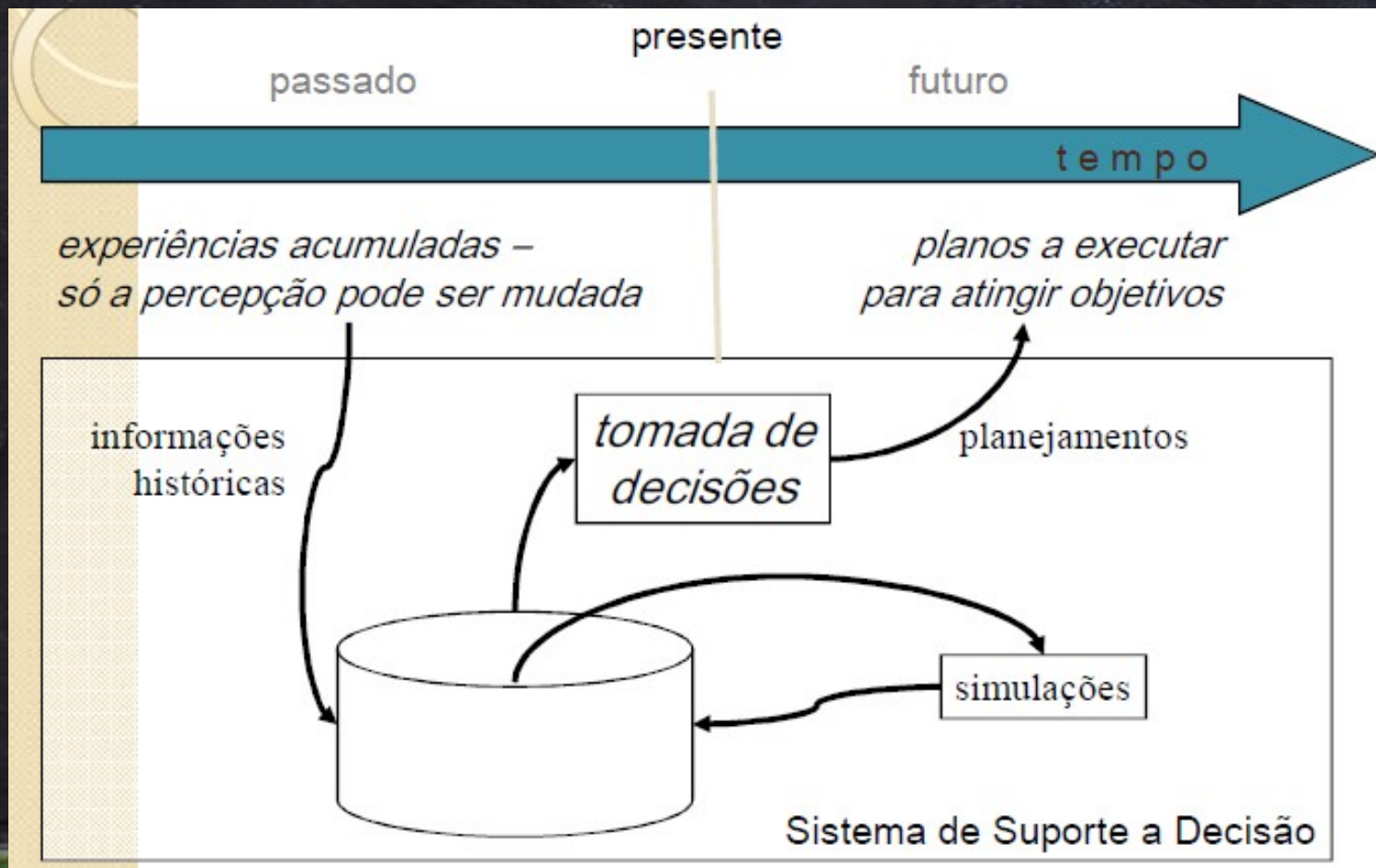
- ✕ Processo decisório e Informação
- ✕ Introdução à Data Warehouse
- ✕ Comparação entre Sistemas Transacionais e Dimensionais
- ✕ Tendências
- ✕ Referências

1.

Processo decisório

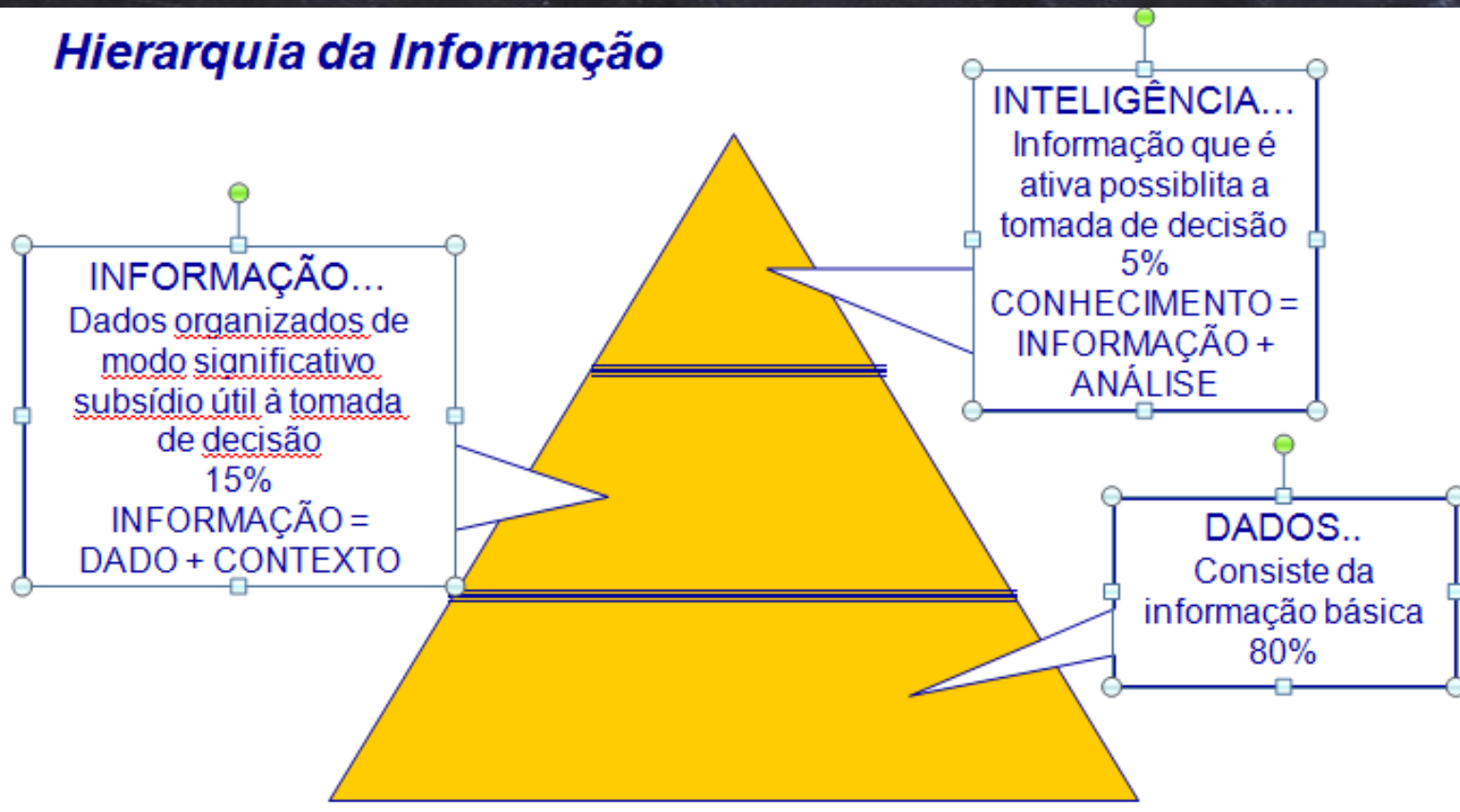
Business need to be firmly in hand!

PROCESSO DECISÓRIO



HIERARQUIA DA INFORMAÇÃO

Hierarquia da Informação





CLASSES DOS SISTEMAS DE INFORMAÇÃO

SISTEMAS TRANSACIONAIS

- Controlam **informações operacionais**, como por exemplo, **vendas, compras, contabilidade, sensoramento**, etc.

- Operações de manipulação de dados **CRUD** (select, insert, update, delete).

- Em geral com **alto nível de detalhe** com **disponibilidade online**.

- Também conhecido como **sistema aplicativo**.

SISTEMAS DE SUPORTE À DECISÃO

- Extraem **informações estratégicas** para a tomada de decisão.

- Envolve a execução de **consultas complexas** sobre **grandes volumes** de dados.

- Em geral, é alimentado com dados de sistemas transacionais, embora não limitado a esse tipo de fonte de dados.



Introdução à DW

Logica and fisical design!

O que
é um
DW?

CLASSES DOS SISTEMAS DE INFORMAÇÃO

- ✗ Armazém de dados;
- ✗ Banco de dados voltado para o suporte à tomada de decisão;
- ✗ Geralmente, derivado de vários bancos de dados operacionais;
- ✗ Pode ser usado com base para executar OLAP e outras tecnologias de análise de informação e extração de conhecimento;
- ✗ Objetivos:
 - Satisfazer necessidades de análise de informações;
 - Monitorar e comparar situações atuais com passadas;
 - Estimar situações futuras;



DEFINIÇÕES

DATA WAREHOUSE - DW

“Coleção de dados orientada a assuntos, integrada, com séries temporais e não volátil, voltada para o apoio à tomada de decisão” - **W. H. Inmon**

“Sistema que extrai, limpa, ajusta e carrega dados de fontes diversas para dentro de uma base de dados dimensional e então suporta e implementa consultas e análises com propósito de tomadas de decisão” - **R. Kimball**

DATA WAREHOUSING

- Processo de construção e suporte de uso de Data warehouses.

BUSINESS INTELLIGENCE - BI

- Refere-se a coleta, organização, análise, compartilhamento e monitoramento de informações para suporte a gestão de negócios.

Características de um dw

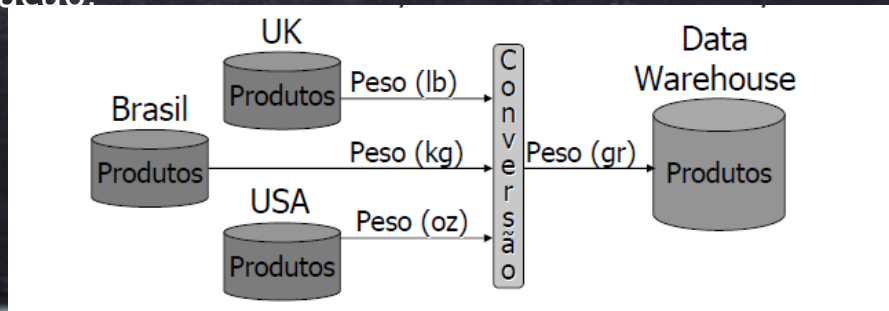
✗ Orientado a assuntos:

- Por exemplo, vendas de produtos, atendimento e diagnóstico de pacientes, rendimento de estudantes, etc.



✗ Integrado:

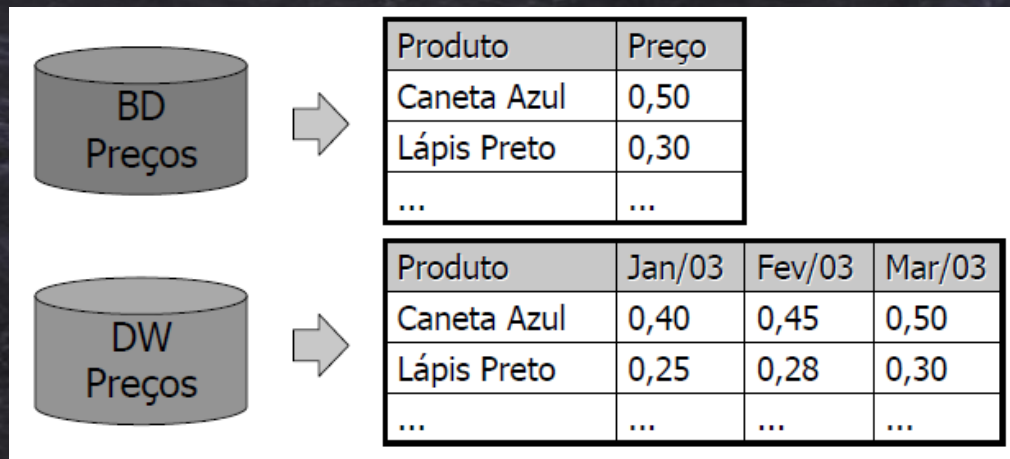
- Diferentes nomenclaturas, formatos e estrutura das fontes de dados precisam ser acomodadas em um único esquema para prover uma visão unificada e consistente da informação.



Características de um dw

✗ Séries temporais:

- O histórico dos dados por um período de tempo superior ao usual em banco de dados transacionais permitem analisar tendências e mudanças.



✗ Não volátil:

- Os dados de um data warehouse não são modificados como em sistemas transacionais (exceto para correções), mas somente carregados e acessados para leitura, com atualizações apenas periódicas.

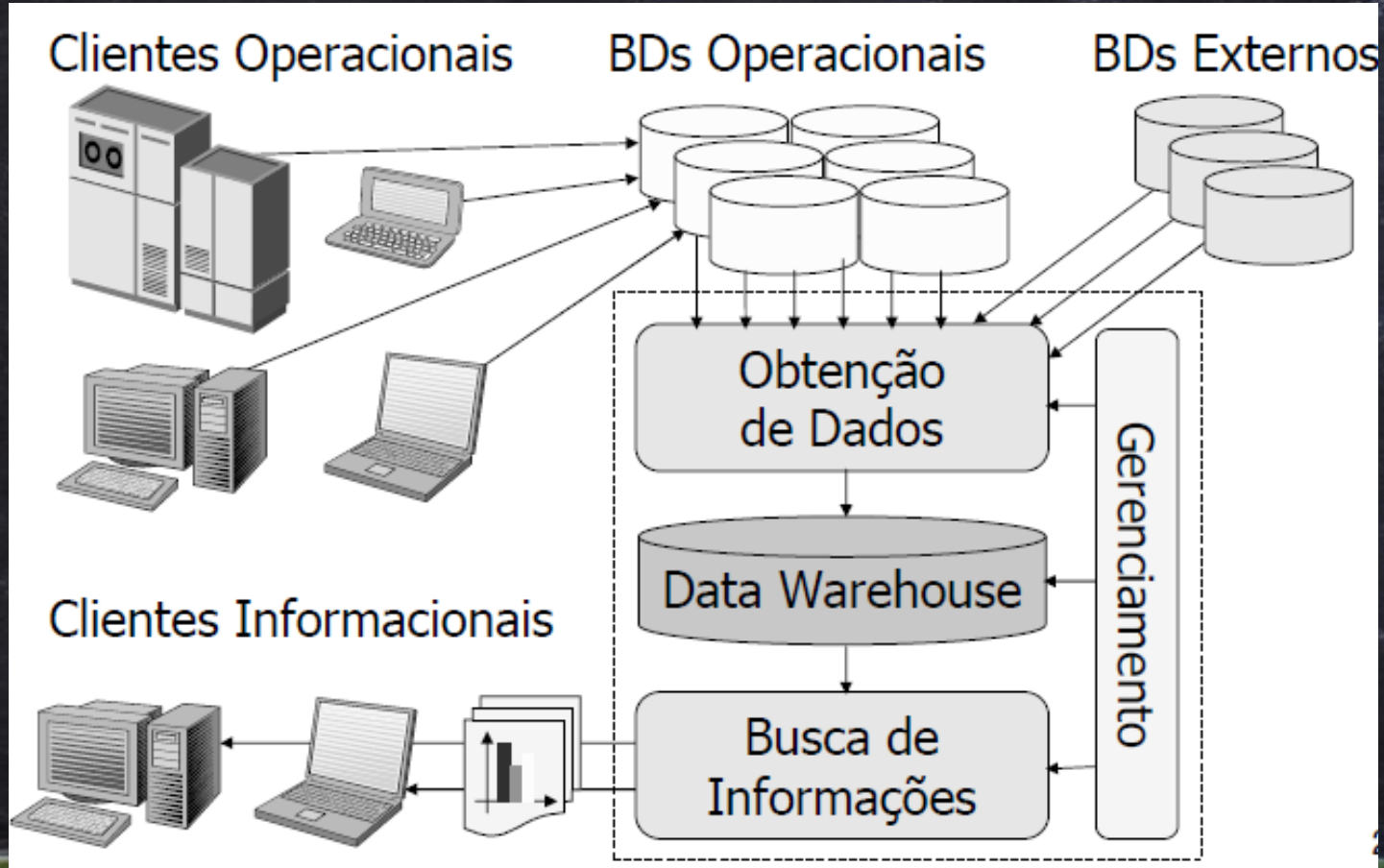
3.

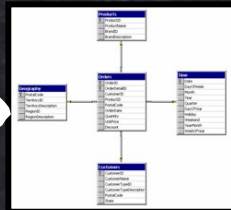
Comparativo entre
sistemas transacionais e
dimensionais

COMPARATIVO – TRANSACIONAL x DIMENSIONAL

Característica	BD Transacional	BD Suporte à Decisão
Objetivo	Atividades cotidianas	Análise do negócio
Uso	Operacional	Informativo
Processamento	OLTP	OLAP
Unidade de trabalho	Inclusão, alteração, exclusão	Carga e consulta
Usuários	Operadores (muitos)	Gerência (poucos)
Interação dos usuários	Ações pré-definidas	Pré-definida e <i>ad-hoc</i>
Dados	Operacionais	Analíticos
Volume	Pode ser alto (MB – GB)	Muito alto (GB – TB)
Histórico	60 a 90 dias	vários anos
Granularidade	Detalhada (baixa)	Detalhada e consolidada (alta)
Redundância	Não ocorre (só p/ eficiência)	Pode ocorrer
Estrutura	Estática	Variável
Manutenção	Mínima é o desejável	Constante
Atualização	Contínua (tempo real)	Periódica (<i>snapshots</i> - retratos)
Integridade	Transação	Cada atualização
Acesso a registros	Poucos - por transação	Muitos - para consolidação
Índices	Poucos/simples	Muitos/complexos
Função dos índices	Localizar um registro	Agilizar consultas

COMPARATIVO – TRANSACIONAL x DIMENSIONAL





Metodologia de projeto que lista **dimensões e fatos relevantes** a cada processo de negócio.

A modelagem dimensional permite-nos **visualizar dados**. A habilidade de visualizar algo tão abstrato quanto um conjunto de dados de forma concreta e tangível é o segredo da compreensibilidade

Ex: Imagine um negócio em que o CEO descreva as atividades da empresa da seguinte forma:

REQUISITOS DE UM DW

- ✗ Manter as **informações acessíveis** na organização:
 - O conteúdo do DW deve ser de **fácil interpretação**, acesso e com altos índices de desempenho para **respostas aos usuários**. Os **nomes dos campos** devem ser óbvios, para garantir uma fácil interpretação na seleção de dados.
- ✗ Garantir a **consistência dos dados** da organização
 - A **mesma informação** divulgada por áreas diferentes dentro da empresa deve ser exatamente igual, garantindo a **unicidade de conceitos**, regras e a confiabilidade do dado.
- ✗ Ser **adaptável e flexível às várias fontes** de informação
 - O DW deve ser desenhado para **suportar contínuas alterações**. Quando **novos questionamentos** são feitos ao DW, tanto a tecnologia quanto os dados existentes não devem ser alterados e devem suportar as respostas
- ✗ Métodos de **controle e segurança dos dados**
 - O DW permite não só o **controle efetivo de acesso aos dados**, mas dá ao proprietário uma visibilidade da forma com que estão sendo utilizados.
- ✗ Base para **suporte à decisão**
 - O DW deve ser definido como **base de dados única para suporte à tomada de decisão** de toda a corporação.

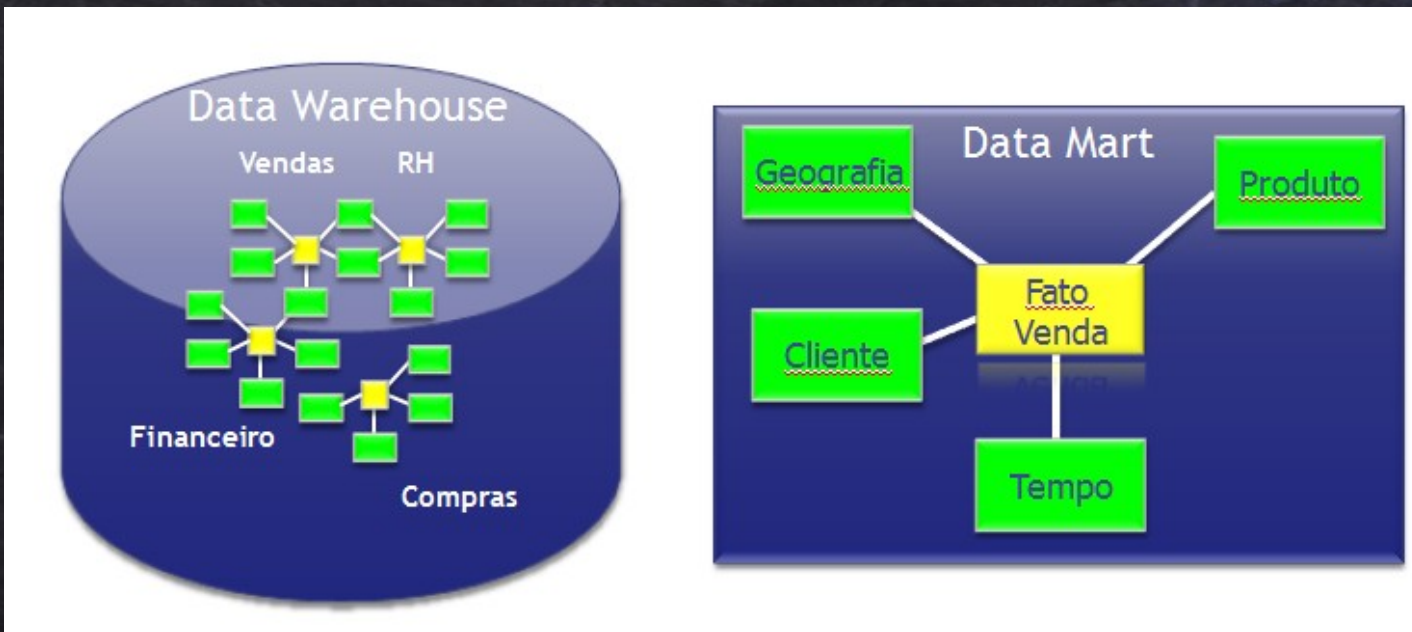
Retomando

...

Um DW é
orientado a
assunto!

DATA WAREHOUSE x DATA MART

- ✗ O **Data Mart** obedece aos mesmos conceitos do data warehouse, diferenciando-se somente no conteúdo, ou seja, os dados são **organizados por assuntos**, permitindo maior independência, **agilidade e desempenho**

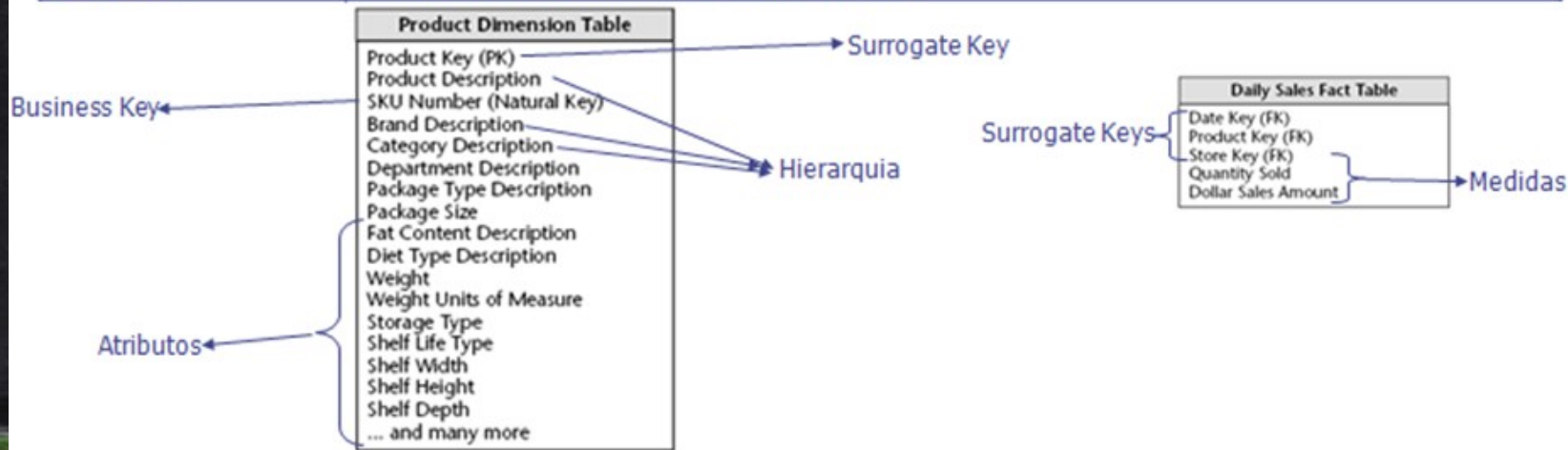


- ✗ O **Data Warehouse** é um conjunto de Data Marts integrados

Conceitos – vocabulário



Medida	Informação numérica proveniente da medição das transações da empresa
Tabela fato	Principal tabela no DW onde as medidas são armazenadas
Tabela dimensão	Contém as informações descritivas e qualificadores do negócio. É a porta de entrada do DW
Grão	Menor nível de informações existente no DW e definida pelas dimensões ligadas às tabelas fato. Define o escopo da medida
Surrogate Key	Chave substituta gerada no DW. É a chave primária das tabelas
Business Key	Chave primária do transacional. Utilizada como chave de negócio
Hierarquia	Conjunto de atributos que possui uma ordem lógica do maior ao menor nível
Atributo	Campo descritivo referente a uma dimensão





CONCEITOS – ANÁLISE MULTIDIMENSIONAL

- ✗ Gerentes pensam de forma multidimensional
- ✗ A tecnologia de BI sobre os data warehouses formatam a informação multidimensionalmente
- ✗ Este formato acelera o processo de uso da informação pelos gerentes
- ✗ Permite reduzir o custo de gerar e manusear a informação. Exemplo:
Analisar 5 indicadores de vendas, de 10 vendedores que vendem 10 produtos para 100 clientes todo mês, em um período de 24 meses:
 $5 \times 10 \times 10 \times 100 \times 24 = 1.200.000$ combinações





MITOS

- ✗ Modelos dimensionais servem somente para dados agregados
- ✗ Deve-se armazenar pouca quantidade de dados históricos
- ✗ Modelos dimensionais são soluções departamentais
- ✗ Modelos dimensionais não são escalonáveis
- ✗ Só devem ser utilizados quando existe um padrão definido de utilização
- ✗ Não é possível integrar modelos dimensionais



4.

Tendências



TENDÊNCIAS PARA DW

DADOS DESESTRUTURADOS

- Documentos, imagens (fotos, diagramas e figuras), áudio (músicas, falas e sons), vídeos (filmes e animações), dados de streaming, textos, e-mails, web sites.
- Como guardar estas informações no DW
- Depois de guardar, como recuperar as informações necessárias para análise

PESQUISA

- Permitir que o conteúdo do DW seja analisado por ferramentas de busca, similares à buscas da internet, exemplo: Google, Bing etc.
- Auxílio no caso de dados desestruturados
- Facilita também encontrar dados estruturados quando os usuários não conhecem a estrutura do DW
- Com pesquisa, tem-se apenas um campo para preencher, onde se coloca o que quiser



TENDÊNCIAS PARA DW

SOA

Em vez de uma grande aplicação contando com todos os componentes do DW, pode-se ter componentes menores utilizados como serviços

Facilita a troca de algum dos componentes no futuro, sem ter que trocar a aplicação inteira

Permite a utilização e comercialização de pequenos “pedaços” do DW

DW em tempo real

Atualização não é mais diária ou semanal por exemplo

Usuários querem ver os dados no DW atualizados a cada dois minutos ou até mesmo em tempo real

Um DW em tempo real é aquele que é atualizado (pelo ETL) no momento em que a transação ocorre no sistema transacional fonte.

TENDÊNCIAS PARA DW



Tópicos/Técnicas:

Algoritmo K-means;

Naive Bayes

Otimização

Grafos e detecção de comunidades

Regressão

Previsões

Detecção de outliers



TENDÊNCIAS PARA DW

Responda as perguntas abaixo:

- 1 – No contexto de um processo decisório, o quê é mais importante: o passado ou o futuro? Explique
- 2 – Entre os sistemas, Transacionais e Dimensional, qual deles deveria apresentar maior desempenho em uma empresa? Explique
- 3 – Posso construir um Data Mart utilizando um banco de dados sqlite3?





Referências

- ✕ KIMBALL, R., ROSS, M. The Data Warehouse Toolkit. 2ª ed., John Wiley Professional, 2002.
- ✕ MACHADO, F. N. R. Tecnologia e Projeto de Data Warehouse. 1ª ed., São Paulo: Ed. Érica, 2004.



Obrigado!

Copyright © 2019 Prof. MSc. Eng. Wakim B. Saba

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).