

# ***DW e OLAP***



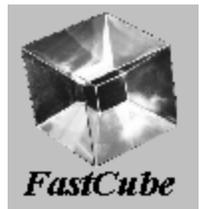
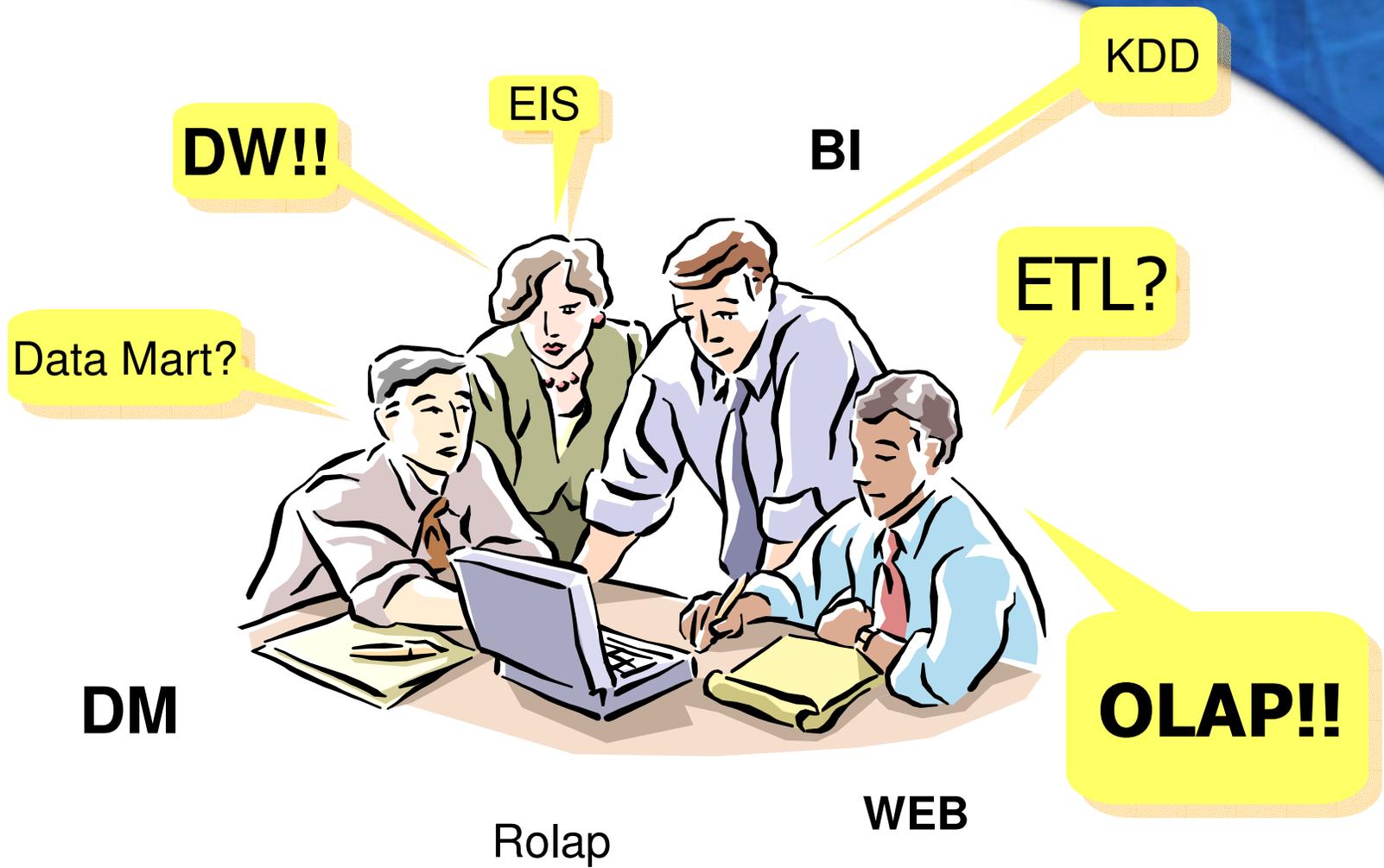


*Roberto Ângelo F. Santos*

*Universidade Federal de Pernambuco*

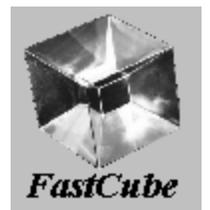
*Email: [rafs@cin.ufpe.br](mailto:rafs@cin.ufpe.br)*



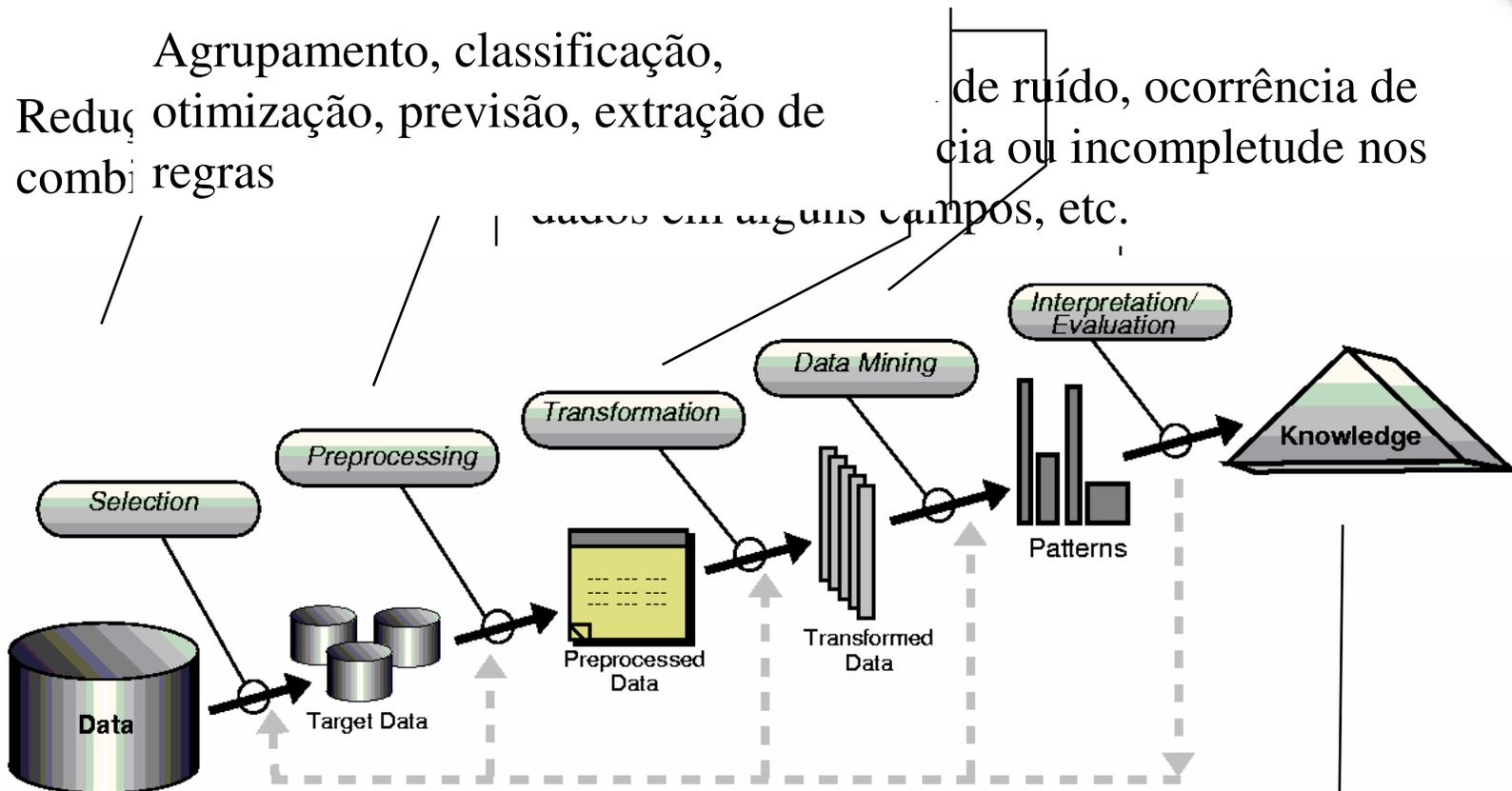


# Descoberta de Conhecimento a Partir de Bancos de Dados (KDD)

A descoberta de conhecimento a partir de bancos de dados (Knowledge Discovery from Databases- KDD) é um processo que envolve desde a preparação da base de dados até a apresentação do conhecimento deles extraído pelas técnicas de mineração.



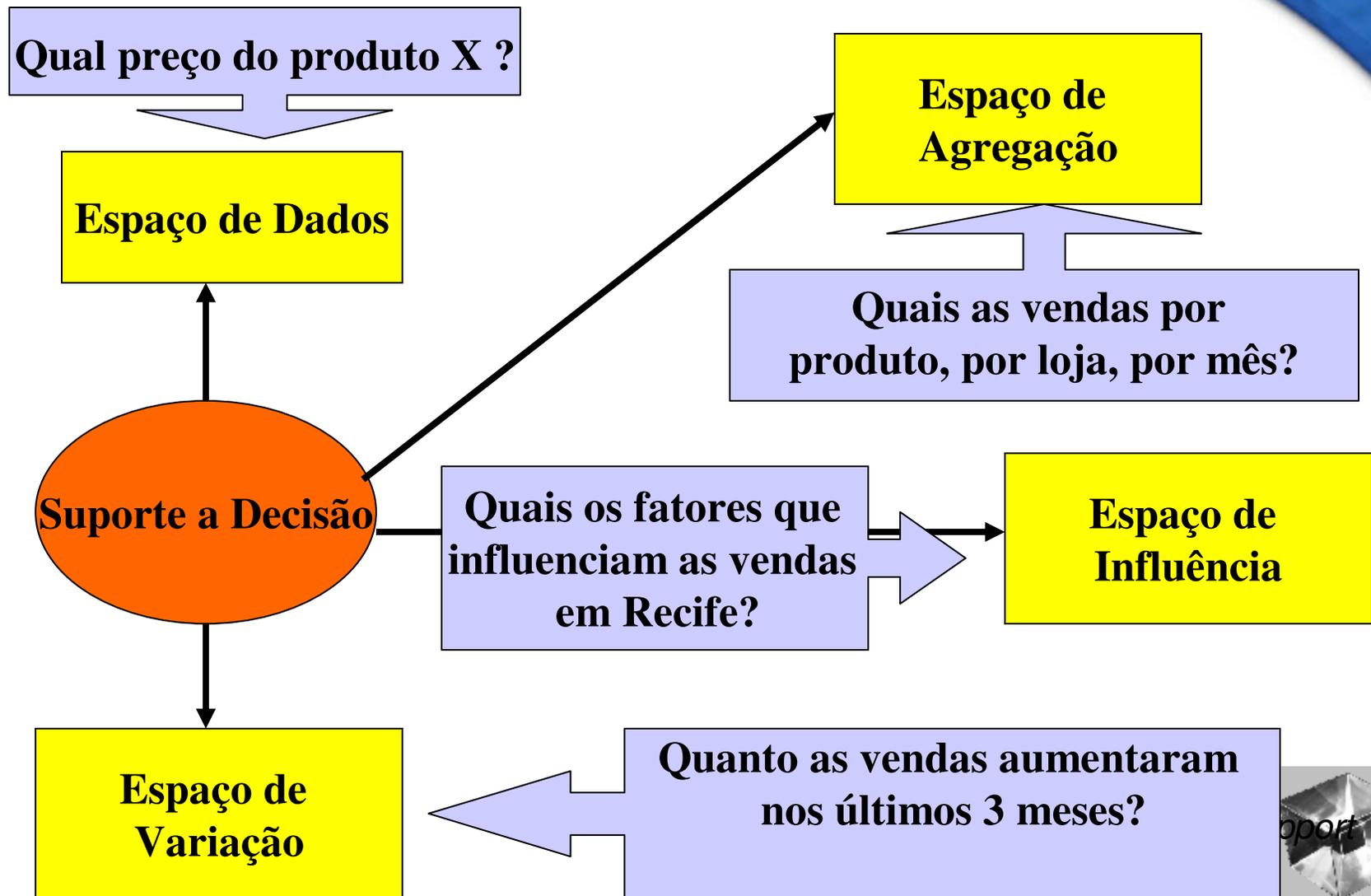
# Processo de Descoberta de Conhecimento



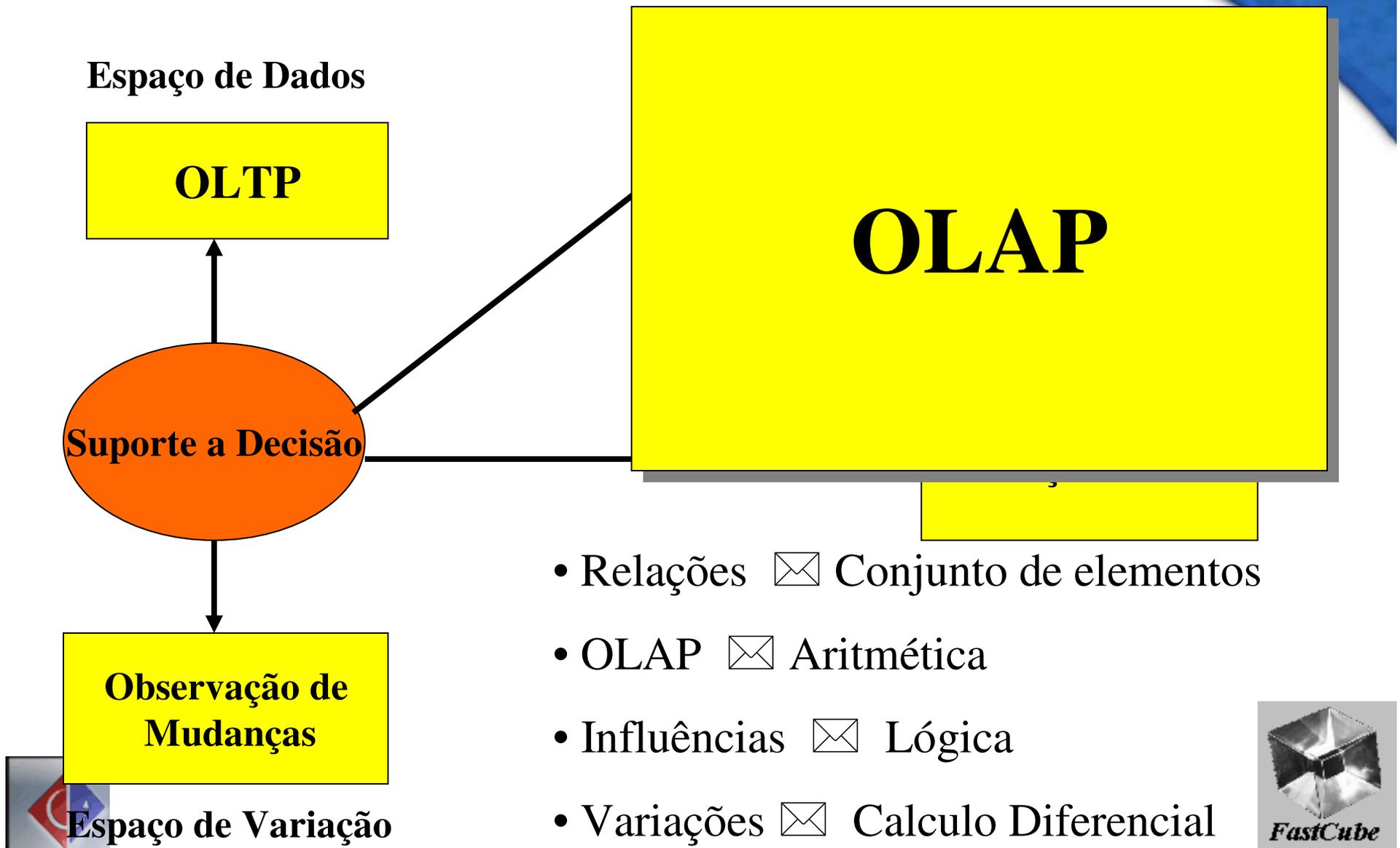
Consolidação: incorporação e documentação do conhecimento e comunicação aos interessados;



# Sistemas de Decisão: 4 Espaços

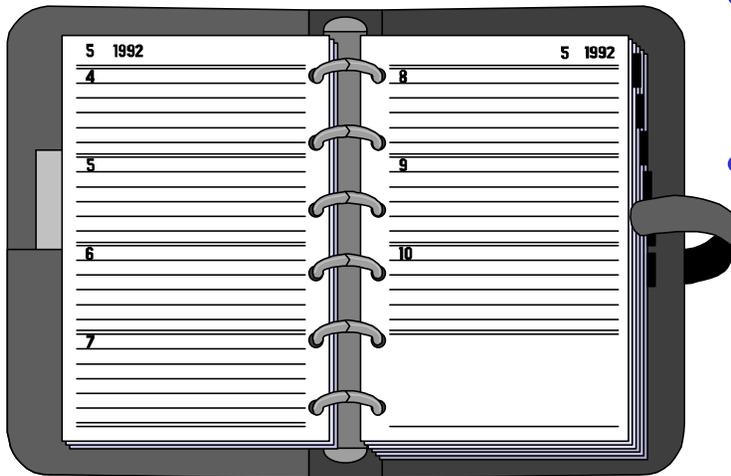


# Sistemas de Decisão: 4 Espaços



# Processamento Operacional

**OLTP (On-line Transaction Processing) =  
Processamento de Transação On-Line**

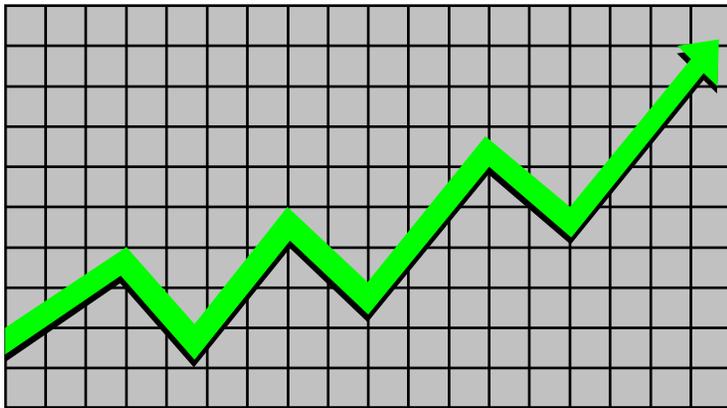


- **Sistemas que processam os negócios de rotina (Produção) das empresas**
- **Ex: Folha de Pagamento, Controle de Pedidos, Inventário, Gerenciamento Financeiro, .....**

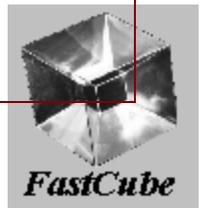


# Processamento Analítico

**OLAP ( On-Line Analytical Processing) =  
Processamento Analítico On-Line**



- **Processamento para Suportar estratégias e gerenciamento para Tomada de Decisão**



# OLAP

OLAP (On-Line Analytical Processing) representa um conjunto de tecnologias projetadas para suportar análise e consultas ad hoc.

Sistemas OLAP ajudam analistas e executivos a sintetizarem informações sobre a empresa, através de comparações, visões personalizadas, análise histórica e projeção de dados em vários cenários de **"e se ..."**.



# Dimensões

- **Visão Multidimensional:** representa a forma como os executivos, analistas de negócio e especialistas analisam as informações. Normalmente, estes usuários não usam informações específicas, mas sim, analisam um cruzamento delas.

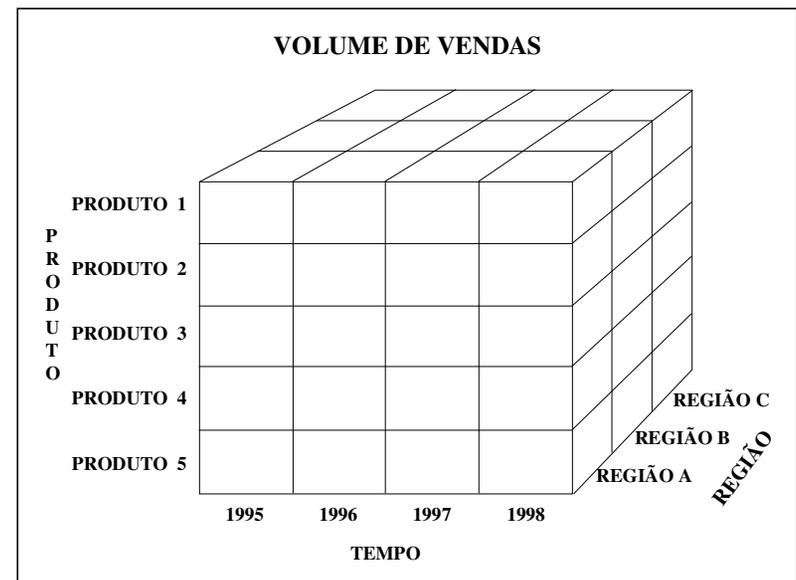
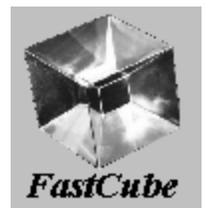


Figura 3.1 - Visão Dimensional para Análise de Volume de Vendas



# Visão Multi-dimensional

DATA MART  
VENDAS

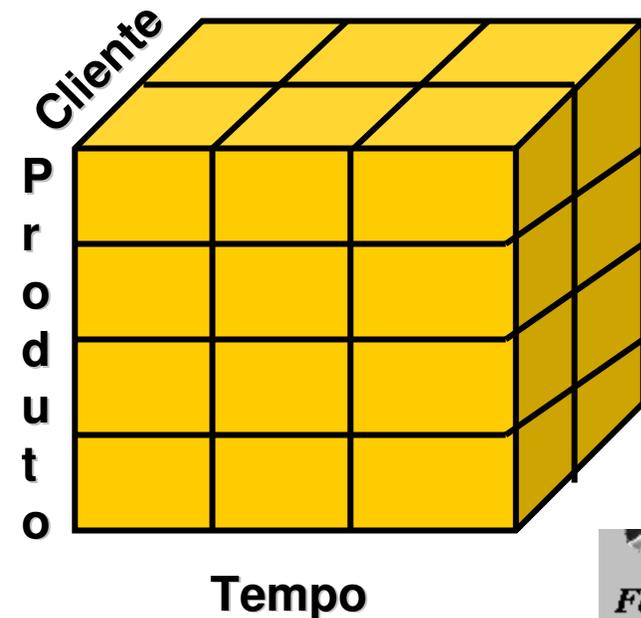
A característica principal dos sistemas OLAP é permitir uma **visão conceitual multi-dimensional** dos dados de uma empresa.

## Volume de Vendas

### 2 dimensões

P r o d u t o	Ferrari	6	5	4
	BMW	3	5	5
	Fusca	4	3	2
		Jan	Fev	Mar
		Tempo		

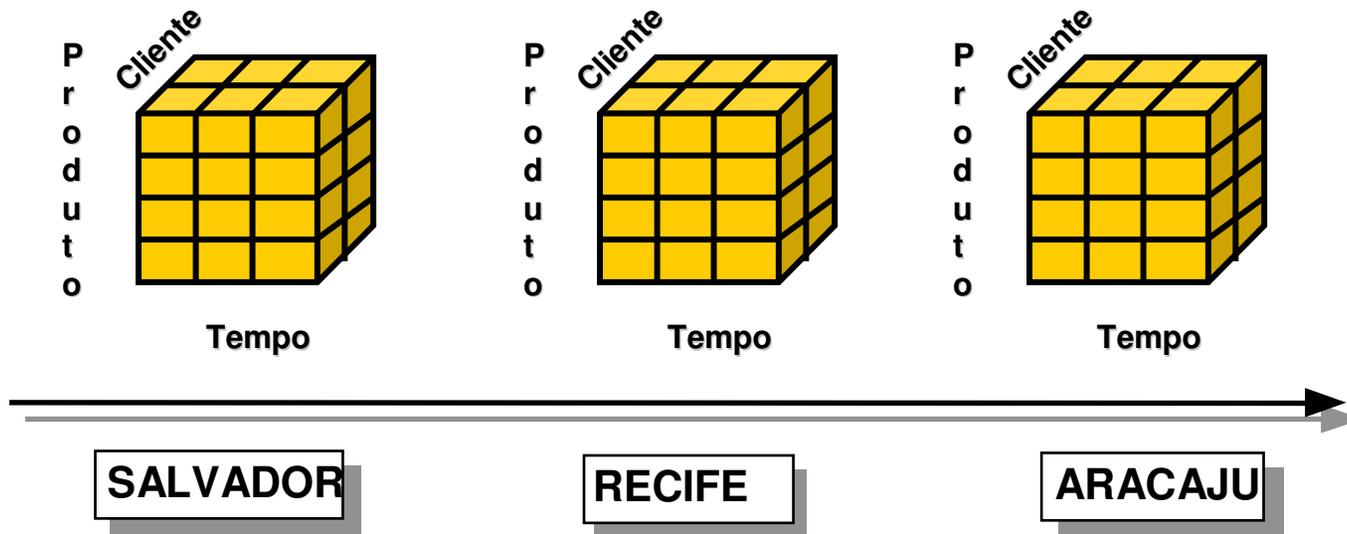
### 3 dimensões



# Adicionando Dimensões

DATA MART  
VENDAS

## Volume de Vendas



# Enfoques

- Operacional
  - Atualizável
  - Processo repetitivo
  - Suporte operacional diário
  - Ciclo de vida clássico
- DSS
  - Não pode atualizar(carga)
  - Processo heurístico
  - Suporte com necessidade gerencial
  - Ciclo de vida particular (Oposto)

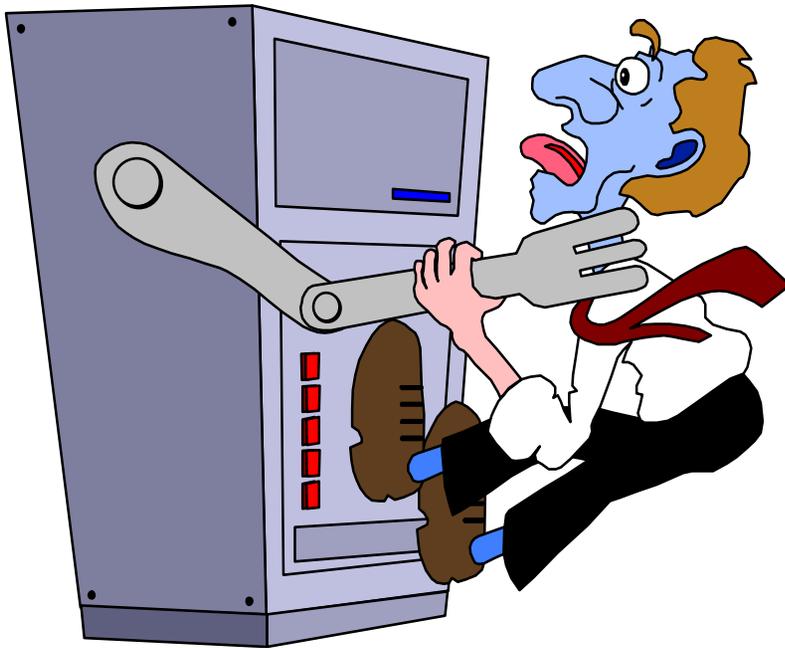


# Data Warehouse

- ***Banco de Dados especialmente modelado para a produção de informações através de cruzamentos ilimitados de variáveis e valores***
- ***Data Warehouse (armazém de dados) é uma coleção de dados, organizados por assunto, integrados, não-voláteis, históricos, cujo propósito é fornecer suporte à tomada de decisão nas organizações - by In***



## Data Warehouse (Inmon -1990)



- Banco de Dados para suporte à decisão
  - Orientado a assunto
  - Integrado
  - Variante no Tempo
  - Não Volátil

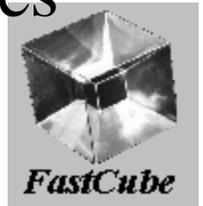


# Características dos Dados no DW

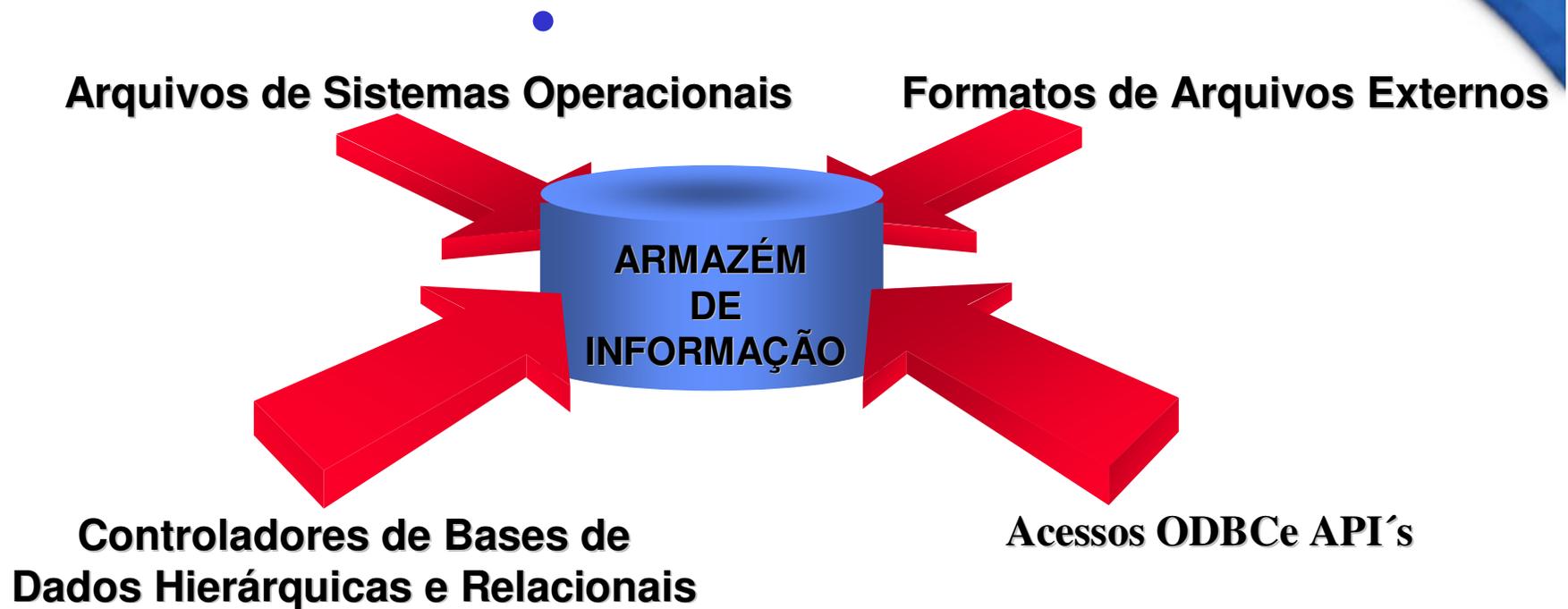


## Orientado a Assunto ou Tema

Os dados são organizados em temas ou áreas específicas, envolvendo diversos sistemas de suporte à operação, provendo informações para os tomadores de decisões

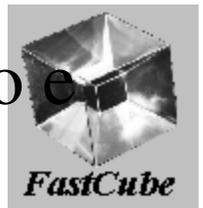


# Características dos Dados no DW

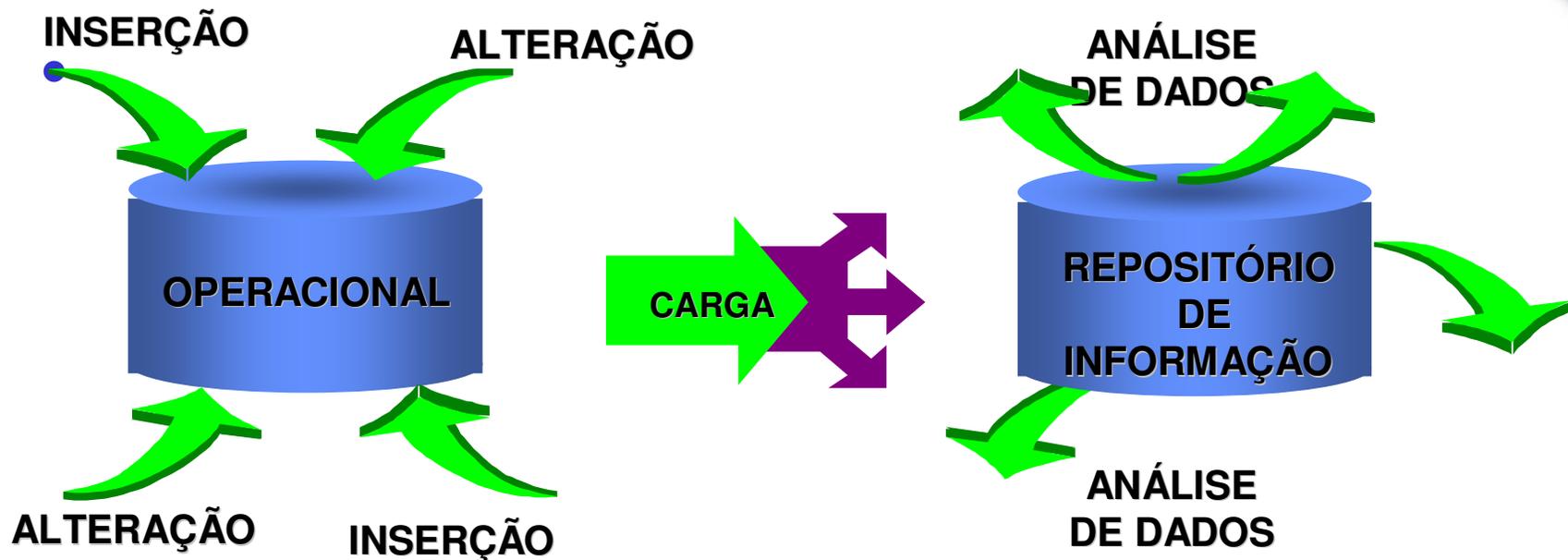


## Integrado

Os dados residentes em diferentes ambientes, podem resultar em inconsistências na forma, conteúdo e conceito



# Características dos Dados no DW

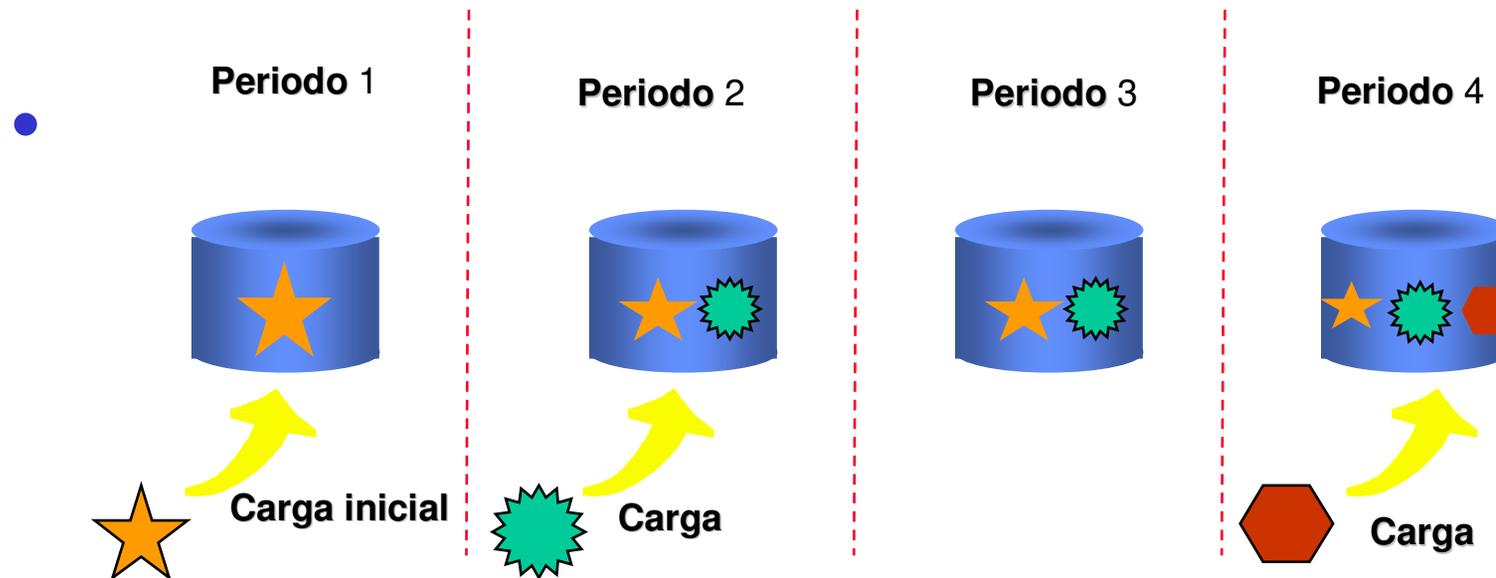


## Não Volátil

No repositório os dados são carregados em volume e não são alterados até a seguinte carga, mantendo assim a integridade das informações

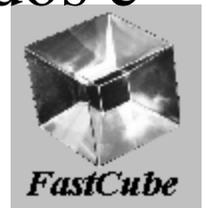


# Características dos Dados no DW



## Variável no Tempo

O repositório cria registros históricos, conservando a informação, para que esta possa ser analisada por períodos e só se adiciona informação na carga do DW

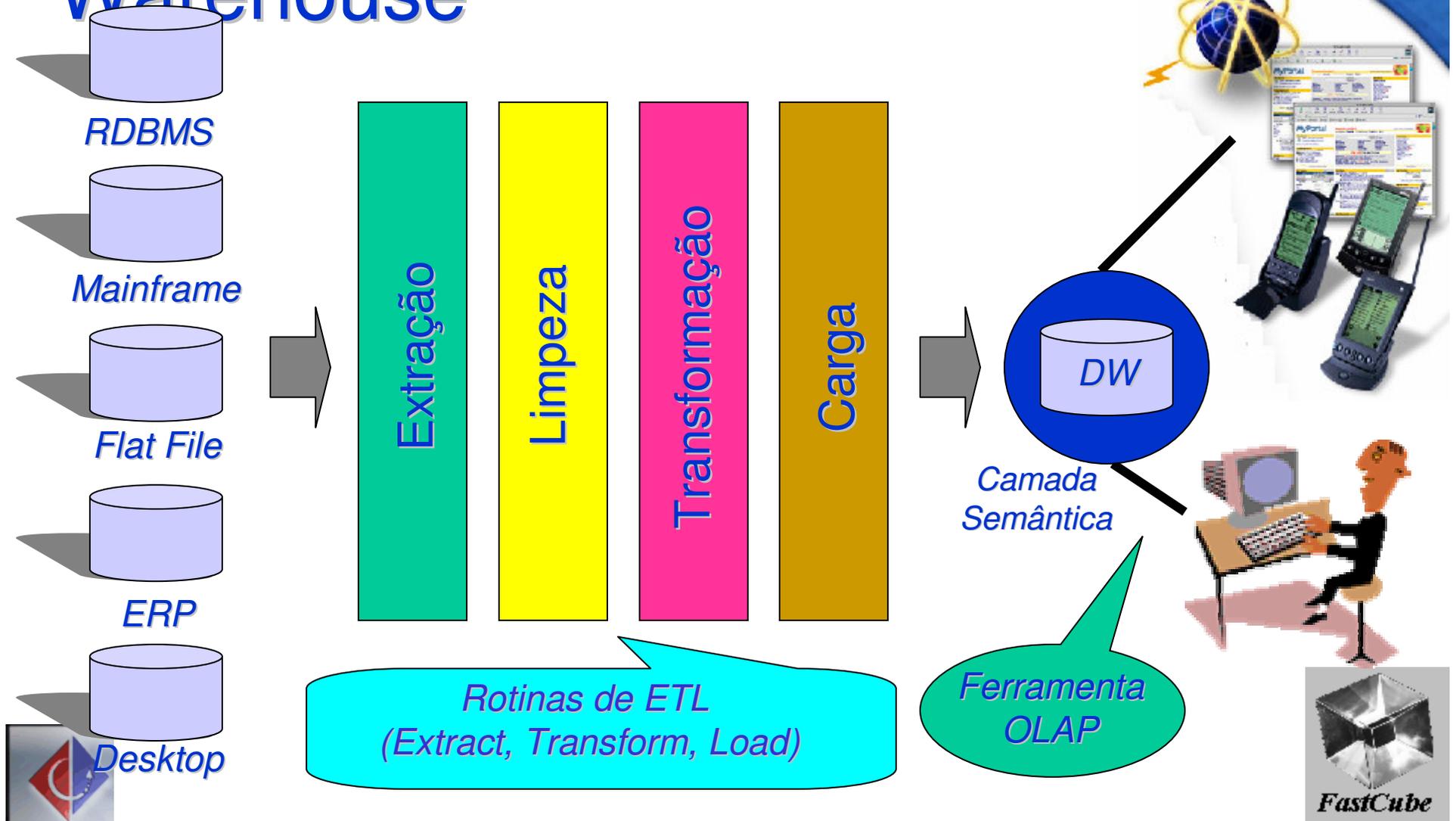


# Características de um DW

- ***Acesso Self-Service e Ad-hoc aos dados pelos Usuários***
- ***Transformações, Limpeza, Comparações e Análises Multidimensionais com os dados***
- ***Construção de Cenários diversos de Simulação***



# Construção de Data Warehouse

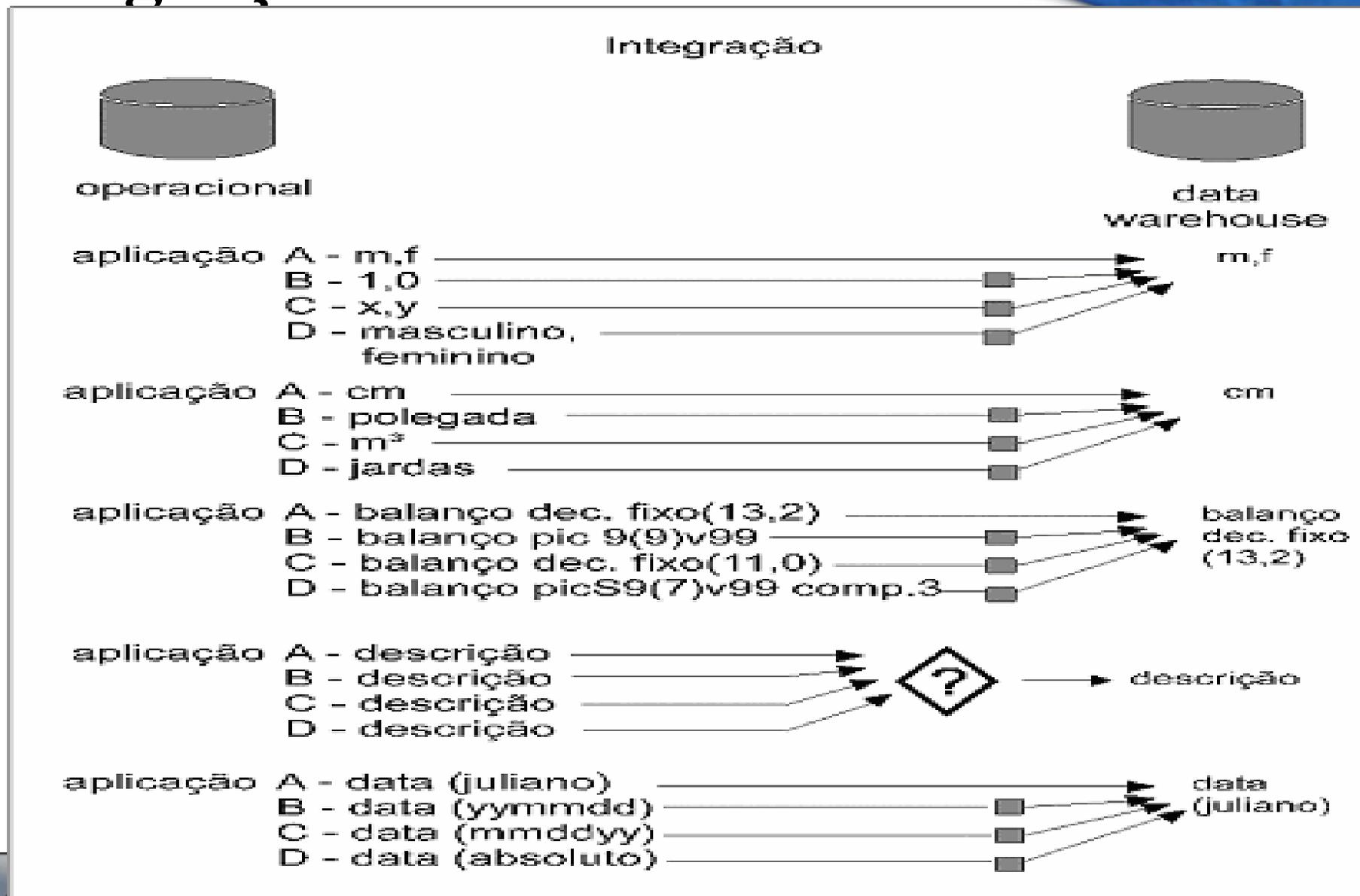


# Integração de dados

- **Objetivo:**
  - fornecer para usuário e software externo interface de consulta e manipulação de dados homogêneo
  - escondendo heterogeneidade subjacente das fontes de dados
- **Dimensões de heterogeneidade:**
  - Modelo de dados: **relacional, O-R, OO, multi-dimensional, semi-estruturado, dedutivo, temporal, ...**
  - Esquema: **relações, atributos, chaves, restrições de integridade**
  - Codificação dos valores: **unidades, nomes**
  - Linguagem de consulta e manipulação
  - SGBD
  - Sistema operacional
  - Hardware



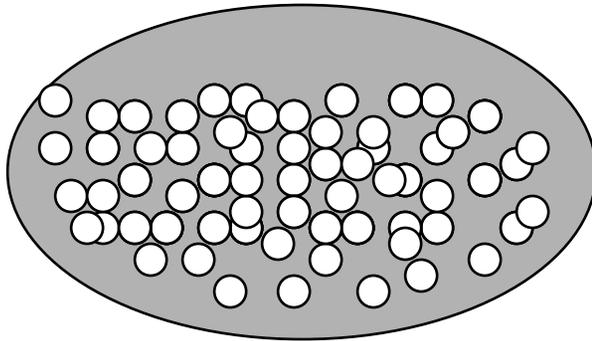
# Integração de dados



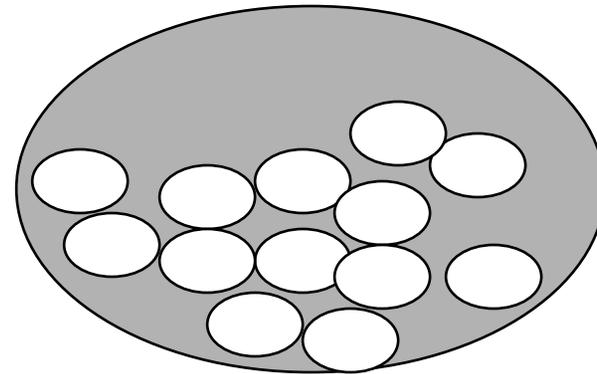
# GRANULARIDADE

**Corresponde ao nível de detalhe ou resumo dos dados em que os dados são mantidas no DW**

**Menor Granularidade  
(Maior Detalhamento)**



**Maior Granularidade  
(Dados Consolidados)**



**Faturamento  
Diário**

**Faturamento  
Mensal**

**Faturamento  
Anual**



# Canularidade

## (Conta Telefônica)



- **Detalhados**

**Maria José Nov/97**

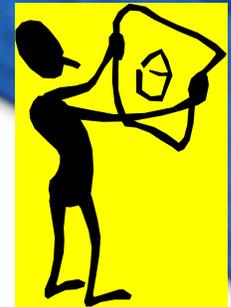
17/11/97 12:45 Recife

10/11/97 12:45 SP

27/10/97 12:45 Rio

07/11/97 12:45 Recife

01/11/97 12:45 Paris



- ◆ **Resumidos**

**Maria José Nov/97**

Resumo Mensal

Ligações: Locais 50

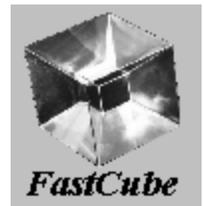
DDD 5

DDI 1

Total 56



**Quem Consome mais Memória?**



# DW - Método de Estruturação

Quanto maior o nível de detalhe menor a granularidade (vice-versa)

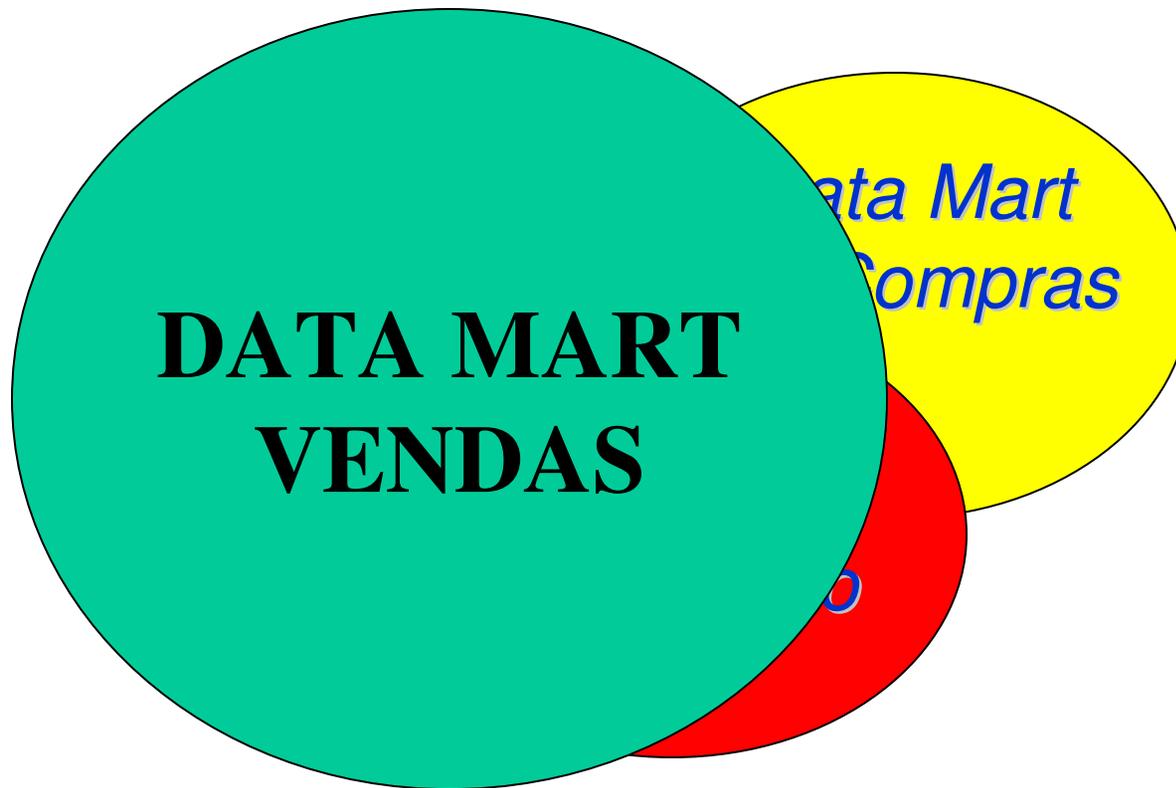
TEM GRANDE IMPACTO NAS CONSULTAS!!!

**GRANULARIDADE**

Conta Telefônica Mensal  
40.000 bytes/mês = 200 reg/mês  
Resumo Conta  
200 bytes e 1 registro por mês



# Dividir para Conquistar



# DATA MART

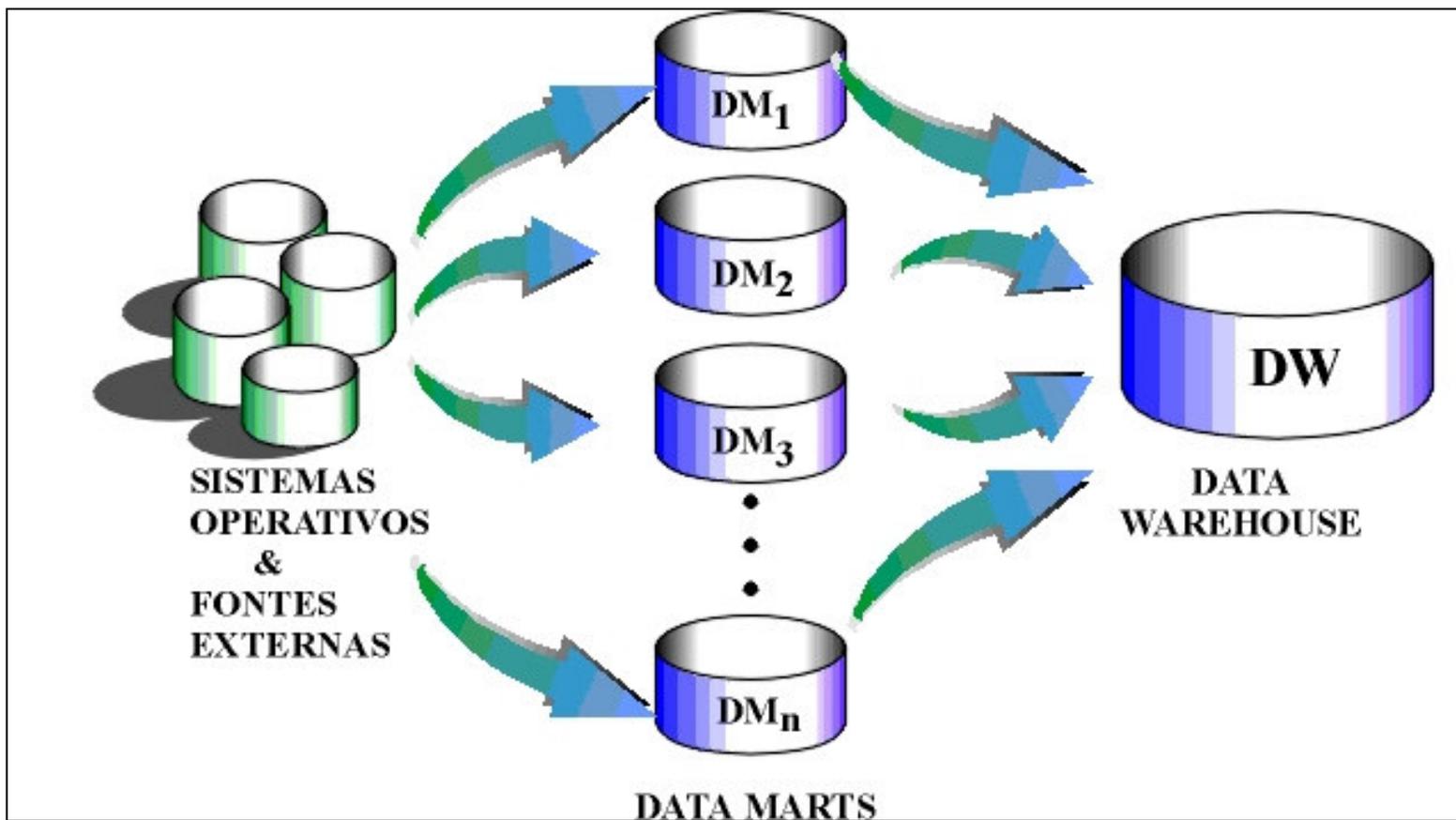
***DATAMARTS são soluções apropriadas a médias e grandes corporações. Representam DW menos complexos, orientados a determinadas áreas/assuntos, sendo por isso mais fáceis de serem gerenciados. Apresentam a desvantagem de serem descentralizados, mas se forem implantados a partir de um DW, este problema se resolve e a organização pode obter o máximo das duas modalidades.***



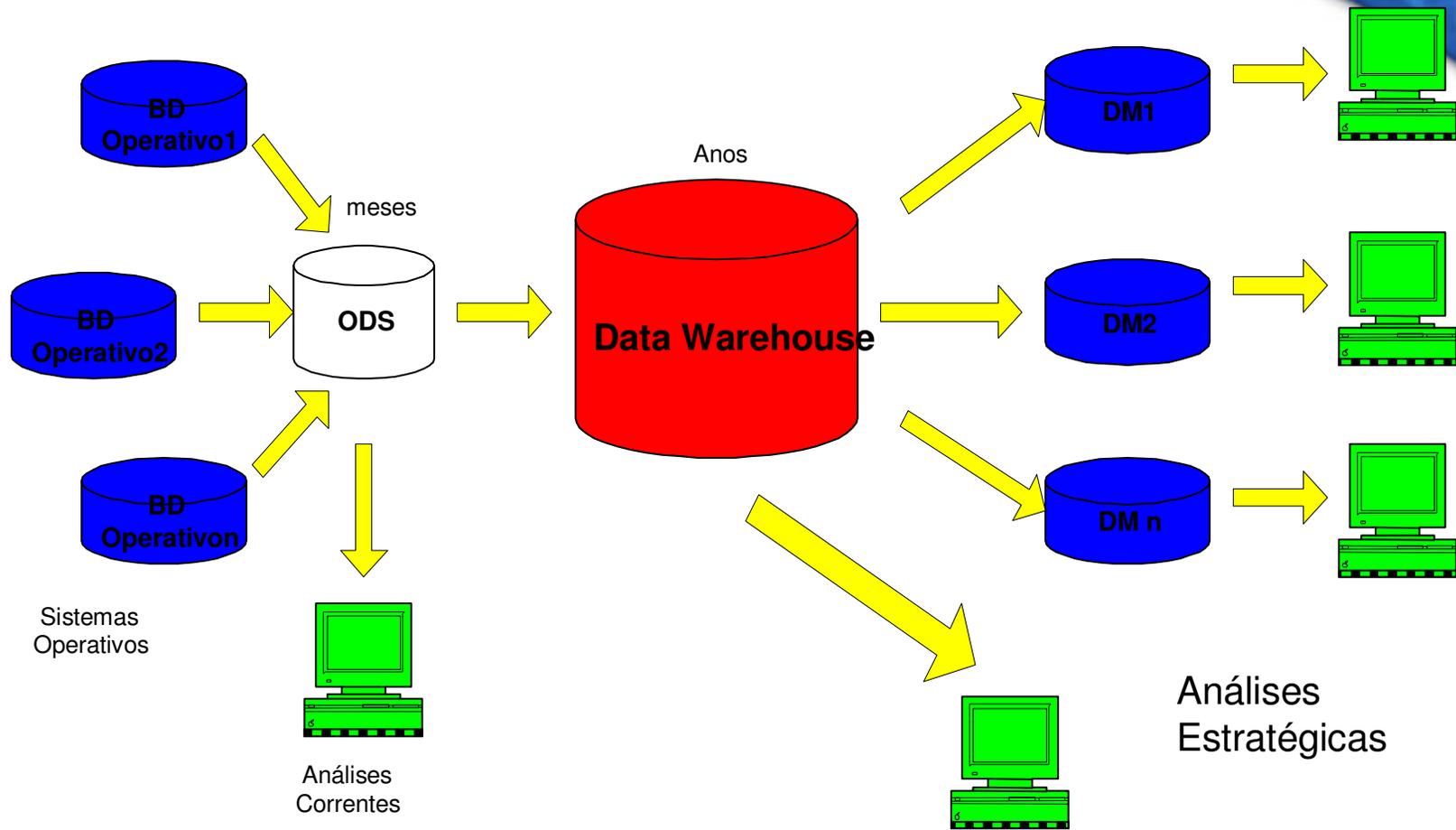
**DATA MART**



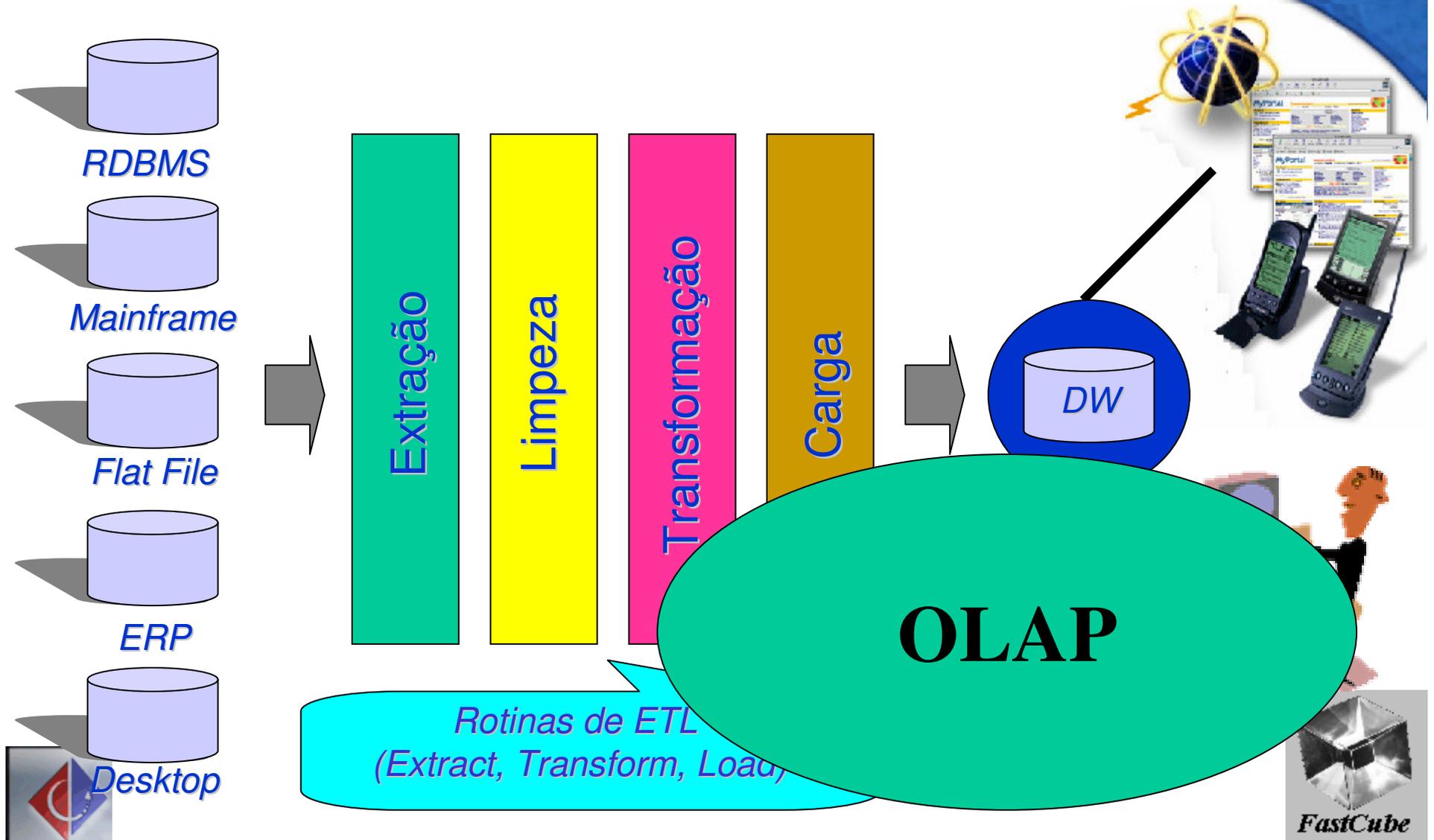
# Arquitetura Caso 1



# Arquitetura Caso 2



# OLAP sobre o DW



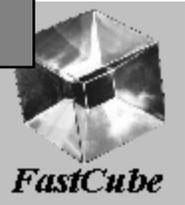
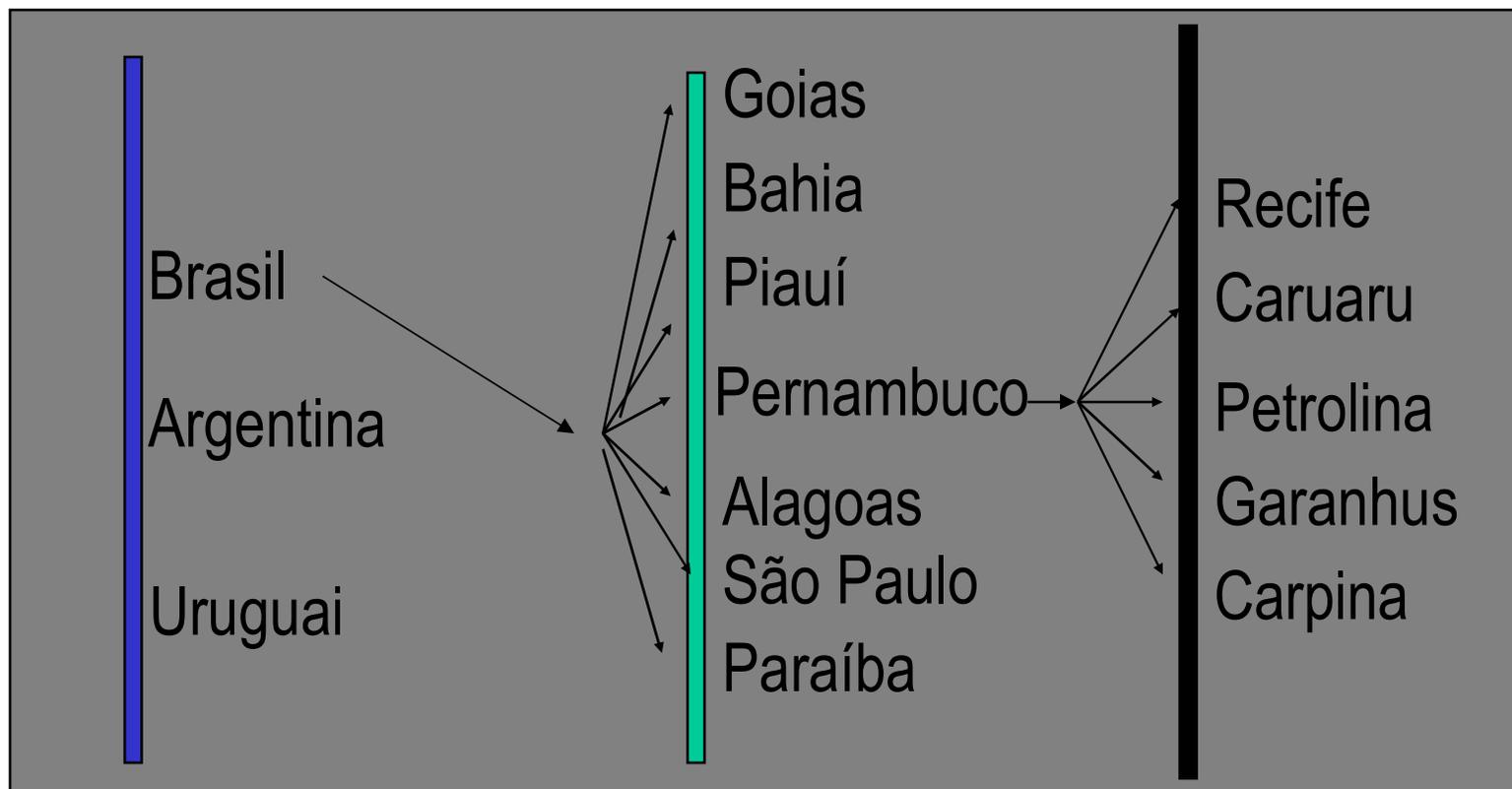
# Modelagem do DM

- É projetado para atender às necessidades de um setor ou departamento
- Requer uma boa definição do escopo
- Pode ser modelado a partir do:
  - DW: garante a consistência e o mapeamento dos dados
  - Sist. Operativos: pode gerar DM Independentes nos casos que não exista um controle e gerência de metadados
- Normalmente, se utiliza o esquema estrela para a modelagem



# Roll-Up, Drill-Down e Slice-Dice

DATA MART  
VENDAS

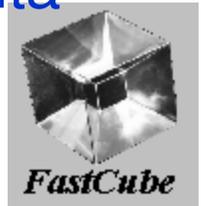


# OLAP -Tipos disponíveis

**ROLAP:** são OLAP's que buscam seus dados de um banco de dados relacional. Utilizadas para analisar um grande volume de dados através de múltiplas dimensões, sendo capazes de lidar bem com dados voláteis e com mudanças frequentes.

**MOLAP:** são OLAP's multidimensionais que utilizam uma estrutura de cubos para realizar a análise de dados. Os dados precisam ser extraídos e carregados para estruturas proprietárias do sistema.

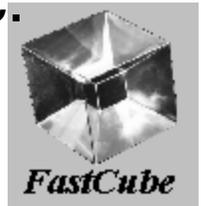
**HOLAP:** é uma OLAP híbrida que apresenta funcionalidades de MOLAP e de ROLAP.



# ROLAP- Esquema Estrela (Star Schema)

O **esquema estrela** é a representação do modelo dimensional em bancos de dados relacionais. No esquema estrela existe uma tabela dominante no centro do esquema. Esta é a única tabela com múltiplos relacionamentos para as outras tabelas. As outras tabelas possuem um único relacionamento para a tabela central. A tabela central é chamada **fato** (*fact table*) e as outras tabelas são chamadas de **dimensão** (*dimension table*).

A tabela **fato** é onde uma ou mais medidas numéricas dos negócios da empresa são armazenadas. Os fatos podem ser "vendas" , "custo" , "unidades vendidas" , etc.

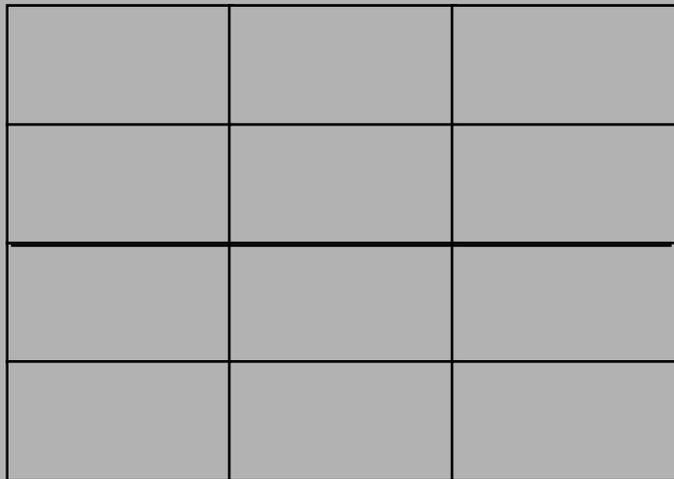


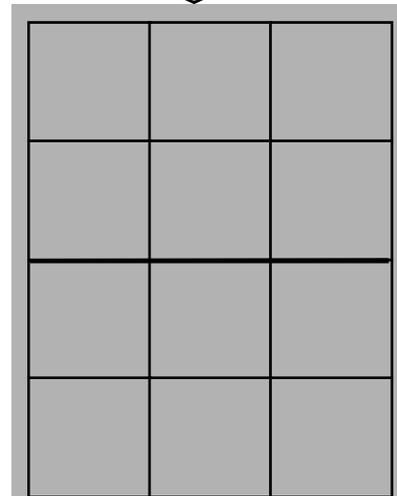
# Conceitos Básicos

**TABELAS**

**FATOS**

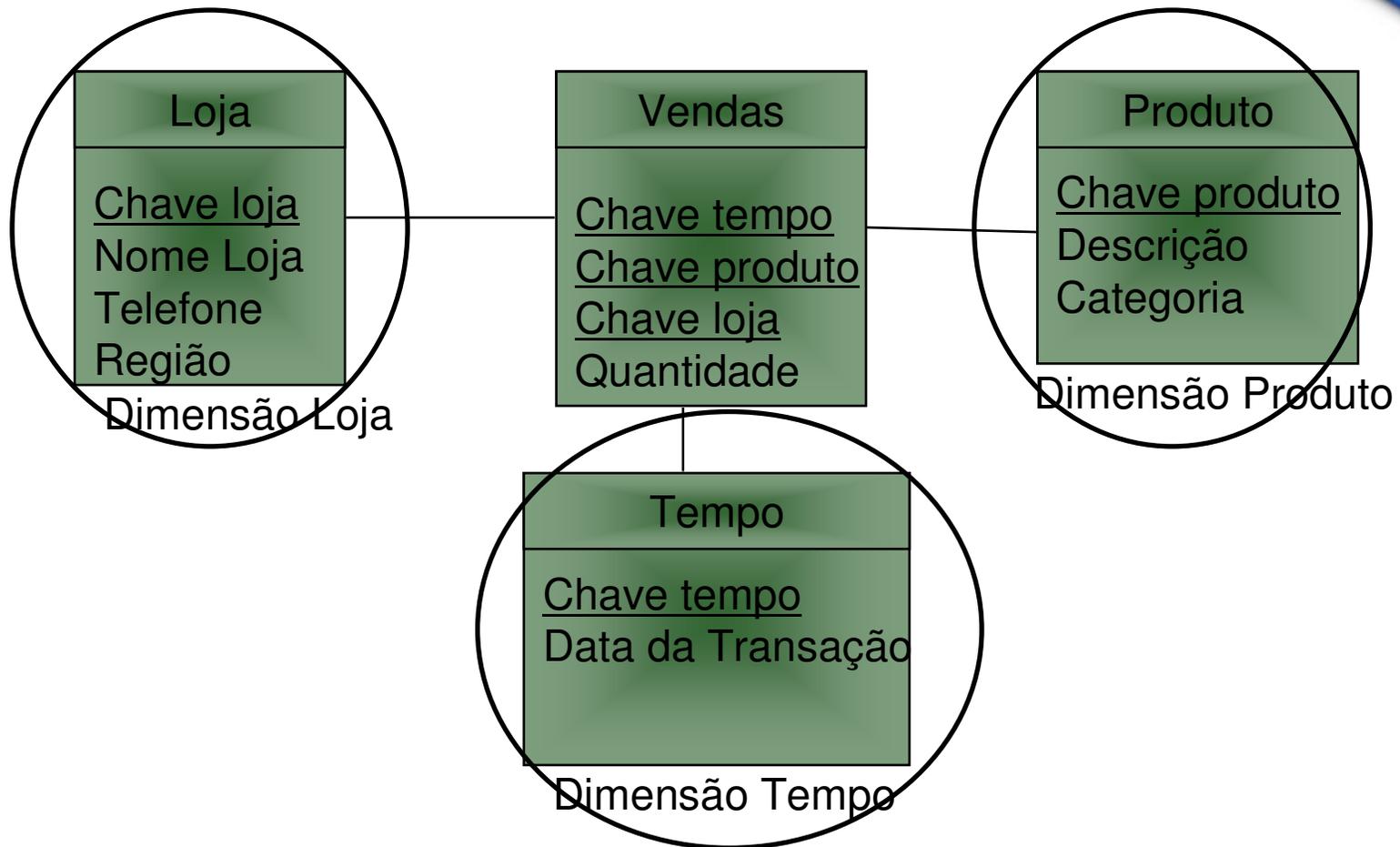
**DIMENSÃO**



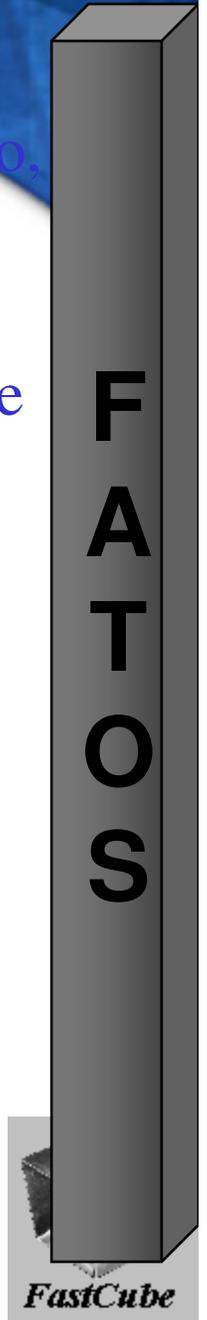



# Dimensão



# Fatos

- Contém dados factuais ou quantitativos de negócio modelado, ou seja, as informações consultadas
- As informações são frequentemente medidas numéricas (aditivas) que podem consistir de várias colunas e milhões de linhas (Longas e Finas)
- Acesso via dimensões
- O Usuário necessita fatos para tomar decisão
- Quando um dado numérico é Fato ou Atributo ?
  - Fato  $\Rightarrow$  varia continuamente a cada amostragem
  - Atributo  $\Rightarrow$  praticamente constante
- Tabelas de fatos são esparsas
- Grãos são os registros da tabela de fatos



# Dimensão

- Contém a informação (textual) usada para restringir uma consulta na tabela FATOS e ser um guia para seleção de linhas
  - Atributos  $\Rightarrow$  descrevem os itens de uma dimensão
  - Os melhores atributos (campos) são: textuais, restritivos
- Contém o tipo de informação que usuário conhece quando acessa BD
- Refletem as dimensões do negócio. Em geral são menores que a tabela de Fatos
- Muito indexadas
- Desnormalizada
- Cada descrição ajuda a definir um componente da dimensão
- Tabela de dimensão possui muitos atributos



The background is a solid blue color with a complex, abstract pattern of overlapping, semi-transparent shapes and lines. A large, white, curved shape, resembling a stylized 'C' or a partial circle, is positioned on the left side of the slide, framing the text.

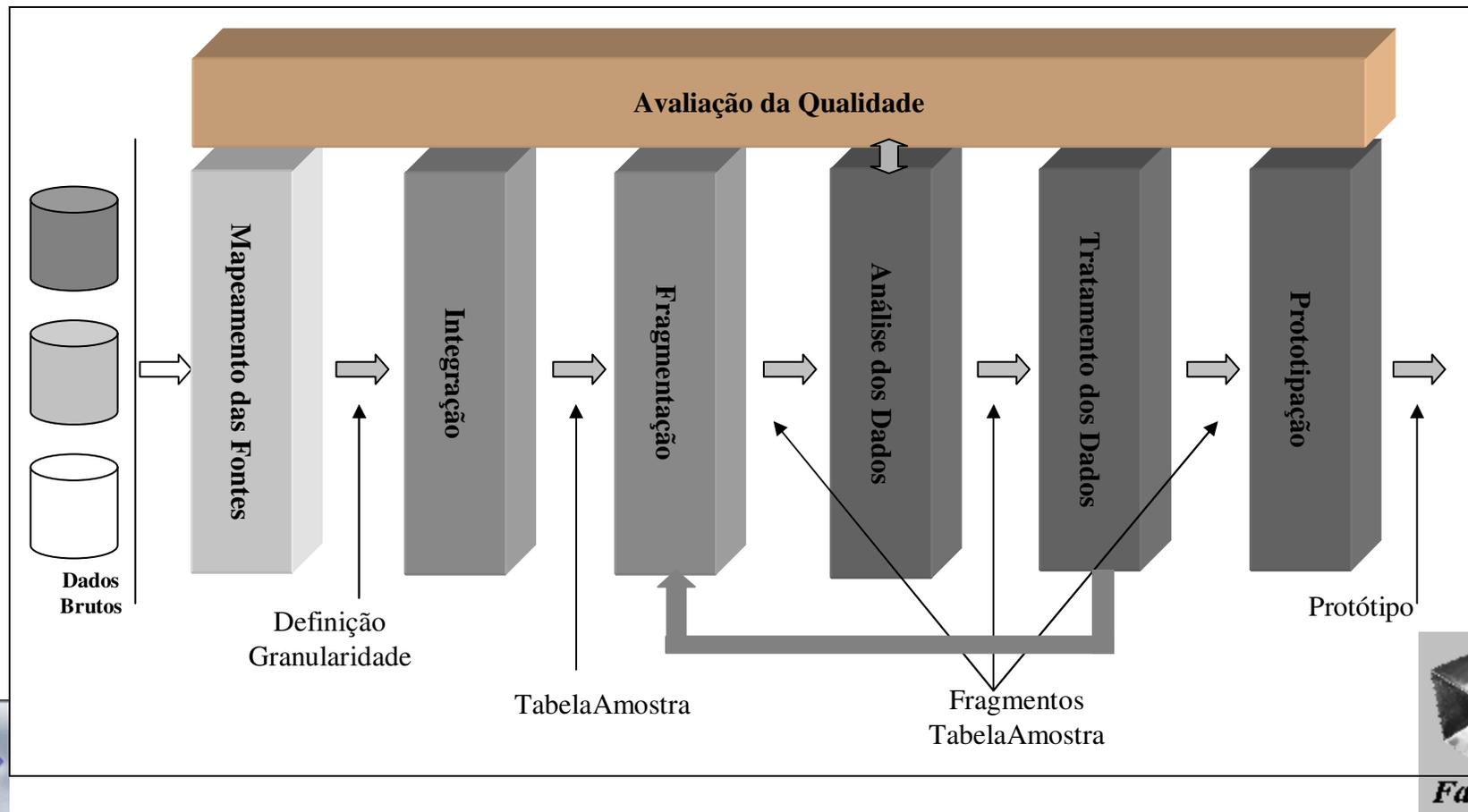
**Uma simples metodologia de  
modelagem de dados para um  
Data Mart**

# Modelagem do DM

- É projetado para atender às necessidades de um setor ou departamento
- Requer uma boa definição do escopo
- Pode ser modelado a partir do:
  - DW: garante a consistência e o mapeamento dos dados
  - Sist. Operativos: pode gerar DM Independentes nos casos que não exista um controle e gerência de metadados
- Normalmente, se utiliza o esquema estrela para a modelagem



# Arquitetura Sugerida

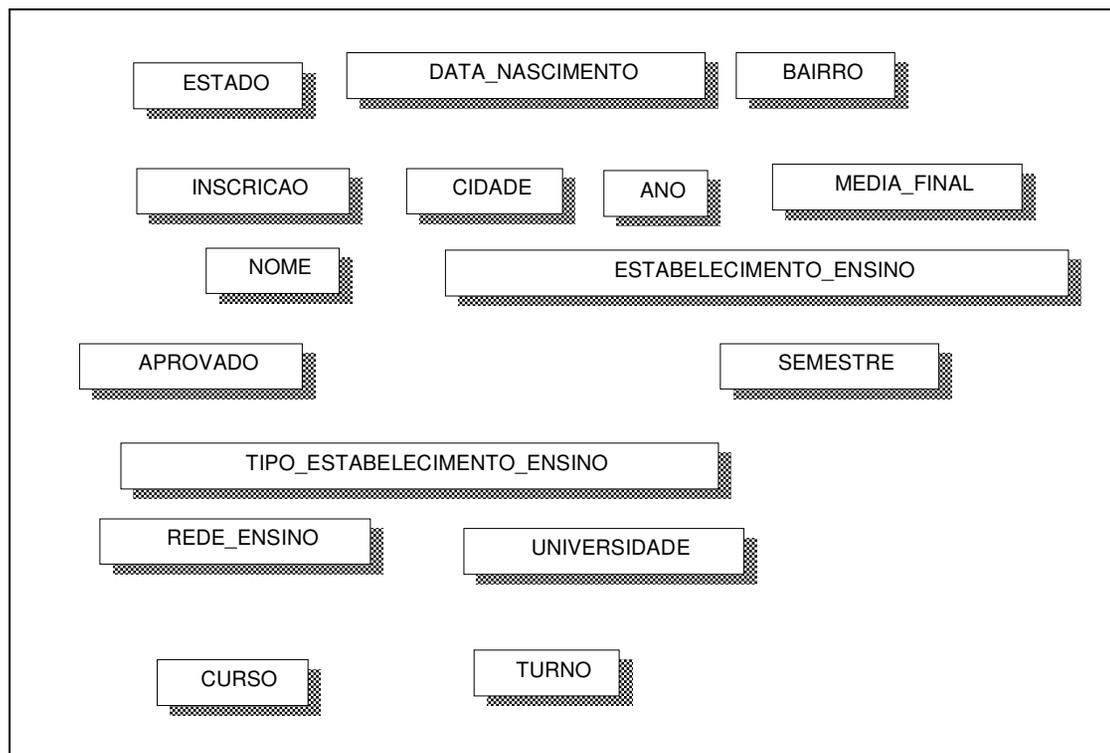


# Resumo Método

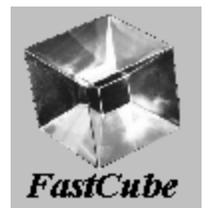
- **Definição dos objetivos e escopo**
- Mapeamento e seleção dos dados que serão utilizados.
- **Definição da granularidade.**
- Integração dos dados de acordo com a granularidade pretendida em uma tabela desnormalizada.
- Fragmentação em colunas com geração de metadados (uso da distribuição de frequência).
- Análise e tratamento das colunas com geração de novas colunas.
- Montagem do modelo dimensional a partir das colunas tratadas e seus metadados.
- Carga no modelo.



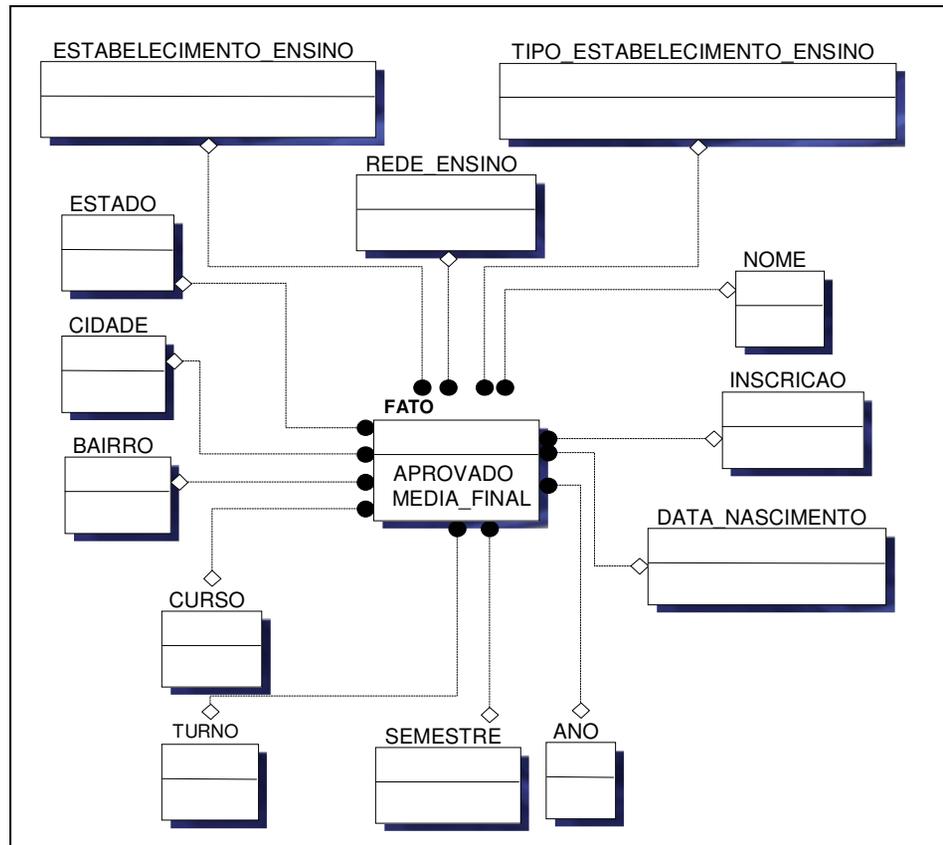
# Montagem do Modelo Dimensional



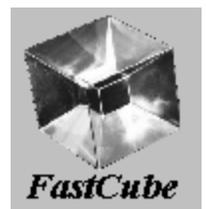
**Seleção dos atributos relevantes para o Data Mart.**



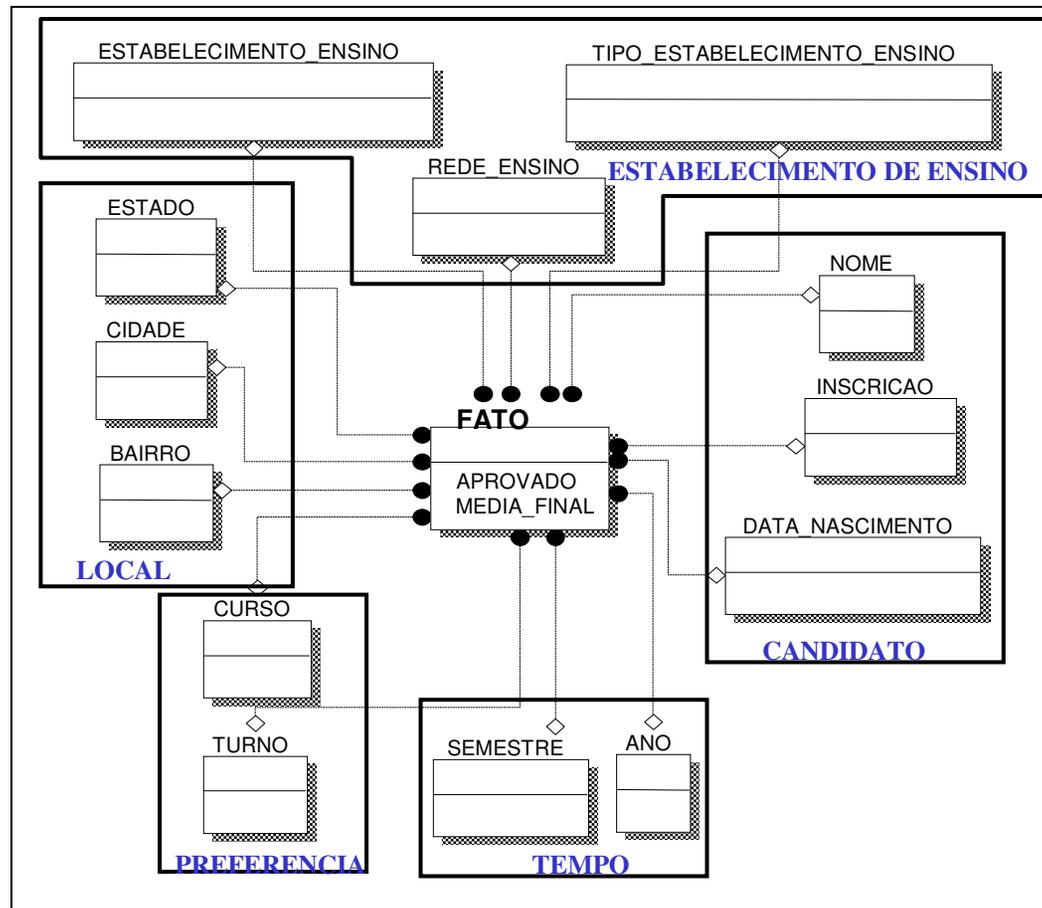
# Montagem do Modelo Dimensional



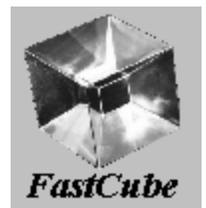
Uma tabela de fato e dimensões com apenas um atributo



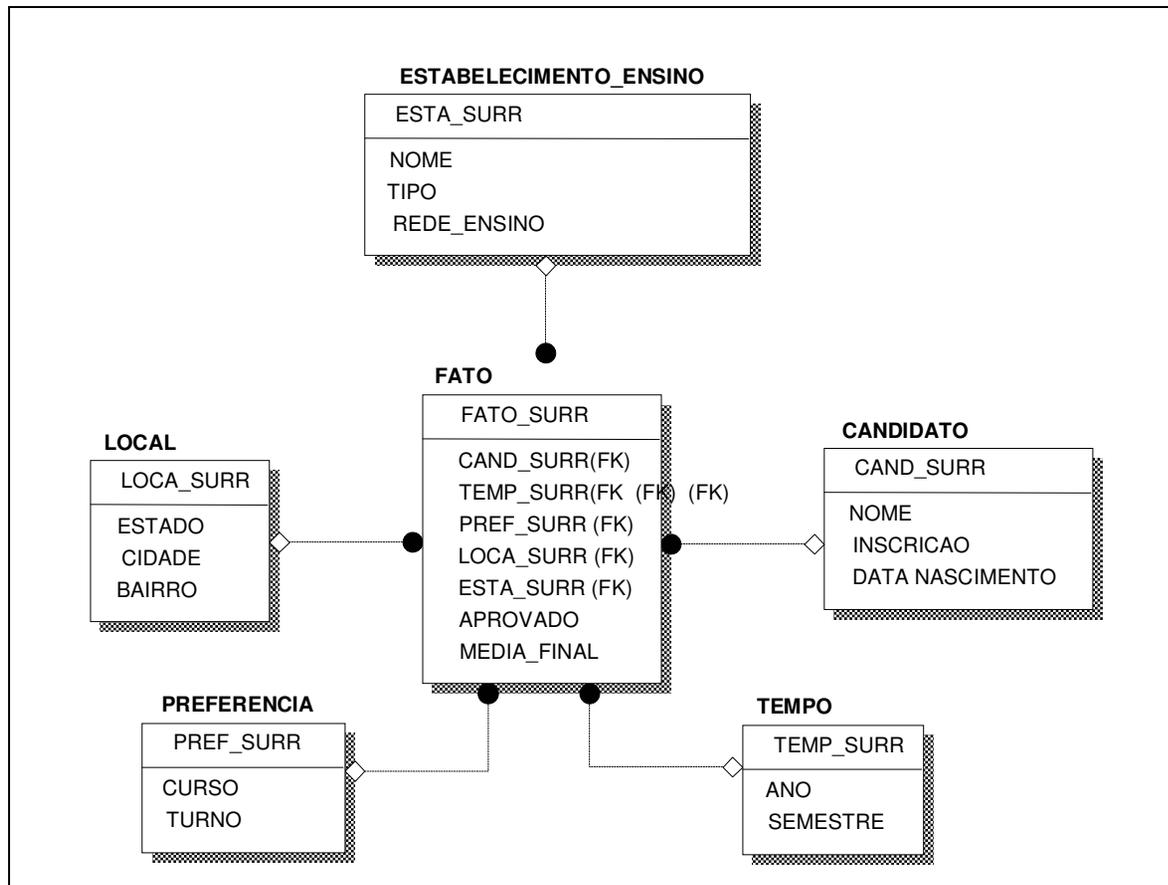
# Montagem do Modelo Dimensional



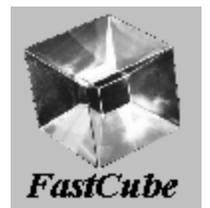
Agrupamento para formação das dimensões.



# Montagem do Modelo Dimensional



Modelo com *surrogate key* e dimensões montadas.



# Algumas Ferramentas DW/OLAP

**DISCOVERER**

**MAESTRO**

**DSS**

**BUSINESS OBJECT**

**MS OLAP**

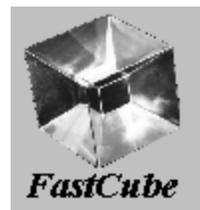
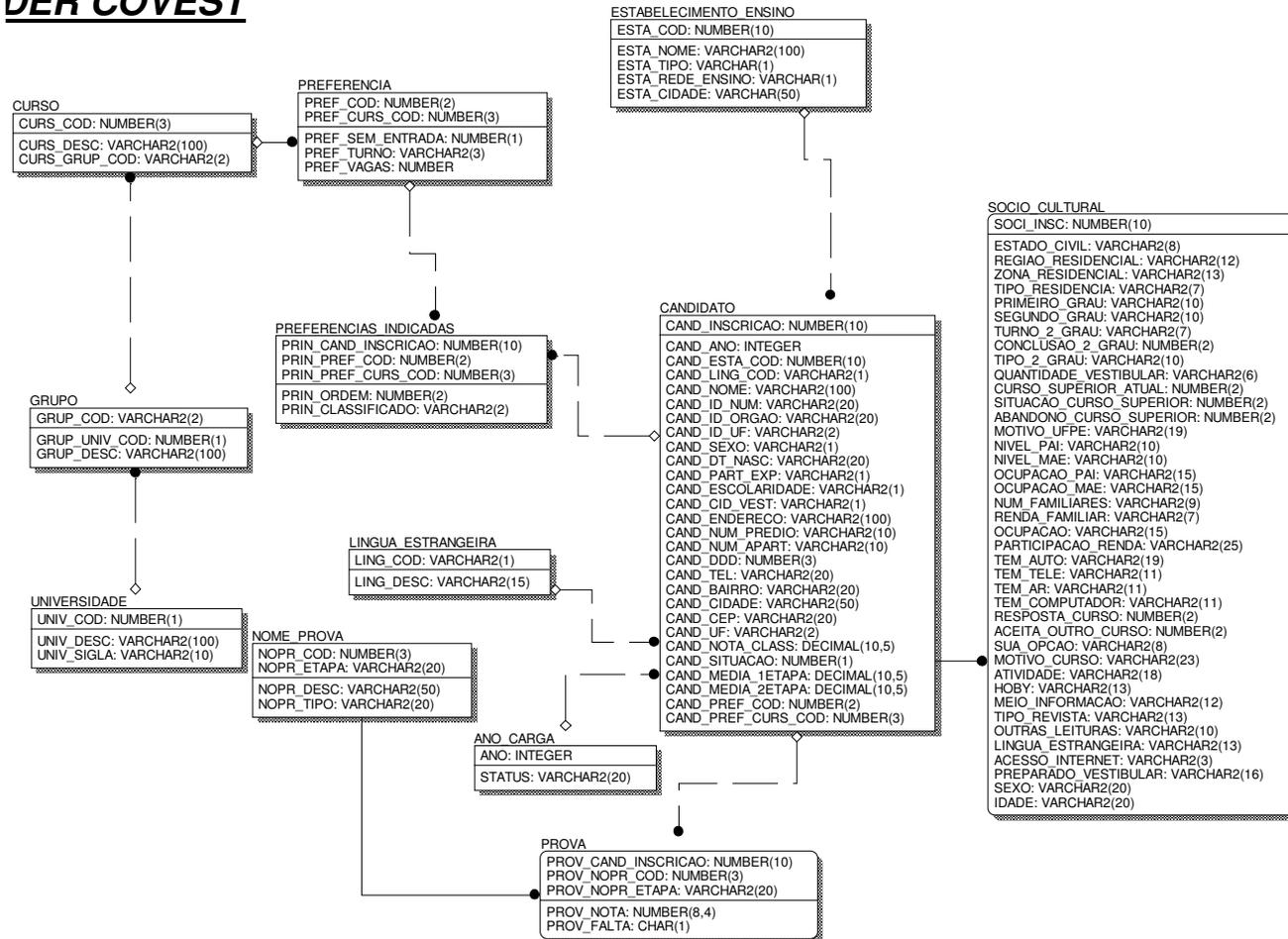
**SAGENT**



# Esquema Relacional Normalizado

DATA MART  
COVEST

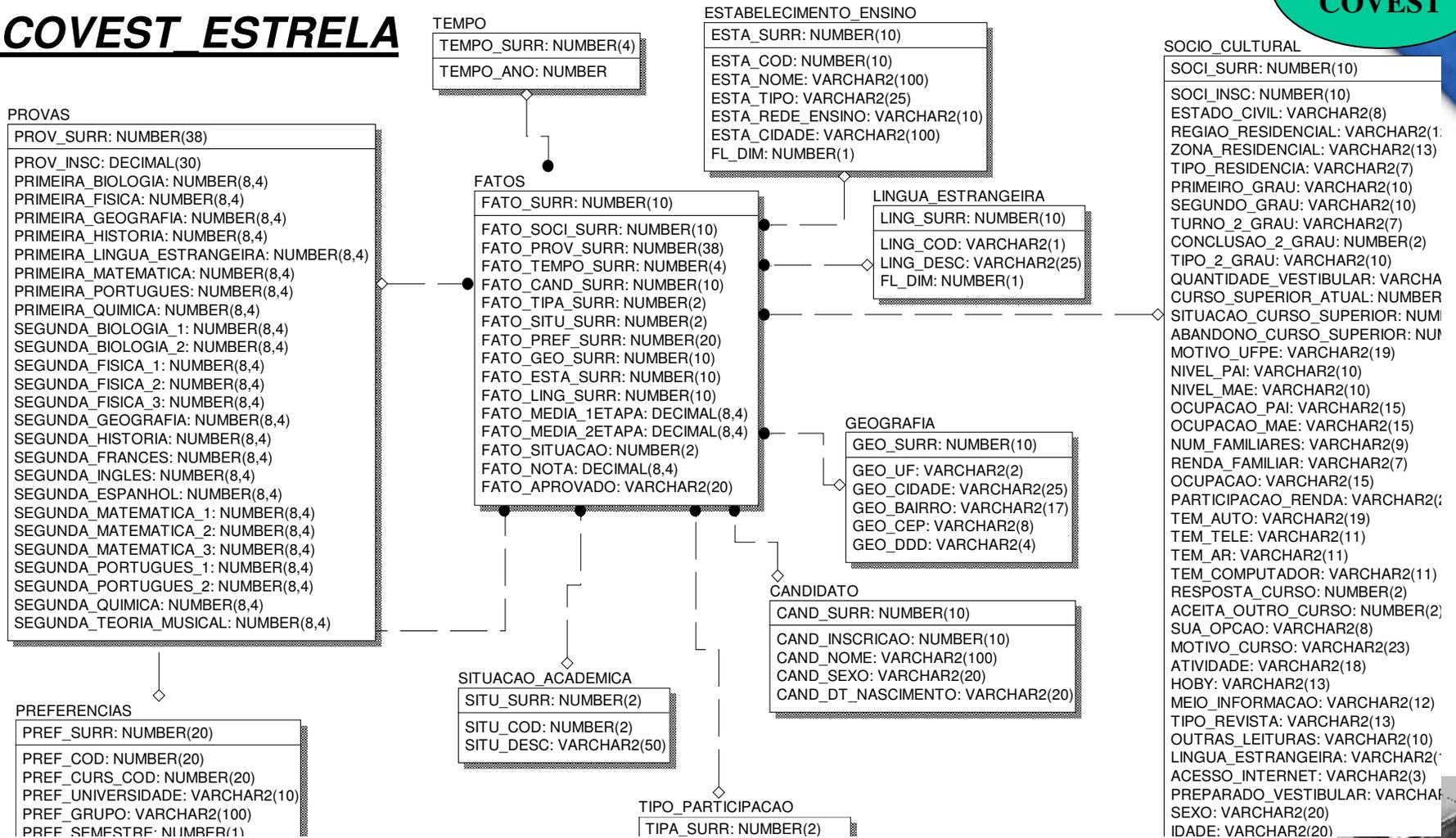
## DER COVEST



# Esquema Estrela (Star Schema)

**DATA MART  
COVEST**

## **COVEST ESTRELA**

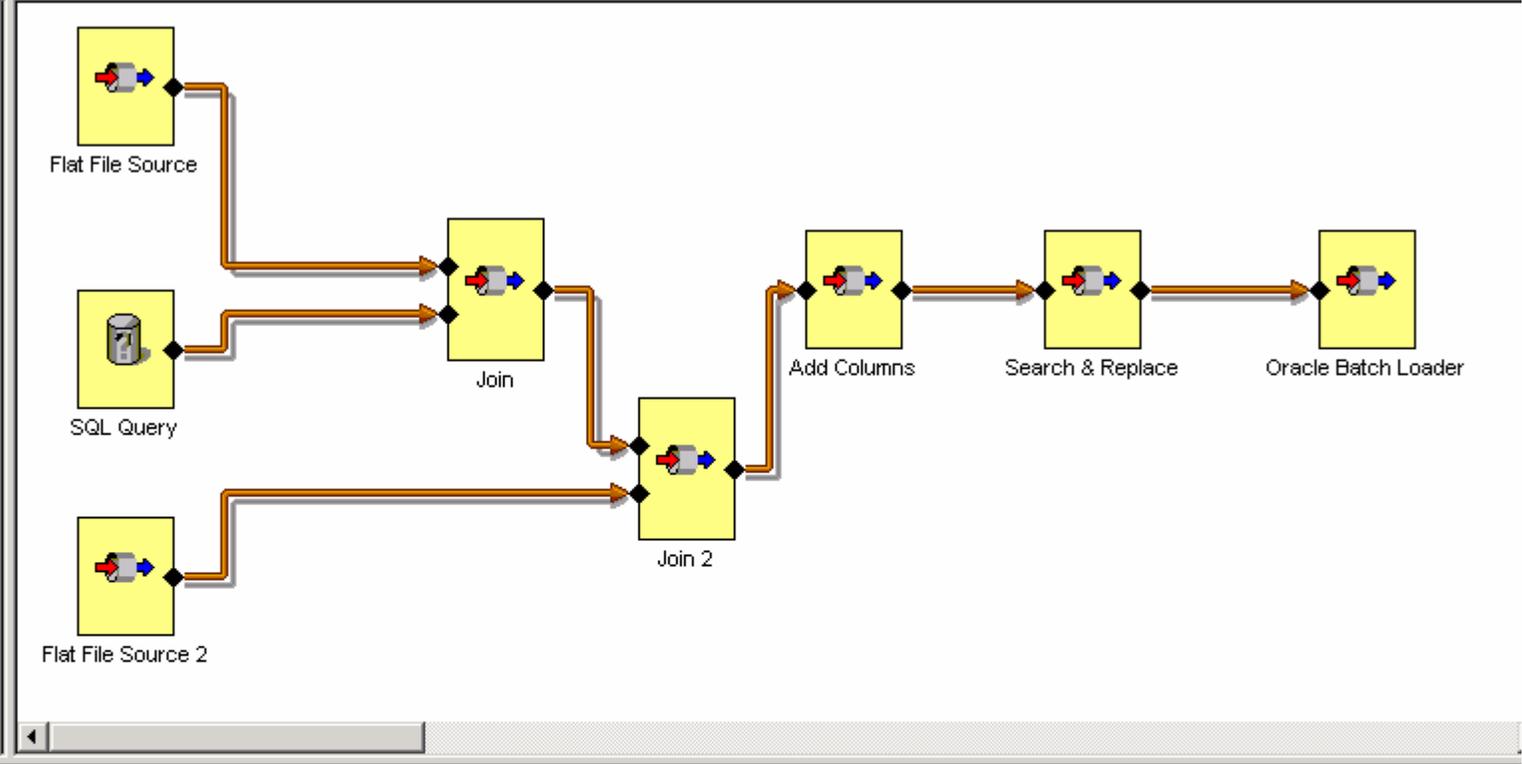


Edit View Tools Help



List all items that I can access

- Consultas - Instrução dos Pais
- Consultas - Micro em Casa
- Consultas - Motivo Curso
- Consultas - Motivo Curso Total
- Consultas - Participação Renda
- Consultas - Rede Ensino
- Consultas - Renda Familiar
- Detail & Impact Analysis - BaseView
- Detail & Impact Analysis - DS Column
- Detail & Impact Analysis - MetaView
- Detail & Impact Analysis - Part
- Detail & Impact Analysis - Plan
- Detail & Impact Analysis - Transform
- Detail & Impact Analysis - User
- Inventory - BaseView
- Inventory - DS Column
- Inventory - MetaView
- Inventory - Part
- Inventory - Plan
- Inventory - Pub/Sub Plans
- Inventory - Transform
- Inventory - User
- Plan Tracking - Load Statistics
- Plan Tracking - SQL Statistics
- Populate Prompt 1
- Populate Prompt 2
- Populate Prompt 3
- Populate Prompt 4
- Populate Prompt 5
- Jsage Statistics - Aggregate Usage
- Jsage Statistics - Table Usage
- Exemplo-Carga

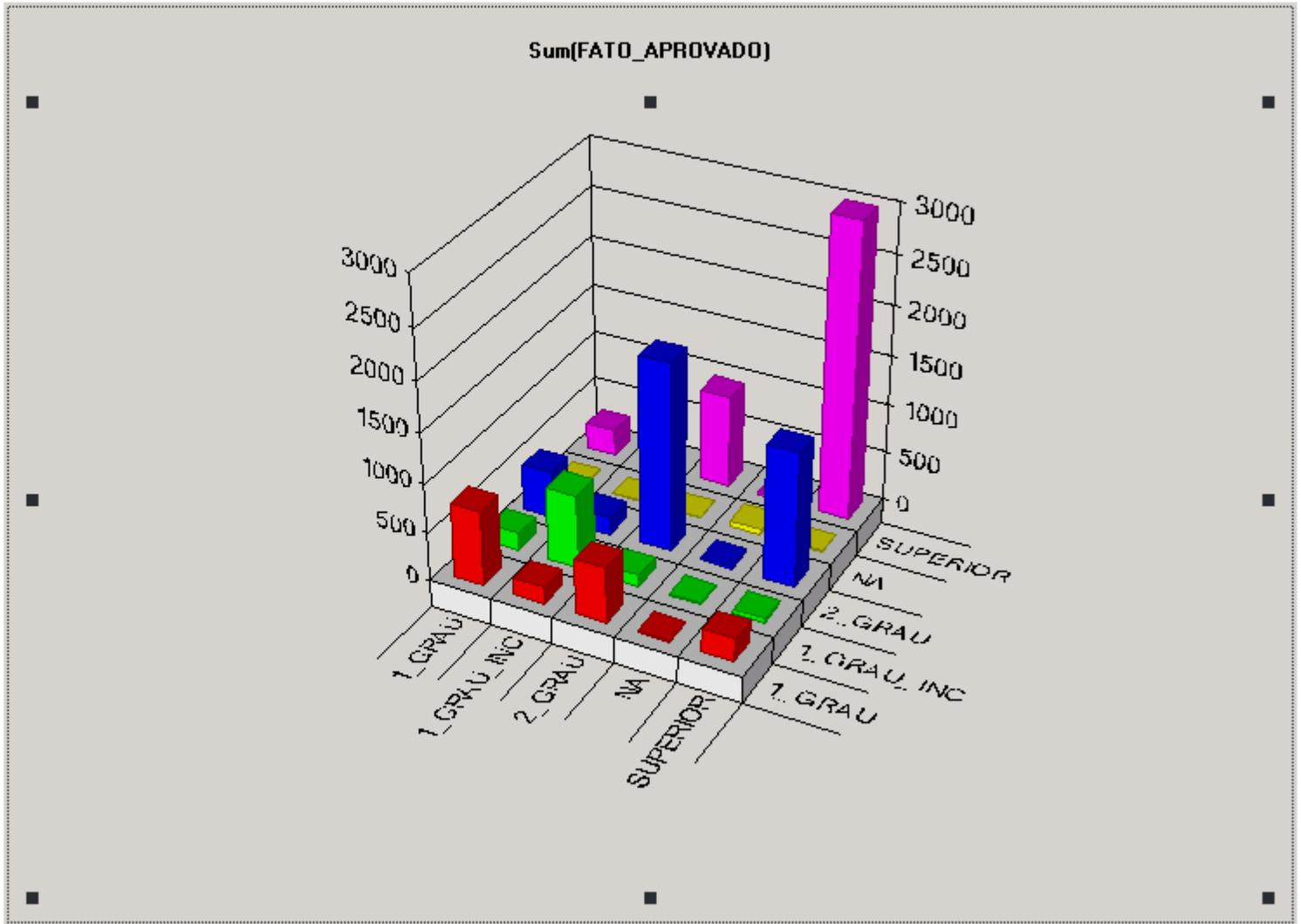


Update Ready



List all items that I can access

- Consultas - Instrução dos Pais
- Consultas - Micro em Casa
- Consultas - Motivo Curso
- Consultas - Motivo Curso Total
- Consultas - Participação Renda
- Consultas - Rede Ensino
- Consultas - Renda Familiar
- Detail & Impact Analysis - BaseView
- Detail & Impact Analysis - DS Column
- Detail & Impact Analysis - MetaView
- Detail & Impact Analysis - Part
- Detail & Impact Analysis - Plan
- Detail & Impact Analysis - Transform
- Detail & Impact Analysis - User
- Inventory - BaseView
- Inventory - DS Column
- Inventory - MetaView
- Inventory - Part
- Inventory - Plan
- Inventory - Pub/Sub Plans
- Inventory - Transform
- Inventory - User
- Plan Tracking - Load Statistics
- Plan Tracking - SQL Statistics
- Populate Prompt 1
- Populate Prompt 2
- Populate Prompt 3
- Populate Prompt 4
- Populate Prompt 5
- Jsage Statistics - Aggregate Usage
- Jsage Statistics - Table Usage
- Exemplo-Carga

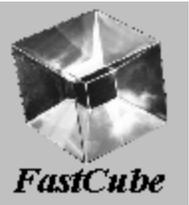
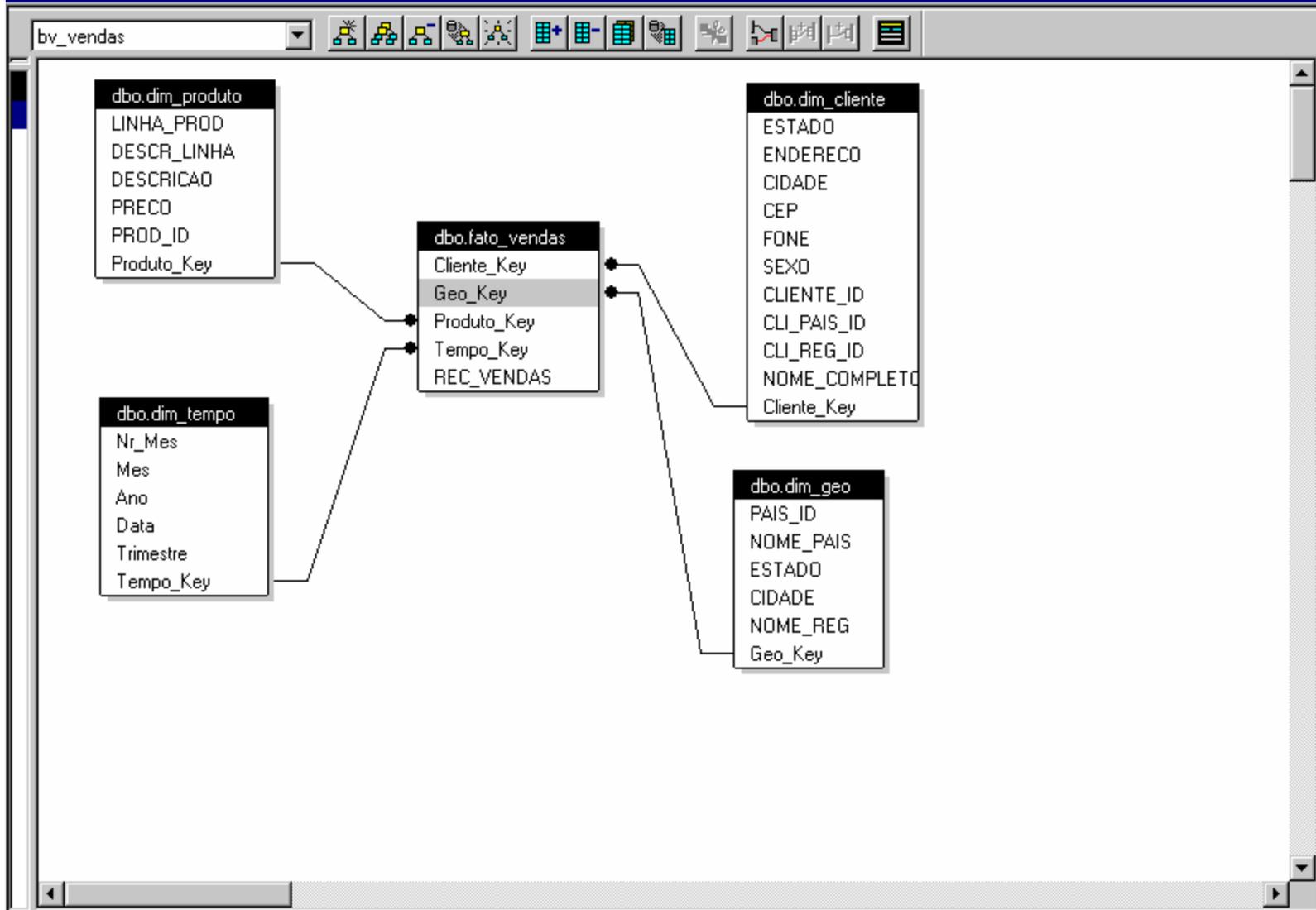


Analysis Excel

Update

Updated (1.01 seconds, 25 records)

bv\_vendas - BaseView Editor



# Sistemas de Suporte à Decisão

- SAGENT - Design Studio

MetaView: Vendas

- Vendas
  - Clientes
    - CEP
    - Cidade
    - Codigo Cliente
    - Endereco
    - Fone
    - Nome
    - Sexo
    - UF
  - Local
    - Cidade
    - Pais
    - Regiao
    - UF Venda
  - Medidas - Fatos
    - Receita Vendas
  - Produtos
    - Codigo Produto
    - Descricao
    - Descricao Linha
    - Preco
  - Referencias Tempo
    - Ano
    - Data
    - Mes
    - Nr. Mes
    - Trimestre

Sum(Receita Vendas)

UF Venda Cidade

Ano Nr. Mes

Receita Vendas

Analysis / Update Ready

Ready



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



### Receita de Vendas

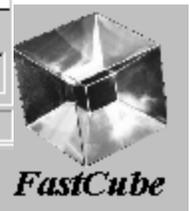
UF Venda	Ano 1997	Trimestre 1998	Mes 1999	Total
BA	1,164,018.00	200,600.00	1,364,224.00	2,728,842.00
MG	461,950.00	1,075,123.00	236,611.00	1,773,684.00
SE	1,126,755.00	422,160.00	124,545.00	1,673,460.00
SP	1,373,761.00	166,615.00	264,942.00	1,805,318.00
Total	4,126,484.00	1,864,498.00	1,990,322.00	7,981,304.00



Update Updated (1.07 seconds, 81 records)

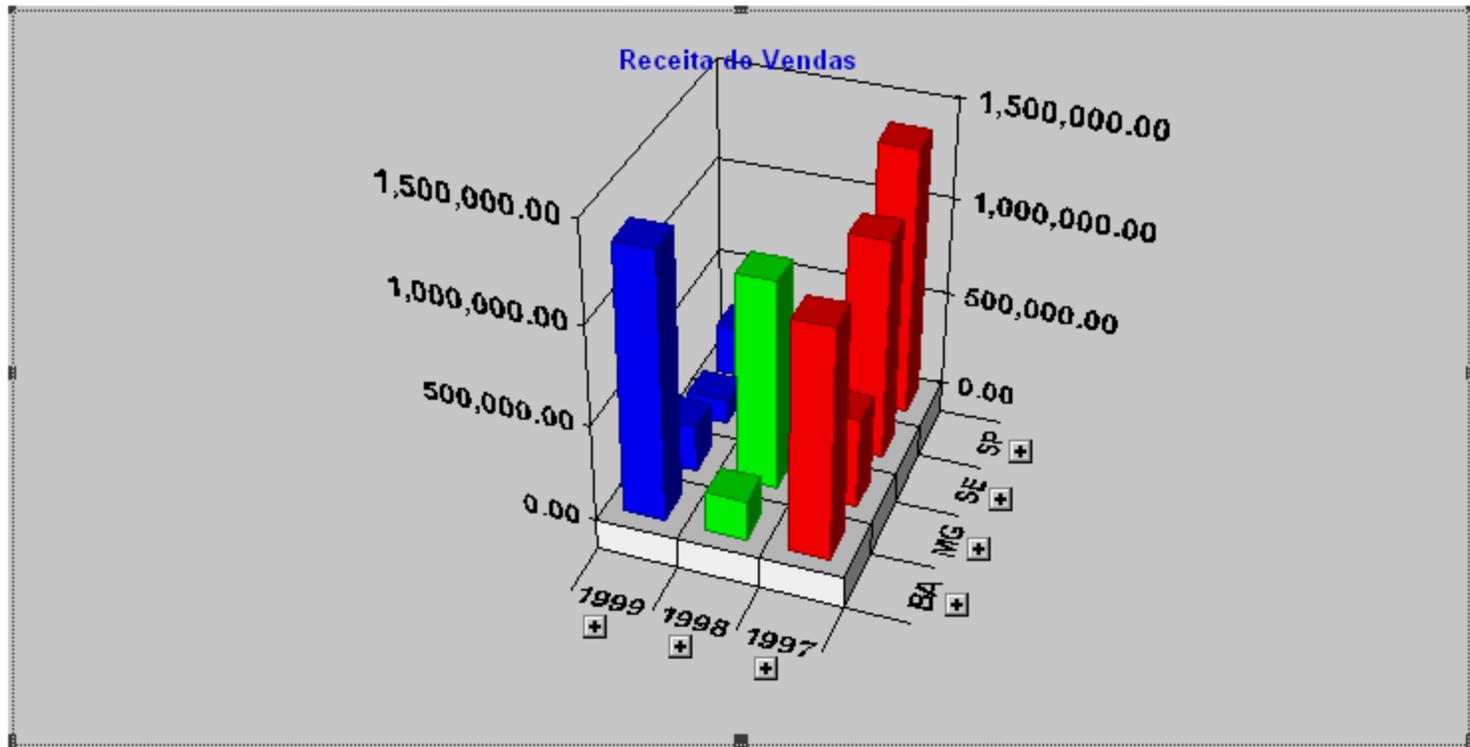
Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



Analysis



Update

Updated (1.07 seconds, 81 records)

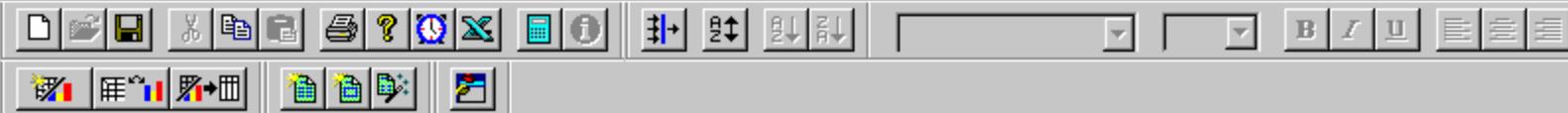
Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



### Receita de Vendas

Ano Trimestre Mes

UF Venda	Cidade	1997				Total	1998	1999	Total
		1	2	3	4				
BA		66,500.00	287,375.00	291,722.00	518,421.00	1,164,018.00	200,600.00	1,364,224.00	2,728,842.00
MG	Belo Horizonte	23,195.00	52,800.00	32,000.00	93,000.00	200,995.00	1,075,123.00	206,111.00	1,482,229.00
	Uberlandia		40,000.00	9,500.00	211,455.00	260,955.00		30,500.00	291,455.00
	Surface Sum	23,195.00	92,800.00	41,500.00	304,455.00	461,950.00	1,075,123.00	236,611.00	1,773,684.00
SE		131,611.00	341,323.00	168,520.00	485,301.00	1,126,755.00	422,160.00	124,545.00	1,673,460.00
SP		62,000.00	97,000.00	136,050.00	1,078,711.00	1,373,761.00	166,615.00	264,942.00	1,805,318.00
Total		283,306.00	818,498.00	637,792.00	2,386,888.00	4,126,484.00	1,864,498.00	1,990,322.00	7,981,304.00



Update Updated (1.07 seconds, 81 records)

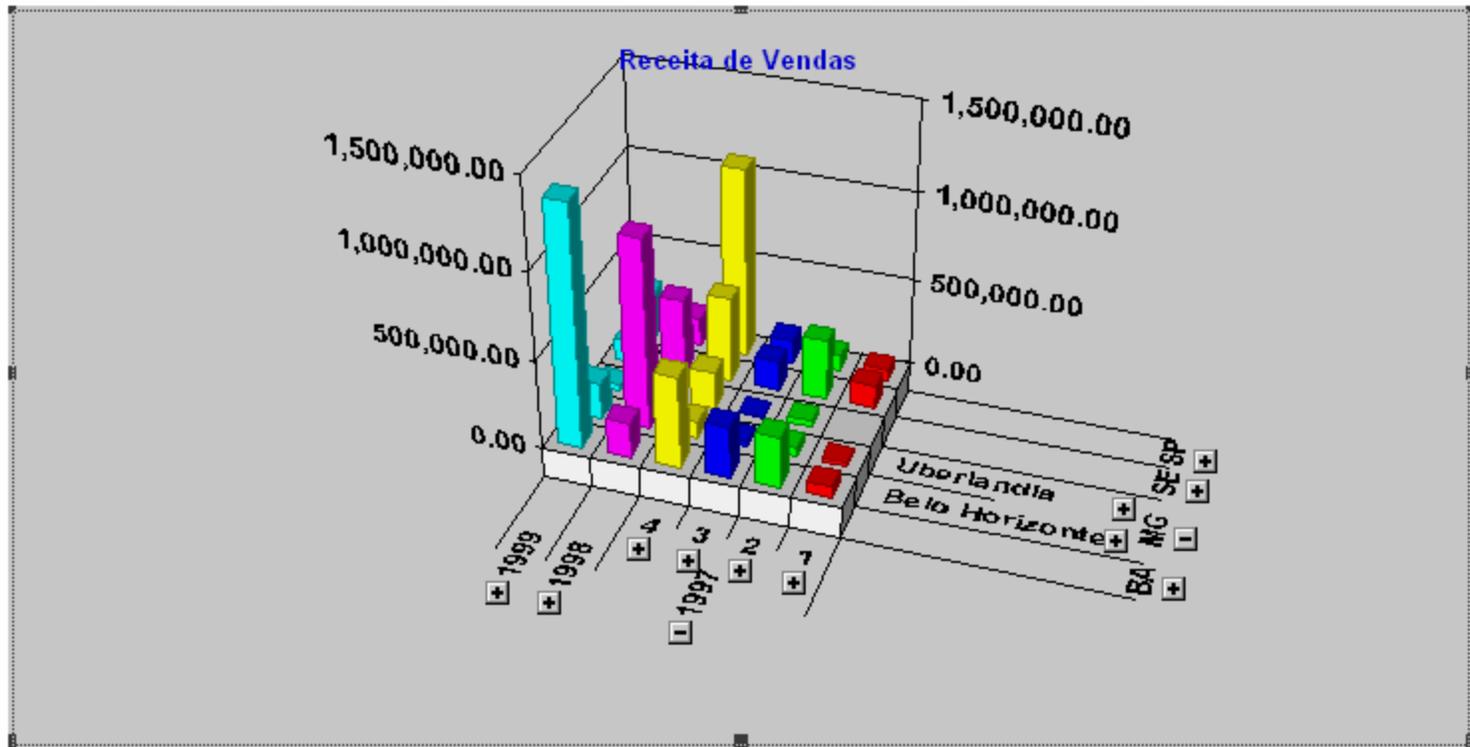
Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



Analysis



Update

Updated (1.07 seconds, 81 records)

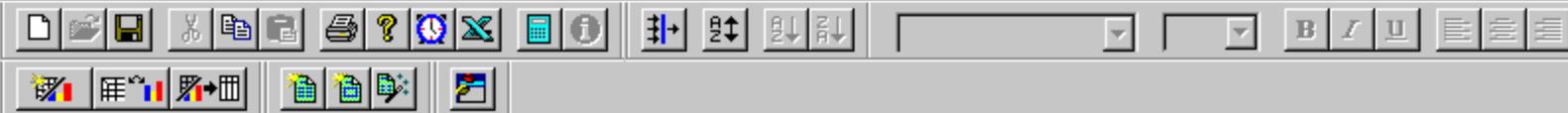
Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



### Receita de Vendas

Ano: 1997 Trimestre: 1 Mes: 1

UF Venda	Cidade	February	March	January	Total	+2	+3	+4	Total	
BA		9,500.00		57,000.00	66,500.00	287,375.00	291,722.00	518,421.00	1,164,018.00	2
MG	Belo Horizonte	21,695.00		1,500.00	23,195.00	52,800.00	32,000.00	93,000.00	200,995.00	1.0
	Uberlandia					40,000.00	9,500.00	211,455.00	260,955.00	
	Surface Sum	21,695.00		1,500.00	23,195.00	92,800.00	41,500.00	304,455.00	461,950.00	1.0
SE		45,000.00	86,611.00		131,611.00	341,323.00	168,520.00	485,301.00	1,126,755.00	4
SP			17,000.00	45,000.00	62,000.00	97,000.00	136,050.00	1,078,711.00	1,373,761.00	1
<b>Total</b>		<b>76,195.00</b>	<b>103,611.00</b>	<b>103,500.00</b>	<b>283,306.00</b>	<b>818,498.00</b>	<b>637,792.00</b>	<b>2,386,888.00</b>	<b>4,126,484.00</b>	<b>1.8</b>

Analysis



Update

Updated (1.07 seconds, 81 records)

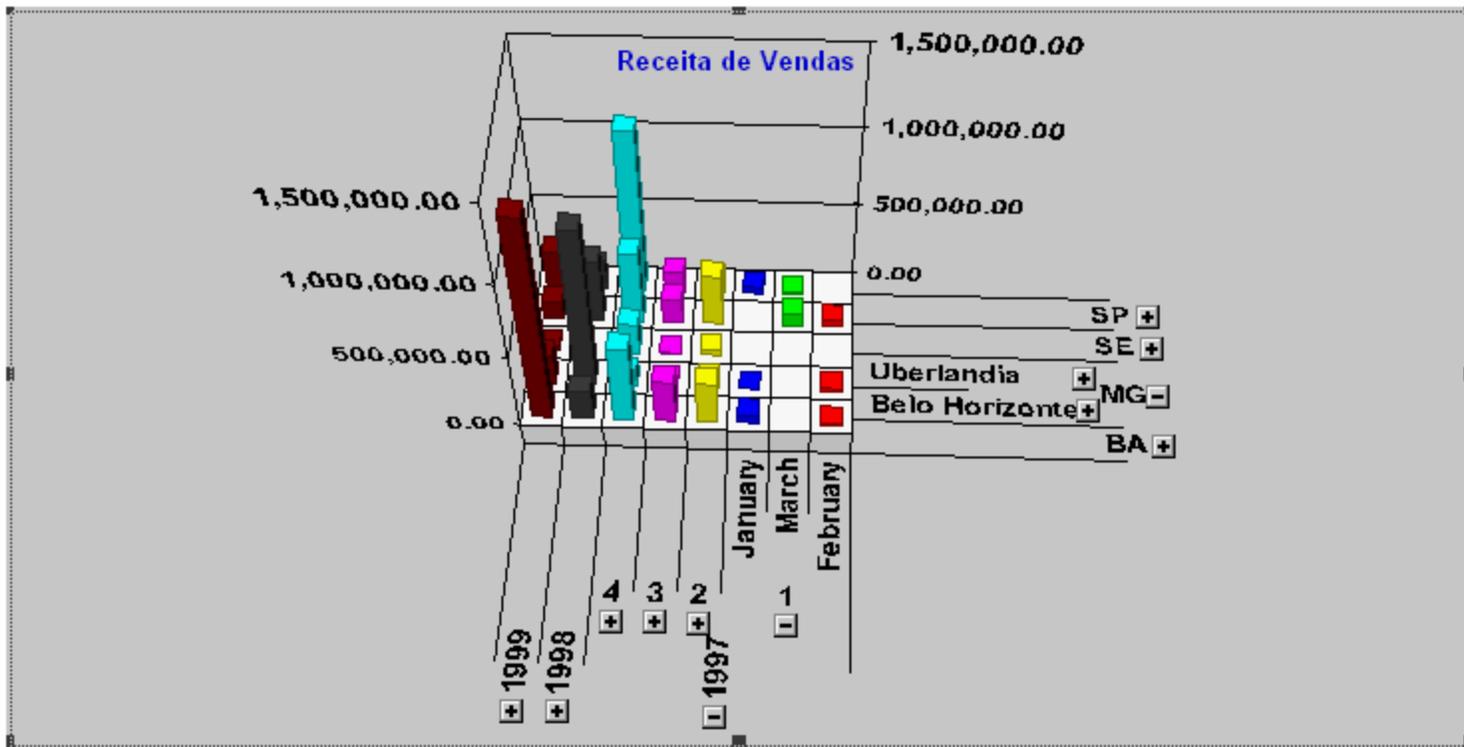
Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



Analysis



Update

Updated (1.07 seconds, 81 records)

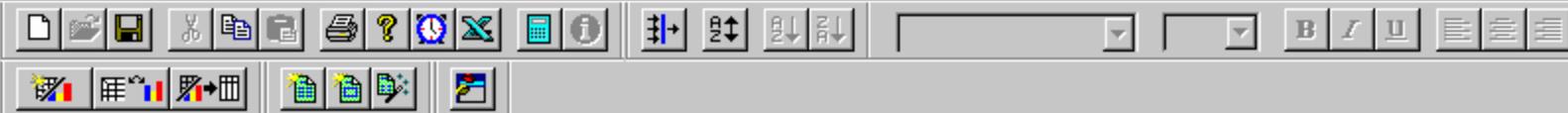
Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



### Receita de Vendas

Ano Trimestre Mes

1997 1998

1 2 3 4 Total

UF Venda	February	March	January	Total	2	3	4	Total	1998
<input checked="" type="checkbox"/> BA	9,500.00		57,000.00	66,500.00	287,375.00	291,722.00	518,421.00	1,164,018.00	200,6
<input checked="" type="checkbox"/> MG	21,695.00		1,500.00	23,195.00	92,800.00	41,500.00	304,455.00	461,950.00	1,075,1
<input type="checkbox"/> SE	45,000.00	86,611.00		131,611.00	341,323.00	168,520.00	485,301.00	1,126,755.00	422,1
<input checked="" type="checkbox"/> SP		17,000.00	45,000.00	62,000.00	97,000.00	136,050.00	1,078,711.00	1,373,761.00	166,6
Total	76,195.00	103,611.00	103,500.00	283,306.00	818,498.00	637,792.00	2,386,888.00	4,126,484.00	1,864,4

Analysis



Update

Updated (1.07 seconds, 81 records)

Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



### Receita de Vendas

Ano ▼ Trimestre ▼ Mes ▼

1997 1998

1 2 3 4 Total

UF Venda ▼	February	March	January	Total	2	3	4	Total	1998
BA	9,500.00		57,000.00	66,500.00	287,375.00	291,722.00	518,421.00	1,164,018.00	200,6
MG	21,695.00		1,500.00	23,195.00	92,800.00	41,500.00	304,455.00	461,950.00	1,075,1
Total	31,195.00		58,500.00	89,695.00	380,175.00	333,222.00	822,876.00	1,625,968.00	1,275,7

Analysis



Update

Updated (1.07 seconds, 81 records)

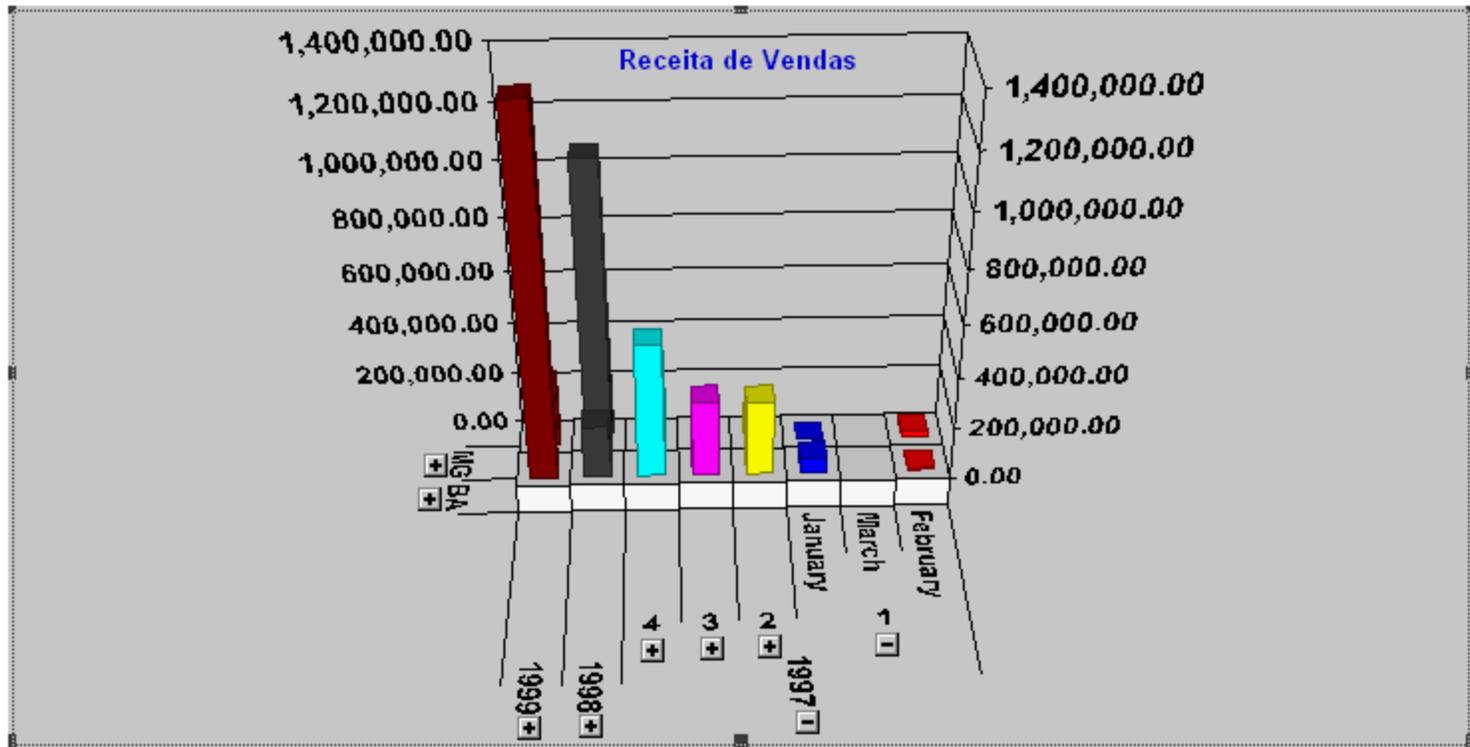
Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



Analysis



Update

Updated (1.07 seconds, 81 records)

Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



### Receita de Vendas

UF Venda	Ano	Trimestre	Mes	1997	1998	1999	Total
+ BA	1997			1,364,224.00	200,600.00		2,728,842.00
+ MG	1997			1,364,224.00	1,075,123.00	236,611.00	1,773,684.00
Total	1997			1,625,968.00	1,275,723.00	1,600,835.00	4,502,526.00

Analysis

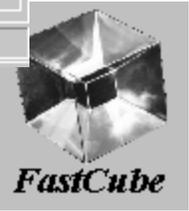


Update

Updated (1.07 seconds, 81 records)

Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



### Receita de Vendas

UF Venda	Ano	Trimestre	Mes	Total
	1999			
BA				1,364,224.00
MG				236,611.00
Total				1,600,835.00

Analysis



Update Updated (1.07 seconds, 81 records)

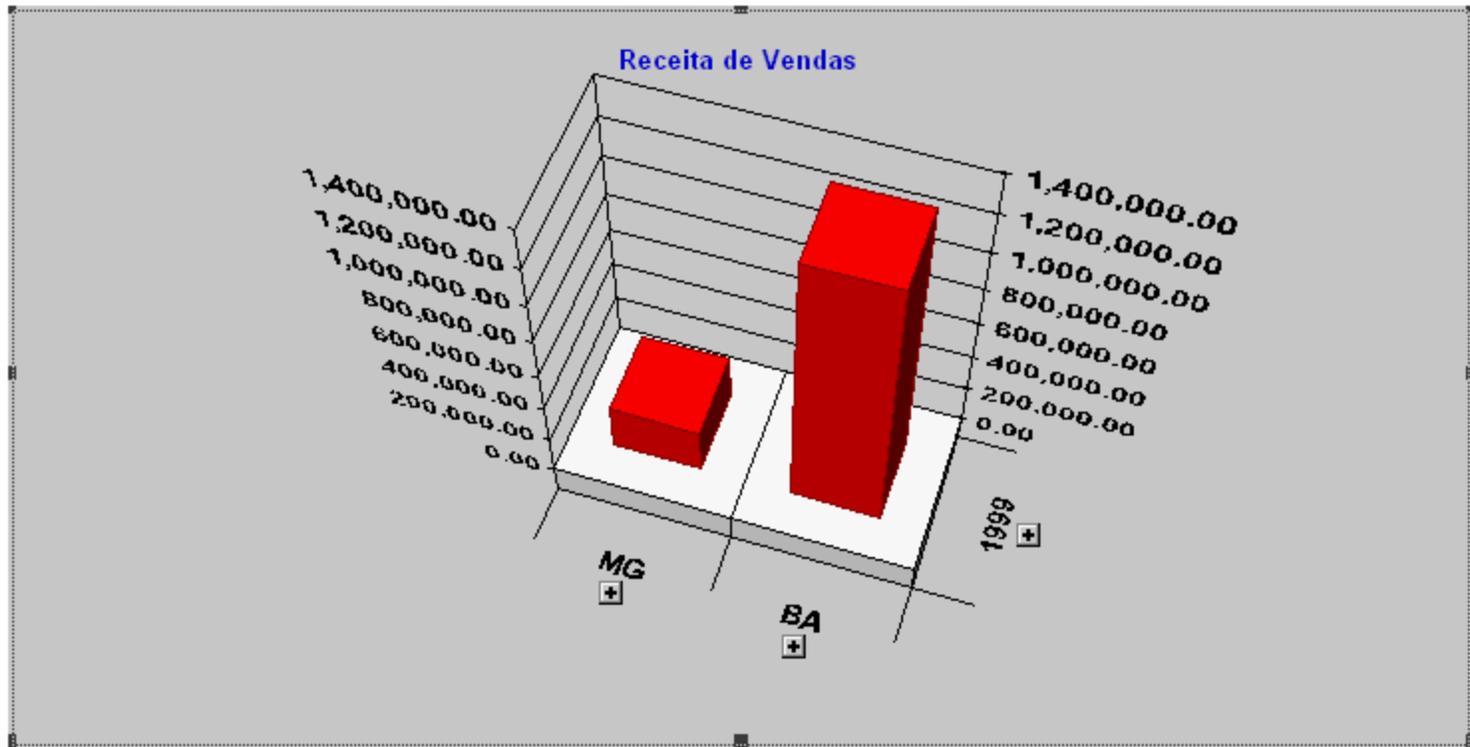
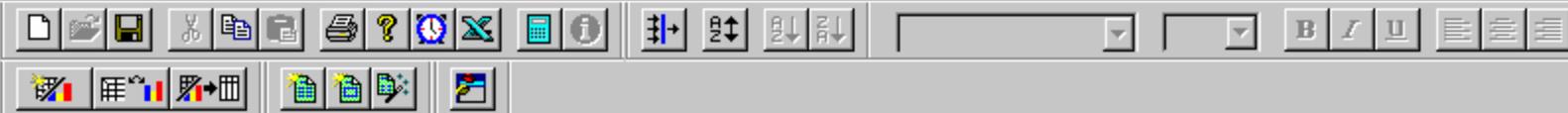
Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



Analysis



Update

Updated (1.07 seconds, 81 records)

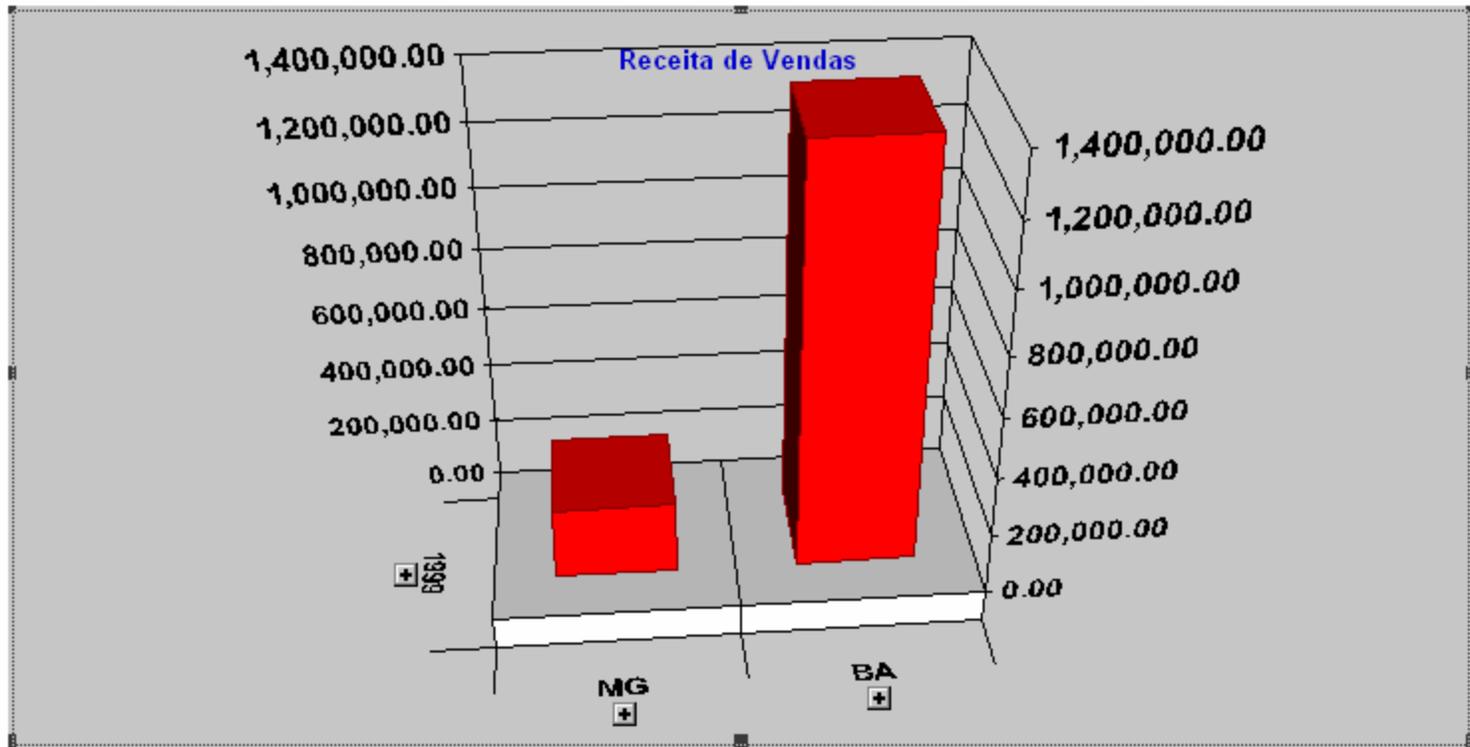
Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



Analysis



Update

Updated (1.07 seconds, 81 records)

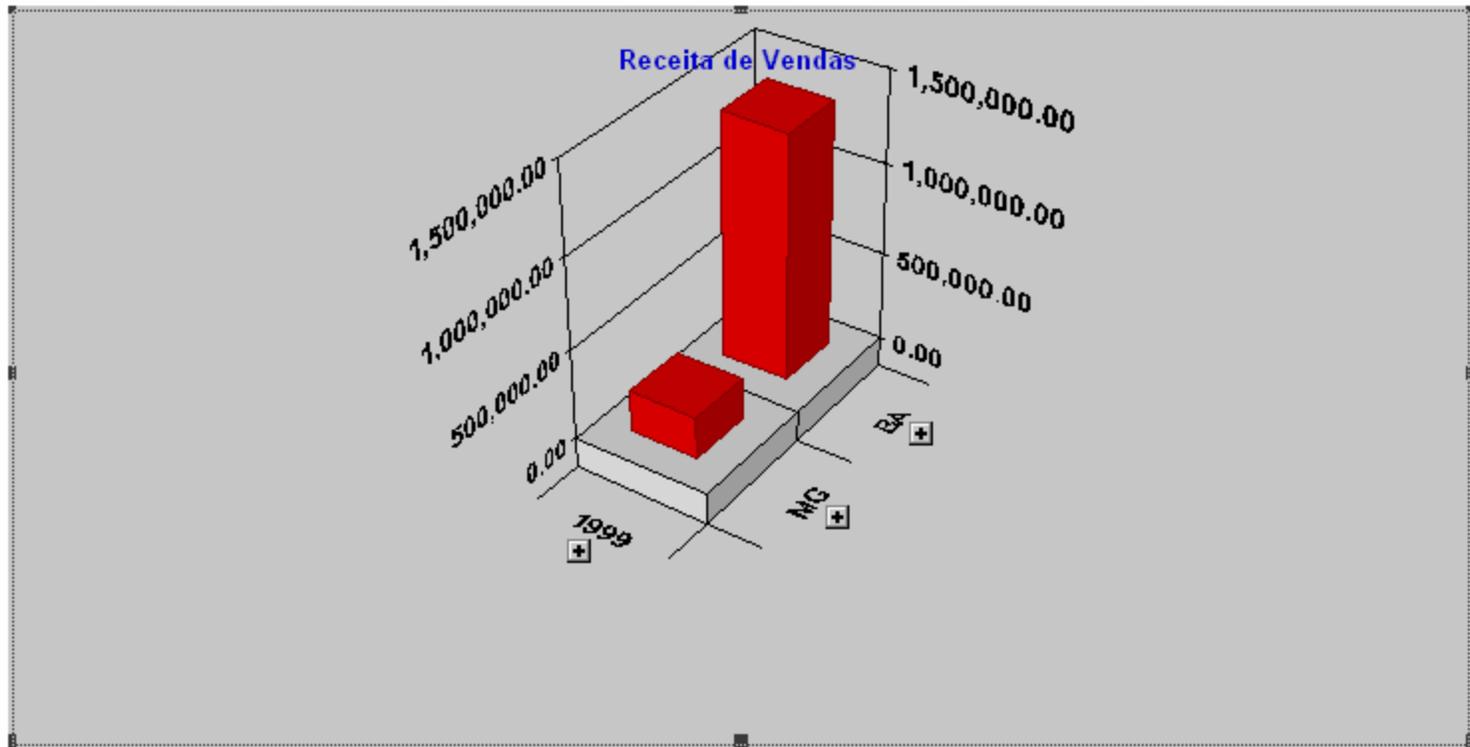
Ready

NUM



Untitled, created 10/05/2001 07:17:01 PM - Design Studio

File Edit View Tools Help



Analysis



Update

Updated (1.07 seconds, 81 records)

Ready

NUM



# Bibliografia

- [1]. Kimball, Ralph, Reeves, Laura, Ross, Margy – The Data Warehouse LifeCycle Toolkit – John Willey & Sons – 1998.
- [2]. Inmon, William – Enterprise Meta Data – DM Review – 1998.
- [3]. Kimball, Ralph – Working in Web Time – Intelligent Enterprise Magazine - 1999.
- [4]. H.Brackett, Michael - The Warehouse Challenge - John Willey & Sons – 1996.
- [5] Help on–line Sagent.
- [6] Documentação técnica do Sagent.

