

Sumário

- 1. Introdução
- 2. Data Warehouse
- 3. Data Mart
- 4. Metadados
- 5. Modelo multidimensional

Pontuando

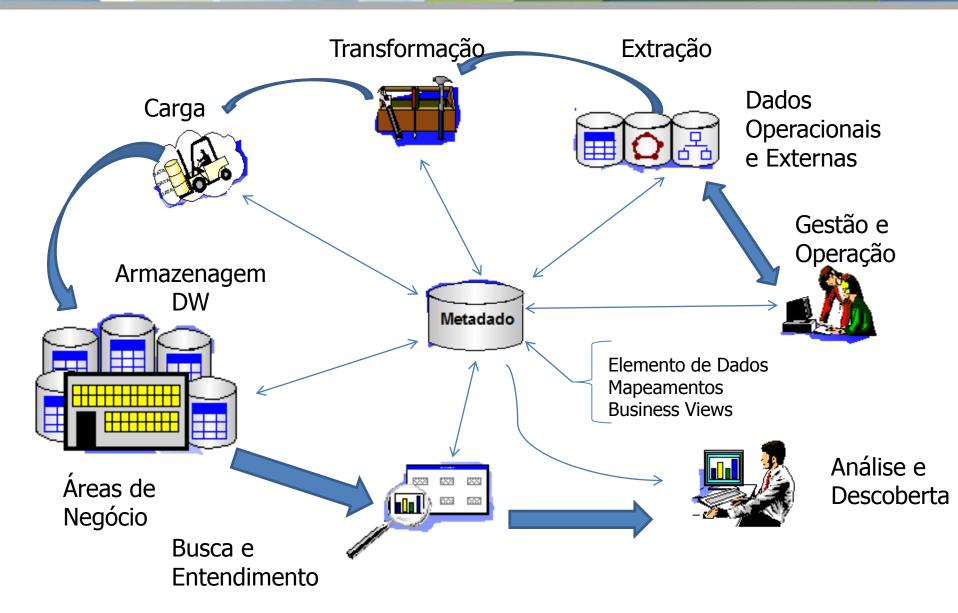
1. Introdução

Fase de armazenagem

- Data warehouse (visão de todo o negócio)
- Data mart (dividido por assunto ou departamentos)
- Metadados (informação sobre os dados)
- Como essas informações estão armazenadas?

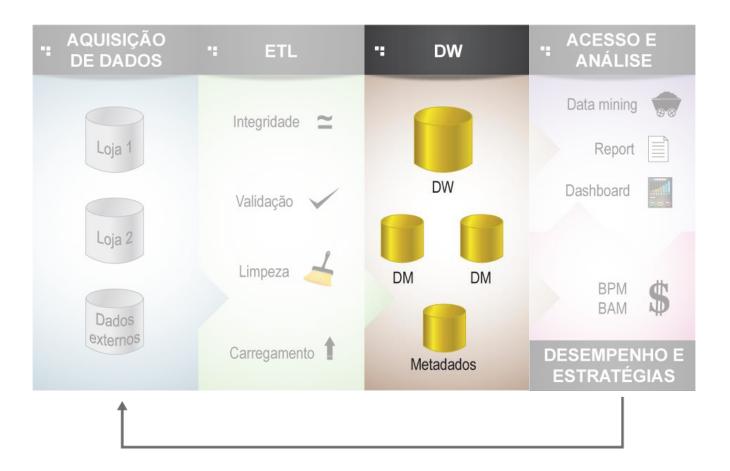


Processo de Datawarehousing



1. Introdução

Fase de armazenagem



Sumário

- 1. Introdução
- 2. Data Warehouse
- 3. Data mart
- 4. Metadados
- 5. Modelo multidimensional

Pontuando

Definição

 O data warehouse pode ser considerado um banco de dados voltado a um conjunto de assuntos. Ele extrai informações valiosas das diversas fontes de dados e as armazena no banco de dados único de um forma estruturada para atender as ferramentas de suporte à decisão.

Definição



- Ralph Kimbal define o data warehouse como um conjunto de data marts.
- Bill Inmon, define o data warehouse como sendo uma estrutura que possui as seguintes características: orientado a assunto, integrado, não volátil e variável no tempo.

Definição

- ✓ Ralph Kimball define o data warehouse como um conjunto de data marts que representa os focos de análise de forma intuitiva, facilitando a filtragem e agregação dos dados gerenciais.
- ✓ Bill Inmon define o data warehouse como uma base de dados orientada por assuntos, integrada, não volátil, com histórico cumulativo indexado no tempo.



Ralph Kimball



Bill Inmon



Características:

 Orientado a assunto: as suas estruturas devem ser organizadas de forma a atender a necessidade de se estudar um determinado assunto.



- Departamento de vendas :
 Produtos x Mês
- Departamento de marketing:
 Avaliar o produto durante o ano

Características:

 Integrado: deve armazenar informações de diferentes fontes de dados de forma consistente.
 Ou seja, deve ter seus dados apurados por ferramentas de ETL.

1	Α	В	С	D
	Nome	Idade	Sexo	CPF
2	Maria	25	F	222
3	João	34	М	333
4	José	30	M	444
5				
6_				

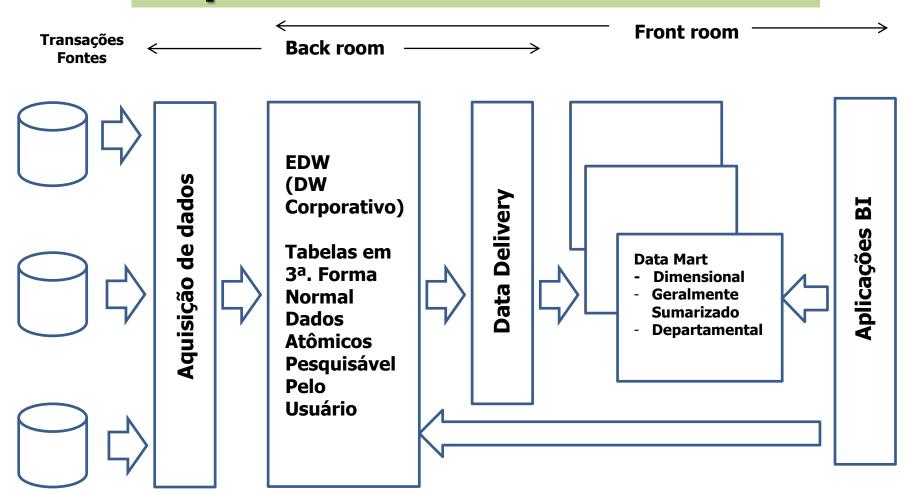
Características:



- Não volátil: isto significa que o banco de dados não pode ser atualizados, eles são somente consultados.
- Variável no tempo: os dados devem ser armazenados de acordo com o momento do carregamento.
 - Permite que a mesma consulta seja retornada com o mesmo dados

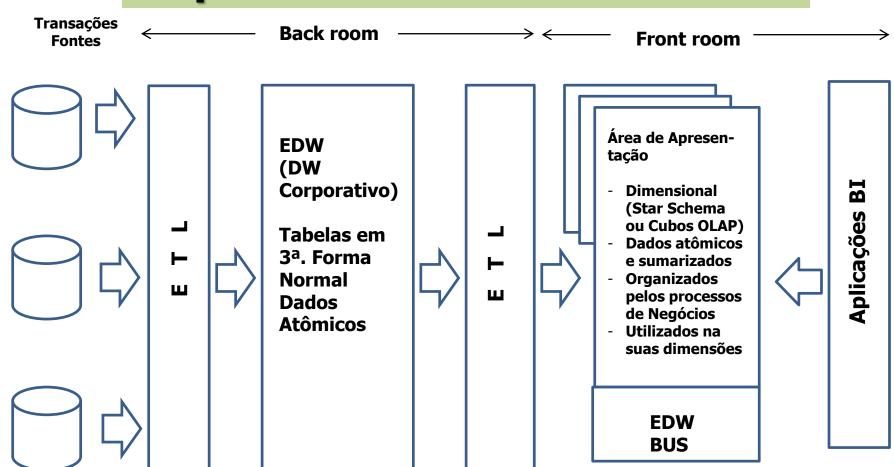
Fonte Imagem: http://Office.com

Arquitetura DW - Bill Inmon



Fonte: The Data Warehouse Toolkit - R. Kimball

Arquitetura DW – R.Kimball



Fonte: The Data Warehouse Toolkit - R. Kimball

Características:

Variável no tempo:

10/01/10

Cliente				
Nome	Est. Civil			
Maria	Casada			
José	Solteiro			

10/01/**11**

Cliente				
Nome	Est. Civil			
Maria	Casada			
José	Solteiro			
Ana	Solteira			
João	Casado			

Estrutura:



Diferenças entre BI e *Data Warehouse*:

- Inteligência de negócios
- Utilização da tecnologia de data warehouse
- Ferramentas de apoio a tomada de decisão
- Voltado a negócios
- Parte organizacional

Sumário

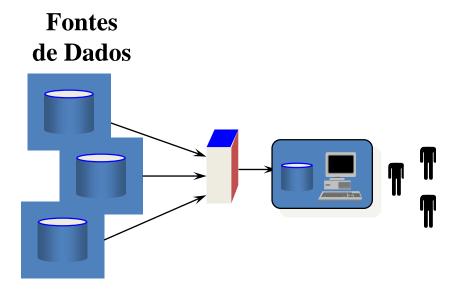
- 1. Introdução
- 2. Data Warehouse
- 3. Data Mart
- 4. Metadados
- 5. Modelo multidimensional

Pontuando

O QUE É UM DATA MART

DATA MART

 "Uma implementação de um DW com um escopo menor e mais restrito de dados e funções, servindo a um único departamento ou parte de uma organização"*



Prof. Dr. Lawrence Chung Koo

Definição:

- É um segmento/parte do Data Warehouse;
- Separado por assunto/departamento;
- Protótipo rápido;
- Restringe o acesso aos dados;
- Possui três tipos: dependente, independente e o híbrido.

DATA MART - CARACTERÍSTICAS

Características de um Data Mart

- Um data mart tem escopo restrito tamanho, usabilidade, ciclo de vida
- Um data mart suporta dados primitivos, sumarizados ou outras derivações, direcionados para resolver um problema de negócio
- Um data mart geralmente suporta análise de dados multi-dimensionais
- O desenho de um data mart tende a ser dependente da ferramenta

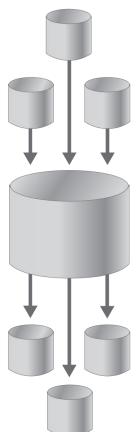


Tipos de *Data Mart*:

Fontes Operacionais / Dados Externos

Data Warehouse

Data Marts Dependentes

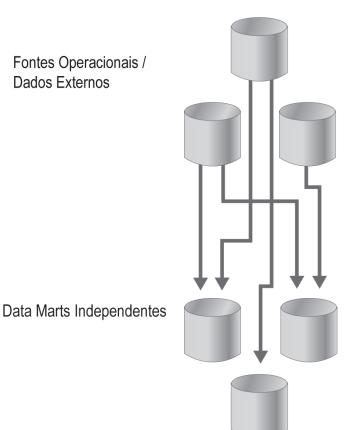


 Dependente: Sua fonte de dados é o próprio data warehouse.

Fonte Imagem: adaptado de LANE et al., 1999

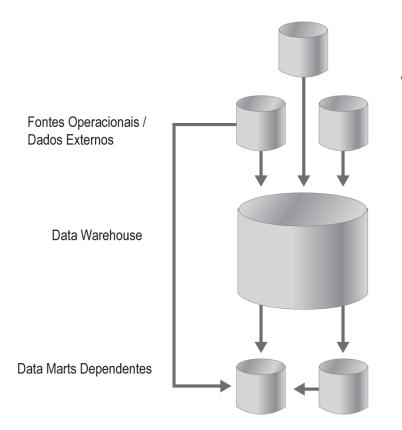
Tipos de *Data Mart*:

Fontes Operacionais / **Dados Externos**



Independente: É criado sem um data warehouse central, busca as suas informações diretamente dos bancos operacionais e dados externos

Tipos de *Data Mart*:



 Híbrido: Não há exigência quanto ao tipo de sua fonte de dados. Pode ser gerado a partir de diversas fontes como data warehouses, data marts, bancos de dados operacionais e dados externos.

Sumário

- 1. Introdução
- 2. Data Warehouse
- 3. Data Mart
- 4. Metadados
- 5. Modelo multidimensional

Pontuando

4. Metadados

Definição:

- Os metadados são informações sobre os dados;
- A ideia principal, é abstrair a informação de forma que ela possa ser compreendida por qualquer sistema;
- Possuem 4 tipos: sintáticos, estruturais, semânticos e negócios/técnicos.

4. Metadados

Definição:

 Os Metadados podem estar dentro encapsulados dentro do mesmo arquivo de dados ou podem estar separados em outros arquivos. Por exemplo, no início de cada arquivo de dados temos a descrição em forma de tags.

<nome>

Conteúdo com a característica da tag

</nome>

 São baseadas nos padrões XML (Extended Markup Language – Linguagem de Marcação Extendida).

4. Metadados

Definição:

- Dicionário de dados
- Catálogo de dados
- Diretório de dados

 Normalmente administrado por um DA (Data Administrator), gestor de informação da empresa



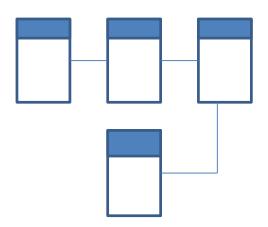
Sumário

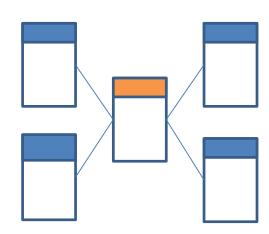
- 1. Introdução
- 2. Data Warehouse
- 3. Data Mart
- 4. Metadados
- 5. Modelo multidimensional

Pontuando

Definição:

 É uma forma de estruturação dos dados é baseada em tabelas dimensão e fato, diferentemente da entidade-relacionamento que nós conhecemos.





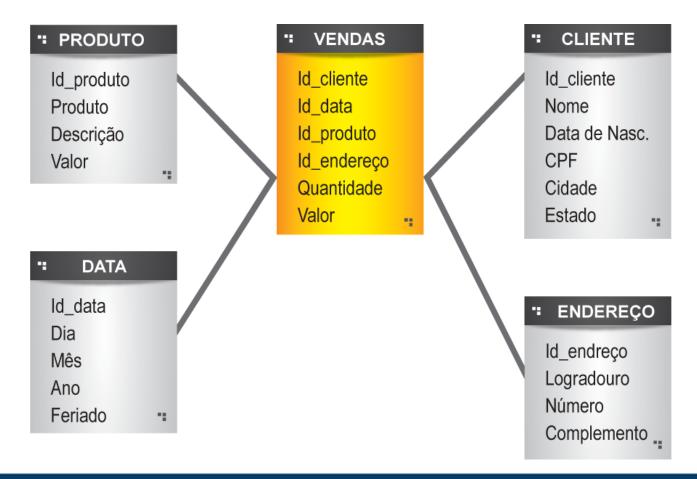
Definição:

 Tabelas dimensão: são as unidades de consultas representadas por tabelas que devem possuir um campo de identificação (chave) e campos interessantes para consulta/junções que podem ser estar organizados de forma hierárquica.

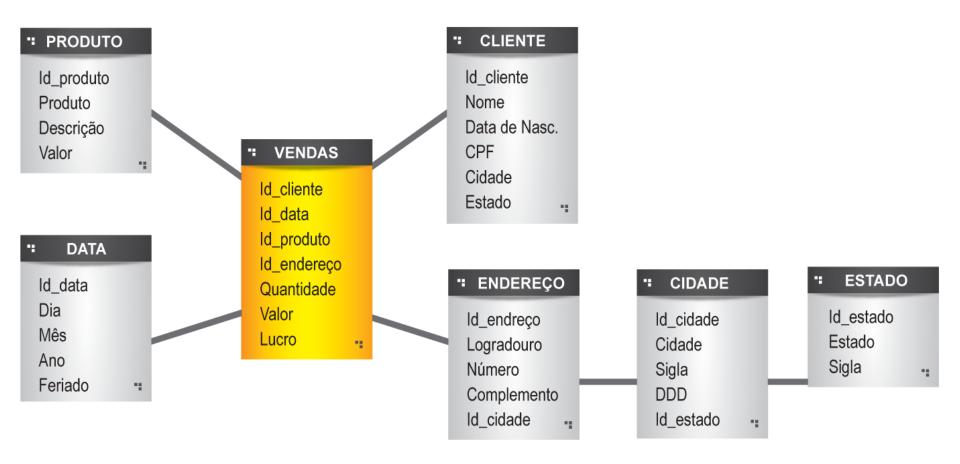
Definição:

 Tabelas fato: são tabelas que relacionam/ agrupam dimensões através de medidas. Ela contém chaves das dimensões e as medidas da relação entre seus membros.

Modelo estrela:

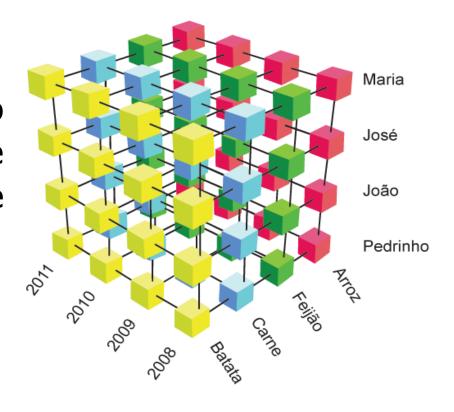


Modelo floco de neve:

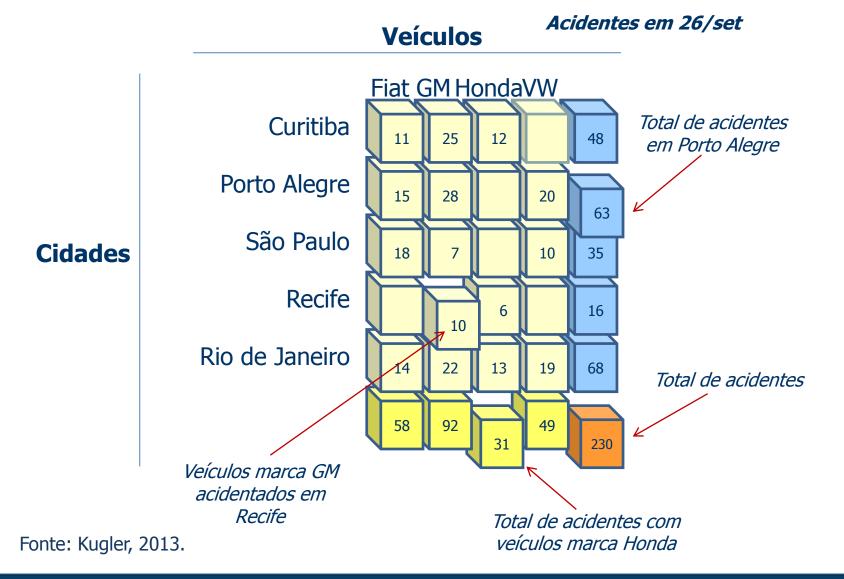


Consultas

 Representação do cubo vendas (fato) para análise da data, produto e cliente (dimensões).

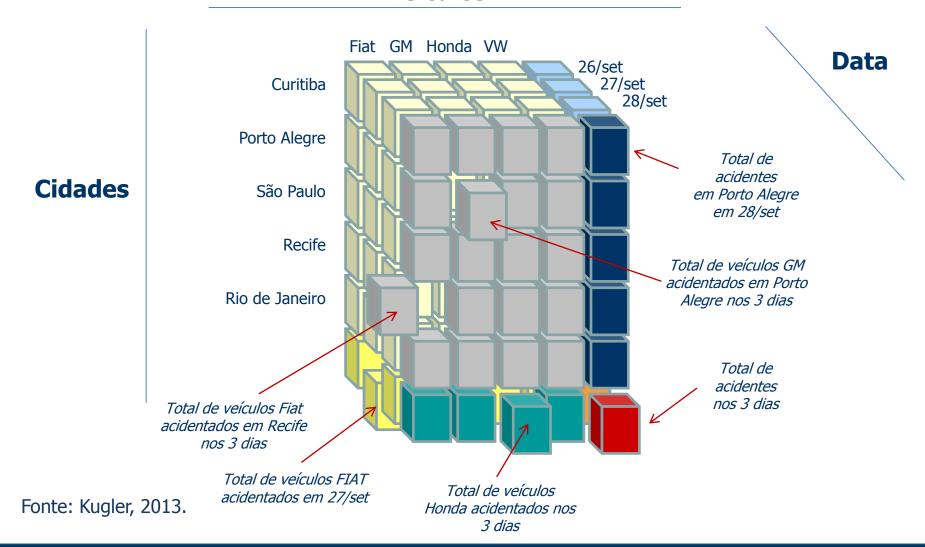


5. Modelo Multidimensional (OLAP)

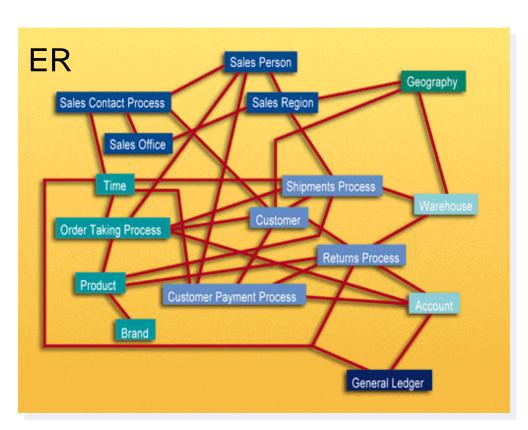


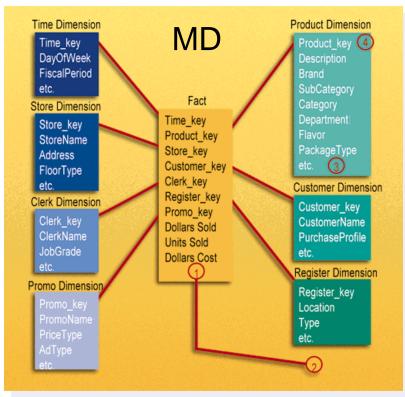
5. Modelo Multidimensional (OLAP)

Veículos



Modelo ER vs. Modelo Dimensional





Fonte: Kugler, 1997.

Modelo Relacional

Voltados para processamento *transacional* (OLTP – *online transaction processing*).

Executam operações específicas, pontuais, predefinidas.

Gerenciam acessos múltiplos ao mesmo banco de dados.

 Muitas pessoas realizando concorrentemente as mesmas operações.

Tempo de resposta precisa ser imediato.

 Transações rápidas e curtas operando sobre poucos dados.

Fonte: Kugler, 2013.

Modelo Relacional

Atualização em tempo real.

 As bases de dados refletem imediatamente as operações realizadas.

Controle rígido de concorrência.

 Mecanismos de *locking, commit* e *rollback* para evitar corrupção dos dados.

Extração e Reporting complicados.

 Muitas solicitações encaminhadas à TI, gerando backlog.

Históricos com horizonte breve.

Em geral, limitados a poucos meses.

Fonte: Kugler, 2013.

Modelo Dimensional

Orientado para *análise* (OLAP – *online analytical processing*).

Consultas para análises táticas e planejamento estratégico.

Menos usuários.

- Relativamente poucas pessoas realizando consultas e análises ao mesmo tempo.
- Tempo de resposta > minutos ou horas.
 - Consultas sobre muitos fatos e grande volume de dados.

