



## **Introdução a DW/OLAP**

Material baseado em originais de Maria Luiza Campos – NCE/UFRJ  
Atualizado com publicações recentes referenciadas no texto

*Licença Creative Commons – Atribuição  
Uso Não Comercial – Compartilhamento pela mesma Licença*



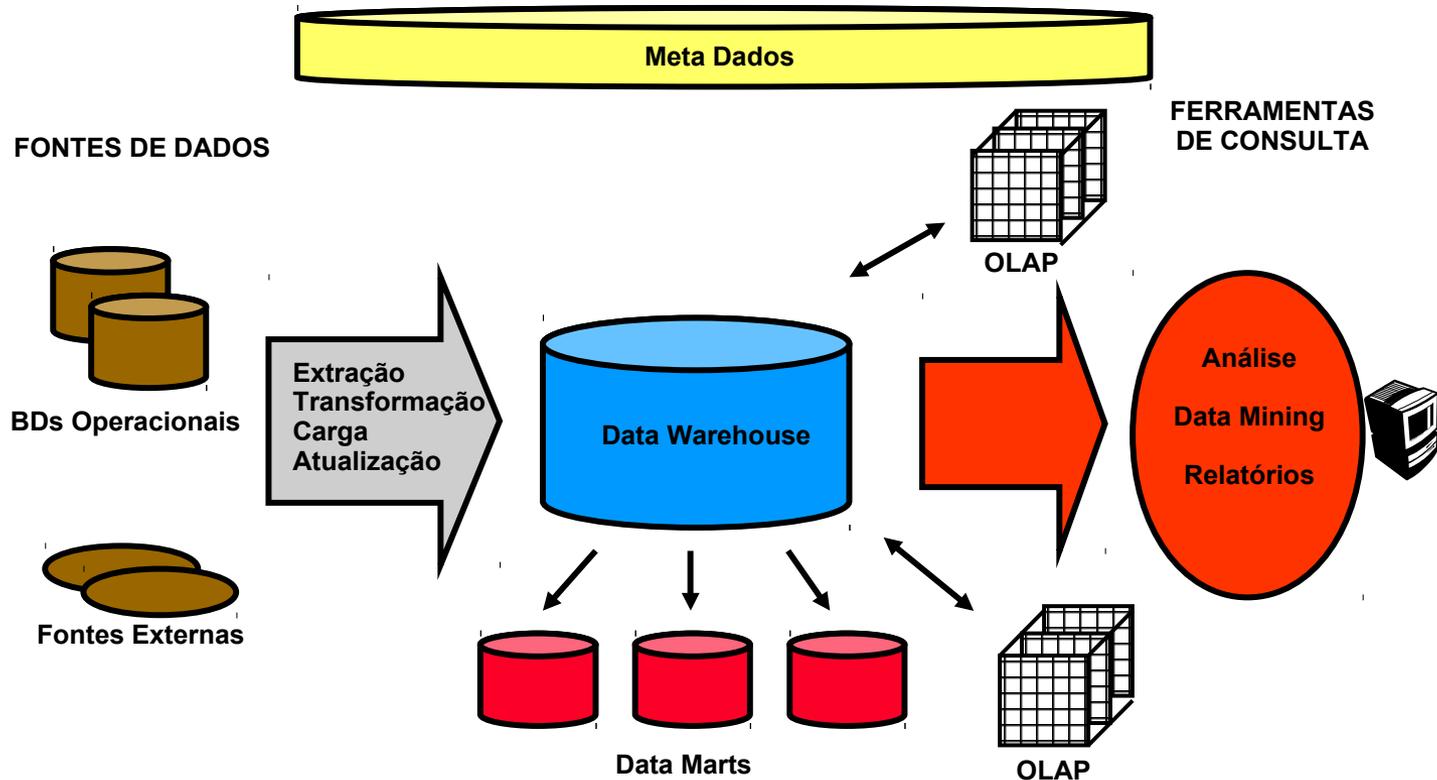
# Data Warehouse e OLAP

- **Data Warehouse: conceitos e arquiteturas**
- **Visão multidimensional de dados**
  - Estrutura relacional x visão dimensional
  - Hierarquias de dimensões
  - Agregados
  - Problema com agregações: armazenar ou calcular?
- **Ferramentas OLAP**
  - As doze regras de Codd
  - Operações dimensionais para OLAP
    - » Slice & Dice, Drilling, Rotação, etc.
  - Tipos de OLAP: MOLAP, ROLAP, HOLAP

# Definições de Data Warehouse

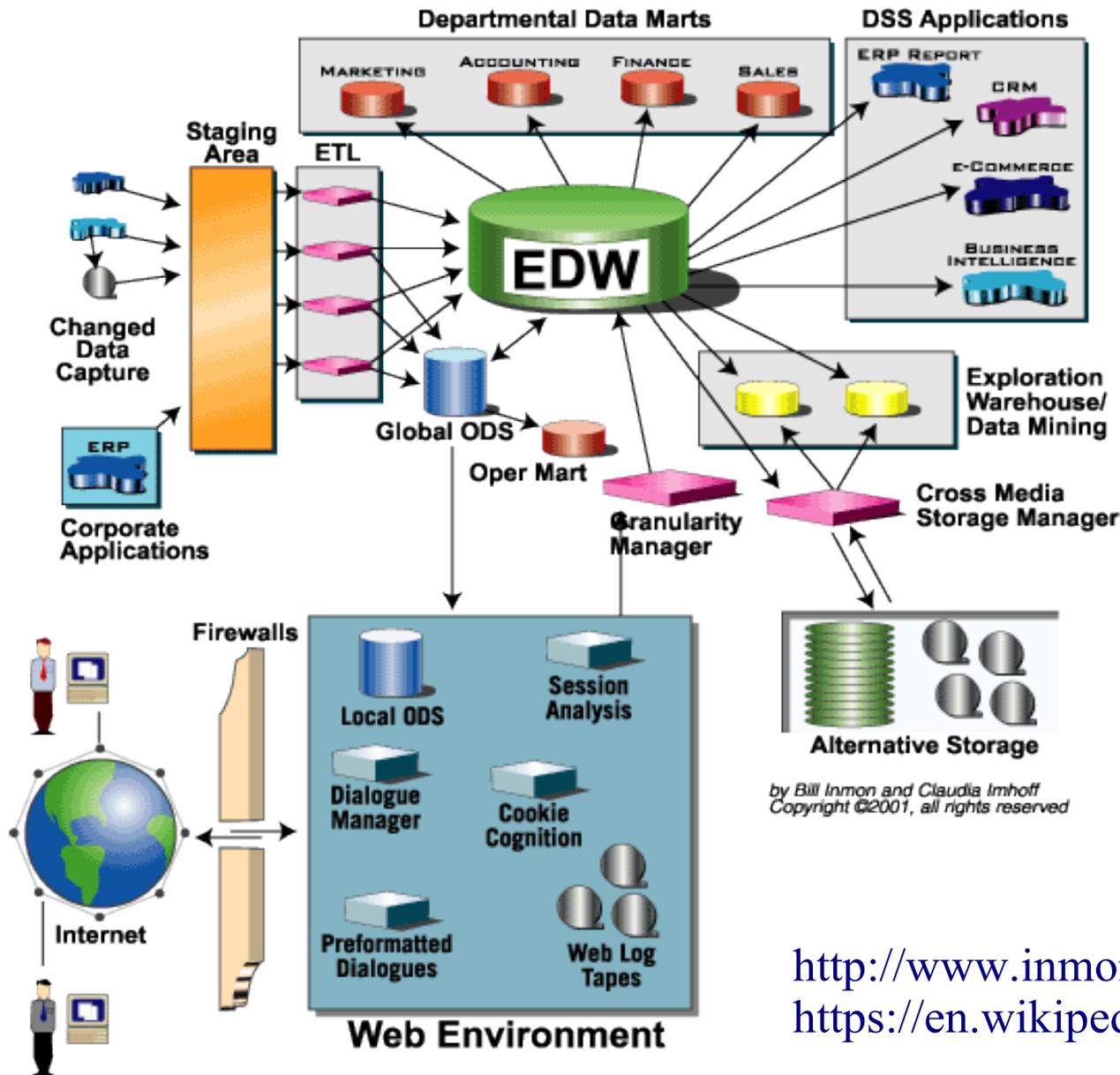
- Uma coleção de dados orientada por assunto, integrada, variante no tempo e não volátil para apoio ao processo de tomada de decisão da gerência (W. Inmon)
- Um banco de dados destinado a sistemas de suporte à decisão, cujos dados são armazenados em estruturas lógicas dimensionais, possibilitando o seu processamento analítico por ferramentas OLAP e de mineração de dados. (R. Kimball)

# Arquitetura Genérica de DW (visão acadêmica)



Chaudhri&Dayal, SIGMOD RECORD 1997

# The Corporate Information Factory and the Web Environment



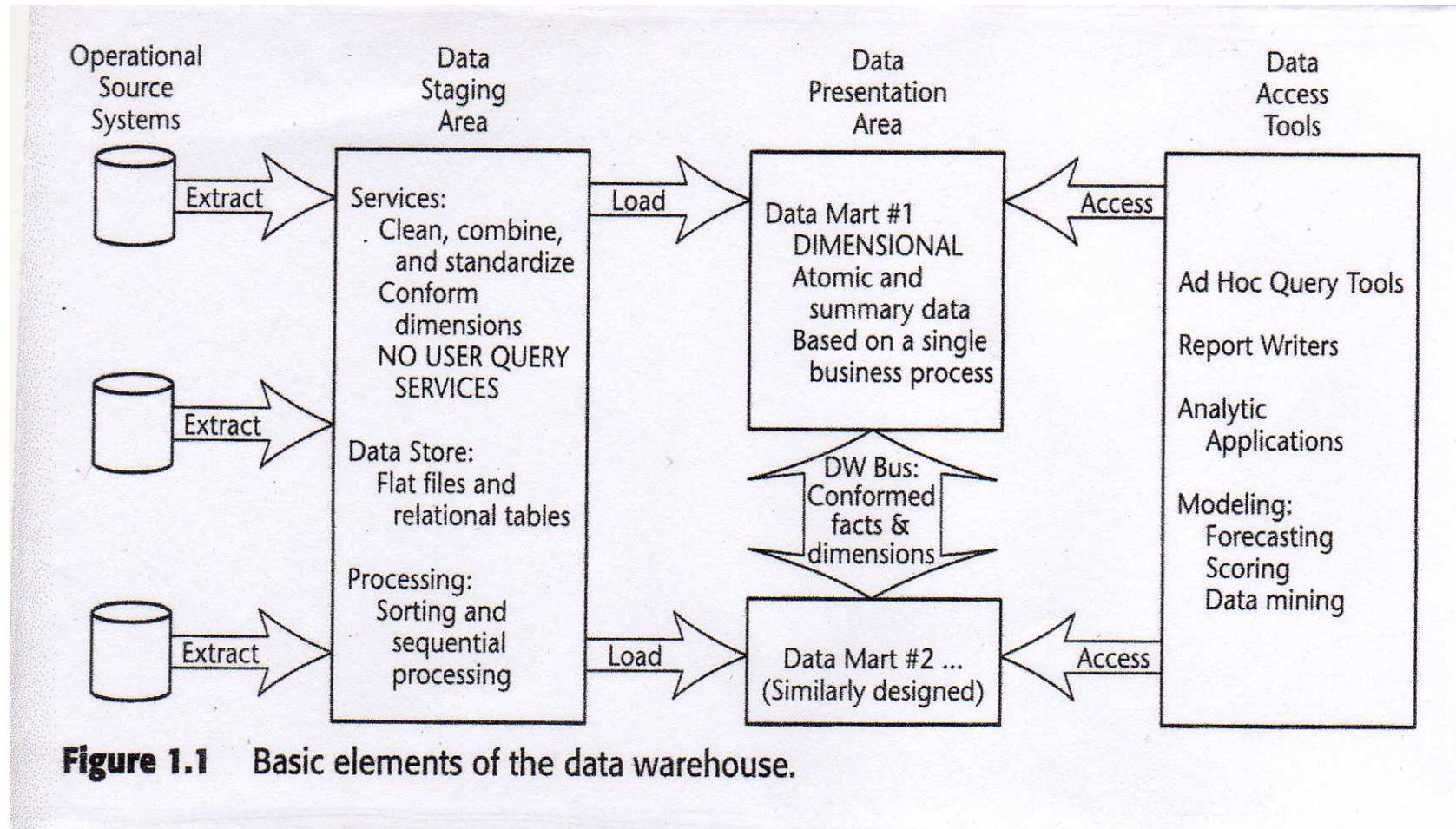
Arquitetura  
de DW  
segundo  
Inmon

Corporate  
Information  
Factory  
de Bill Inmon

*by Bill Inmon and Claudia Imhoff  
Copyright ©2001, all rights reserved*

<http://www.inmoncif.com/library/cif/>  
[https://en.wikipedia.org/wiki/Bill\\_Inmon](https://en.wikipedia.org/wiki/Bill_Inmon)

# Arquitetura de DW segundo Kimball (2002)

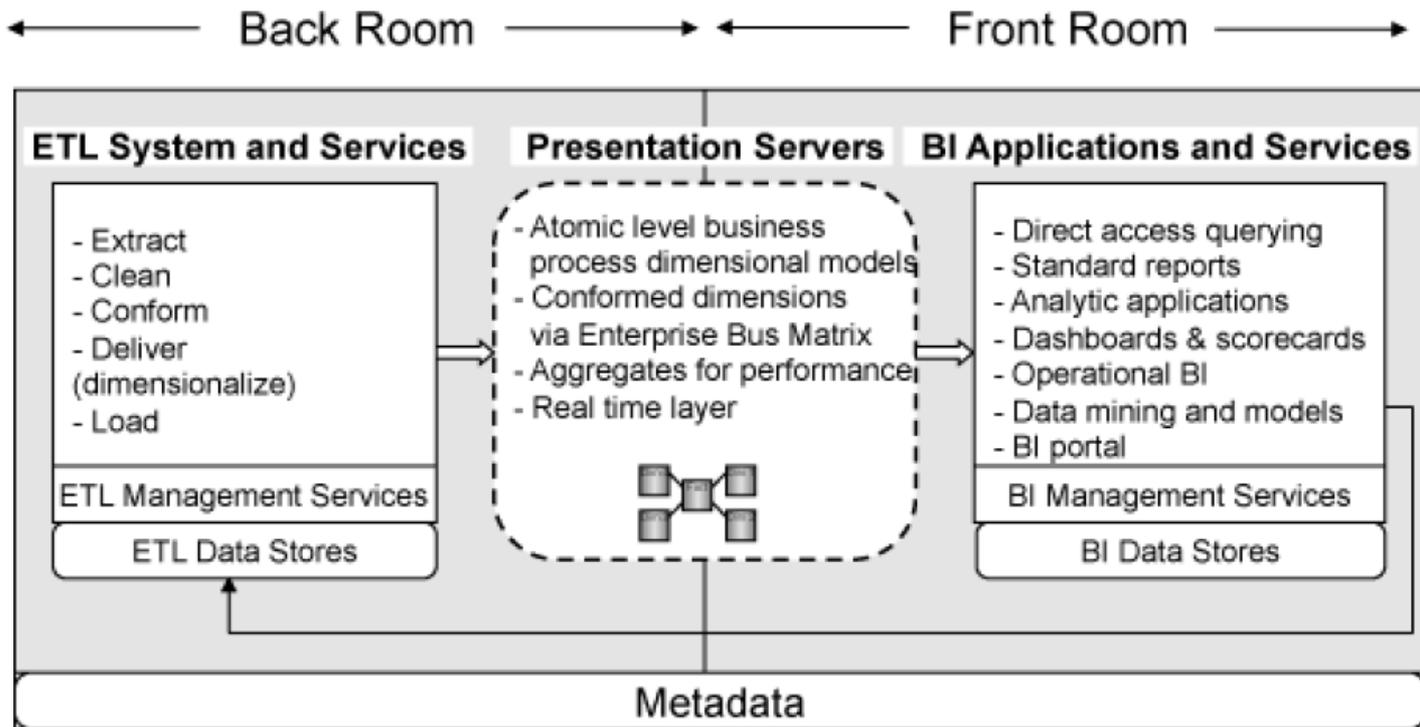


**Figure 1.1** Basic elements of the data warehouse.

Ralph Kimball, Margy Ross: The Data Warehouse Toolkit, 2ª Edição, Wiley, 2002

# Data Warehouse (Kimball 2013)

Figure 1: Kimball technical system architecture diagram.



# OLTP x OLAP

<b>Características</b>	<b>Dados operacionais (transacionais)</b>	<b>Dados informacionais (analíticos)</b>
Conteúdo	Valores correntes, detalhados	Valores históricos, sumarizados
Organização dos dados	Por aplicação, sistema de informação, disponível para poucos usuários; abrangência restrita	Por assunto, negócio; abrangência ampla.
Natureza dos dados	Dinâmica, sujeita a atualizações frequentes	Estática, atualização apenas com “refresh”
Estrutura de dados	Relacional, própria para processamento transacional	Dimensional, própria para processamento analítico (segundo Kimball)
Uso	Estruturado, repetitivo, solução para requisitos conhecidos	Desestruturado, consultas sob diferentes perspectivas, permite descoberta de conhecimento.
Desempenho	Otimizado para tempo de resposta (máximo 2 a 3 segundos); inviável para análises complexas.	Otimizado para análises complexas, com tempos de respostas viáveis.

# Visão multidimensional de dados

- Facilita o entendimento e visualização de problemas típicos de suporte à decisão
- Mais intuitiva para o processamento analítico
- Utilizada por 100% pelas ferramentas OLAP

*A visão lógica é multidimensional, embora a estrutura física quase sempre tenha a mesma visão tabular do modelo relacional, por força de mercado dos bancos de dados relacionais.*

# Estrutura Relacional

**Quantidade de vendas de carros por Modelo e Cor  
(do revendedor MAICAR no mês de Janeiro/2014)**

MODELO	COR	QUANTIDADE
COUPE	AZUL	2
COUPE	BRANCO	4
COUPE	PRATA	7
MINI VAN	AZUL	6
MINI VAN	BRANCO	2
MINI VAN	PRATA	3
SEDAN	AZUL	6
SEDAN	BRANCO	2
SEDAN	PRATA	1

# Visão matricial ou (bi)dimensional

**Quantidade de vendas de carros por Modelo e Cor  
(do revendedor MAICAR no mês de Janeiro/2014)**

Soma - QUANTIDADE	COR			
MODELO	AZUL	BRANCO	PRATA	Total Resultado
COUPE	2	4	7	13
MINI VAN	6	2	3	11
SEDAN	6	2	1	9
<b>Total Resultado</b>	<b>14</b>	<b>8</b>	<b>11</b>	<b>33</b>

- Estrutura de dados: array multidimensional com um número fixo de dimensões.
- Os valores são armazenados em posições do array.
- Como tabela dinâmica (Pivot Table) em planilhas.

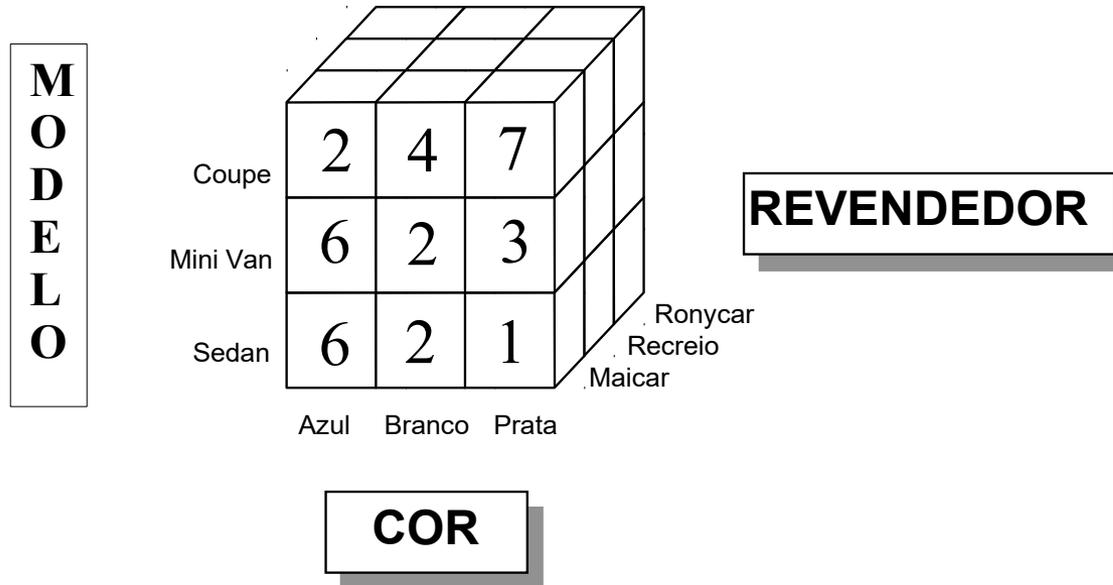
# Acrescentando mais uma coluna...

**Quantidade de Vendas de Carros  
Por Modelo, Cor e Revendedor  
no mês Janeiro/2014**

MODELO	COR	REVENDEDOR	QUANTIDADE
COUPE	AZUL	MAICAR	2
COUPE	AZUL	RECREIO	3
COUPE	AZUL	RONYCAR	2
COUPE	BRANCO	MAICAR	4
COUPE	BRANCO	RECREIO	5
COUPE	BRANCO	RONYCAR	1
COUPE	PRATA	MAICAR	7
COUPE	PRATA	RECREIO	5
COUPE	PRATA	RONYCAR	2
MINI VAN	AZUL	MAICAR	6
MINI VAN	AZUL	RECREIO	6
MINI VAN	AZUL	RONYCAR	2
MINI VAN	BRANCO	MAICAR	2
MINI VAN	BRANCO	RECREIO	4
MINI VAN	BRANCO	RONYCAR	3
MINI VAN	PRATA	MAICAR	3
MINI VAN	PRATA	RECREIO	5
MINI VAN	PRATA	RONYCAR	5
SEDAN	AZUL	MAICAR	6
SEDAN	AZUL	RECREIO	4
SEDAN	AZUL	RONYCAR	2
SEDAN	BRANCO	MAICAR	2
SEDAN	BRANCO	RECREIO	2
SEDAN	BRANCO	RONYCAR	3
SEDAN	PRATA	MAICAR	1
SEDAN	PRATA	RECREIO	3
SEDAN	PRATA	RONYCAR	4

# Visão (tri)dimensional - Cubo

Quantidade de Vendas de Carros por Modelo, Cor e Revendedor  
no mês Janeiro/2014



- O cubo é, de fato, apenas uma metáfora visual.
- É uma representação intuitiva do fato, porque todas as dimensões coexistem para todo ponto no cubo e são independentes umas das outras.

# Visão (tri)dimensional - Cubo

Quantidade de Vendas de Carros por Modelo, Cor e Revendedor  
no mês Janeiro/2014

Soma - QUANTIDADE	REVENDEDOR				
MODELO	COR	MAICAR	RECREIO	RONYCAR	Total Resultado
COUPE	AZUL	2	3	2	7
	BRANCO	4	5	1	10
	PRATA	7	5	2	14
MINI VAN	AZUL	6	6	2	14
	BRANCO	2	4	3	9
	PRATA	3	5	5	13
SEDAN	AZUL	6	4	2	12
	BRANCO	2	2	3	7
	PRATA	1	3	4	8
<b>Total Resultado</b>		<b>33</b>	<b>37</b>	<b>24</b>	<b>94</b>

# Mais colunas de dimensões ...

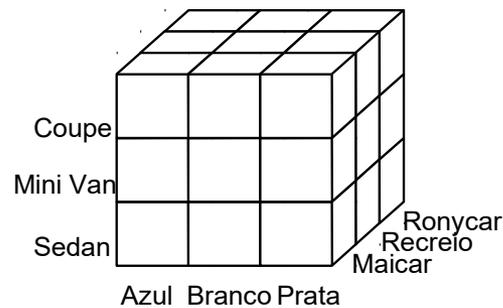
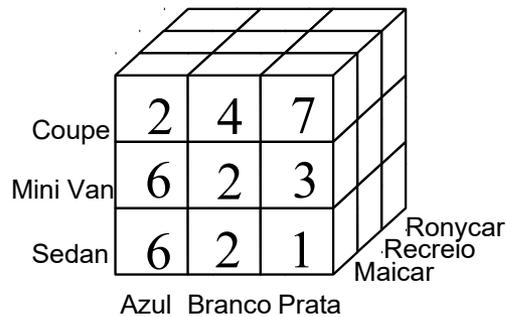
**Quantidade de Vendas de Carros  
por Modelo, Cor, Revendedor e Mês**

Vide Planilha [PivotTable.ods](#)

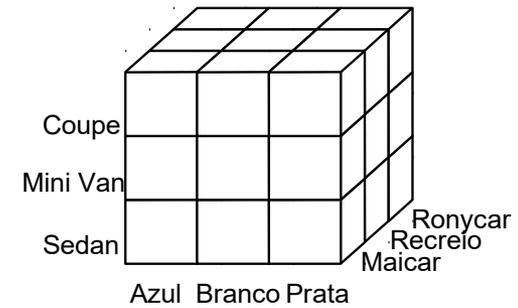
# Mais de Três Dimensões? Hipercubos

Quantidade de Vendas de Carros  
por Modelo, Cor, Revendedor e Mês

**M  
O  
D  
E  
L  
O**



**COR**



**COR**

**Janeiro**

**Fevereiro**

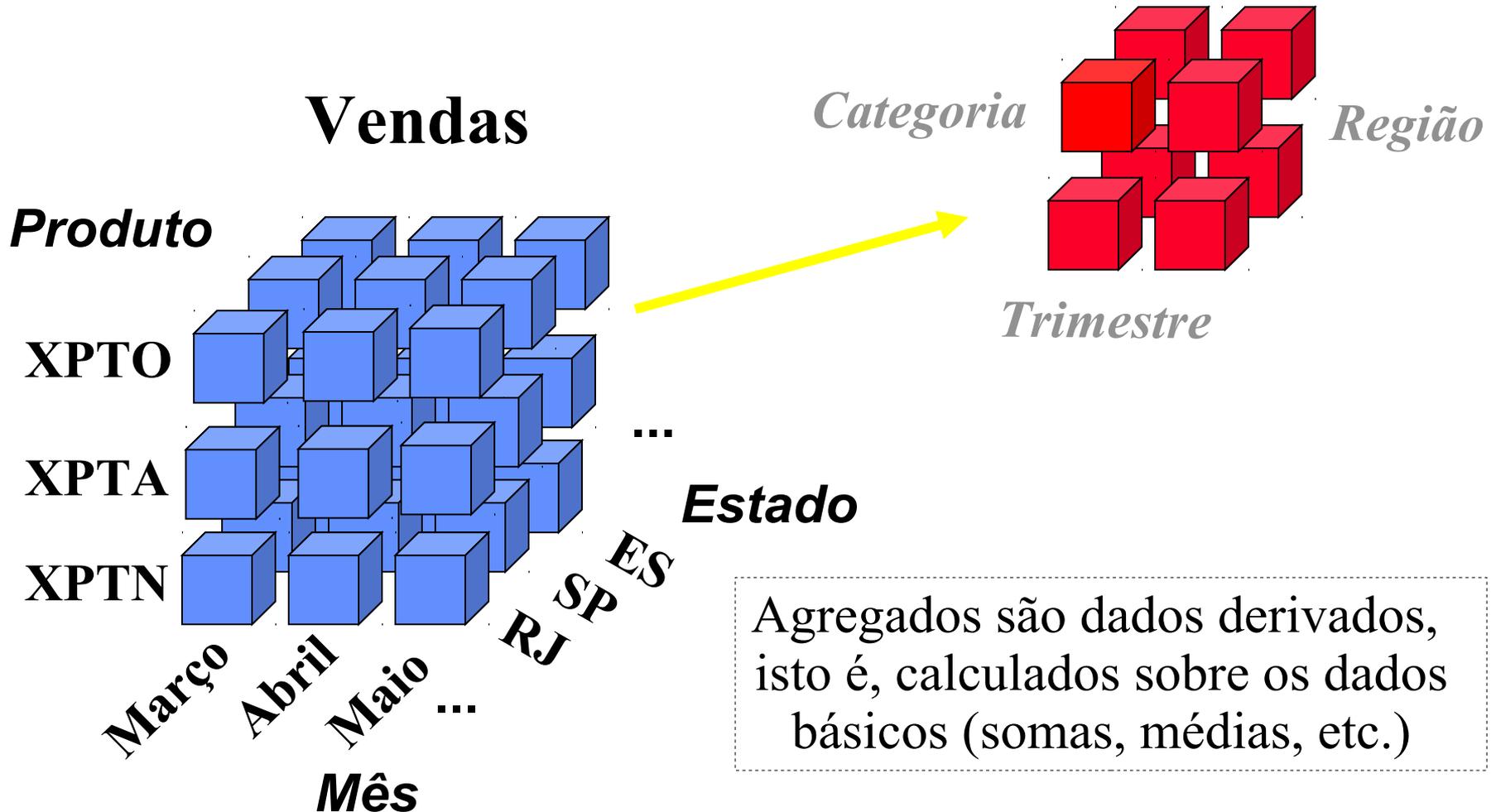
**Março**

# Níveis nas dimensões ou Hierarquias

Total de Vendas do Produto Alfa			Brasil									
			Sul			Sudeste				Nordeste		
			RS	SC	PR	SP	MG	RJ	ES	BA	SE	---
2013	Trimestre 1	Janeiro	1	5	5	2		5	1	---	---	---
		Fevereiro	5		9	8	5	6	4	---	---	---
		Março		2		6	9	4	5	---	---	---
	Trimestre 2	Abril	6		3	2	5			---	---	---
		Maio		3	4			3	5	---	---	---
		Junho	8	4		5	3	5		---	---	---
	Trimestre 3	Julho	9		2		4	7	4	---	---	---
		Agosto		2		6				---	---	---
		Setembro	5	7	5		6		3	---	---	---
	Trimestre 4	Outubro	2		3	7		9	6	---	---	---
		Novembro	1	8	4	8	8			---	---	---
		Dezembro	2	9		9		2	7	---	---	---
2014	Trimestre 1	Janeiro	---	---	---	---	---	---	---	---	---	
		Fevereiro	---	---	---	---	---	---	---	---	---	
		Março	---	---	---	---	---	---	---	---	---	
	---	---	---	---	---	---	---	---	---	---		

Hierarquias servem de base para as agregações (sumarizações, por exemplo, somas e médias)

# Agregados



Agregados são dados derivados, isto é, calculados sobre os dados básicos (somam, médias, etc.)

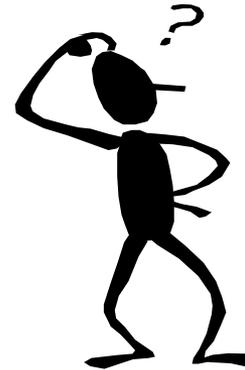
# Problema com Agregações (1)

*Calcular os agregados no momento da recuperação ou armazená-los?*

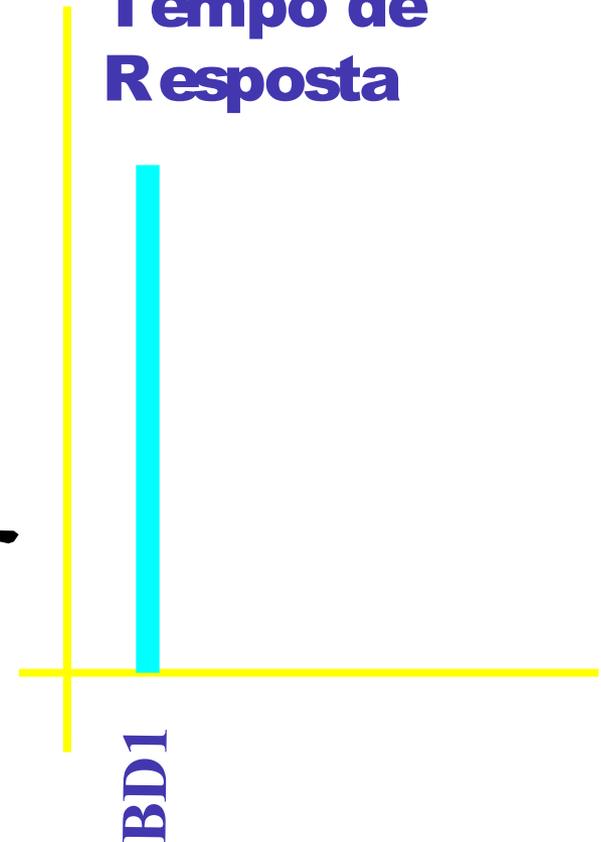
**Armazenamento**



**X**



**Tempo de Resposta**

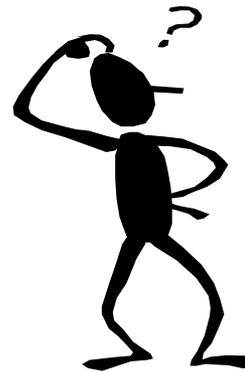


# Problema com Agregações (2)

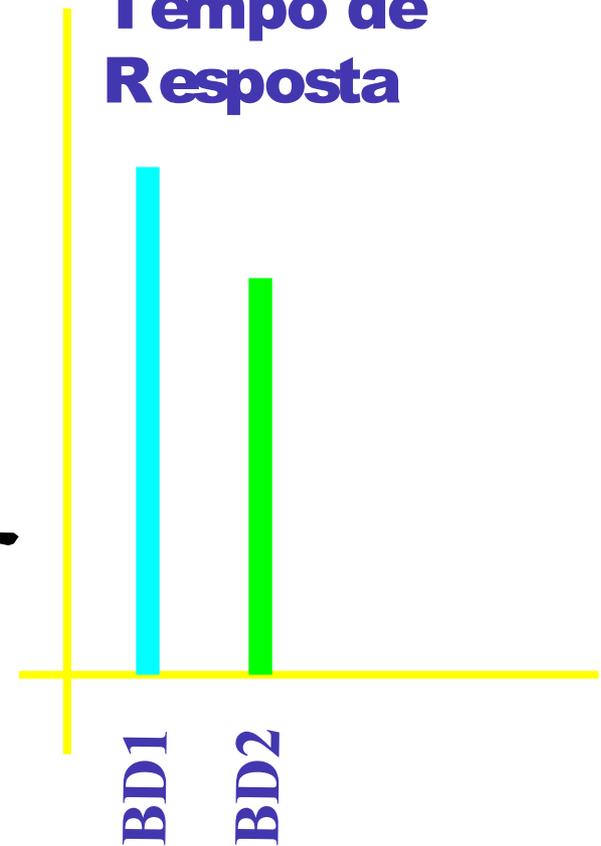
Armazenamento



X

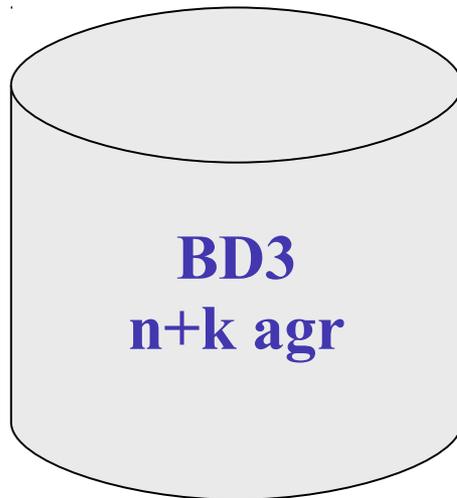


Tempo de Resposta

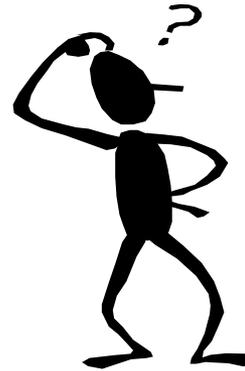


# Problema com Agregações (3)

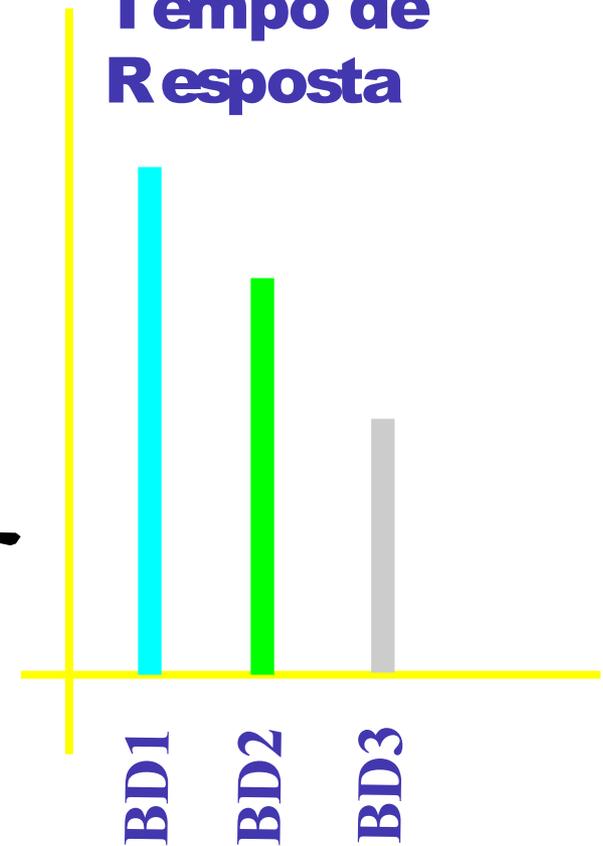
**Armazenamento**



**X**



**Tempo de Resposta**

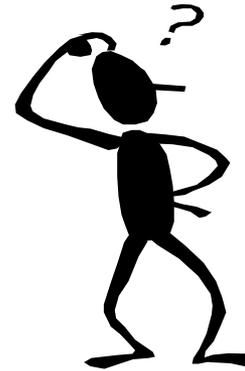


# Problema com Agregações (4)

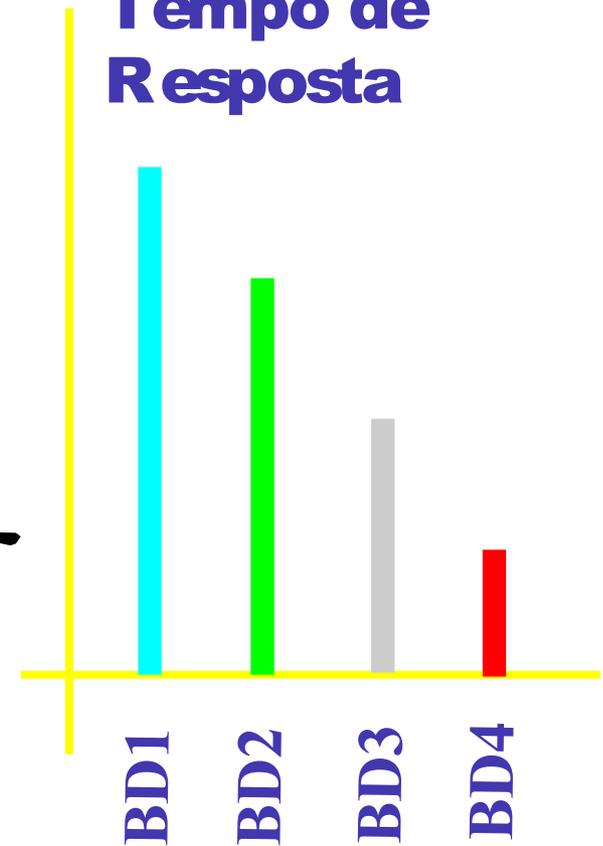
**Armazenamento**



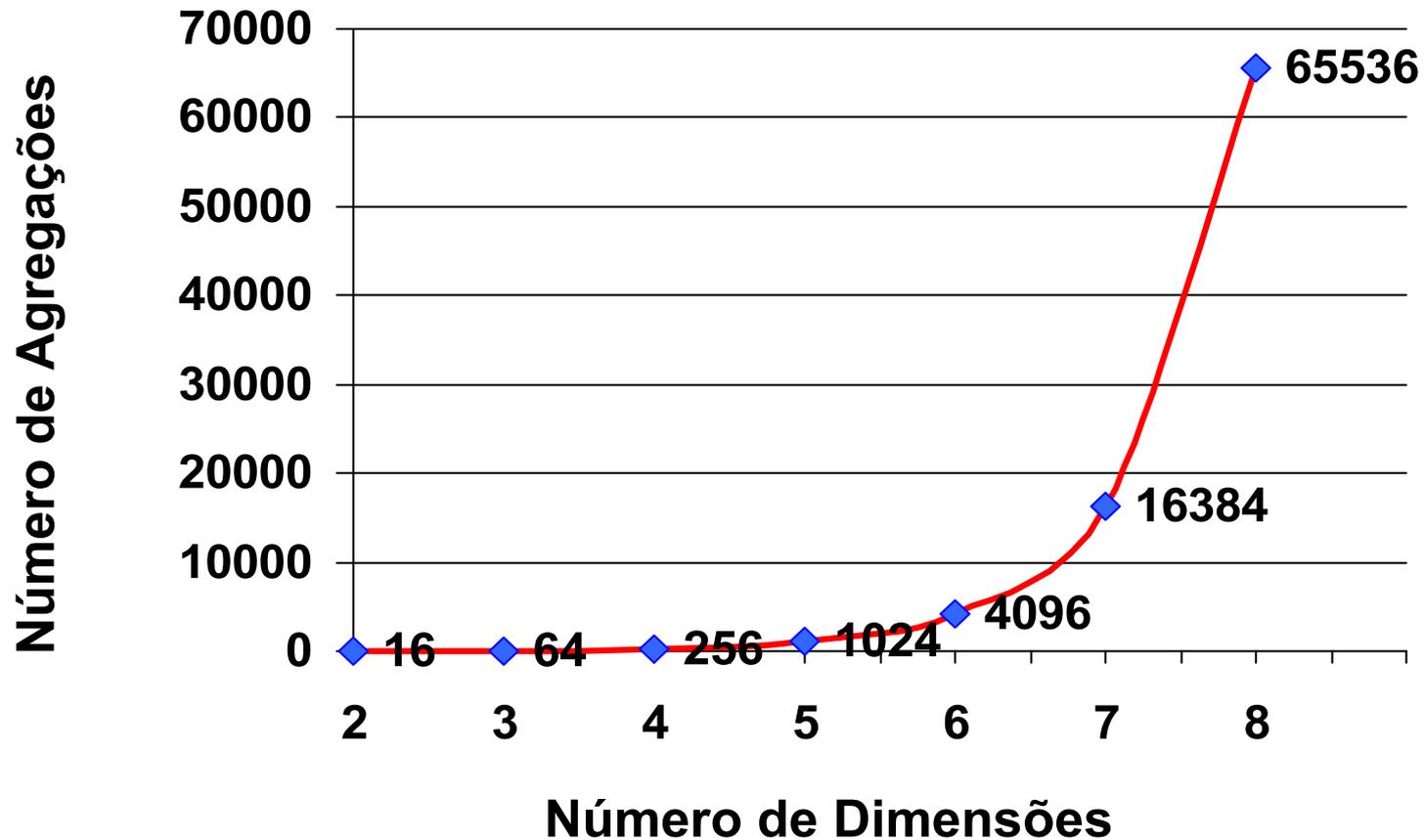
**X**



**Tempo de Resposta**

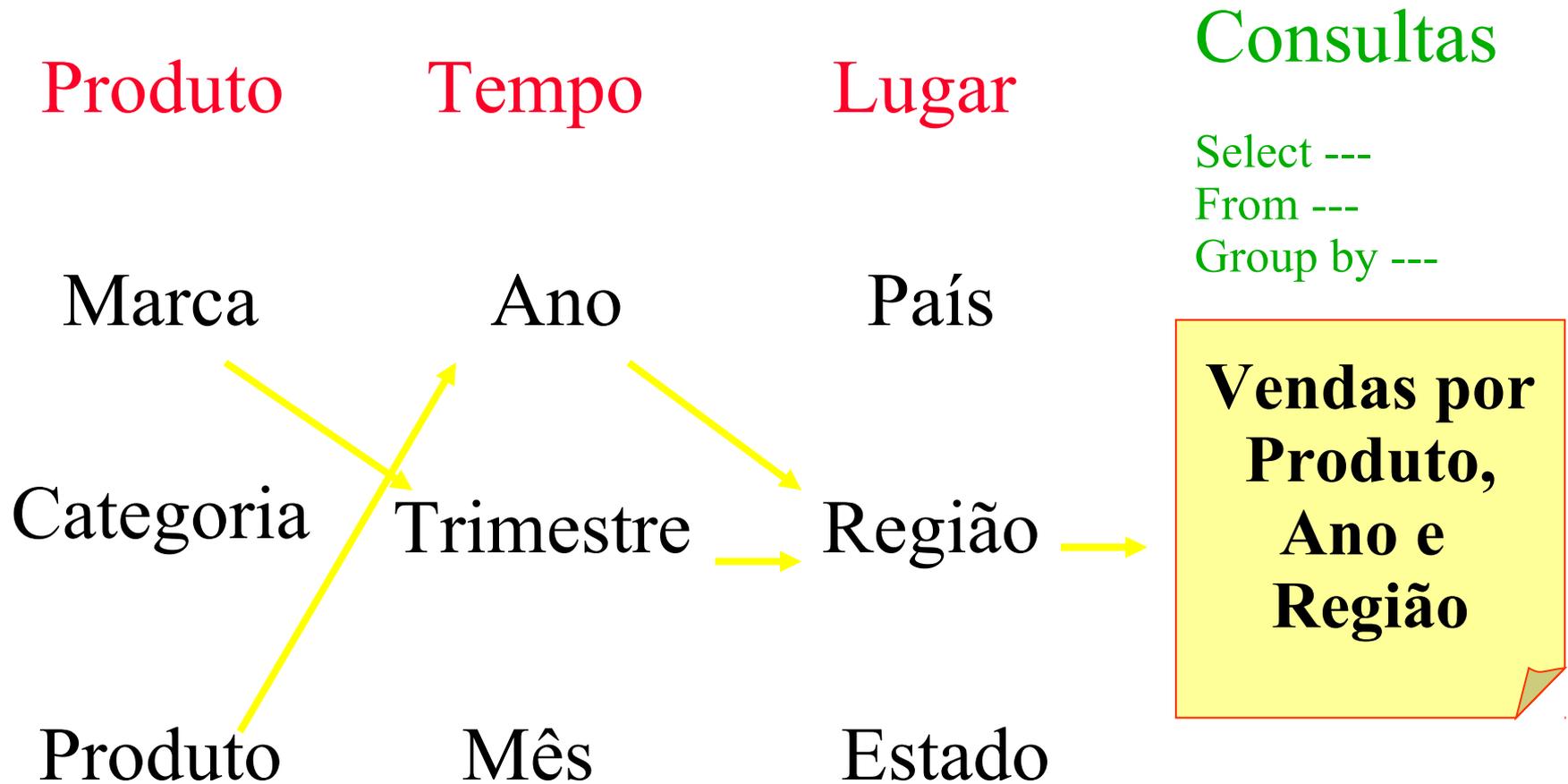


# A Síndrome da Explosão no Volume de Dados



(supondo 4 níveis em cada dimensão)

# Hierarquias e Agregados



# Ferramentas OLAP

- **OLAP: On Line Analytical Processing**
  - Conjunto de técnicas para tratar informações contidas em DW.
  - Visão Multidimensional dos Dados
- Termo proposto por **E.F. Codd**, em 1993
  - *Providing OLAP to User-Analysts: An IT Mandate.*
- **“Doze Regras de Codd” para ferramentas OLAP:**
  - Visão conceitual multidimensional
  - Transparência
  - Acessibilidade
  - Desempenho de Informações consistentes
  - Arquitetura Cliente Servidor
  - Dimensionalidade genérica
  - Manipulação de dados dinâmicos
  - Suporte a multiusuários
  - Operações ilimitadas em dimensões cruzadas
  - Manipulação intuitiva de dados
  - Flexibilidade nas consultas
  - Níveis de dimensão e agregação ilimitados

# “Doze Regras de Codd” para ferramentas OLAP

- **Visão conceitual multidimensional**
  - Os dados são modelados em diversas dimensões podendo haver cruzamento de todos os tipos de informações
- **Transparência**
  - OLAP deve atender a todas as solicitações do analista, não importando de onde os dados virão. Todas as implicações devem ser transparentes para os usuários finais.
- **Acessibilidade**
  - As ferramentas OLAP devem permitir conexão com todas as bases de dados legadas. A distribuição de informações deve ser mapeada para permitir o acesso a qualquer base.
- **Desempenho de Informações consistentes**
  - As ferramentas OLAP devem possuir conhecimento sobre todas as informações armazenadas que possa disponibilizar, sem complexidade para o usuário final, qualquer tipo de consulta.

# “Doze Regras de Codd” para ferramentas OLAP

- **Arquitetura Cliente Servidor**
  - OLAP deve ser construída em arquitetura C/S para que possa atender a qualquer usuário em qualquer ambiente operacional
- **Dimensionalidade genérica**
  - Deve ser capaz de tratar informações em qualquer quantidade de dimensões
- **Manipulação de dados dinâmicos**
  - Devido ao grande volume de informações armazenadas nas diversas dimensões de um modelo multidimensional, é comum a esparsidade dos dados, e então essas células nulas devem ser tratadas para evitar custos com memória.
- **Suporte a multiusuários**
  - Nas grandes organizações, é comum vários analistas trabalharem com a mesma massa de dados.

# “Doze Regras de Codd” para ferramentas OLAP

- **Operações ilimitadas em dimensões cruzadas**
  - As ferramentas OLAP devem ser capazes de navegar nas diversas dimensões existentes.
- **Manipulação intuitiva de dados**
  - O usuário deve ser capaz de manipular os dados livremente, sem necessitar de qualquer tipo de ajuda.
- **Flexibilidade nas consultas**
  - O usuário deve ter a flexibilidade para efetuar qualquer tipo de consulta.
- **Níveis de dimensão e agregação ilimitados**
  - Devido às várias dimensões existentes, deve haver vários níveis de agregação dos dados.

# Operações OLAP

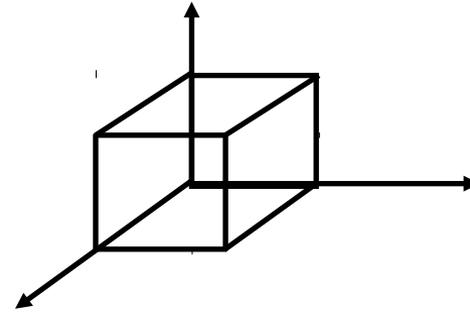
- Ferramentas OLAP fornecem suporte para funções de análise de dados, típicas de aplicações avançadas de planilhas eletrônicas.
- Operações dimensionais de ferramentas OLAP:
  - *Slice and Dice (Ponto, Plano, Cubo)*
  - *Rotation (Rotação ou Pivotamento)*
  - *Drilling*
    - » *Drill Down*
    - » *Drill Up (Roll Up)*
    - » *Drill Across (Drill Around)*
    - » *Drill Through*
  - *Ranking (Classificação por uma coluna - “Top N”)*

## Vide Wikipedia

[http://en.wikipedia.org/wiki/OLAP\\_cube](http://en.wikipedia.org/wiki/OLAP_cube)

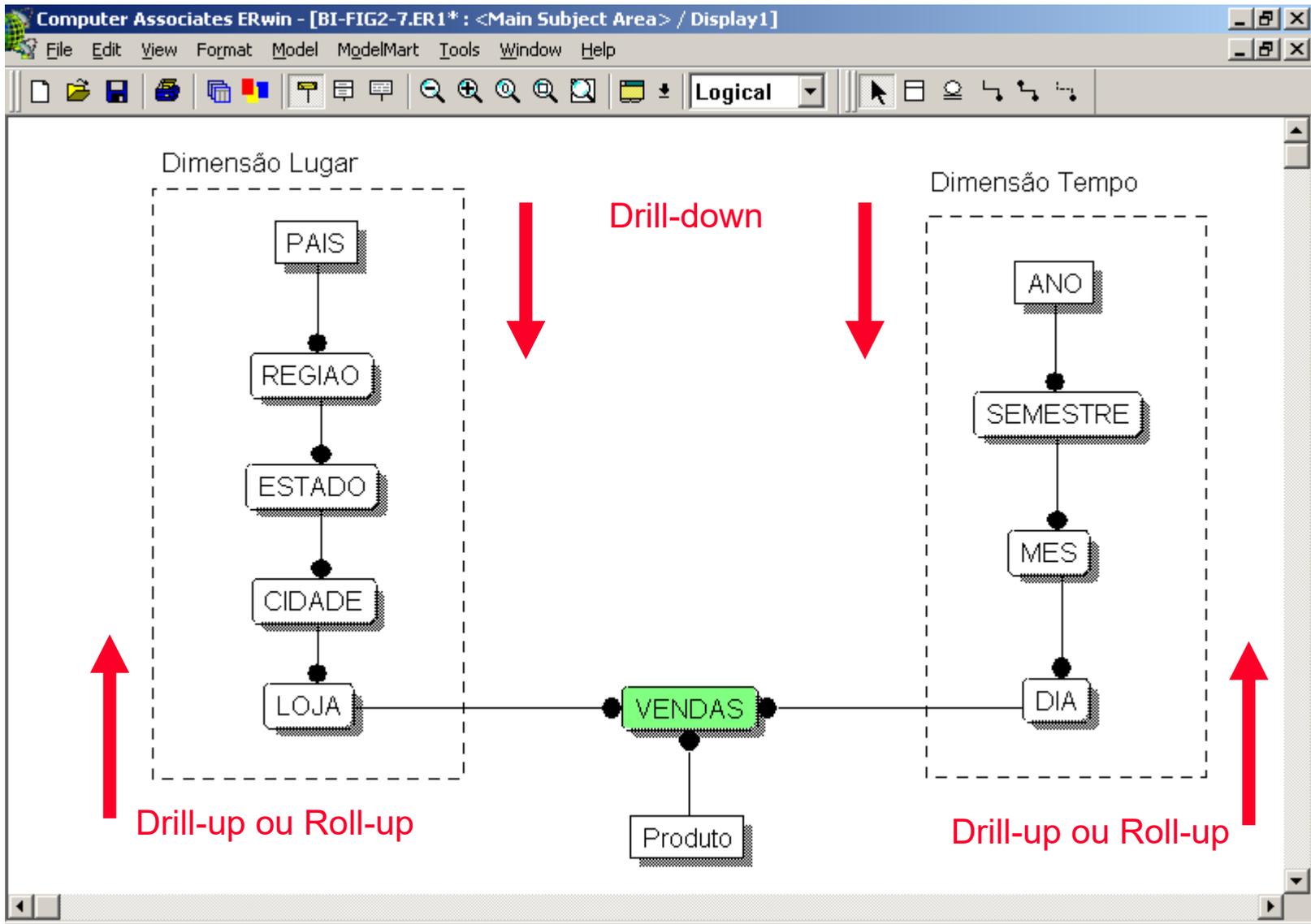
[http://en.wikipedia.org/wiki/Online\\_analytical\\_processing](http://en.wikipedia.org/wiki/Online_analytical_processing)

# Operadores Dimensionais (sobre “cubos de dados”)

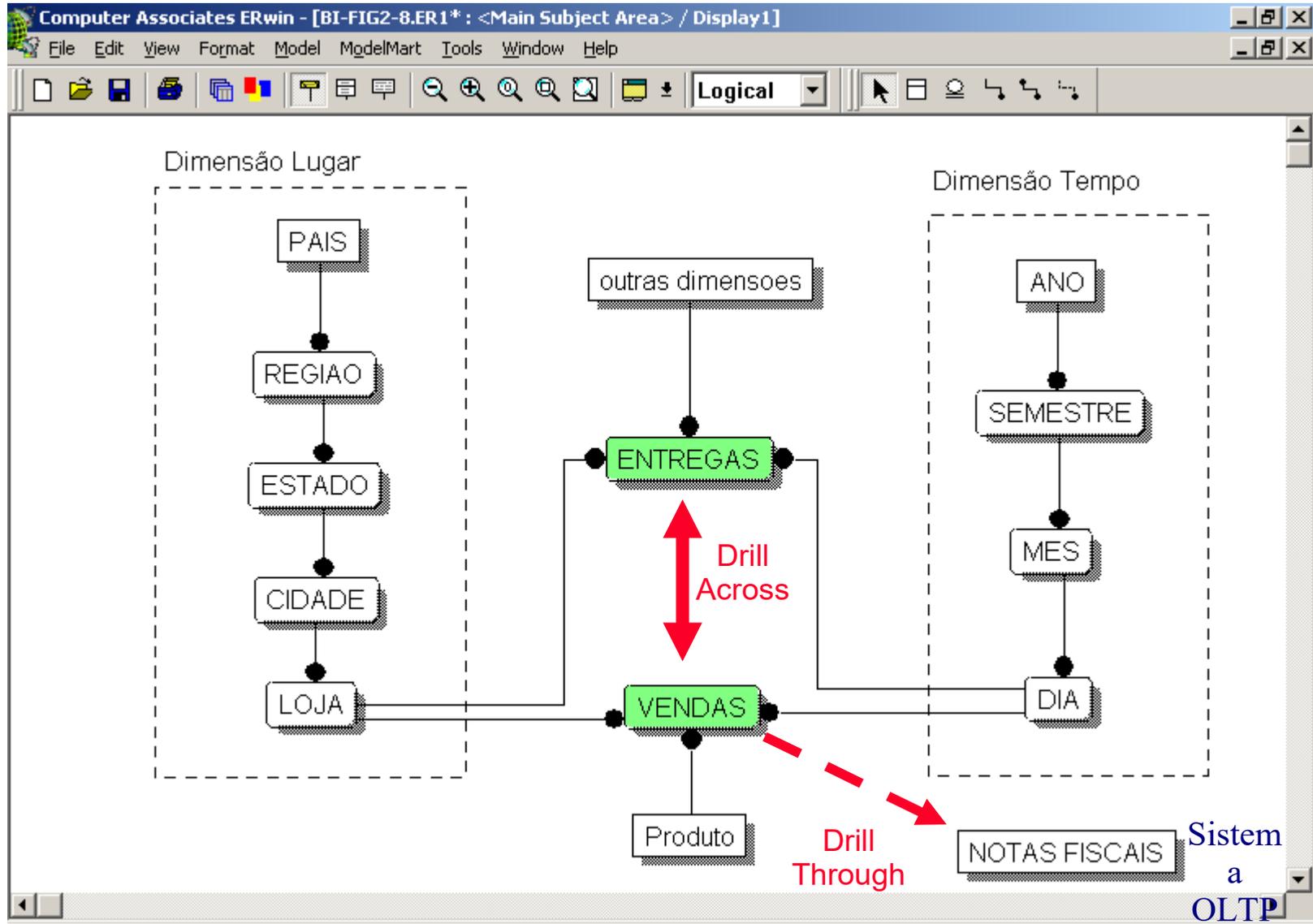


- Ponto - Valor pontual
  - Interseção de valores (Medida do Fato) com relação aos eixos (Dimensões)
- Plano – Slicing
  - Duas dimensões variando com outras fixas.
- Cubo – Dicing
  - Todas as dimensões variando (“subcubo”)
- Rotação – Pivotamento
  - Mudança dos eixos das dimensões, para fins de visualização
- Operações sobre hierarquias de dimensões
  - Drill down: “descer a hierarquia”, “desagregar”, “dar mais detalhes”
  - Drill up ou Roll up: “subir a hierarquia”, “agregar”
  - Drill across, drill around: navegar entre fatos (cubos) diferentes
  - Drill through: navegar entre sistemas diferentes (OLAP-OLTP)

# Operadores Drill Down e Drill Up (Roll Up)



# Operadores Drill Across e Drill Through



<http://www.kimballgroup.com/1996/03/01/drilling-down-up-and-across/>

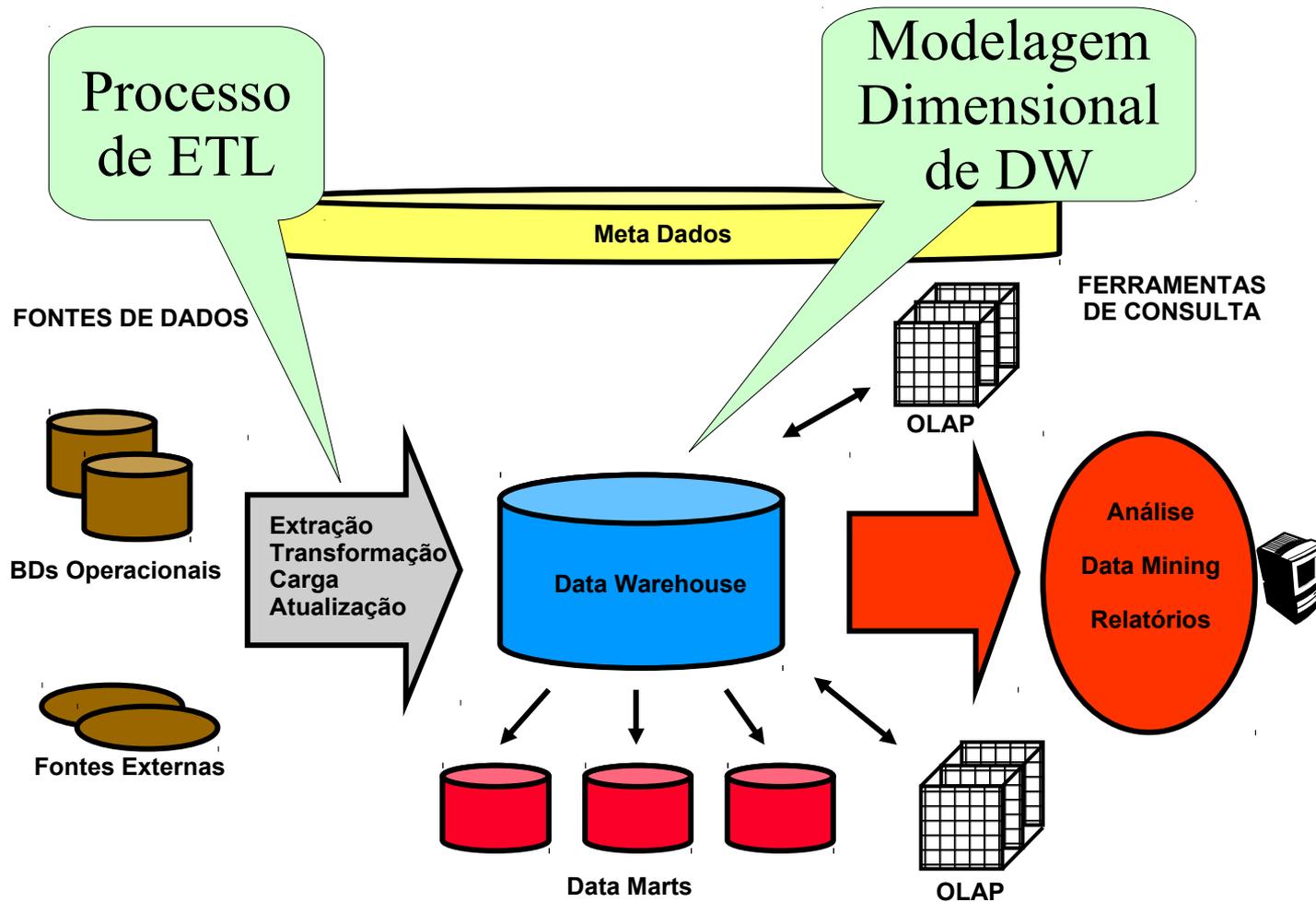
# Tipos de ferramentas OLAP

- **OLAP Multidimensional (MOLAP)**
  - Utilizam estrutura de dados multidimensional e permitem a navegação pelos níveis de detalhamento em tempo real.
  - O BD e o SGBD são multidimensionais
  - Estrutura de dados é um array com um número fixo de dimensões. O (hiper)cubo é uma metáfora visual, onde as dimensões coexistem para todo ponto e são independentes entre si.
- **OLAP RELACIONAL (ROLAP)**
  - Decorrência do uso consagrado de SGBDs relacionais nos BDs operacionais (transacionais), com as vantagens da tecnologia aberta e padronizada (SQL).
  - Utiliza os metadados no apoio à descrição do modelo de dados e na construção de consultas. Através de uma camada semântica acima do esquema relacional, os dados são apresentados ao usuário com visão multidimensional.
- **OLAP HÍBRIDO (HOLAP)**
  - Tendência dos modernos SGBDs relacionais de adicionar uma arquitetura multidimensional para prover facilidades a ambientes de suporte a decisão.
  - Proporciona o desempenho e flexibilidade de um BD multidimensional e mantém a gerenciabilidade, escalabilidade, confiabilidade e acessibilidade conquistadas pelos BDs relacionais.

[http://en.wikipedia.org/wiki/Comparison\\_of\\_OLAP\\_Servers](http://en.wikipedia.org/wiki/Comparison_of_OLAP_Servers)

Product Reviews: <http://www.bi-verdict.com/the-bi-verdict-content/products-covered/>

# A seguir ???



An Overview of Data Warehousing and OLAP Technology  
Surajit Chaudhuri, Umeshwar Dayal SIGMOD Record 1997