

Business Intelligence



PROF. ANTONIO GUARDADO

Agenda

- 1- Business Intelligence
- 2 - Introdução ao Data Warehouse
- 3 - Arquitetura do Data Warehouse
- 4 – Modelagem Multidimensional

1- Business Intelligence

- Inteligência de Negócios
- Um conjunto de conceitos, métodos e recursos tecnológicos que habilitam a obtenção e distribuição de informações geradas a partir de dados operacionais, históricos e externos, visando proporcionar subsídios para a tomada de decisões gerenciais e estratégicas

1- BI - Características

- Extrai e integra dados de múltiplas fontes;
- Fazer uso da experiência (dados históricos);
- Analisa dados contextualizados;
- Trabalha com hipóteses;
- Procura relações de causa e efeito;
- Transformar os registros obtidos em informação útil para o conhecimento empresarial;
- Conceitos de BI são aplicáveis a todo tipo de empresa, independente de seu porte, faturamento ou segmento.

1.1 - Sistemas de Apoio à Decisão



1.1 - Sistemas de Apoio à Decisão

- Informação
 - Melhor recurso do qual uma empresa pode dispor para tomar decisões estratégicas
 - Obtida analisando **dados históricos** sobre vendas, produção, clientes, etc.
- Análise dos dados
 - Fornece **informações** vitais para a empresa
 - Pode aumentar a competitividade da empresa
 - Era feita intuitivamente pelos gerentes

1.1-Sistemas de Apoio à Decisão

- Dificuldades para obter informação
 - Quantidade de dados a serem analisados cresce com a expansão do negócio e com o passar dos anos
 - Dados conflitantes vindos de fontes diferentes podem gerar informações desencontradas
 - Impossível para um ser humano manter e analisar todos os dados
 - Informação não é mais mantida por gerentes devido à mobilidade no mercado de trabalho

1.1 - Sistemas de Apoio à Decisão

- Usam dados históricos mantidos em um banco de dados convencional
- Dados históricos são analisados usando técnicas de mineração de dados para obter informações usadas na tomada de decisões
- Estatísticas de venda, produção, clientes, etc. podem ser levantadas e consideradas para tomar decisões estratégicas de negócio

1.1 - Sistemas de Apoio à Decisão

- Benefícios dos Sistemas de Apoio à Decisão
 - Determinar o mercado-alvo de um produto
 - Definir o preço de um produto, criar promoções e condições especiais de compra
 - Verificar a eficácia de campanhas de marketing
 - Otimizar a quantidade de produtos no estoque
 - Responder rapidamente a mudanças no mercado e determinar novas tendências
- ... ou seja, ganhar **eficiência** e **lucratividade**

1.1 - Sistemas de Apoio à Decisão

- Problema: **dados históricos não são mantidos nos BDs da empresa**
 - Volume de dados seria muito grande
 - Desempenho seria insatisfatório
- Solução: criar um BD exclusivamente para manter os dados históricos
 - Especializado para realizar poucas consultas sobre um grande volume de dados
 - Surge o **Data Warehouse** (DW)

2 - Data Warehouse

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

2 - Data Warehouse

- O que é?
 - Sistema que armazena dados históricos usados no processo de tomada de decisão
 - Integra os dados corporativos de uma empresa em um único repositório
- Para que serve?
 - Para criar uma visão única e centralizada dos dados que estavam dispersos em diversos BDs
 - Permite que usuários finais executem consultas, gerem relatórios e façam análises

2 - Data Warehouse

- BDs usados nas aplicações de negócio são chamados **BDs operacionais**
- DW é um **BD informacional alimentado com dados dos BDs operacionais da empresa**
- Disponibiliza **dados atuais** e **dados históricos**
- Dados podem ser **sumarizados** (condensados) para que sejam analisados
- Contém também **metadados**, que são dados sobre os dados armazenados no DW

2 - Data Warehouse

- Então o Data Warehouse é apenas um BD que contém também dados históricos?
- Para que seja considerado um Data Warehouse, um banco de dados deve:
 - **Coletar dados de várias fontes**
 - **Dados coletados devem ser transformados para que haja uma visão única dos dados**
 - **Dados devem ser usados por aplicativos para obter informações que dêem apoio à decisão**

2 - Data Warehouse

	BD Operacional	Data Warehouse
Usuários	Funcionários	Alta administração
Utilização	Tarefas cotidianas	Decisões estratégicas
Padrão de uso	Previsível	Difícil de prever
Princípio de funcionamento	Com base em transações	Com base em análise de dados
Valores dos dados	Valores atuais e voláteis	Valores históricos e imutáveis
Detalhamento	Alto	Sumarizado
Organização dos dados	Orientado a aplicações	Orientado a assunto

15

2.1 – DW- Principais Características

- De acordo com a definição dada por Inmon, um Data Warehouse deve ser:
 - Orientado a assunto
 - Integrado
 - Não-volátil
 - Variável com o tempo

16

2.1 – DW - Principais Características

- **Orientação a assunto**
 - Os dados em um DW são organizados de modo a facilitar a análise dos dados
 - Dados são organizados por assunto e não por aplicação, como em BDs operacionais

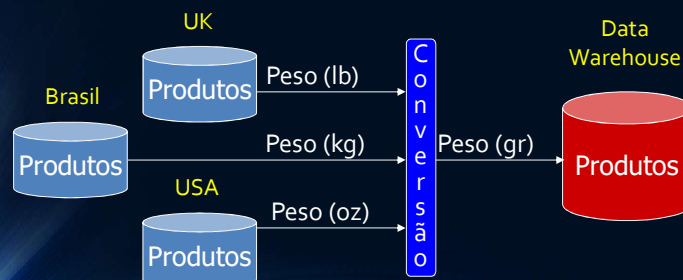


Business Intelligence - Prof. Antonio Guardado

17

2.1 – DW - Principais Características

- **Integração**
 - Dados de um DW provém de diversas fontes
 - Dados podem ser sumarizados ou eliminados
 - Formato dos dados deve ser padronizado para uniformizar nomes, unidades de medida, etc.



Business Intelligence - Prof. Antonio Guardado

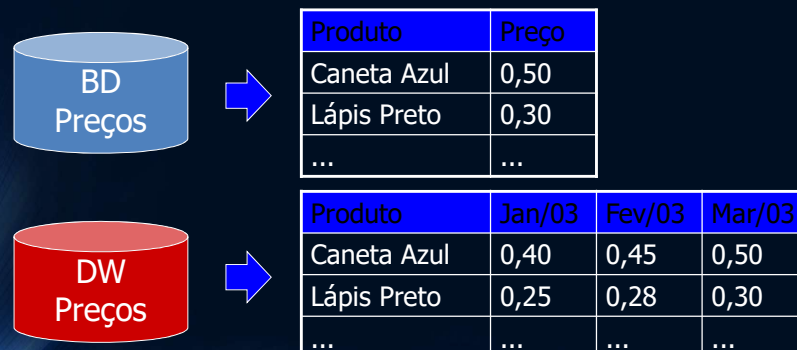
18

2.1 – DW - Principais Características

- **Não-Volátil**
 - Dados não são mais alterados depois de incluídos no DW
- Operações no DW
 - Em um BD operacional é possível incluir, alterar e eliminar dados
 - Já no DW é possível apenas incluir dados
- Garante que consultas subseqüentes a um dado produzirão o mesmo resultado

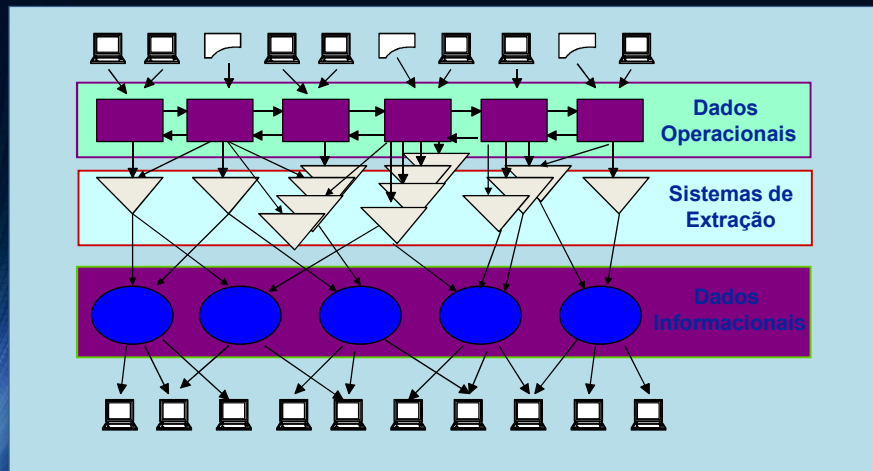
2.1 – DW - Principais Características

- **Variável com o Tempo**
 - Os dados no DW são relativos a um determinado instante de tempo



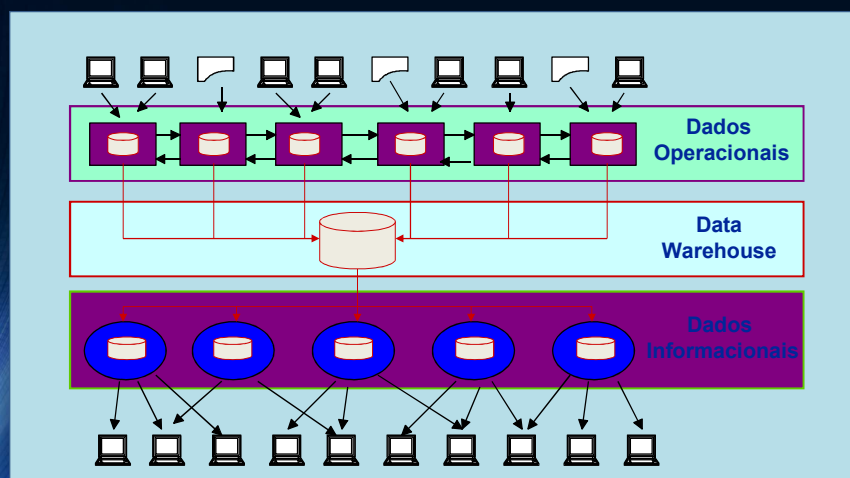
3- Arquitetura do Data Warehouse

- Sistemas de Extração Tradicionais



3 - Arquitetura do Data Warehouse

- Sistemas baseados em Data Warehouse



3 - Arquitetura do Data Warehouse

- Principais tarefas efetuadas pelo DW
 - Obter dados dos BDs operacionais e externos
 - Armazenar os dados
 - Fornecer informações para tomada de decisão
 - Administrar o sistema e os dados
- Principais componentes do DW
 - Mecanismos para acessar e transformar dados
 - Mecanismo para armazenamento de dados
 - Ferramentas para análise de dados
 - Ferramentas de gerência

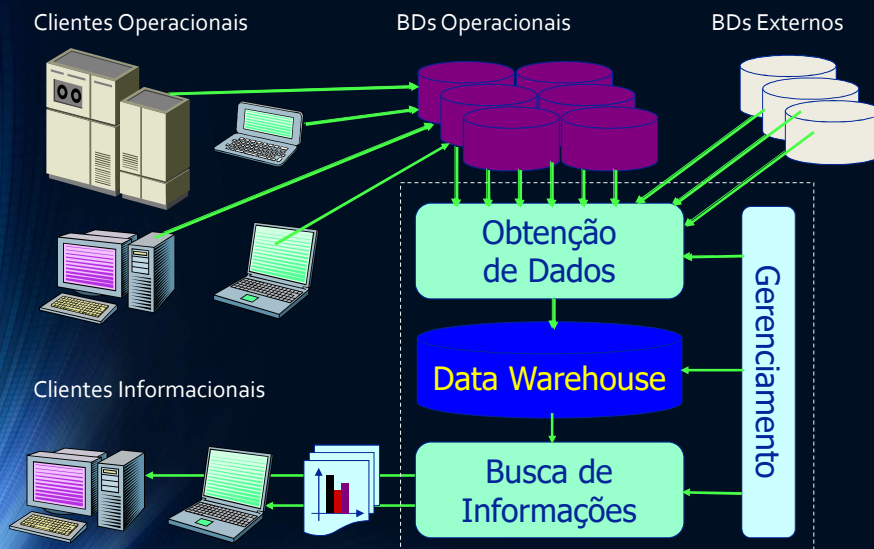
3.1 - Estrutura Interna

- Requisitos do DW
 - Eficiente
 - Grande volume de dados imutáveis
 - Processamento paralelo e/ou distribuído
 - Confiável
 - Funcionamento do sistema
 - Resultado das análises
 - Expansível
 - Crescente volume de dados
 - Maior número de fontes de dados

3.1 - Estrutura Interna

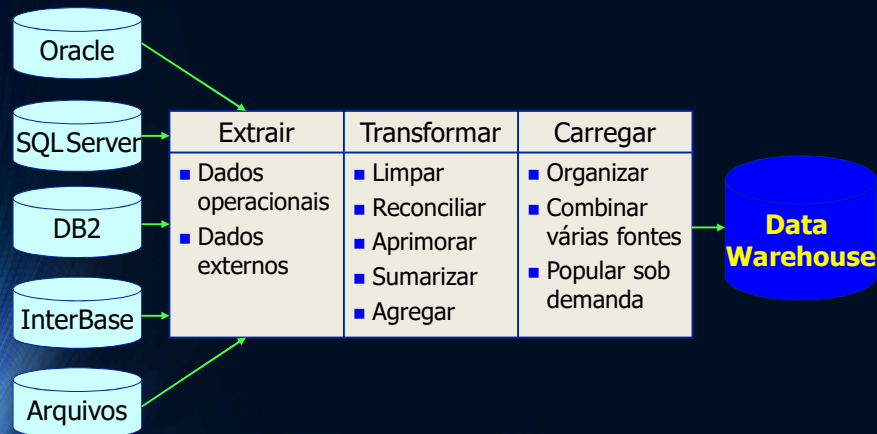
- Em geral são **usados BDs relacionais** para armazenar os dados do DW
- Capazes de manter e processar grandes volumes de dados
- Otimizados para lidar com dados imutáveis
- As ferramentas de análise empregam:
 - Técnicas de **mineração de dados**
 - **Inteligência artificial**: redes neurais, *fuzzy*, etc.
 - A Internet: *Web mining*, agentes móveis, etc.

3.1 - Estrutura Interna



3.1 - Estrutura Interna

• Obtenção de Dados



3.1 - Estrutura Interna

• Busca de Informações



3.1 - Estrutura Interna

- Modelo de Camadas



Business Intelligence - Prof. Antonio Guardado

29

3.1 - Estrutura Interna

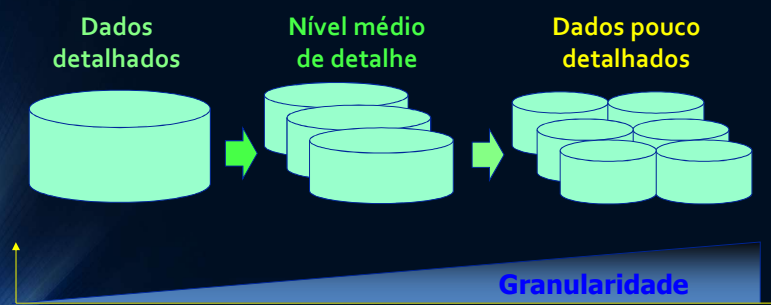
- Funções das Camadas do DW
 - Dados Operacionais/Externos: fontes de dados
 - Acesso aos Dados: extrair dados dos BDs
 - Data Staging: transformar e carregar dados
 - Data Warehouse Físico: armazenar dados
 - Acesso aos Dados: localizar dados para análise
 - Acesso à Informação: analisar dados
 - Troca de Mensagens: transportar dados
 - Gerenc. de Processos: controlar atividades

Business Intelligence - Prof. Antonio Guardado

30

3.2 - Granularidade

- Granularidade
 - Nível de detalhe dos dados
 - De extrema importância no projeto do DW



Business Intelligence - Prof. Antonio Guardado

31

3.2 - Granularidade

- Definir a granularidade adequada é vital para que o DW atenda seus objetivos
 - Mais detalhes → Mais dados → Análise mais longa → Informação mais detalhada
 - Menos detalhes → Menos dados → Análise mais curta → Informação menos detalhada
- Para evitar que se perca informação são criados vários níveis de granularidade

Business Intelligence - Prof. Antonio Guardado

32

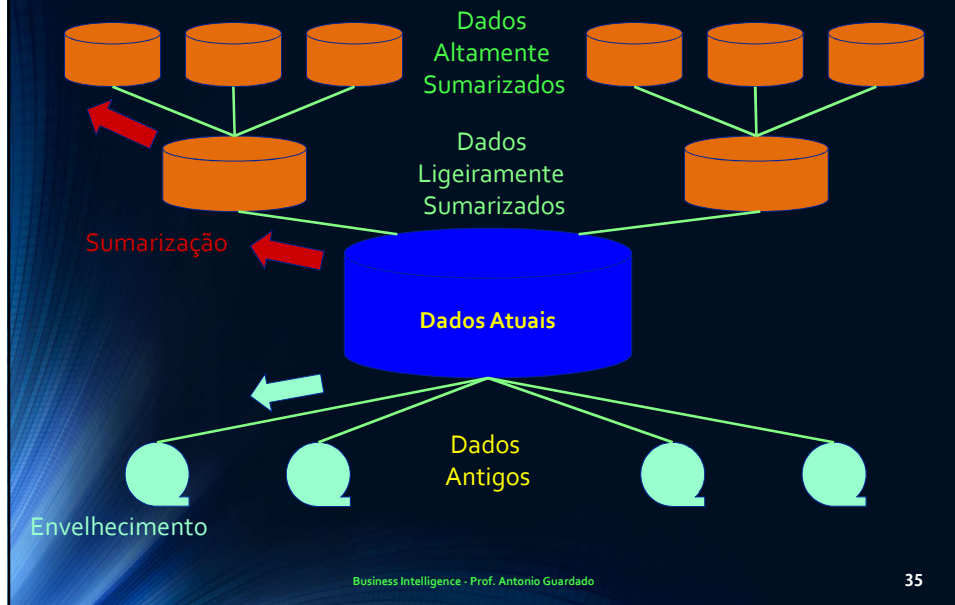
3.2 - Granularidade

- Dados x Granularidade
 - **Dados Atuais**
 - Refletem acontecimentos recentes
 - Alto nível de detalhe (baixa granularidade)
 - **Dados Sumarizados** (1 ou + níveis)
 - Dados históricos condensados
 - Menor nível de detalhe (maior granularidade)
 - **Dados Antigos**
 - Dados históricos mantidos em fita, CD, etc
 - Alto nível de detalhe (baixa granularidade)

3.2 - Granularidade

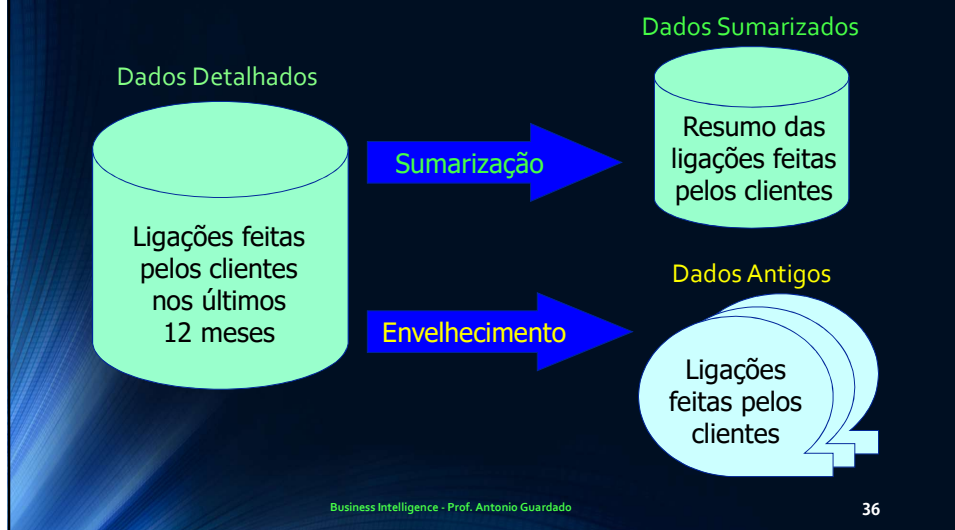
- Processo de **sumarização**
 - Aplica um novo esquema de modo a condensar os dados
 - Ex.: armazenar totais, médias, etc.
- Processo de **envelhecimento**
 - Transfere os dados antigos do HD para fita, CD, etc.
 - Mantém o nível de detalhe para que nenhuma informação seja perdida

3.2 - Granularidade



3.2 - Granularidade

- Exemplo: Companhia Telefônica



3.2 - Granularidade

• Exemplo: Companhia Telefônica (cont.)

Dados Detalhados

Ligações
Origem
Destino
Início
Fim
Tarifa
Status

Nº de registros:
ligações nos
últimos 12 meses

Dados Sumarizados

Ligações
Cliente
Mês
Pulsos
LongaDist
ValConta

Nº de registros:
contas emitidas
pela empresa

Dados Antigos

Ligações
Origem
Destino
Início
Fim
Tarifa
Status

Nº de registros:
ligações efetuadas
pela empresa

3.2 - Granularidade

- Quanto **menor a granularidade**, **mais detalhada é a informação** disponível
- No exemplo anterior, poderíamos determinar se o cliente A ligou para B na semana passada
- Também poderíamos verificar se A faz muitas chamadas de longa distância
- Durante o processo de sumarização, algumas informações podem ser perdidas
 - Não seria possível saber se A ligou para B
 - É possível verificar o padrão de consumo de A

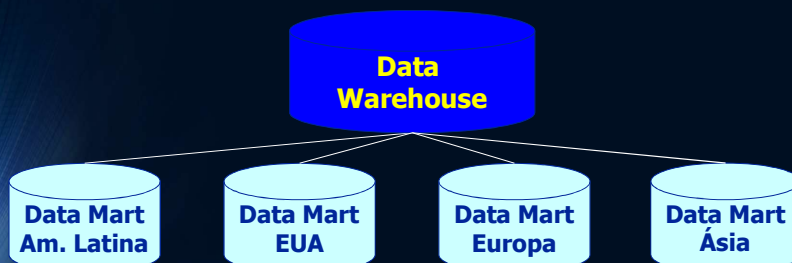
3.3 - Data Marts

- Dados mantidos no DW são separados por assunto em subconjuntos de acordo com:
 - A estrutura interna da empresa
 - O processo de tomada de decisão
- Estes subconjuntos dos dados são chamados de **Data Marts**



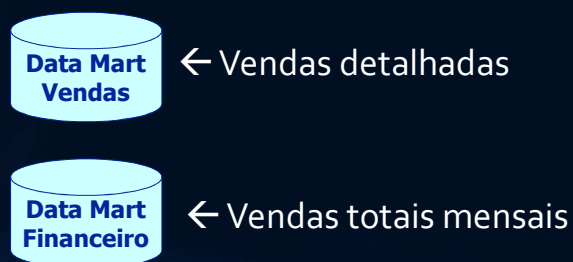
3.3 - Data Marts

- Um Data Mart desempenha o papel de um **DW departamental**, regional ou funcional
- Uma empresa pode construir seus Data Marts gradativamente a partir do DW



3.3 - Data Marts

- Dados podem ser repetidos em dois ou mais Data Marts
- Os mesmos dados podem estar representados com granularidade diferente
- Ex:



3.4 - Metadados

- Os Metadados são dados sobre os dados
- Para cada atributo mantido no DW há uma entrada no dicionário de dados
- Os dados são processados, atualizados e consultados partindo dos metadados
- Usuários ficam conhecendo a estrutura e o significado dos dados
- No BD operacional, a estrutura e o significado dos dados estão embutidos nas aplicações

3.4 - Metadados

- Camadas de Metadados
 - Metadados Operacionais
 - Definem a estrutura dos dados operacionais
 - Metadados do DW
 - Orientados por assunto
 - Informam como os dados do DW foram calculados e como devem ser interpretados
 - Metadados do Usuário
 - Organizam os metadados do DW com base em conceitos familiares ao usuário final

3.4 - Metadados

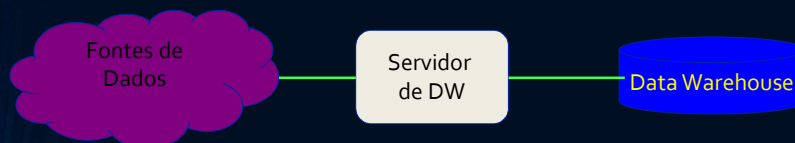
- Classificação em função dos dados descritos
 - Metadados de Mapeamento
 - Como BDs operacionais são mapeados no DW
 - Metadados de Sumarização
 - Como os dados foram sumarizados no DW
 - Metadados Históricos
 - Como a estrutura dos dados vem mudando
 - Metadados de Padrões de Acesso
 - Como os dados do DW vem sendo acessados
 - Metadados de Miscelânea

3.4 - Metadados

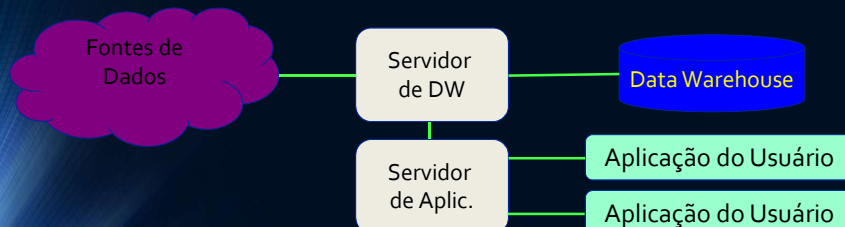
- Fontes de Metadados
 - Código fonte dos SGBDs operacionais
 - Diagramas CASE de BDs operacionais e do DW
 - Documentação dos BDs operacionais e do DW
 - Entrevistas com usuários, administradores e programadores dos BDs e do DW
 - O ambiente de DW
 - Frequência de acesso aos dados, tempo de resposta, controle de usuários, etc.

3.5 - Acesso aos Dados

- Acesso em Duas Camadas



- Acesso em Três Camadas



3.5 - Tipos de Data Warehouse

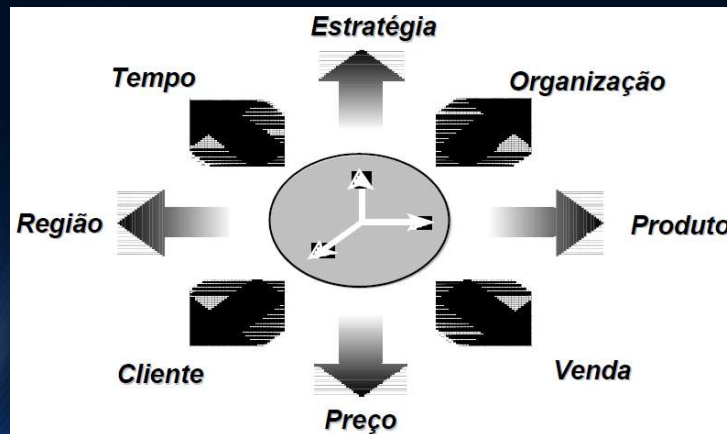
- DW baseado em Servidor
 - Mainframe ou servidor de rede local (LAN)
- DW Virtual
 - Reúne dados operacionais e dados históricos mantidos em BDs – não há um DW central
- DW Distribuído
 - DW global reúne dados de vários DWs locais
- DW baseado na Web
 - Dados provenientes da World Wide Web

4- Modelagem Multidimensional

- Dados operacionais visualizados sob **diversos ângulos**
- Os ângulos são chamados **dimensões do negócio**
 - Produto
 - Loja
 - Tempo
 - Região
- Uma das dimensões é “sempre” o Tempo

4 - Modelagem Multidimensional

- Outros Exemplos de Dimensões



4 - Modelagem Multidimensional

- Atributos de Dimensão
 - São atributos **qualitativos** que caracterizam os ramos do negócio envolvidos na medida de desempenho de determinado fato
- Exemplo
 - Dimensão "Produto"
 - Descrição, embalagem, preço, etc.

4- Modelagem Multidimensional

- Granularidade
 - Define o nível de detalhe das dimensões
 - Influencia o tamanho das dimensões
- Exemplo
 - Dimensão “Loja”
 - País **Maior Granularidade**
 - Região
 - Estado
 - Cidade **Menor Granularidade**

4 - Modelagem Multidimensional

- Hierarquias entre os **atributos** das dimensões
 - Úteis para a geração de relatórios
- Dimensão “Região”
 - Fornecedor → Cidade → Estado → Região
- Dimensão “Tempo”
 - Dia → Semana → Mês → Trimestre → Ano

4 - Modelagem Multidimensional

- Associados a cada dimensão existem **fatos**:
 - Vendas
 - Compras
 - Sinistro
- Fatos
 - São valores **quantitativos** referentes ao desempenho de um grupo de dimensões
- Exemplo
 - Fato "**Vendas**" (Loja, Produto e Tempo)
 - Quantidade, lucro, valor, etc.

4.1 - Processamento OLAP

- OLAP (*On-line Analytical Processing*)
- Aplicado sobre estruturas dimensionais
- O termo "On-line" significa que
 - os resumos solicitados são obtidos rapidamente
- Difere substancialmente daquele utilizado por aplicações do nível operacional (OLTP)

4.1 - Processamento OLAP x OLTP

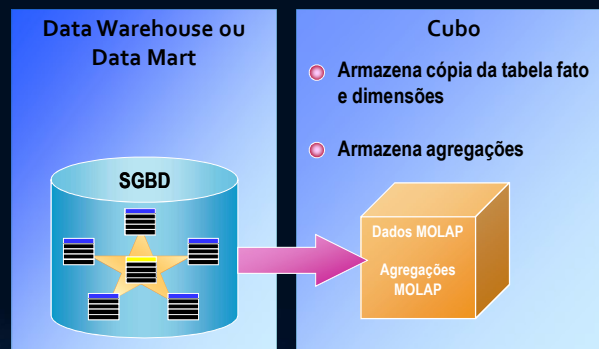
Propriedade	OLTP	OLAP
Tempo de Resposta	Milisegundos para segundos	Segundos para horas
Operações	DML	Leitura
Natureza dos dados	30 – 60 dias	Snapshots no tempo
Organização dos Dados	Aplicação	Assunto, tempo
Tamanho	Pequeno para grande	Grande para Muito Grande
Fontes de Dados	Operacional, Interna	Operacional, Interna, Externa
Atividades	Processos	Análise

4.2 – Tipos Modelagem Multidimensional

- Principais vertentes para Modelagem Multidimensional
 - Multidimensional OLAP (MOLAP)
 - Relational OLAP (ROLAP)
 - Hybrid OLAP (HOLAP)
- Outras variações
 - Web-based OLAP (WOLAP)
 - Desktop OLAP (DOLAP)
 - Real-Time OLAP (RTOLAP)

4.2.1 -MOLAP (Multidimensional OLAP)

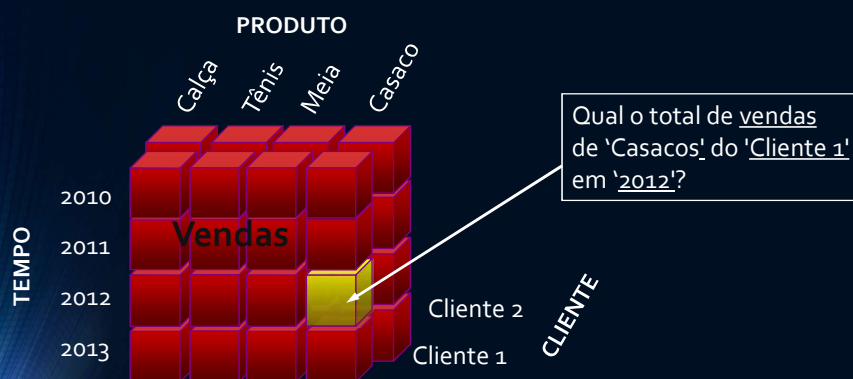
- Baseia-se nos Bancos de Dados Multidimensionais
- Estrutura utilizada: Cubo Multidimensional
- Cada aresta representa uma dimensão do negócio
- As células do cubo são preenchidas com valores quantitativos (medidas ou fatos)



Business Intelligence - Prof. Antonio Guardado

57

4.2.1 -MOLAP (Multidimensional OLAP)

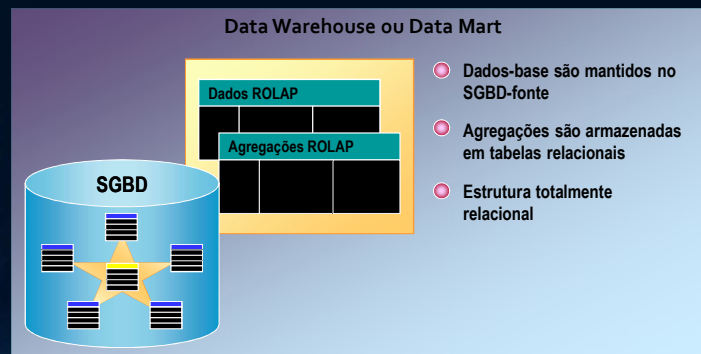


Business Intelligence - Prof. Antonio Guardado

58

4.2.2. - ROLAP (Relational OLAP)

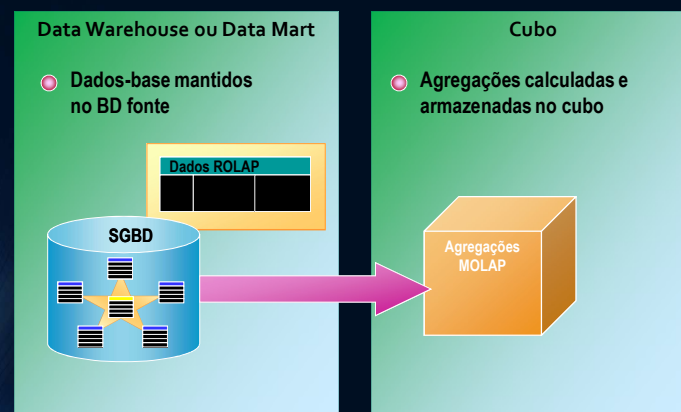
- Tabelas relacionais simulam o cubo multidimensional
- O esquema proposto concentra numa tabela as medidas ou fatos do negócio (tabela de fatos)
- Os componentes das dimensões são armazenados em outras tabelas (tabelas de dimensão)



Business Intelligence - Prof. Antonio Guardado

59

4.2.3 - HOLAP (Hybrid OLAP)



Business Intelligence - Prof. Antonio Guardado

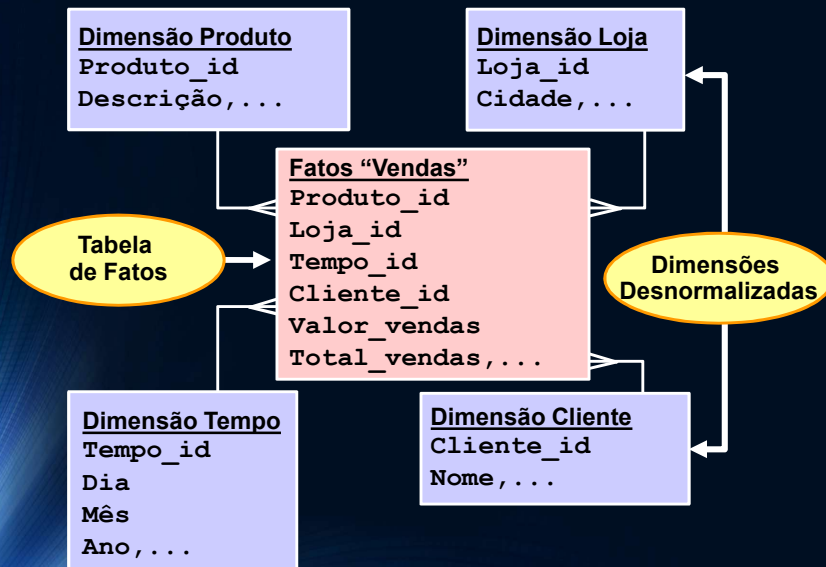
60

4.3- MOLAP vs. ROLAP vs. HOLAP

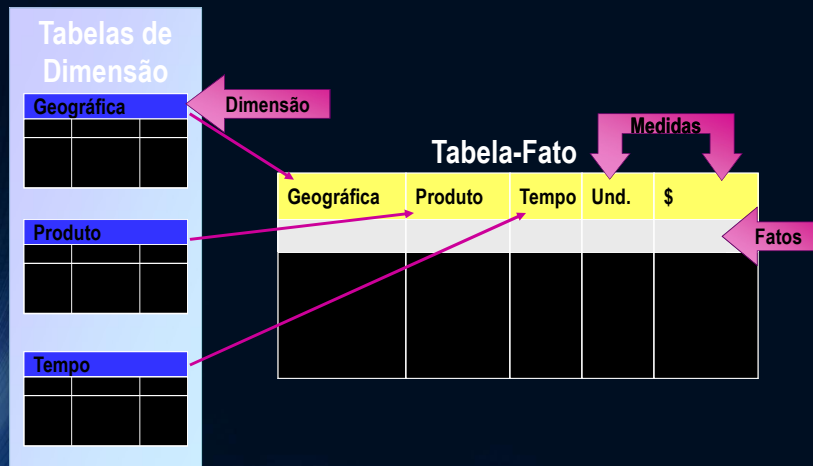
Armazenamento	MOLAP	HOLAP	ROLAP
Dados de base	Cubo	Tabela Relacional	Tabela Relacional
Agregações	Cubo	Cubo	Tabela Relacional

Perspectiva do Cliente	MOLAP	HOLAP	ROLAP
Desempenho de consulta	Imediato	Mais rápido	Rápido
Consumo em disco	Alto	Médio	Baixo
Manutenção do cubo	Alto	Médio	Baixo

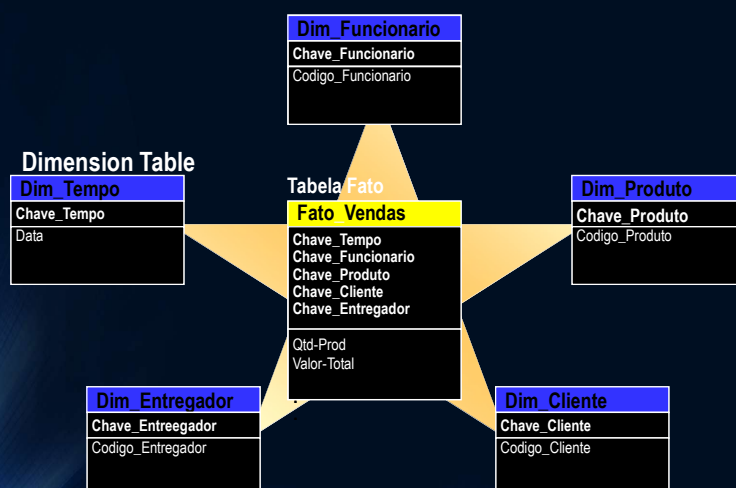
4.4 - Modelo Estrela ("Star Schema")



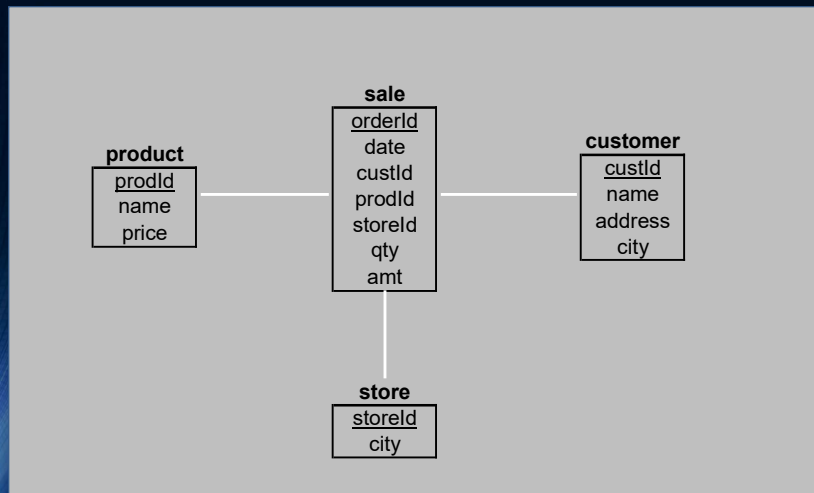
4.4 - Modelo Estrela ("Star Schema")



4.4 - Modelo Estrela ("Star Schema")



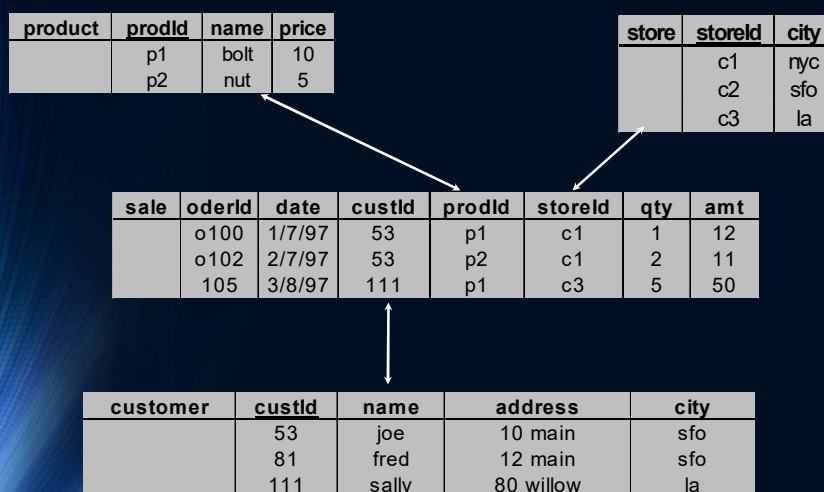
4.4 - Modelo Estrela ("Star Schema")



Business Intelligence - Prof. Antonio Guardado

65

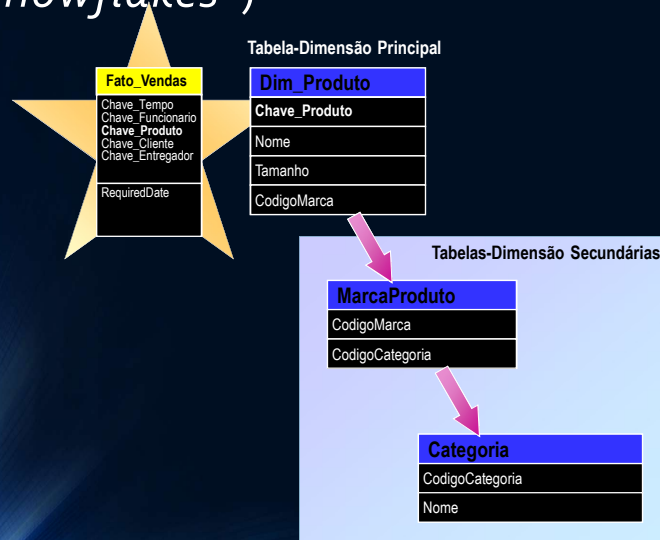
4.4 - Modelo Estrela ("Star Schema")



Business Intelligence - Prof. Antonio Guardado

66

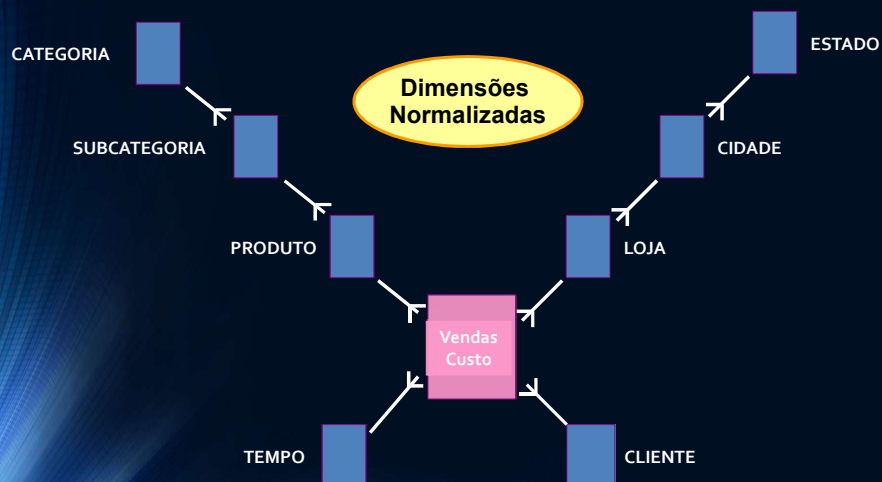
4.5 - Modelo Flocos de Neve ("Snowflakes")



Business Intelligence - Prof. Antonio Guardado

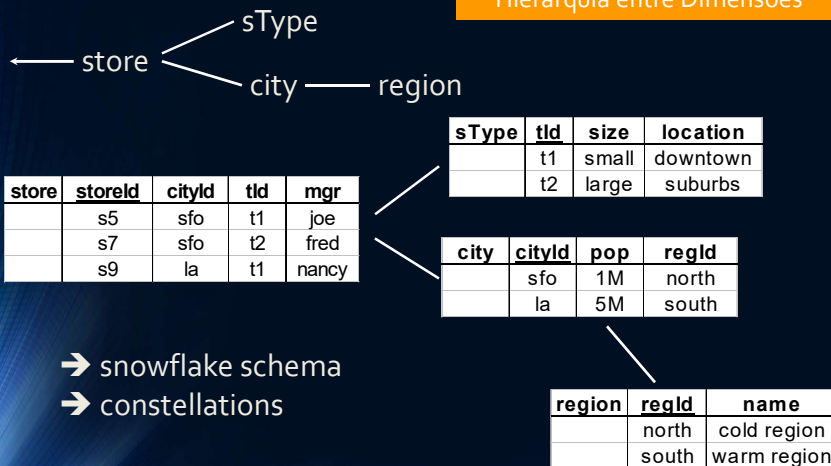
67

4.5- Modelo Flocos de Neve ("Snowflakes")



4.5 - Modelo Flocos de Neve ("Snowflakes")

Hierarquia entre Dimensões



4.6 - Escolhendo uma Modelagem Multidimensional

	Star Schema	Snowflake
Clareza	+ fácil	+ difícil
Número de tabelas	<	>
Complexidade de consultas	+ simples	+ complexo
Desempenho de consulta	+ rápido	+ lento